

## A COVID-19 risk score combining chest CT radiomics and clinical characteristics to differentiate COVID-19 pneumonia from other viral pneumonias

Zuhua Chen<sup>1,2,\*</sup>, Xiadong Li<sup>3,4,\*</sup>, Jiawei Li<sup>5</sup>, Shirong Zhang<sup>4</sup>, Pengfei Zhou<sup>3</sup>, Xin Yu<sup>3</sup>, Yao Ren<sup>3</sup>, Jiahao Wang<sup>3</sup>, Lidan Zhang<sup>3</sup>, Yunjiang Li<sup>1</sup>, Baoliang Wu<sup>1</sup>, Yanchun Hou<sup>1</sup>, Ke Zhang<sup>3</sup>, Rongjun Tang<sup>3</sup>, Yongguang Liu<sup>1</sup>, Zhongxian Ding<sup>4</sup>, Bin Yang<sup>4</sup>, Qinghua Deng<sup>3</sup>, Qin Lin<sup>8</sup>, Ke Nie<sup>6</sup>, Zhaobin Cai<sup>1,2</sup>, Shenglin Ma<sup>3,4</sup>, Yu Kuang<sup>7</sup>

<sup>1</sup>Department of Radiology, Hangzhou Xixi Hospital, Hangzhou 310000, Zhejiang, China

<sup>2</sup>Department of Radiology, Hangzhou 6th People's Hospital, the Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou 310000, Zhejiang, China

<sup>3</sup>Department of Radiation Oncology, Hangzhou Cancer Hospital, Zhejiang University Cancer Centre, Hangzhou First People's Hospital Group, Hangzhou 310000, Zhejiang, China

<sup>4</sup>Department of Radiation Oncology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou 310000, Zhejiang, China

<sup>5</sup>Department of Radiology, The Fourth Clinical Medical College, Zhejiang Chinese Medical University, Hangzhou 310000, Zhejiang, China

<sup>6</sup>Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ 07097, USA

<sup>7</sup>Medical Physics Program, University of Nevada, Las Vegas, NV 89154, USA

<sup>8</sup>Department of Radiation Oncology, Xiamen Cancer Hospital, The First Affiliated Hospital of Xiamen University, Teaching Hospital of Fujian Medical University, Xiamen 361003, Fujian, China

\*Equal contribution

**Correspondence to:** Xiadong Li, Zhaobin Cai, Shenglin Ma, Yu Kuang; **email:** [lixiadong2019@outlook.com](mailto:lixiadong2019@outlook.com), <https://orcid.org/0000-0002-9956-9111>; [caizhaobin2020@sina.com](mailto:caizhaobin2020@sina.com), <https://orcid.org/0000-0002-5545-569X>; [mashenglin@medmail.com.cn](mailto:mashenglin@medmail.com.cn), <https://orcid.org/0000-0003-4138-0305>; [yu.kuang@unlv.edu](mailto:yu.kuang@unlv.edu)

**Keywords:** coronavirus disease 2019, COVID-19, severe acute respiratory syndrome coronavirus 2, chest CT, radiomics, nomogram

**Received:** September 24, 2020

**Accepted:** January 4, 2021

**Published:** March 13, 2021

**Copyright:** © 2021 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

With the continued transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) throughout the world, identification of highly suspected COVID-19 patients remains an urgent priority. In this study, we developed and validated COVID-19 risk scores to identify patients with COVID-19. In this study, for patient-wise analysis, three signatures, including the risk score using radiomic features only, the risk score using clinical factors only, and the risk score combining radiomic features and clinical variables, show an excellent performance in differentiating COVID-19 from other viral-induced pneumonias in the validation set. For lesion-wise analysis, the risk score using three radiomic features only also achieved an excellent AUC value. In contrast, the performance of 130 radiologists based on the chest CT images alone without the clinical characteristics included was moderate as compared to the risk scores developed. The risk scores depicting the correlation of CT radiomics and clinical factors with COVID-19 could be used to accurately identify patients with COVID-19, which would have clinically translatable diagnostic and therapeutic implications from a precision medicine perspective.

## INTRODUCTION

The novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has been identified as the cause of coronavirus disease 2019 (COVID-19) in Wuhan, Hubei Province, China in late 2019 [1]. It spread rapidly, resulting in a global pandemic, with over 23,342,798 confirmed cases and 807,383 deaths globally as of August 2020 [2]. COVID-19 developed through person-to-person spread of SARS-CoV-2 via respiratory droplets is associated with adverse outcomes, and increased short- and long-term morbidity and mortality [1]. The identification of suspected patients with COVID-19 is urgently needed so that we can evaluate patients at greater risk and/or more vulnerable to COVID-19 and facilitate appropriate clinical decision making for earlier quarantine and interventions that could minimize the severity of COVID-19, thus substantially improving patient outcome.

Currently, the standard for the diagnosis of COVID-19 is the use of the reverse transcription polymerase chain reaction (RT-PCR) to detect SARS-CoV-2 in lower throat respiratory tract secretions, sputum, swabs, or blood samples [3]. However, the sensitivity of the RT-PCR varies within a range of 60–71% because its accuracy could be compromised by the quality of the RT-PCR kit, the varying lowest limit of detection (LOD) of virus RNA copies per mL with the kits of different vendors, the quality and location of specimens collected (upper vs. lower respiratory tract), the low viral load in test specimens collected, and/or sampling timing (different phases of the disease), thus easily leading to false negative results [3, 4].

Recently, the European Society for Radiotherapy and Oncology (ESTRO) and the American Society for Radiation Oncology (ASTRO) jointly issued an ESTRO-ASTRO consensus statement to recommend the use of simulation-CT in clinical practice as a COVID-19 screening tool during the SARS-CoV-2 pandemic [5]. The consensus statement suggests that the CT imaging techniques used in radiotherapy are a potential screening opportunity and may be an added value to identify asymptomatic COVID-19 patients that are not identified by standard screening in hospitals (e.g., temperature screening and questions regarding COVID-19-related symptoms) [5]. The consensus was based on the fact that studies using CT imaging have identified patients with COVID-19 with negative RT-PCR results [6, 7]. In particular, thoracic CT screening allows early diagnosis of COVID-19, when patients are still in the asymptomatic phase [5–8].

Chest radiography and CT imaging have a sensitivity of 56–98% to identify suspected patients before the occurrence of positive RT-PCR detection results as well

as use to assess disease extent and follow-up [4]. The principal CT manifestations include ground-glass opacification (GGO) with or without consolidative abnormalities and a bilateral, peripheral, and diffuse distribution with or without an involvement of the lower lobes [4, 9]. Especially, asymptomatic patients with initially negative RT-PCR results also showed early CT changes [9].

However, the identification of CT manifestations highly relies on radiologists' clinical experience due to the qualitative CT features used, which might pose a challenge to resource-limited clinics with health care disparities for COVID-19 diagnosis. Meanwhile, COVID-19 shares similar manifestations with severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and other viral pneumonias in the images of qualitative CT, thus significantly reducing the specificity of qualitative CT for COVID-19 detection [10]. As such, chest CT may be helpful in making the diagnosis, but no finding can completely confirm or exclude the possibility of COVID-19 without including other clinical characteristics due to the extremely low specificity of 25% of the chest CT alone for diagnostic purposes [4]. These aspects of qualitative CT emphasize limitations of the current imaging model for diagnosing COVID-19 before the occurrence of its clinical symptoms and have compelled radiologists to call for new, imaging-based methods to answer this critical clinical question.

Digital biopsy techniques have evolved to use high-throughput processes to extract quantifiable radiomic features from medical images and have the potential to facilitate disease characterization and assessment. The aim of this study was to develop and validate clinically translatable COVID-19 risk scores encompassing chest CT radiomics with or without clinical characteristics included for distinguishing COVID-19 from other viral pneumonia. As a reference, we also compared the prediction performance of the risk scores with that of 130 well-experienced radiologists from the epicenters of COVID-19 outbreak and non-epicenters in China as well as that of other machine learning methods in this study. The risk scores integrating the spatial information derived from chest CT radiomic features and/or clinical characteristics could better characterize the SARS-CoV-2 infection landscape, which still significantly overlaps with other virus-induced pneumonias in visual inspection of CT manifestations.

## RESULTS

### Patient characteristics

The clinical characteristics of patient data used are shown in Table 1. The COVID-19 patients had significantly

**Table 1. Clinical characteristics of the COVID-19 and non-COVID-19 (viral-induced pneumonias) patient cohorts.**

Characteristics	COVID-19 patients (n = 108)	Non-COVID-19 patients (viral-induced pneumonias) (n = 77)	P-value
Age, years			
>50	41	27	0.002
≤50	67	50	0.234
Lesion number			
1 ≤ n < 3	14	72	<0.001
3 ≤ n < 5	12	5	0.729
5 ≤ n < 10	64	0	<0.001
10 ≤ n	18	0	<0.001
Sex			
Male	44	28	0.322
Female	64	49	0.876
Epidemiologic contact			
Travel history to Hubei Province, China <sup>‡</sup>	12	—	—
Travel history to Wenzhou city, Zhejiang Province, China <sup>‡</sup>	22	—	—
Unknown exposure	74	—	—
Symptoms			
Fever	89	57	0.437
Dyspnea	51	55	0.051
Chest tightness	17	14	0.121
Cough	67	75	0.532
Sputum	23	54	0.067
Rhinorrhea	37	65	0.213
Asymptomatic	17*	2	—
Laboratory results			
D-dimers, mg/L	0.51 ± 0.44	0.52 ± 0.34	0.897
C-reactive protein, mg/L	12.32 ± 18.7	7.49 ± 14.27	0.055
White blood cells, 10 <sup>9</sup> /L	3.31 ± 2.13	3.58 ± 1.94	0.112
Creatine kinase isoenzyme, μg/L	9.12 ± 5.56	13.93 ± 5.69	<0.001
Lactate dehydrogenase, U/L	245.91 ± 75.35	167.35 ± 42.88	<0.001
CT manifestations			
Location			
Unilateral	1	75	<0.001
Bilateral	107	2	<0.001
Distribution			
Central	1	3	0.233
Peripheral	73	72	0.191
Central + peripheral	34	2	0.013
Main features			
Ground-glass opacity	67	43	0.278
Consolidation	11	9	0.055
Linear opacity	23	12	0.231
Mixed type	7	13	0.101
Interstitial change			
Septal thickening	37	25	0.062
Fine reticular opacity	11	39	0.012
Other features			
Vascular thickening	17	39	0.054
Crazy-paving pattern	45	39	0.123
Pleural thickening	13	2	0.053
Pleural effusion	0	0	—

<sup>‡</sup>Two epicenters of the COVID-19 outbreak in China.

\*These were tested as close contacts with confirmed COVID-19 patients.

higher lesion numbers, CK-MB activity, LDH activity, and bilateral, peripheral, or mixed central and peripheral pulmonary distribution than the non-COVID-19 viral-induced pneumonia patients. Most of the symptoms, laboratory results, and CT manifestations had no significant differences between COVID-19 and non-COVID-19 patients (Table 1). Representative images of COVID-19 pneumonia, adenovirus pneumonia, cytomegalovirus pneumonia, and influenza virus pneumonia are shown in Figure 1.

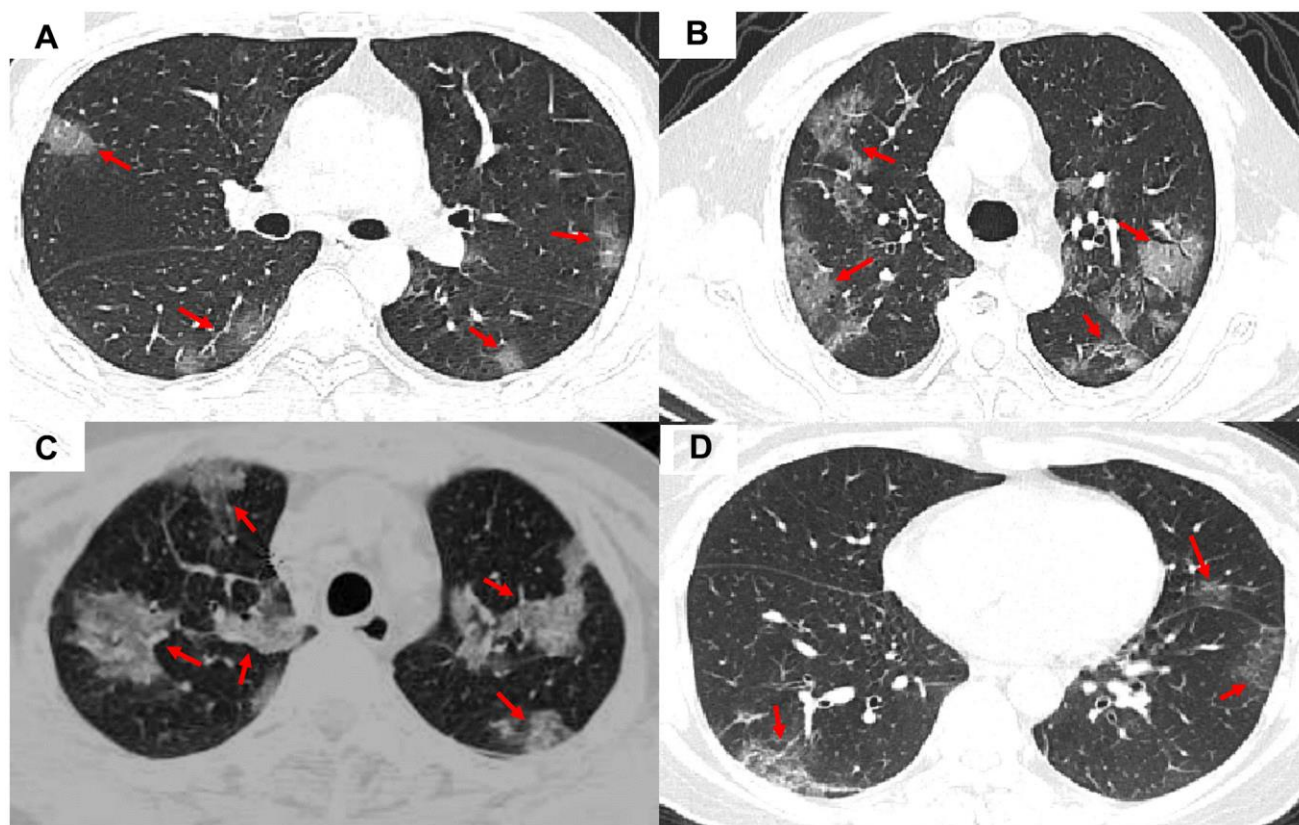
### Human diagnosis of COVID-19

Supplementary Figure 5 shows the geographic distribution of 130 radiologists from 10 provinces in China, including Hubei. The performance of the 130 radiologists based on the chest CT images only (without providing patients' clinical information and laboratory results) was moderate due to the overlap of CT manifestations between COVID-19 lesions and

non-COVID-19 viral pneumonia lesions using a supervised human learning format (Table 2). Notably, the radiologists from Hubei Province, China, the epicenter of the COVID-19 outbreak in China, had a better performance than the radiologists from outside of Hubei Province ( $P < 0.05$ ).

### Patient-based risk scores

The patient-based risk scores using radiomic features only, clinical factors only, and a combination of radiomic features and clinical factors are shown in Figure 2A–2C and Equations (2)–(4), respectively. The utility of the risk scores achieved area under the receiver operating characteristic curve (AUC) values of 0.791 (95% confidence interval [CI]: 0.651–0.932), 0.813 (95% CI: 0.682–0.944), and 0.915 (95% CI: 0.841–0.991), respectively, in the validation set (Table 3 and Figure 3), suggesting a high performance of COVID-19 classification using the COVID-19 risk scores.



**Figure 1. Representative images of COVID-19 pneumonia, adenovirus pneumonia, cytomegalovirus pneumonia, and influenza virus pneumonia. (A)** A transverse CT image from a 35-year-old man with adenovirus pneumonia showing bilateral ground-glass opacities in the upper lobes with a rounded morphology (arrows). **(B)** COVID-19: A transverse CT image from a 57-year-old man with COVID-19 showing more limited ground-glass opacities in the bilateral upper lobes with an elliptical morphology (arrows). **(C)** A transverse CT image obtained in a 45-year-old female with cytomegalovirus pneumonia showing bilateral ground-glass and burr-like, denser, and less transparent distribution (arrows). **(D)** A transverse CT image of a 61-year-old man diagnosed with influenza virus pneumonia showing bilateral ground-glass opacities in the upper lobes (arrows).



**Table 2. Performance of the radiologists to diagnose COVID-19 from chest CT images.**

Performance		Radiologists from Hubei Province, China* (n = 40)	Radiologists from outside of Hubei Province, China (n = 90)	P-value
	<i>n</i>	13	42	—
Assistant attending radiologists	Average time for reviewing each CT image (sec)	3.2 ± 2.3	3.3 ± 2.1	0.132
	Precision	0.49	0.3	0.153
	Recall	0.30	0.23	0.121
	Specificity	0.57	0.39	0.058
	F1	0.37	0.28	0.131
	Accuracy	0.41	0.30	0.154
	<i>n</i>	15	34	—
Associate attending radiologists	Average time for reviewing each CT image (sec)	3.3 ± 2.6	3.6 ± 2.8	0.053
	Precision	0.49	0.35	0.002
	Recall	0.29	0.24	0.042
	Specificity	0.57	0.39	0.051
	F1	0.37	0.28	0.032
	Accuracy	0.41	0.30	0.021
	<i>n</i>	12	14	—
Attending radiologists	Average time for reviewing each CT image (sec)	3.1 ± 2.4	3.2 ± 2.9	0.129
	Precision	0.47	0.34	0.215
	Recall	0.29	0.22	0.055
	Specificity	0.54	0.40	0.067
	F1	0.35	0.26	0.042
	Accuracy	0.39	0.29	0.102
Overall	Average time for reviewing each CT image (sec)	3.2 ± 0.1	3.4 ± 0.2	0.054
	Precision	0.48	0.35	<0.001
	Recall	0.29	0.23	0.027
	Specificity	0.56	0.39	0.002
	F1	0.36	0.28	0.034
	Accuracy	0.41	0.30	0.029

\*Hubei Province was the epicenter of the COVID-19 outbreak in China.

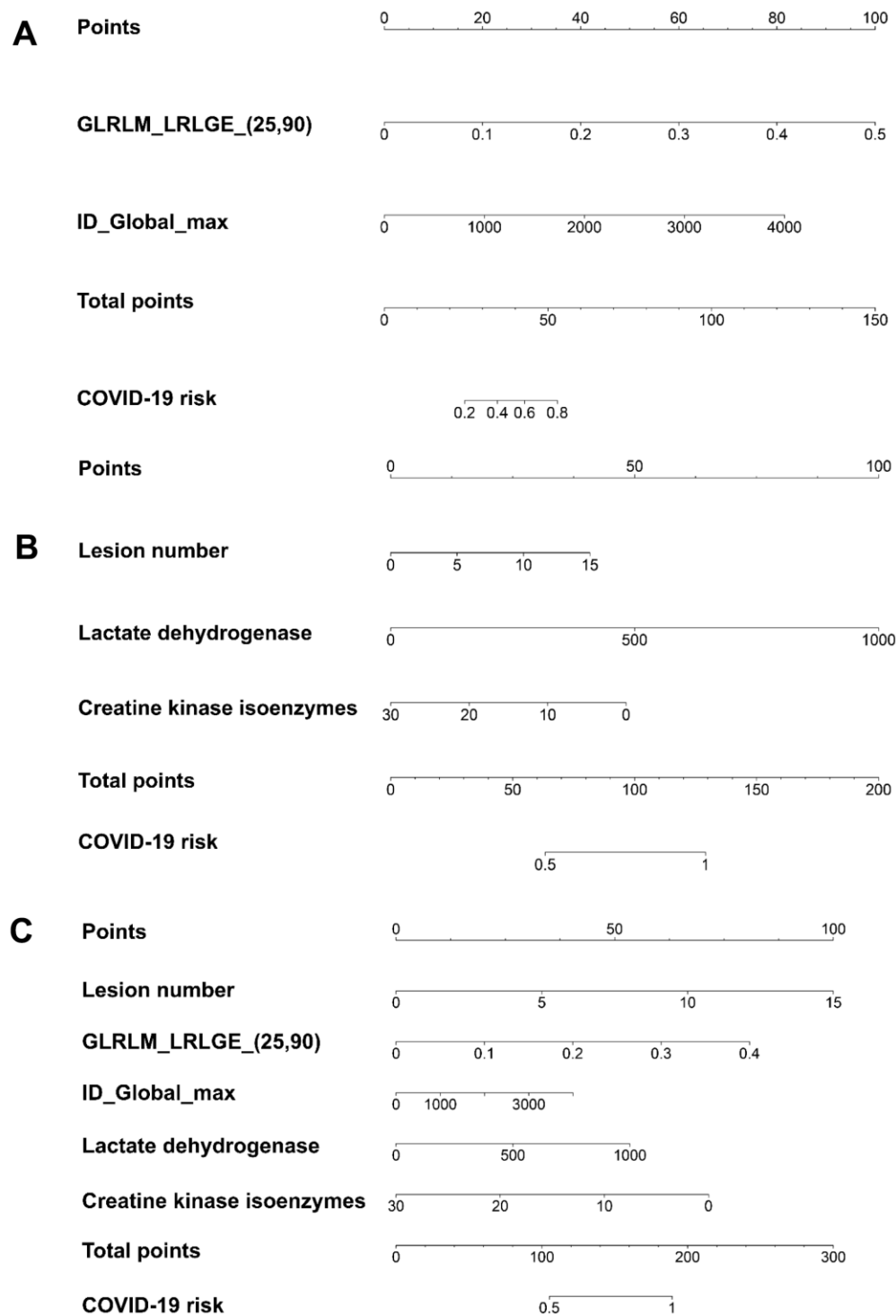
The patient – based risk score using radiomic features only  
 $= -3.785 + 19.563 \times \text{GLRLM\_LRLGE}_{(25,90)} + 0.002 \times \text{ID\_Global\_Max}$  (2)

The patient – based risk score using clinical factors only  
 $= -15.680 + 2.833 \times \text{lesion number} + 0.104 \times \text{lactate dehydrogenase} - 1.674 \times \text{creatinine kinase isoenzymes}$  (3)

The patient – based risk score combining radiomics and clinical features  
 $= -114.053 + 9.911 \times \text{lesion number} + 122.045 \times \text{GLRLM\_LRLGE}_{(25,90)} + 0.0196 \times \text{ID\_Global\_Max} + 0.334 \times \text{lactate dehydrogenase} - 7.593 \times \text{creatinine kinase isoenzymes}$  (4)

where  $\text{GLRLM\_LRLGE}_{(25, 90)}$  represents the radiomic feature, long-run, low gray-level emphasis, which describes the distribution of the long homogeneous runs with low gray-levels within the image. The numbers in the brackets represent the parameters used to calculate that particular radiomic feature.  $\text{ID\_Global\_max}$  represents the radiomic feature intensity direct global max, which describes that the binary mask was preprocessed for the features derived directly from the image intensity. A detailed description of the parameters used is shown in Figure 2.

In contrast, the developed patient-based random forest models demonstrate comparable AUC values, precision, recall, specificity, F1, and accuracy as compared to the patient-based risk scores in the validation set (Table 3). The results of the decision curve analysis (DCA) to evaluate the clinical utility of the risk scores and the



**Figure 2. The patient-based COVID-19 risk scores demonstrated by nomograms. (A)** The risk score using radiomic features only. **(B)** The risk score using clinical factors only. **(C)** The risk score combining radiomic features and clinical factors. GLRLM\_LRLGE\_(25, 90) represents the radiomic feature long run low gray-level emphasis, which describes the distribution of the long homogeneous runs with low gray-levels within the image. The numbers in the bracket represents the parameters used to calculate that particular radiomic feature. The parameters of 25 and 90 in GLRLM\_LRLGE represent the binary mask in 2.5D and 90 degrees, which describes that the GLRLM was computed in 2D slice by slice; then, the occurrence of run length from 90 degrees from all 2D image slices was summed. ID\_Global\_max represents the radiomic feature intensity direct global max, which describes that the binary mask was preprocessed for the features derived directly from the image intensity. The binary mask in ID\_Global\_max can be modified through intensity thresholding, by binary erosion, and using only the binary slice with the maximum area. The unit for lactate dehydrogenase is U/L. The unit for creatine kinase isoenzymes is  $\mu\text{g/L}$ . Supplementary Appendix 2 explains how to use the nomograms.

**Table 3. The classification performance using patient-based COVID-19 risk scores and random forest models.**

Signature	AUC	Precision	Recall	Specificity	F1 score	Accuracy	Delong test for AUC values
<b>Training set</b>							
COVID-19 risk score using radiomic features only	0.807 (0.717–0.853)	0.823 (0.703–855)	0.792 (0.711–0.834)	0.843 (0.781–0.867)	0.807 (0.723–0.897)	0.811 (0.703–0.966)	
COVID-19 risk score using clinical variables only	0.882 (0.847–0.921)	0.877 (0.712–0.903)	0.897 (0.784–0.922)	0.901 (0.879–0.944)	0.887 (0.813–0.922)	0.892 (0.824–0.927)	Z = 7.241, P = 0.000
COVID-19 risk score using combined radiomic and clinical variables	0.935 (0.913–0.978)	0.902 (0.878–0.966)	0.942 (0.807–0.989)	0.921 (0.893–0.964)	0.899 (0.812–0.962)	0.923 (0.879–0.978)	
Random forest using radiomic features only	0.837 (0.775–0.901)	0.712 (0.645–0.834)	0.896 (0.812–0.934)	0.877 (0.812–0.921)	0.793 (0.743–0.854)	0.845 (0.798–0.939)	
Random forest using clinical variables only	0.925 (0.892–0.963)	0.867 (0.772–0.934)	0.914 (0.854–0.963)	0.937 (0.879–0.987)	0.890 (0.807–0.919)	0.955 (0.913–0.977)	Z = 8.574, P = 0.000
Random forest using radiomic and clinical variables	0.958 (0.911–0.989)	0.886 (0.719–0.968)	0.934 (0.812–0.977)	0.954 (0.903–0.987)	0.909 (0.855–0.950)	0.966 (0.923–0.989)	
<b>Validation set</b>							
COVID-19 risk score using radiomic features only	0.791 (0.651–0.932)	0.804 (0.723–0.902)	0.733 (0.693–0.854)	0.822 (0.734–0.876)	0.767 (0.717–0.856)	0.797 (0.701–0.892)	
COVID-19 risk score using clinical variables only	0.813 (0.682–0.944)	0.821 (0.721–0.876)	0.934 (0.877–0.965)	0.917 (0.832–0.989)	0.874 (0.793–0.941)	0.882 (0.769–0.923)	Z = 9.307, P = 0.000
COVID-19 risk score using combined radiomic and clinical variables	0.915 (0.841–0.991)	0.855 (0.744–0.913)	0.945 (0.897–0.988)	0.934 (0.899–0.989)	0.898 (0.844–0.953)	0.919 (0.87–0.955)	
Random forest using radiomic features only	0.872 (0.771–0.973)	0.809 (0.723–0.881)	0.913 (0.856–0.956)	0.896 (0.859–0.931)	0.858 (0.739–0.907)	0.868 (0.792–0.899)	
Random forest using clinical variables only	0.949 (0.894–0.956)	0.902 (0.843–0.977)	0.965 (0.913–0.998)	0.967 (0.943–0.997)	0.932 (0.899–0.956)	0.956 (0.933–0.979)	Z = 7.896, P = 0.000
Random forest using radiomic and clinical variables	0.979 (0.949–0.997)	0.943 (0.879–0.987)	0.987 (0.897–0.999)	0.934 (0.917–0.986)	0.964 (0.889–0.981)	0.963 (0.892–0.992)	

The values in the brackets represent the 95% confidence interval. Abbreviations: AUC, area under the ROC curve.

random forest models built in this study are shown in Figure 3E, 3F. The risk scores show a comparable clinical utility as compared to the random forest models.

**Lesion-wise COVID-19 risk score with radiomic features only**

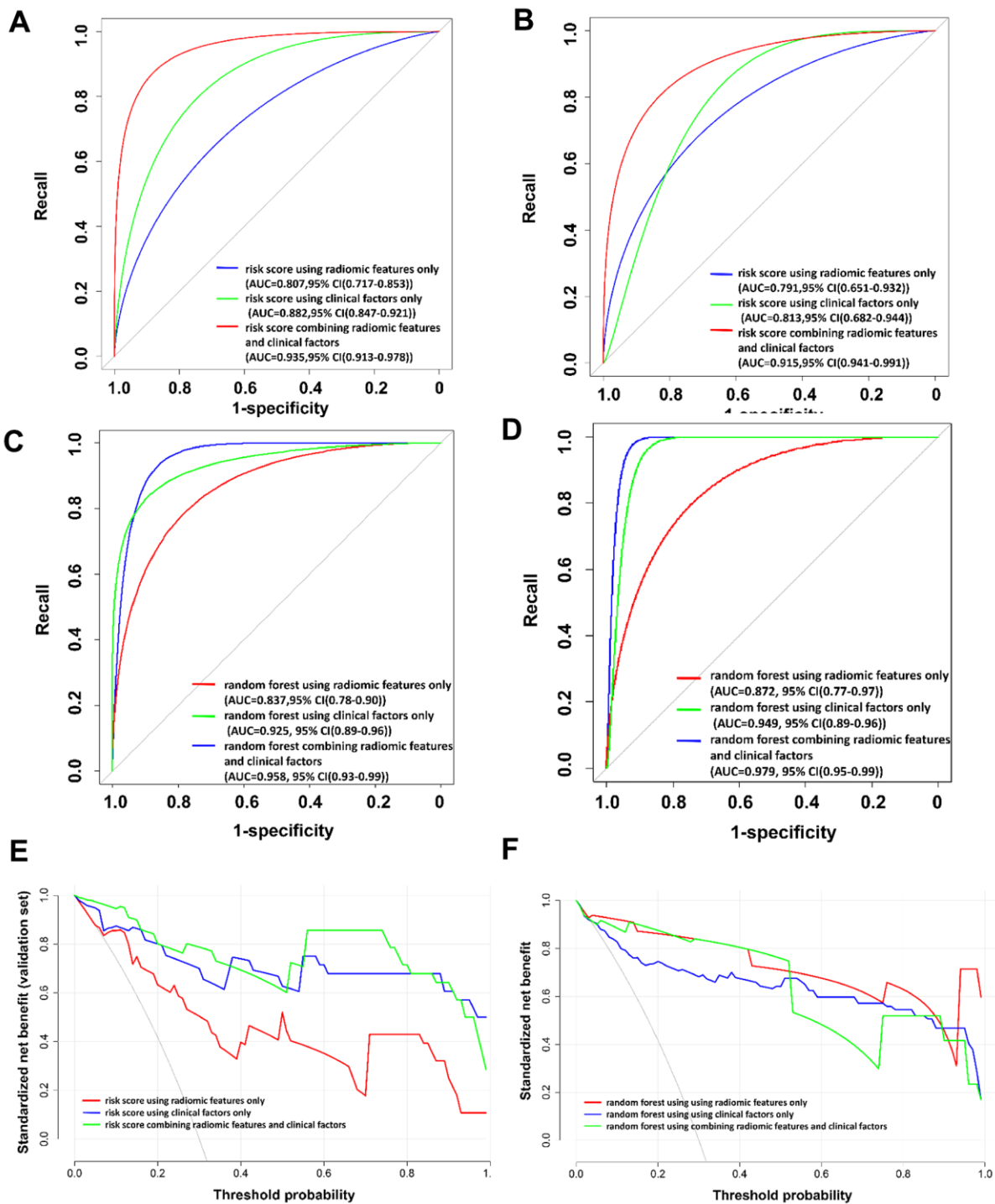
To characterize different infectious lesions within the same patient, a lesion-based risk score using three radiomic features alone was also constructed (Figure 4 and Equation (5)). The utility of the risk score achieved an AUC value of 0.931 (95% CI: 0.898–0.956) (Table 4 and Figure 5A).

The lesion  
 –based risk score using radiomics features alone  

$$= -55.389 - 6.769 \times \text{GLCM\_Correlation\_}(25,0,1) + 0.33 \times \text{ID\_Local\_Range\_Std} + 0.136 \times \text{GOH\_Percentile\_}(15)$$
 (5)

where  $\text{GLRLM\_Correlation\_}(25,0,1)$  represents the radiomic feature gray-level co-occurrence matrix with statistical measurement of correlation between a pixel and its neighbor over the whole image. The numbers in the brackets represent the parameters used to calculate that particular radiomic feature.  $\text{ID\_Local\_Range\_Std}$  represents the radiomic feature of intensity direct in the neighborhood region, which describes the standard deviation among all the voxels.  $\text{GOH\_Percentile\_}(15)$  represents the radiomic feature gradient orient histogram, which describes the percentiles of the occurrence probability values in the histogram of the image. A detailed description of the parameters used is shown in Figure 4.

In contrast, the lesion-based weighted support vector machine (WSVM) model using the radiomic features only demonstrates a comparable AUC value, precision, recall, specificity, F1, and accuracy as compared to the lesion-based risk score (Table 4). The results of the DCA analysis to evaluate the clinical utility of the risk



**Figure 3. The receiver operating characteristic (ROC) curves and the decision curve analysis (DCA) for the patient-based risk scores and random forest models.** (A) ROC curve for patient-based risk scores in the training set. (B) ROC curve for patient-based risk scores in the validation set. (C) ROC curve for patient-based random forest models in the training set. (D) ROC curve for patient-based random forest models in the validation set. (E) DCA for patient-based risk scores in the validation set. (F) DCA for patient-based random forest models in the validation set. In (E) and (F), the x-axis of the decision curve is the threshold of the predicted probability using the risk score to classify COVID-19 and non-COVID-19 patients. The y-axis shows the clinical decision net benefit for patients based on the classification result in this threshold. The decision curves of the treat-all scheme (the monotonically decreasing dash-line curve in the figure) and the treat-none scheme (the line when x equals zero) are used as references in the DCA. In this study, the treat-all scheme assumes that all the patients had COVID-19; the treat-none scheme assumes that none of the patients had COVID-19. Abbreviations: AUC, area under the ROC curve; 95% CI, 95% confidence interval.



score and the WSVM model using radiomic features only are shown in Figure 5B.

## DISCUSSION

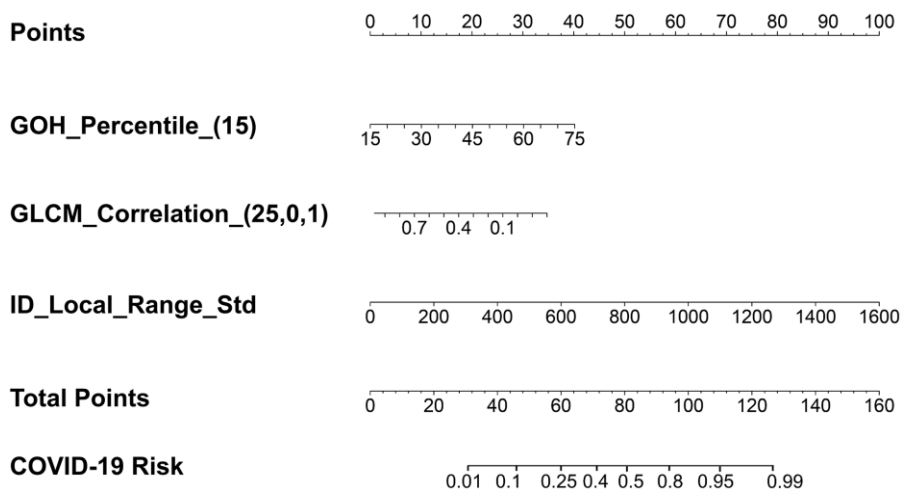
During the incubation period of SARS-CoV-2, before the onset of clinical symptoms confirmed by positive nucleic acid detection, about 96% of patients would have non-specific CT imaging changes similar to other viral pneumonias in the lungs, i.e., GGO, patchy consolidation, and sub-solidification [4, 9, 11, 12]. In this study, the performance of the radiomics-based risk scores was compared to that of human diagnosis in differentiating COVID-19 from viral pneumonia. We demonstrated that both the patient-based risk score using radiomic features only and the lesion-based risk score using radiomic features only have significantly better classification abilities than the human diagnosis at the patient- and lesion-wise levels. This can partially be attributed because without the aid of other clinical information, radiologists might achieve a relatively low sensitivity and specificity in differentiating COVID-19 from viral pneumonias based only on the chest CT manifestations.

The risk score could provide a quantitative measure to appropriately adjust the cut-off value based on desired levels of recall and specificity to reduce the adverse consequences of false negatives in the differentiation of COVID-19. In addition, with the quantitative

measurements used, it might be useful to longitudinally monitor disease progress over time or recurrence in the recovered COVID-19 patients using delta radiomics methods, although this possibility is still under investigation.

In the patient-based COVID-19 risk scores, three clinical variables, i.e., lesion number, LDH activity, and CK-MB activity, show discriminative abilities for COVID-19 detection. Notably, the imaging pattern showing a multifocal appearance with a lesion number larger than 3–5 could be used as a rapid cut-off in the case of a strong suspicion of SARS-CoV-2 infection. Meanwhile, LDH serves as an inflammatory predictor in many pulmonary diseases, such as obstructive disease, microbial pulmonary disease, and interstitial pulmonary disease [13, 14]. A recent study showed that refractory COVID-19 patients had increased blood LDH and CRP levels. Moreover, another study demonstrated that COVID-19 patients treated in the ICU had higher levels of LDH and CRP than those not treated in the ICU [15]. These observations suggest that LDH levels might reflect the acute severe systemic inflammatory response involved in cell-mediated immunity and cytokine storms caused by SARS-CoV-2 infection, which is a distinguishable biochemical parameter for inflammation in the risk score.

Furthermore, a previous study suggested that the increases in LDH and CK-MB levels were correlated



**Figure 4. The lesion-based risk score using three radiomic features only.** GOH\_Percentile\_(15) represents the radiomic feature gradient orient histogram, which describes the percentiles of the occurrence probability values in the histogram of the image. The numbers in the brackets represent the parameters used to calculate that particular radiomic feature. The parameter of 15 in GOH\_Percentile represents the histogram percentile. GLCM\_Correlation\_(25,0,1) represents the radiomic feature gray-level co-occurrence matrix with statistical measurement of correlation between a pixel and its neighbor over the whole image, which describes that the gray-level co-occurrence matrix was computed from the image inside the binary mask in 2.5D with the direction of the angle of intensity pair at 0 degrees and the distance between the intensity pairs at 1. ID-Local\_Range\_Std represents the intensity direct in the neighborhood region, which describes the standard deviation among all the voxels.

**Table 4. The diagnosis performance using the lesion-based risk score and weighted support vector machine.**

Signature	AUC	Precision	Recall	Specificity	F1 score	Accuracy	Delong test for AUC
COVID-19 risk score	0.931 (0.898–0.956)	0.976 (0.944–0.996)	0.891 (0.831–0.927)	0.921 (0.872–0.965)	0.927 (0.901–0.966)	0.902 (0.834–0.981)	Z = 4.371, P < 0.000
Weighted support vector machine	0.949 (0.925–0.971)	0.969 (0.923–0.981)	0.904 (0.824–0.936)	0.942 (0.899–0.966)	0.935 (0.876–0.964)	0.987 (0.886–0.995)	

The values in the brackets represent the 95% confidence interval (95% CI).

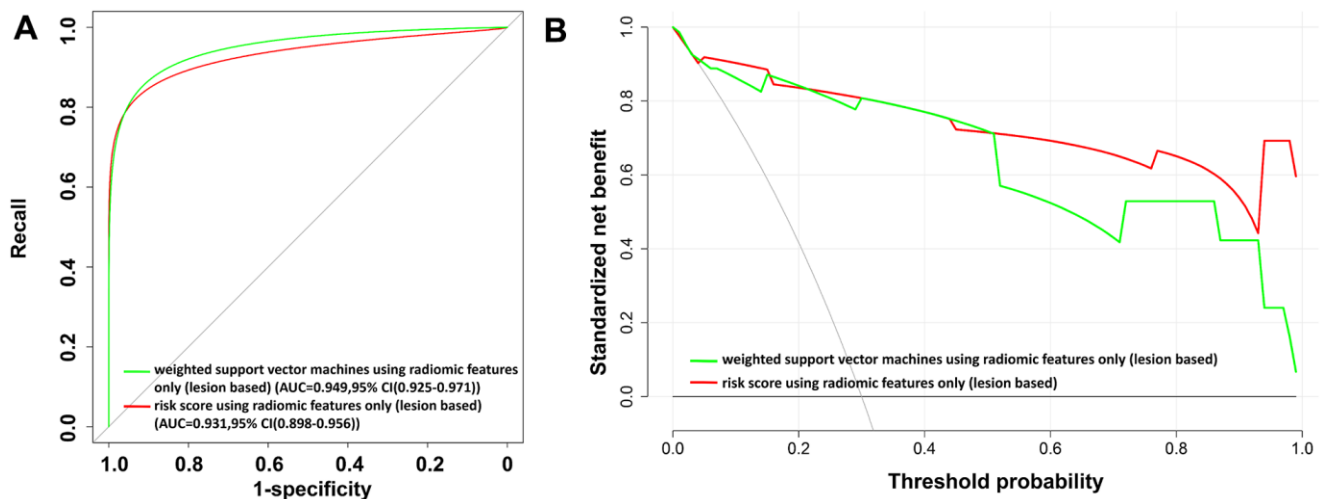
Abbreviations: AUC, area under the curve.

with SARS-CoV-2 mRNA levels in RT-PCR positive patients [16]. As such, all three clinical variables in the risk scores might emphasize the underlying biological mechanism(s) related to COVID-19. The immunological mechanism of SARS-CoV-2 infection still requires further investigation.

In the patient-based COVID-19 risk scores, two radiomic features, GLRLM\_LRLGE\_(25,90) and ID\_Global\_max, were selected to build the risk score with significantly strong discriminative abilities for COVID-19 detection (i.e., the features with  $P < 0.001$  in the multivariable logistic regression). GLRLM\_LRLGE had a higher weight (larger coefficient) in the risk score compared to other radiomics and clinical features. GLRLM\_LRLGE analyzes the spatial information within chest CT image runs in the upper right quadrant of the GLRLM with long run lengths and low gray-levels. The longer runs with different gray-level intensities are closely linked with

coarse texture and regional heterogeneity as compared to fine texture [17]. Therefore, GLRLM\_LRLGE might be associated with the coarseness of COVID-19 [18]. Consequently, a higher GLRLM\_LRLGE, i.e., a coarser texture on chest CT images, may be associated with a higher risk of occurrence of COVID-19 [19].

Interestingly, three different radiomic features, GOH\_Percentile\_(15), GLCM\_Correlation\_(25,0,1), and ID\_Local\_Range\_Std, were identified in our lesion-based analysis. In particular, previous studies suggested that a GLCM\_correlation value might be inversely related with the levels of vascular endothelial growth factor (VEGF), which controls critical physiological functions in the lung [20–22]. For example, a decrease in VEGF expression is believed to be associated with acute lung injury and alkaloid monocrotaline-pulmonary hypertension [23], which is one of the most common comorbidities in COVID-19 [24]. However, the relationship between



**Figure 5. The receiver operating characteristic (ROC) curves and the decision curve analysis (DCA) for the lesion-based risk score and weighted support vector machine model using radiomic features alone. (A) ROC curve. (B) DCA analysis. In (B), the x-axis of the decision curve is the threshold of the predicted probability using the risk score to classify COVID-19 and non-COVID-19 patients. The y-axis shows the clinical decision net benefit for patients based on the classification result in this threshold. The decision curves of the treat-all scheme (the monotonically decreasing dash-line curve in the figure) and the treat-none scheme (the line when x equals zero) are used as references in the DCA. In this study, the treat-all scheme assumes that all patients had COVID-19; the treat-none scheme assumes that none of the patients had COVID-19. Abbreviations: AUC, area under the curve; 95% CI, 95% confidence interval.**

radiomic features and the phenotypes linked to COVID-19 is not well understood at present.

There are different processing methods for patient-based analysis. Some studies selected the largest lesion and/or the most metabolically active lesion as a representative lesion for that patient based on a method reported previously [25, 26]. However, the large heterogeneous lesions (often necrotic and/or with multiple uptake peaks) may underestimate image texture measurements [27]. Nevertheless, all the lesions could be used for radiomics analysis, which enriches the analysis through the use of the information derived from all lesions. However, the method of averaging the radiomic values of all lesions as the characteristic value for one particular radiomic feature could dilute the feature value of large lesions by other small lesions. As such, in this study, a weighted power mean method was used in the patient-based analysis to emphasize that the lesions with relatively large volume represent the main characteristics of the biological behavior and characteristics of the disease type, while still retaining the other small lesions representing a certain kind of disease progression. In contrast, the lesion-based analysis allowed us to examine each individual lesion with a consideration of different infectious lesions within the same patient.

There could be certain bias introduced in the boundary and volume contoured in the manual delineating process by different radiologists, which could certainly affect the radiomics values calculated. However, this kind of inter-observer variation mainly influences the shape-related radiomic features. It has relatively limited influence on the features of GLRLM\_LRLGE\_(25, 90), ID\_Global\_Max, GOH\_Percentile\_(15), GLCM\_Correlation\_(25,0,1), and ID\_Local\_range\_std selected in this study. A previous study conducted by eight research centers in the United States and one medical imaging center in Canada suggested that the segmentation mainly affects the global shape descriptors features, but has relatively little effect on the texture and intensity features of the entire three-dimensional volume [28]. Also, GLRLM\_LRLGE\_(25, 90), ID\_Global\_Max GOH\_Percentile\_(15), GLCM\_Correlation\_(25,0,1), and ID\_Local\_range\_std are five important features in the texture and intensity features category. A verification study is described in Supplementary Appendix 3.

As a retrospective study, this study has several limitations. First, the patient cohort in this study is relatively small. The use of digital biopsy technologies with promising retrospective radiomics analyses must still be further evaluated in prospective clinical trials, thus facilitating a better personalized patient management. Second, all patients are from Zhejiang

Province, China, and might not fully represent the spectrum of COVID-19 phenotypes. The relationships between radiomic features and their underlying immune interactions and biological mechanism(s) directing COVID-19 progression at the early stage of SARS-CoV-2 infection also remain to be explored. Third, although the radiomic features and clinical variables associated with disease progression were not evaluated in this study, the findings of this research may still provide useful insights for future studies to identify the underlying mechanism(s) and relevant radiomic features for disease severity, prognosis, and patient outcome of SARS-CoV-2 infection.

## CONCLUSIONS

The point-of-care COVID-19 risk scores could be an easy-to-use tool to quantitatively differentiate COVID-19 from other viral pneumonias. The risk scores using chest CT radiomic features and/or clinical characteristics could better characterize the SARS-CoV-2 infection landscape, which still significantly overlaps with other virus-induced pneumonias in visual inspection of CT manifestations. The risk scores developed could potentially afford a clinically translatable means to improve the diagnostic confidence using chest CT for COVID-19 detection in the future.

## MATERIALS AND METHODS

### Patients

This study was approved by the Hangzhou Xixi Hospital Institutional Review Board. As this is a retrospective study, the need for written informed consent from patients was waived. A total of 193 patients confirmed with COVID-19 or other types of viral pneumonia were enrolled in this study. Eight patients with negative chest CT imaging result were excluded. A total of 108 patients with COVID-19 confirmed by RT-PCR between December 2019 and March 2020 in the Hangzhou Xixi Hospital were retrospectively included into this study. Another group of 77 patients with influenza virus-induced, adenovirus-induced, syncytial virus-induced, and cytomegalovirus-induced pneumonias from Hangzhou First People's Hospital (19 cases) and Hangzhou Xixi Hospital (58 cases) were used as controls.

The patients' electronic medical data were retrieved from the Hospital Information System (HIS). The high-resolution CT images were retrieved from the picture archiving and communication system of the hospitals. The patients' RT-PCR results were retrieved from the electronic medical records in the HIS. The patients with negative chest CT results or

lacking both chest CT and RT-PCR examinations were excluded from this study. Figure 6 summarizes the study workflow and methods.

Baseline clinical data, including patient's age, gender, lesion number, and five biochemical indicators recommended in the "Handbook of COVID-19 Prevention and Treatment," [29] including white blood cell count (WBC), C-reactive protein (CRP) levels, creatine kinase isoenzyme (CK-MB) activity, lactate dehydrogenase (LDH) activities, and plasma D-dimer (DD) levels, were collected by reviewing the medical records and data of serial CT imaging, including baseline, mid-treatment, and post-treatment CT scans, were also recorded to monitor the disease progression. The patients' daily basic status, daily examination results, and complications were also analyzed to check how the disease progressed. Based on the RT-PCR results for COVID-19 confirmation, the enrolled patients were divided into two groups, i.e., the COVID-19 group and the non-COVID-19 viral pneumonia group.

### **CT image acquisition**

All patients included underwent chest CT imaging using a two multi-detector row CT system (GE Revolution Evo CT, Chicago, USA; Siemens SOMATOM Emotion 16, Erlangen, Germany). The acquisition parameters were as follows: 120/130 kV, 100/10–240 mA, 0.35- or 0.8-second rotation time, a layer spacing of 5 mm, an acquisition layer thickness of 5 mm, high-resolution reconstruction with a lung window layer thickness of 1.25/1.50 mm, a detector collimation of 16×0.5 mm or 64×0.625 mm, a field of view of 350×350 mm, and an image matrix of 512×512. The CT scans before onset of symptoms or CT scans done ≤1 week after symptom onset were used as baseline. The baseline CTs of highly suspected patients were used in this study. GGO and/or consolidation are the main manifestations in the CT images at this early stage. The other CT imaging patterns included linear opacity, mixed type and interstitial change patterns including septal thickening and fine reticular opacity, and other features including vascular thickening, crazy paving pattern, pleural thickening, and pleural effusion.

### **Human diagnosis of COVID-19 using a human supervised learning fashion**

To compare the classification performance between the COVID-19 risk scores developed and radiologists, 147 radiologists were invited to differentiate COVID-19 from the virus-induced pneumonias based on the CT manifestations only. The diagnosis was performed using a human supervised learning fashion.

A total of five COVID-19 CT images and five influenza virus-induced, adenovirus-induced, syncytial virus-induced, and cytomegalovirus-induced pneumonia images were randomly drawn from the total patient cohort to form a learning sample set. These learning sample images along with the CT manifestations described in the China Clinical Consensus on Radiological Diagnosis on COVID-19 [29] were used to train 130 radiologists in China with thoracic CT diagnosis experiences ranging from assistant attending radiologists or associate attending radiologists to attending radiologists using a human supervised learning fashion. The radiologists were then given the remaining 176 CT images without the clinical and follow-up information provided. Based on the CT manifestations they learned from the 10 image samples provided, the radiologists diagnosed whether these 176 CT images were COVID-19 or influenza virus-/adenovirus-/syncytial virus-/cytomegalovirus-induced pneumonias. The accuracy and the average time for diagnosis per CT image were used for statistical analysis. To rule out the random diagnosis, two equal CT images were mixed within the 176 CT images, and if the radiologist's answers were not consistent for these two images, his or her answers were excluded from the statistical analysis. A total of 130 radiologists' diagnoses were eligible for statistical analysis (40 radiologists from Hubei Province, the epicenter of the COVID-19 outbreak in China, and 90 radiologists from outside of Hubei Province).

### **Radiomic feature extraction**

Before extracting all chest CT radiomic features, 3D adaptive histogram equalization enhancement (AHEE-3D) and edge preserve smooth 3D (EPS-3D) methods were used to remove random noise in the images. The lesions of pneumonias on CT images were reviewed and manually delineated by two experienced attending radiologists who were blind to the clinical and follow-up information. The final contour for each lesion was agreed upon by both radiologists. The patient-based and lesion-based analyses were performed. The lesion region of interest (ROI) was segmented on the CT image as the only input for radiomics analysis of pneumonia.

A total of 1766 radiomic features were extracted from each ROI delineated using the image biomarker explorer (IBEX) public platform developed by the University of Texas MD Anderson Cancer Center for feature extraction and classification of radiomic features [30, 31]. The radiomic features extracted include seven categories: shape, intensity direct, intensity histogram, gray-level co-occurrence matrix (2.5D and 3D), neighbor intensity difference (2.5D and 3D), gray-level run length matrix (2.5D), and intensity histogram Gauss fit.



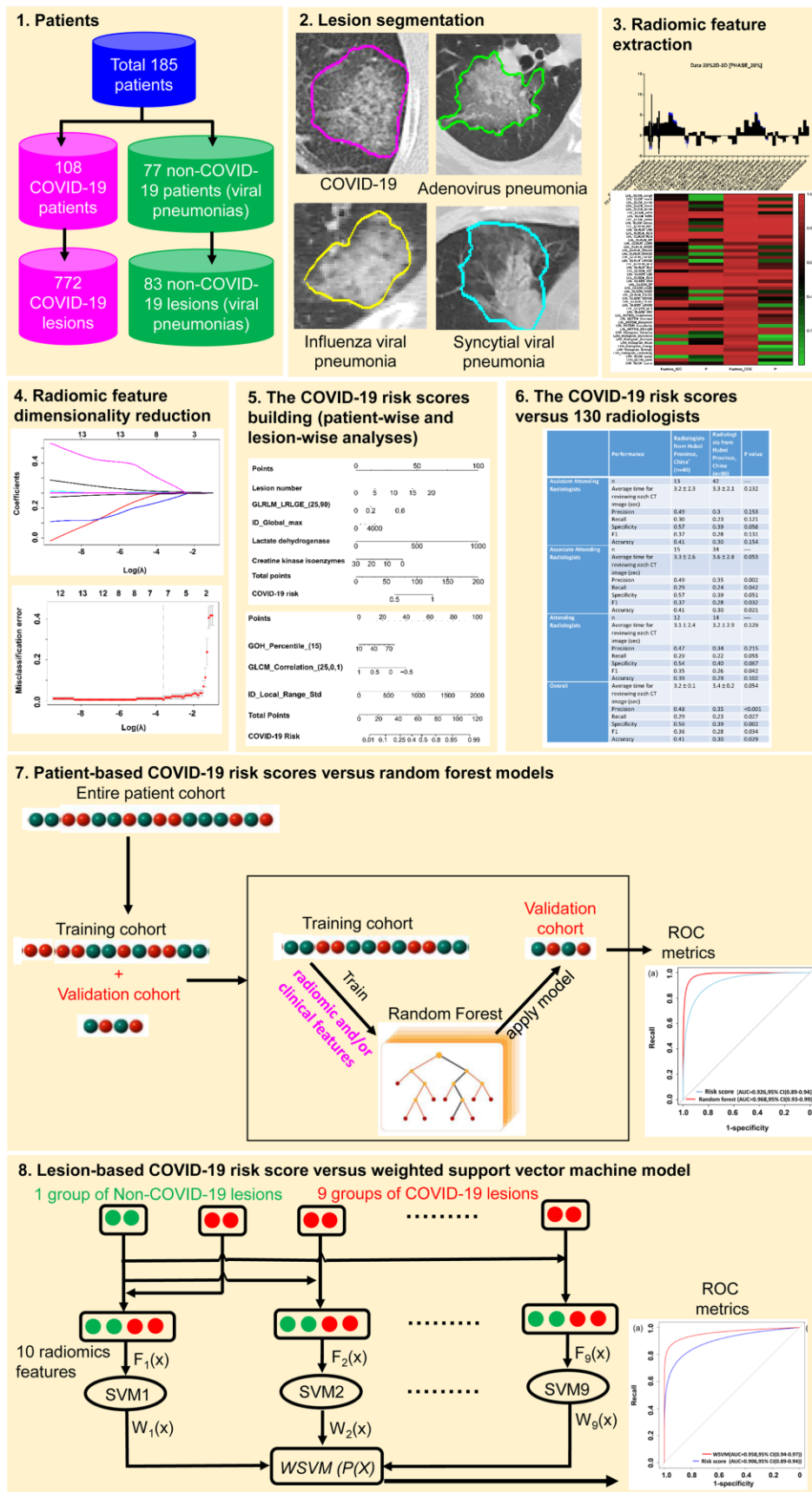


Figure 6. The workflow for the development and validation of COVID-19 risk scores.



It is believed that some of the radiomic features are sensitive to each step of the data processing procedure, including image acquisition settings, image reconstruction algorithm, and the digital image preprocessing procedure, so that the repeatability and reproducibility of the extraction of these radiomic features are easily compromised [32]. To account for the potential impact of the accuracy of radiomic feature extraction, the radiomic feature extraction procedure was repeated twice and Lin's concordance correlation coefficient (CCC) tests were performed to assess the feature reproducibility in repeated feature extraction [33, 34]. Only the 1237 radiomic features showing high CCC values (CCC > 0.99) were used. With 1237 radiomic features selected, 510 features with null value were eliminated and the remaining 727 radiomic features plus 9 clinical variables (lesion number, age, gender, WBC, LC, CRP level, LDH activity, CK-MB activity, and plasma DD level) were used for further analysis.

### Patient-based risk scores

Of the patient data, 25% was randomly selected as an independent validation set ( $n = 46$ ) and the remaining 75% of the patient data were used for the training set ( $n = 139$ ). The ratio of COVID-19 to non-COVID-19 patients was about 1.41:1 in the training and validation sets (in the training set, COVID-19:non-COVID-19 = 81:58 patients; in the validation set, COVID-19:non-COVID-19 = 27:19 patients).

For our patient-based analysis, the same features extracted from multiple lesions within one single patient were combined using a weighted power mean method [35]. Briefly, all lesions of the patient were delineated and the radiomic features were extracted. A weighting calculation was performed to combine the same feature from different lesions within the same patient as described in the following equations (Equation 6):

$$F(j) = \sum_{i=1}^n \frac{V_i}{V_T} \cdot f_j(i) \quad (6)$$

$$V_T = \sum_{i=1}^n V_i,$$

where  $F(j)$  represents the value of the  $j$ -th radiomic feature of the patient,  $i$  represents the  $i$ -th lesion of the patient,  $V(i)$  represents the volume of the  $i$ -th lesion,  $n$  represents the number of lung lesions in the patient,  $V_T$  represents the total volume of all lesions in the patient, and  $f_j(i)$  represents the value of feature  $j$  in the  $i$ -th

lesion. The weight assigned was based on the volume of the lesion. The larger the lesion volume, the greater the weight value of the features extracted from that lesion. Thus the contribution of the features extracted from this lesion to the patient's radiomic feature was also greater.

Owing to the imbalanced sample distribution between COVID-19 and non-COVID-19 patients (the number of non-COVID-19 patient is lower than the number of COVID-19 patients), synthetic minority over-sampling technology [36–40] was used to generate synthetic non-COVID-19 patient samples in the training set so that a synthetically class-balanced training set could be achieved prior to training the models in this study. Briefly, for each minority sample “a” in the non-COVID-19 patient group, the synthesis strategy was applied to randomly select a minority sample “b” from its nearest neighbors. And then one point was randomly selected as the newly synthesized non-COVID-19 patient sample on the line between “a” and “b,” so that the ratio of COVID-19 and non-COVID-19 patients in the training set was close to 1:1.

Three signatures, including a risk score using radiomic features only, a risk score using clinical factors only, and a risk score combining radiomic features and clinical variables, were built in this study (Figure 7). For the construction of the risk score using radiomic features only and the risk score combining radiomic features and clinical variables, principal component analysis (PCA), the Mann–Whitney U test, and least absolute shrinkage and selection operator (LASSO) regression with a four fold cross-validation method and a 100 times iterative selection process were successively applied to eliminate redundant features and irrelevant variables to establish the COVID-19 risk scores. A multivariate logistic regression method was used to build these two risk scores. For the construction of the risk score using clinical factors only, because the number of clinical factors is much lower than the number of radiomic features, a multivariate logistic regression method was directly applied to build the clinical signature. The model performance of the three risk scores was evaluated in the training and independent validation sets.

For the construction of the risk score using radiomic features only and the risk score combining radiomic features and clinical factors, PCA was used to reduce the feature dimensionality and select the radiomic features and radiomic features plus clinical variables that accounted for 90% of the significant feature subset variability to increase the discriminative ability. After PCA analysis, the feature dimensionality was reduced from 727 radiomic features to 32 features for the risk score using radiomic features only (Supplementary Table 1), and from 727 radiomic features plus 9 clinical

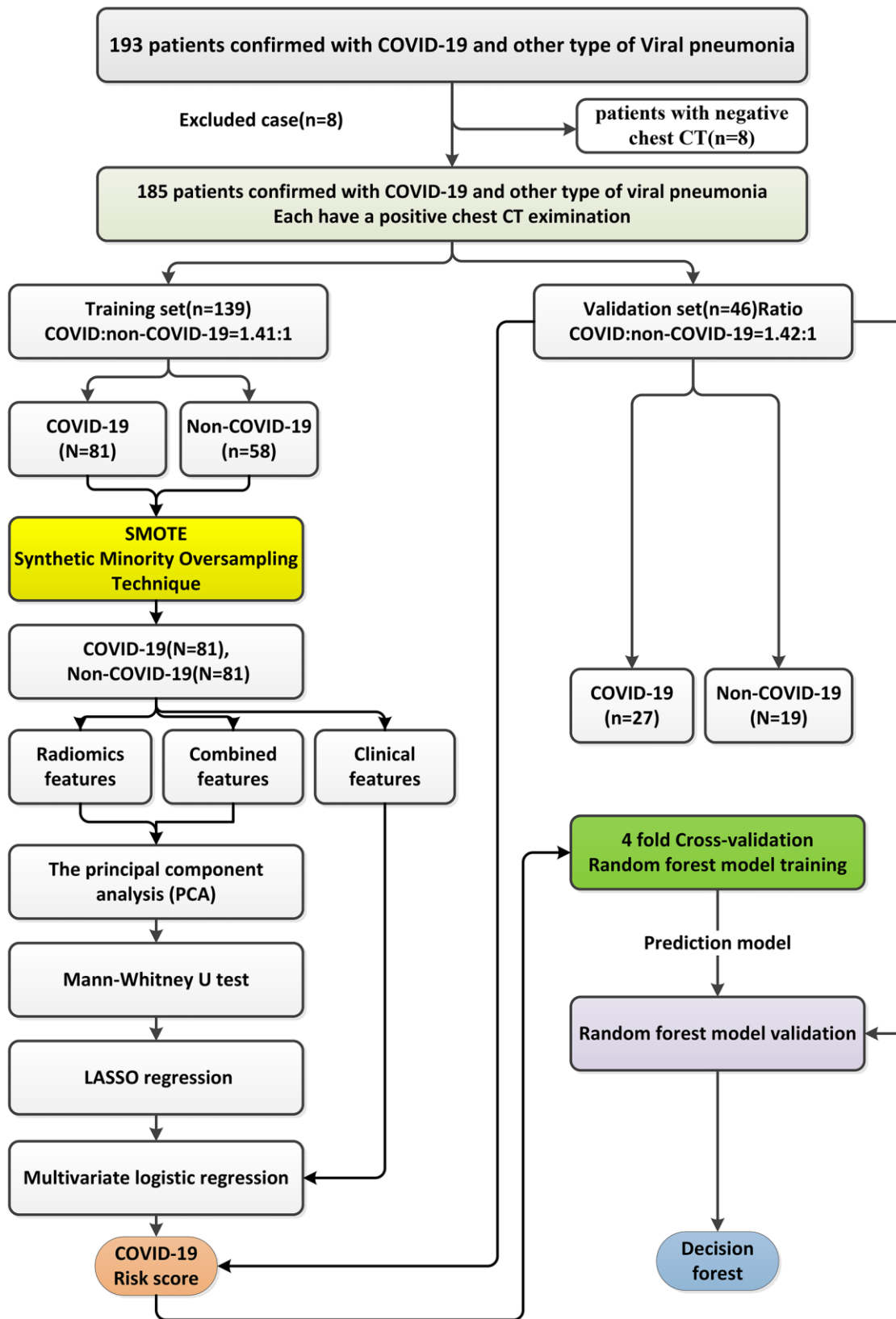


Figure 7. The workflow of the construction of the patient-based risk scores using radiomic features only, the risk score using clinical factors only, and the risk score combining radiomic features and clinical variables using a multivariate logistic regression method and a random forest model.

variables to 26 features for the risk score combining radiomic features and clinical factors (20 radiomic features plus 6 clinical variables including lesion number, gender, WBC, CRP level, LDH activity, CK-MB activity, and plasma DD level) (Supplementary Table 2).

The Mann–Whitney U test was used to further explore the potential association of the features/variables selected from PCA with COVID-19 and further reduce the feature dimensions. For the risk score using radiomic features only (Supplementary Table 3), the feature dimensions were reduced from 32 to 20 radiomic features. For the risk score combining radiomic features and clinical factors, the feature dimensions were reduced from 26 to 17 features (11 radiomic features plus 6 clinical variables including lesion number, gender, WBC, CRP level, LDH activity, and CK-MB activity) (Supplementary Table 4).

To select the most suitable features for classification of COVID-19, LASSO regression with a four fold cross-validation method and a 100 times iterative selection process was used to continually choose non-redundant and the most robust radiomic features and radiomic features and clinical variables, respectively [41, 42]. The coefficient of each variable was controlled by the parameter  $\lambda$  in the LASSO method and only the features with non-zero coefficients were selected. The misclassification error was calculated to minimize the binary classification error and maintain a balance of optimal classification performance and the optimal number of radiomic features needed for binary classification (COVID-19 vs. non-COVID-19).

As such, for the risk score using radiomic features only, only those 16 features with non-zero coefficients were selected via the LASSO process (Supplementary Table 5 and Supplementary Figure 1). For the risk score combining radiomic features and clinical variables, only those five features with non-zero coefficients were selected via the LASSO process (two radiomic features, GLRLM\_LRLGE\_(25,90) and ID\_Global\_Max, plus three clinical variables, lesion number, LDH activity, and CK-MB activity) (Supplementary Table 6 and Supplementary Figure 2).

After the feature dimensionality was reduced, a multivariable logistic regression analysis was employed and only the features with  $P < 0.001$  in this process were selected to build the COVID-19 risk scores. For the risk score using radiomic features only, two radiomic features, GLRLM\_LRLGE\_(25,90) and ID\_Global\_Max, were finally preserved (Supplementary Table 7). For the risk score combining radiomic features and clinical variables, seven features were further reduced to

five features (two radiomic features, GLRLM\_LRLGE\_(25,90) and ID\_Global\_Max, plus three clinical variables, lesion number, LDH activity, and CK-MB activity) (Supplementary Table 8). The COVID-19 risk scores using radiomic features only and using radiomics and clinical variables were built as the final classifiers by summing these features multiplied with their respective coefficients.

The COVID-19 risk scores developed were also represented by nomograms. The threshold of using the risk score using radiomic features only to classify COVID-19 is 0.2. The threshold of using the risk score with combined radiomic and clinical variables to classify COVID-19 is 0.5. DCA was applied to evaluate the clinical decision utility of the COVID-19 risk scores developed [34, 35]. The definition of net benefit in the DCA is described in Supplementary Appendix 1.

For the risk score using clinical factors only, the multivariable logistic regression analysis was directly employed and only the variables with  $P < 0.001$  in this process were selected to build the COVID-19 risk score, including lesion number, LDH activity, and CK-MB activity (Supplementary Table 9). The COVID-19 risk score using clinical factors only was also represented by a nomogram. The threshold using a nomogram to classify COVID-19 is 0.5.

### Patient-based random forest models

As a comparison to the patient-based COVID-19 risk scores developed, three random forest classifiers using radiomic features only, clinical factors only, and a combination of radiomic features and clinical factors were also constructed using grid search with fourfold cross-validation with the following parameters: the number of trees in the forest (ntree) = 500 and the maximum depth of the tree (mtry) = 3.

### Lesion-wise COVID-19 risk score with radiomic features only

A lesion-based COVID-19 risk score using radiomic features alone was also built so that potentially different infectious lesions could be characterized individually. In total, 772 COVID-19 lesions were extracted from COVID-19 patients and 83 non-COVID-19 lesions were extracted from related viral pneumonia patients in this study.

The feature dimensionality reduction was conducted to select the optimal radiomic features. Briefly, a total of 1766 radiomic features were extracted from each lesion individually. After eliminating the radiomic features with null values and employing PCA, 32 radiomic

features were used for the Mann–Whitney U test (Supplementary Table 10). The Mann–Whitney U test further reduced the feature dimensions from 32 to 20 features (Supplementary Table 11). The LASSO regression further selected the 10 non-redundant and most robust radiomic features (Supplementary Table 12 and Supplementary Figure 3).

After the feature dimensionality was reduced, the multivariable logistic regression analysis was employed to choose the radiomic features with  $P < 0.001$  to build the lesion-based COVID-19 risk score with radiomic features alone so that 10 features were further reduced to 3 features: GLCM\_Correlation\_(25,0,1), ID\_Local\_Range\_Std, and GOH\_Percentile\_(15) (Supplementary Table 13). The lesion-based COVID-19 risk score based on three features only was also represented by a nomogram. The threshold of using the risk score to classify COVID-19 is 0.5. DCA was also employed to evaluate the clinical decision utility of the nomogram developed [43, 44].

### **Lesion-based weighted support vector machine analysis**

As a comparison to the lesion-based COVID-19 risk score using radiomic features alone, a lesion-based WSVM analysis was also conducted using the 10 radiomic features (Supplementary Table 12) selected by the LASSO. The data distribution between the COVID-19 and non-COVID-19 lesions (approximately 9.3:1) was extremely imbalanced. To ensure the models with predictive power were equally balanced between COVID-19 and non-COVID-19, a previously described strategy was used to adjust the distribution imbalance between COVID-19 and non-COVID-19 lesions and construct the WSVM [45].

Briefly, the strategy is to separate the major class (i.e., the COVID-19 lesion group in this study) into small subset groups size-comparable to the minor class (i.e., the non-COVID viral pneumonia lesion group in this study) to achieve a balanced distribution between the major class and the minor class; the COVID-19 lesion groups was randomly decomposed into nine partitions, and all the non-COVID-19 lesions were combined with each partition of COVID-19 lesions to form an individual subset so that the ratio of the COVID-19 and non-COVID-19 lesions was nearly 1:1 in each individual subset. In each individual subset, the total lesions were randomly separated into the training set (70%) and the validation set (30%).

The support vector machine (SVM) was trained independently with 10 radiomic features selected by the LASSO process within the training set of each subset.

The weight for the SVM was determined via the recall value of the prediction using the validation set to reduce the false negative rate. In each individual subset, the SVM generated was validated with the validation set of each subset (i.e., the balanced data of each subset) as well as the validation set of the entire data to evaluate the classification performances (Supplementary Table 14 and Supplementary Figure 4).

Finally, all constituent SVMs were combined by summing constituent SVMs multiplied by weights determined, divided by the sum of the weights. The classical (metric) multidimensional scaling matrix (CMDScale) was used to demonstrate the correlation of features and COVID-19 for each constituent SVM.

### **Performance evaluation and statistical analysis**

The AUC between the risk scores and the random forest models and the WSVM model was compared using the Delong test. Six metrics, including precision, recall (sensitivity), specificity, F1, accuracy, and AUC, were calculated from the receiver operating characteristic (ROC) curve with the model output.

The classification performances of COVID-19 by the developed risk scores were assessed by ROC analysis. For numeric variables, mean and standard deviation were calculated and the differences between COVID-19 and non-COVID-19 patient groups were compared using rank-sum tests. A two-tailed  $P$  value less than 0.05 was regarded as statistically significant.

### **Data sharing**

The datasets analyzed in this study will be available from the corresponding author (Xiadong Li, email: lixiadong2019@outlook.com) at the time of publication. Per institutional policy, the datasets are designated limited access. Upon receiving access, the investigator may only use them for the purposes outlined in the request to the data provider, and redistribution of the data is prohibited.

### **AUTHOR CONTRIBUTIONS**

XL, SM, and YK designed the study. ZC, JL, SZ, PZ, XY, YR, JW, LQ, LZ, YL, BW, YH, KZ, RT, YL, KN, ZD, BY, and QD collected and interpreted data. XL and YK processed the data and performed the programming. JL and SZ did the statistical analyses. ZC, YK, XL, SM, and ZC interpreted and analyzed data. All authors vouch for the veracity of the data, analyses, and interpretations. YK and XL wrote the manuscript, and all authors reviewed, contributed to, and approved the manuscript.

## ACKNOWLEDGMENTS

We thank our colleagues in the Radiotherapy, Radiology, Finance, and Science and Education departments of the Hangzhou Cancer Hospital for their helpful assistance during this study. This work was supported by the key project of Hangzhou Health Science and Technology Plan in 2020 (Grant No. ZD20200112). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## CONFLICTS OF INTEREST

We declare no conflicts of interest.

## FUNDING

This work was supported by Science and Technology Development Project of Hangzhou (grant. 20180417A01).

## REFERENCES

1. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020; 395:1054–62. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) PMID:[32171076](https://pubmed.ncbi.nlm.nih.gov/32171076/)
2. Johns Hopkins University of Medicine. Coronavirus Resource Center. <https://coronavirus.jhu.edu/>
3. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C, Hoelscher M, Bleicker T, Brünink S, et al. Virological assessment of hospitalized patients with COVID-2019. *Nature*. 2020; 581:465–69. <https://doi.org/10.1038/s41586-020-2196-x> PMID:[32235945](https://pubmed.ncbi.nlm.nih.gov/32235945/)
4. Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, Ji W. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020; 296:E115–17. <https://doi.org/10.1148/radiol.2020200432> PMID:[32073353](https://pubmed.ncbi.nlm.nih.gov/32073353/)
5. Nevens D, Billiet C, Weytjens R, Joye I, Machiels M, Vermylen A, Chiari I, Bauwens W, Vermeulen P, Dirix L, Hugot P, Verellen D, Dirix P, Meijnders P. The use of simulation-CT's as a coronavirus disease 2019 screening tool during the severe acute respiratory syndrome coronavirus 2 pandemic. *Radiother Oncol*. 2020; 151:17–19. <https://doi.org/10.1016/j.radonc.2020.07.006> PMID:[32679307](https://pubmed.ncbi.nlm.nih.gov/32679307/)
6. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020; 296:E32–40. <https://doi.org/10.1148/radiol.2020200642> PMID:[32101510](https://pubmed.ncbi.nlm.nih.gov/32101510/)
7. Boldrini L, Dinapoli N, Valentini V. Radiotherapy imaging: an unexpected ally in fighting COVID 19 pandemic. *Radiother Oncol*. 2020; 148:223–24. <https://doi.org/10.1016/j.radonc.2020.04.036> PMID:[32342876](https://pubmed.ncbi.nlm.nih.gov/32342876/)
8. Vitullo A, De Santis MC, Marchianò A, Valdagni R, Lozza L. The simulation-CT: radiotherapy's useful tool in the race against COVID-19 pandemic. A serendipity approach. *Radiother Oncol*. 2020; 147:151–52. <https://doi.org/10.1016/j.radonc.2020.05.028> PMID:[32447034](https://pubmed.ncbi.nlm.nih.gov/32447034/)
9. Shi H, Han X, Jiang N, Cao Y, Alwalid O, Gu J, Fan Y, Zheng C. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis*. 2020; 20:425–34. [https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4) PMID:[32105637](https://pubmed.ncbi.nlm.nih.gov/32105637/)
10. Hosseiny M, Kooraki S, Gholamrezanezhad A, Reddy S, Myers L. Radiology perspective of coronavirus disease 2019 (COVID-19): lessons from severe acute respiratory syndrome and middle east respiratory syndrome. *AJR Am J Roentgenol*. 2020; 214:1078–82. <https://doi.org/10.2214/AJR.20.22969> PMID:[32108495](https://pubmed.ncbi.nlm.nih.gov/32108495/)
11. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, Diao K, Lin B, Zhu X, Li K, Li S, Shan H, Jacobi A, Chung M. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology*. 2020; 295:200463. <https://doi.org/10.1148/radiol.2020200463> PMID:[32077789](https://pubmed.ncbi.nlm.nih.gov/32077789/)
12. Halit Nahit S. Debate of Chest CT and RT-PCR Test for the Diagnosis of COVID-19. *Radiology*. 2020; 297:E341–E342. <https://doi.org/10.1148/radiol.2020203627> PMID:[7543097](https://pubmed.ncbi.nlm.nih.gov/7543097/)
13. Mo P, Xing Y, Xiao Y, Deng L, Zhao Q, Wang H, Xiong Y, Cheng Z, Gao S, Liang K, Luo M, Chen T, Song S, et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin Infect Dis*. 2020. [Epub ahead of print]. <https://doi.org/10.1093/cid/ciaa270> PMID:[32173725](https://pubmed.ncbi.nlm.nih.gov/32173725/)
14. Zhou Y, Zhang Z, Tian J, Xiong S. Risk factors associated with disease progression in a cohort of patients infected with the 2019 novel coronavirus. *Ann Palliat Med*. 2020; 9:428–36. <https://doi.org/10.21037/apm.2020.03.26> PMID:[32233642](https://pubmed.ncbi.nlm.nih.gov/32233642/)



15. Mardani R, Ahmadi Vasmehjani A, Zali F, Gholami A, Mousavi Nasab SD, Kaghazian H, Kaviani M, Ahmadi N. Laboratory parameters in detection of COVID-19 patients with positive RT-PCR; a diagnostic accuracy study. *Arch Acad Emerg Med*. 2020; 8:e43. PMID:[32259132](https://pubmed.ncbi.nlm.nih.gov/32259132/)
16. Yuan J, Zou R, Zeng L, Kou S, Lan J, Li X, Liang Y, Ding X, Tan G, Tang S, Liu L, Liu Y, Pan Y, Wang Z. The correlation between viral clearance and biochemical outcomes of 94 COVID-19 infected discharged patients. *Inflamm Res*. 2020; 69:599–606. <https://doi.org/10.1007/s00011-020-01342-0> PMID:[32227274](https://pubmed.ncbi.nlm.nih.gov/32227274/)
17. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging*. 2012; 3:573–89. <https://doi.org/10.1007/s13244-012-0196-6> PMID:[23093486](https://pubmed.ncbi.nlm.nih.gov/23093486/)
18. Dasarathy BV, Holder EB. Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters*. 1991; 12:497–502. [https://doi.org/10.1016/0167-8655\(91\)80014-2](https://doi.org/10.1016/0167-8655(91)80014-2)
19. Nielsen B, Albrechtsen F, Danielsen HE. Statistical nuclear texture analysis in cancer research: a review of methods and applications. *Crit Rev Oncog*. 2008; 14:89–164. <https://doi.org/10.1615/critrevoncog.v14.i2-3.10> PMID:[19409060](https://pubmed.ncbi.nlm.nih.gov/19409060/)
20. Mosconi C, Cucchetti A, Bruno A, Cappelli A, Bargellini I, De Benedittis C, Lorenzoni G, Gramenzi A, Tarantino FP, Parini L, Pettinato V, Modestino F, Peta G, et al. Radiomics of cholangiocarcinoma on pretreatment CT can identify patients who would best respond to radioembolisation. *Eur Radiol*. 2020; 30:4534–44. <https://doi.org/10.1007/s00330-020-06795-9> PMID:[32227266](https://pubmed.ncbi.nlm.nih.gov/32227266/)
21. Yu YX, Wang XM, Shi C, Hu S, Hu CH. [The value of CT radiomics in the prediction of EGFR mutation in lung cancer]. *Zhonghua Yi Xue Za Zhi*. 2020; 100:690–95. <https://doi.org/10.3760/cma.j.issn.0376-2491.2020.09.009> PMID:[32187913](https://pubmed.ncbi.nlm.nih.gov/32187913/)
22. Li K, Sun H, Lu Z, Xin J, Zhang L, Guo Y, Guo Q. Value of [18F]FDG PET radiomic features and VEGF expression in predicting pelvic lymphatic metastasis and their potential relationship in early-stage cervical squamous cell carcinoma. *Eur J Radiol*. 2018; 106:160–66. <https://doi.org/10.1016/j.ejrad.2018.07.024> PMID:[30150039](https://pubmed.ncbi.nlm.nih.gov/30150039/)
23. Tuder RM, Yun JH. Vascular endothelial growth factor of the lung: friend or foe. *Curr Opin Pharmacol*. 2008; 8:255–60. <https://doi.org/10.1016/j.coph.2008.03.003> PMID:[18468486](https://pubmed.ncbi.nlm.nih.gov/18468486/)
24. Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging (Albany NY)*. 2020; 12:6049–57. <https://doi.org/10.18632/aging.103000> PMID:[32267833](https://pubmed.ncbi.nlm.nih.gov/32267833/)
25. Ben Bouallègue F, Tabaa YA, Kafrouni M, Cartron G, Vauchot F, Mariano-Goulart D. Association between textural and morphological tumor indices on baseline PET-CT and early metabolic response on interim PET-CT in bulky Malignant lymphomas. *Med Phys*. 2017; 44:4608–19. <https://doi.org/10.1002/mp.12349> PMID:[28513853](https://pubmed.ncbi.nlm.nih.gov/28513853/)
26. Tatsumi M, Isohashi K, Matsunaga K, Watabe T, Kato H, Kanakura Y, Hatazawa J. Volumetric and texture analysis on FDG PET in evaluating and predicting treatment response and recurrence after chemotherapy in follicular lymphoma. *Int J Clin Oncol*. 2019; 24:1292–300. <https://doi.org/10.1007/s10147-019-01482-2> PMID:[31165310](https://pubmed.ncbi.nlm.nih.gov/31165310/)
27. El-Galaly TC, Villa D, Gormsen LC, Baech J, Lo A, Cheah CY. FDG-PET/CT in the management of lymphomas: current status and future directions. *J Intern Med*. 2018; 284:358–76. <https://doi.org/10.1111/joim.12813> PMID:[29989234](https://pubmed.ncbi.nlm.nih.gov/29989234/)
28. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SK, et al. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography*. 2016; 2:430–37. <https://doi.org/10.18383/j.tom.2016.00235> PMID:[28149958](https://pubmed.ncbi.nlm.nih.gov/28149958/)
29. Liang T. Handbook of COVID-19 Prevention and Treatment. 2020. <https://covid-19.alibabacloud.com/>
30. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015; 42:1341–53. <https://doi.org/10.1118/1.4908210> PMID:[25735289](https://pubmed.ncbi.nlm.nih.gov/25735289/)
31. Ger RB, Cardenas CE, Anderson BM, Yang J, Mackin DS, Zhang L, Court LE. Guidelines and experience using imaging biomarker explorer (IBEX) for radiomics. *J Vis Exp*. 2018; 131:57132. <https://doi.org/10.3791/57132> PMID:[29364284](https://pubmed.ncbi.nlm.nih.gov/29364284/)

32. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys*. 2018; 102:1143–58.  
<https://doi.org/10.1016/j.ijrobp.2018.05.053>  
PMID:[30170872](https://pubmed.ncbi.nlm.nih.gov/30170872/)
33. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989; 45:255–68.  
PMID:[2720055](https://pubmed.ncbi.nlm.nih.gov/2720055/)
34. Barnhart HX, Haber M, Song J. Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics*. 2002; 58:1020–7.  
<https://doi.org/10.1111/j.0006-341x.2002.01020.x>  
PMID:[12495158](https://pubmed.ncbi.nlm.nih.gov/12495158/)
35. Bullen PS. *Handbook of means and their inequalities*: Springer Science and Business Media. 2013.
36. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002; 16:321–57.  
<https://doi.org/10.1613/jair.953>
37. Batuwita R, Palade V. Efficient resampling methods for training support vector machines with imbalanced datasets. *The 2010 International Joint Conference on Neural Networks (IJCNN)*: IEEE. 2010; 1–8.  
<https://doi.org/10.1109/IJCNN.2010.5596787>
38. Xie C, Du R, Ho JW, Pang HH, Chiu KW, Lee EY, Vardhanabhuti V. Effect of machine learning re-sampling techniques for imbalanced datasets in <sup>18</sup>F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur J Nucl Med Mol Imaging*. 2020; 47:2826–35.  
<https://doi.org/10.1007/s00259-020-04756-4>  
PMID:[32253486](https://pubmed.ncbi.nlm.nih.gov/32253486/)
39. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2017; 18:559–63.
40. Wu L, Gao C, Xiang P, Zheng S, Pang P, Xu M. CT-imaging based analysis of invasive lung adenocarcinoma presenting as ground glass nodules using peri- and intra-nodular radiomic features. *Front Oncol*. 2020; 10:838.  
<https://doi.org/10.3389/fonc.2020.00838>  
PMID:[32537436](https://pubmed.ncbi.nlm.nih.gov/32537436/)
41. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC Med Res Methodol*. 2016; 16:154.  
<https://doi.org/10.1186/s12874-016-0254-8>  
PMID:[27842498](https://pubmed.ncbi.nlm.nih.gov/27842498/)
42. Kumamaru KK, Saboo SS, Aghayev A, Cai P, Quesada CG, George E, Hussain Z, Cai T, Rybicki FJ. CT pulmonary angiography-based scoring system to predict the prognosis of acute pulmonary embolism. *J Cardiovasc Comput Tomogr*. 2016; 10:473–79.  
<https://doi.org/10.1016/j.jcct.2016.08.007>  
PMID:[27591768](https://pubmed.ncbi.nlm.nih.gov/27591768/)
43. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016; 34:2534–40.  
<https://doi.org/10.1200/JCO.2015.65.5654>  
PMID:[27247223](https://pubmed.ncbi.nlm.nih.gov/27247223/)
44. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016; 352:i6.  
<https://doi.org/10.1136/bmj.i6>  
PMID:[26810254](https://pubmed.ncbi.nlm.nih.gov/26810254/)
45. Rong Y, Yan L, Rong J, Hauptmann A. On predicting rare classes with SVM ensembles in scene classification. *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 Proceedings (ICASSP '03)*. 2003; III–21.  
<https://doi.org/10.1109/ICASSP.2003.1199097>

## SUPPLEMENTARY MATERIALS

### Supplementary Appendix 1

#### Definition of net benefit in the decision curve analysis

The net benefit [1, 2] was defined by the following equation:

$$\text{Net Benefit} = \text{TPR} * \omega - \frac{P_t}{1 - P_t} * \text{FPR} * (1 - \omega) \quad (1)$$

$P_t$  is the “threshold possibility” to stratify the patients into high-risk COVID-19 or low-risk non-COVID-19 groups. Patients with a probability of having COVID-19 higher than  $P_t$  are high-risk patients. TPR is the true positive rate, defined as the proportion of high-risk patients in the patients having COVID-19. FPR is the false positive rate, defined as the proportion of high-risk patients in the patients having non-COVID-19.  $\omega$  is the prevalence of having COVID-19, calculated by dividing

the total patients number by the number of patients with COVID-19. In this study, the treat-all scheme assumes that all the patients were COVID-19; the treat-none scheme assumes that all the patients were non-COVID-19. In the condition of “treat none”, no patient is classified as high risk, both the TPR and FPR are zero, so the Net Benefit is zero. In the condition of “treat all”, all patients are classified as high risk COVID-19 (TPR=FPR=1), so the Net Benefit is calculated as

$$\begin{aligned} \text{Net Benefit}_{\text{Treat Everyone}} &= \omega - \frac{P_t}{1 - P_t} * (1 - \omega) \\ &= \frac{1 - \omega}{P_t - 1} + 1 \end{aligned} \quad (2)$$

, which is a monotonically decreasing dash-line curve in the figure.

## Supplementary Appendix 2

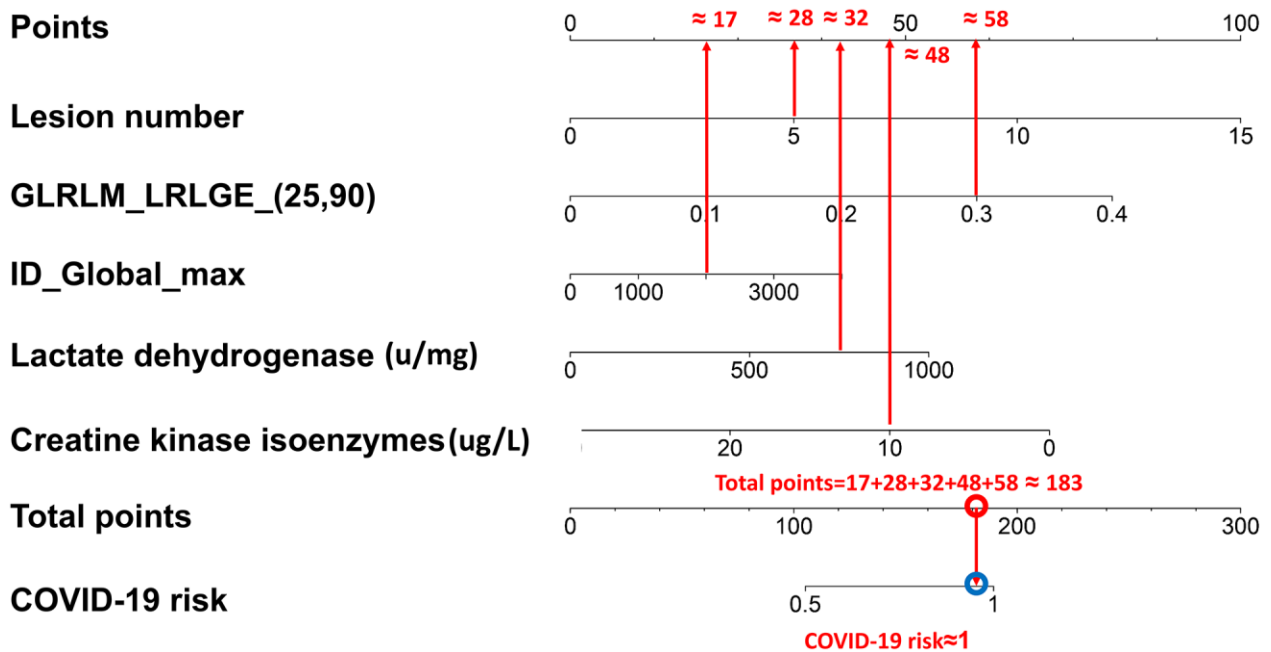
### The procedure of using the nomogram of the risk score combining radiomic features and clinical factors

This is the patient-based risk score integrating 2 radiomic features and 3 clinical variables. For example, a suspicious patient was found having the following radiomic features and clinical factors detected/calculated: lesions numbers = 5, GLRLM\_LRLGE\_(25,90) = 0.3, ID\_Global\_Max = 2000, lactate dehydrogenase = 750 u/mg, creatine kinase isoenzymes = 10 ug/L. The values in the Points line in the 1st row corresponding to these radiomic features and clinical factors are 28, 58, 17, 32, 48. As such, the total point adding all the values in the Points line is 183 in the Total points line in the 7th row. So, the patient's risk of COVID-19 can be calculated from this nomogram with a risk score close to 0.95. Alternatively, all radiomic features and clinical

factors detected/calculated can be plugged into the risk score equation to get the score value:

$$\begin{aligned}
 &\text{The patient – based risk score combining radiomics} \\
 &\text{and clinical features} \\
 &= -114.053 + 9.529 \times \text{lesion number} + 122.045 \\
 &\quad \times \text{GLRLM\_LRLGE\_}(25,90) + 0.0196 \\
 &\quad \times \text{ID\_Global\_Max} + 0.334 \\
 &\quad \times \text{lactate dehydrogenase} - 7.593 \\
 &\quad \times \text{creatine kinase isoenzymes} \\
 &= -114.053 + 9.529 \times 5 + 122.045 \\
 &\quad \times 0.3 + 0.0196 \times 2000 + 0.334 \times 750 \\
 &\quad - 7.593 \times 10 = 183.9755
 \end{aligned} \tag{3}$$

The score of 183.9755 is corresponding to 0.95 on the COVID-19 Risk line in the 8<sup>th</sup> row of the nomogram.



## Supplementary Appendix 3

### The influence of bias induced in the boundary and volume contoured in the manual delineating process on the radiomics values calculated

To further verify the repeatability of two features GLRLM\_LRLGE\_ (25, 90) and ID\_Global\_Max selected in the construction of patient-based risk scores in this study, a verification study was conducted. A CT data set from one COVID-19 patient was used for delineation by five different radiologists and the differences of volumes and surface areas caused by different delineations were calculated (Supplementary Figure 6). The 5 volumes-of-interest (VOIs) delineated were also used for the GLRLM\_LRLGE\_ (25, 90) and ID\_Global\_Max feature extraction and calculation. Three different tools (2 open sources ((image biomarker explorer (IBEX) [3] and Pyradiomics [4]) and 1 in-house Matlab codes) were used to extract and calculate the radiomic feature values of the GLRLM\_LRLGE\_ (25, 90) and ID\_Global\_Max. The patient-based COVID-19 risk score using radiomic features only was calculated using the radiomic features extracted from 5 VOIs contoured according to the formula developed in this study (Equation 5). The results are shown in Supplementary Table 15 as follows.

The patient-based risk score using radiomic features only  
 $= -3.785 + 19.563 \times \text{GLRLM\_LRLGE\_}(25,90) + 0.002$   
 $\times \text{ID\_Global\_Max}$

As shown in Supplementary Table 15, the VOI delineation biases induced by different radiologists had a relatively small impact on the radiomic feature values of GLRLM\_LRLGE\_ (25, 90) and ID\_Global\_Max as well as the COVID-19 risk score values calculated. In addition, the feature extractions on the VOIs contoured using different tools had not significantly affected the calculation of COVID-19 risk score, which could be considered to be a good repeatability.

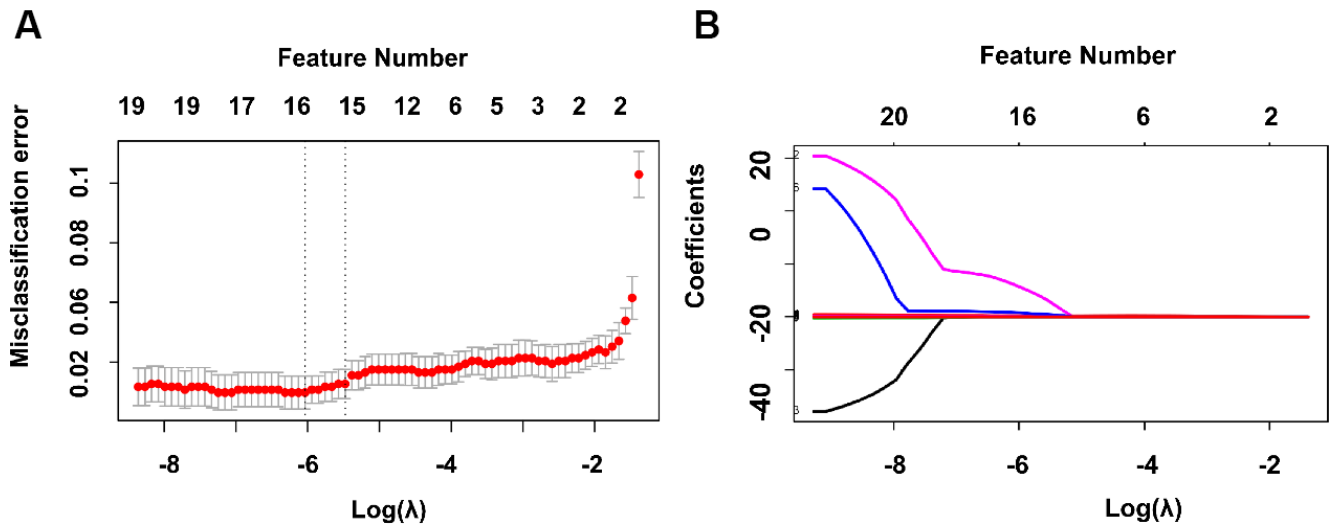
Furthermore, the same verification study was also repeated on three radiomic features identified in the lesion based analysis and the results were shown in Supplementary Table 16. Similarly, the VOI delineation biases induced by different radiologists had a relatively small impact on the radiomic feature values of GOH\_Percentile\_(15), GLCM\_Correlation\_(25,0,1) and ID\_Local\_range\_std as well as the COVID-19 risk score values calculated. In addition, the feature extractions on the VOIs contoured using different tools had not significantly affected the calculation of COVID-19 risk score.

### Supplementary References

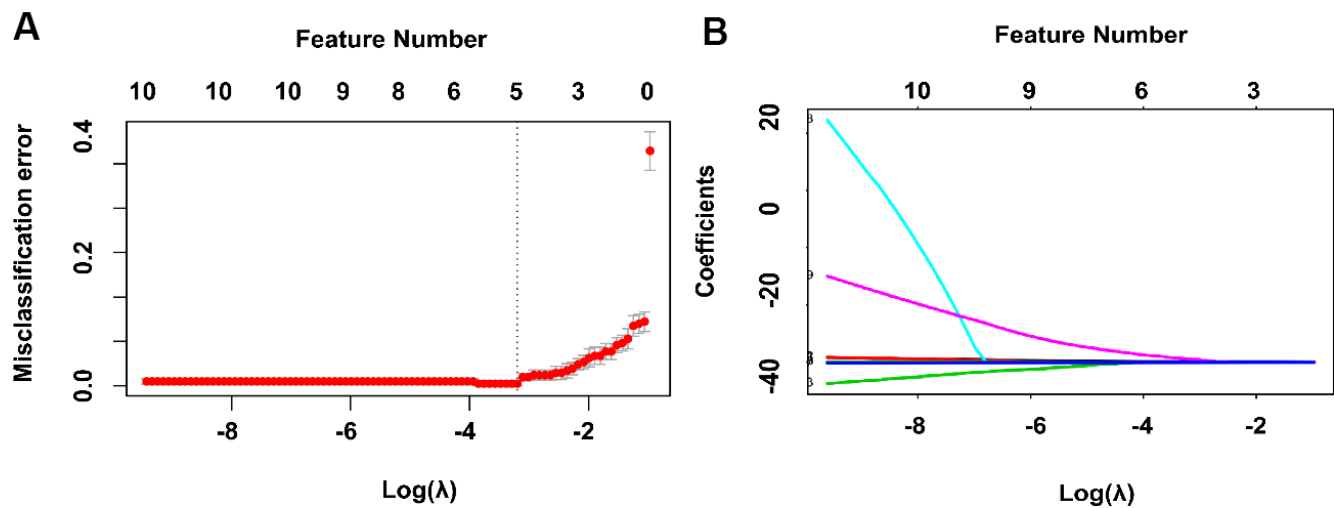
1. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *J Clin Oncol*. 2016; 34:2534–40. <https://doi.org/10.1200/JCO.2015.65.5654> PMID:[27247223](https://pubmed.ncbi.nlm.nih.gov/27247223/)
2. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016; 352:i6. <https://doi.org/10.1136/bmj.i6> PMID:[26810254](https://pubmed.ncbi.nlm.nih.gov/26810254/)
3. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015; 42:1341–53. <https://doi.org/10.1118/1.4908210> PMID:[25735289](https://pubmed.ncbi.nlm.nih.gov/25735289/)
4. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, Echegaray S, Rubin D, McNitt-Gray M, Lo P, Sieren JC, Uthoff J, Dilger SK, et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography*. 2016; 2:430–37. <https://doi.org/10.18383/j.tom.2016.00235> PMID:[28149958](https://pubmed.ncbi.nlm.nih.gov/28149958/)



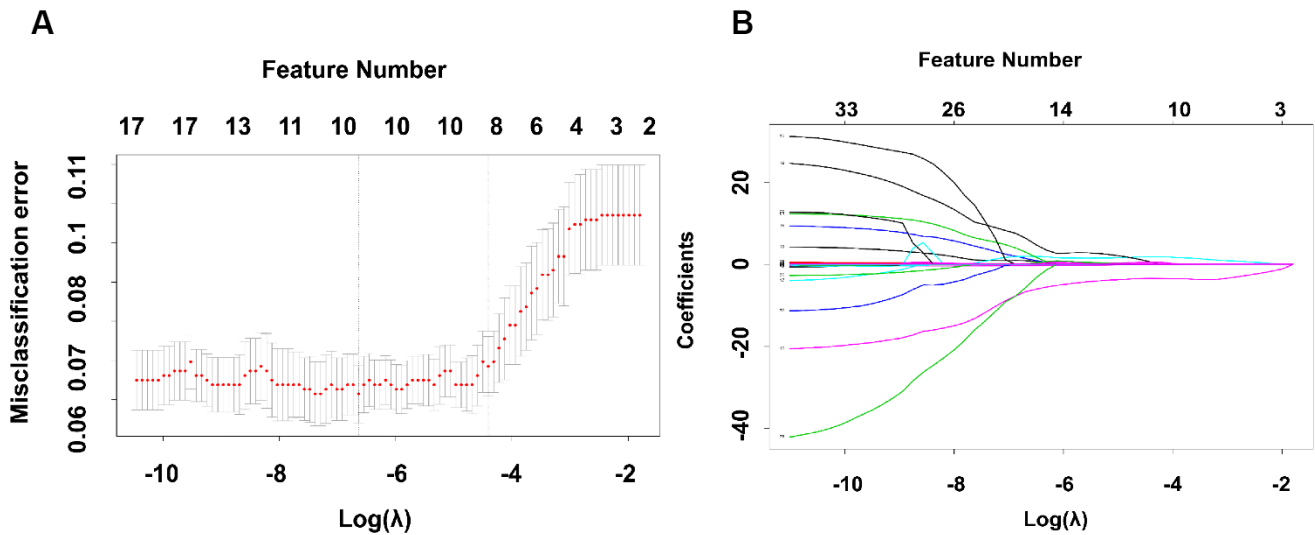
## Supplementary Figures



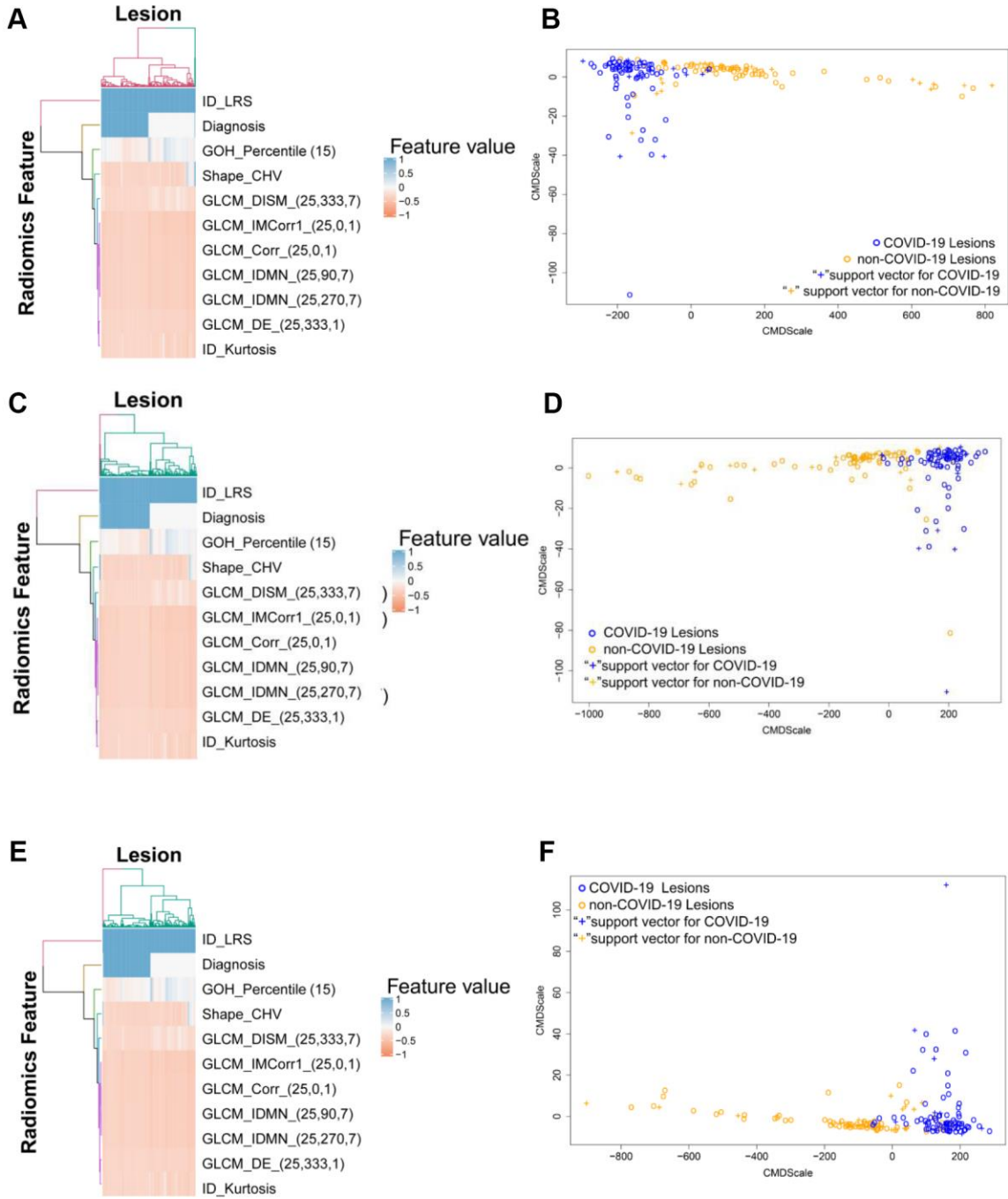
**Supplementary Figure 1. The LASSO process to select the radiomic features for the construction of the patient-based risk score using radiomic features only.** (A) The radiomic features selection procedure using LASSO regression method. To determine the best features combination for building the risk score, the control parameter  $\lambda$  value in the LASSO model was selected via 4-fold cross-validation with minimum criteria. The x-axis is the value of  $\log(\lambda)$  and the y-axis is the binomial deviance in the 4-fold cross validation method 100 times. The upper x-axis is the number of non-zero-coefficient features with a given  $\lambda$ . The red curve indicated the average binomial deviance value with the vertical bars showing the upper and lower boundaries. The left vertical dotted line defined the  $\lambda$  with the least binomial deviance. The right vertical dotted line indicates the largest value of  $\lambda$  such that the binomial deviance is within one standard error of the minimum binomial deviance. (B) The LASSO coefficient profiles of the radiomic features. The figure shows the feature coefficient change with the tuning of  $\lambda$  value. The dotted line was plotted at the  $\lambda$  value determined in (A) resulting 16 non-zero-coefficient radiomic features.

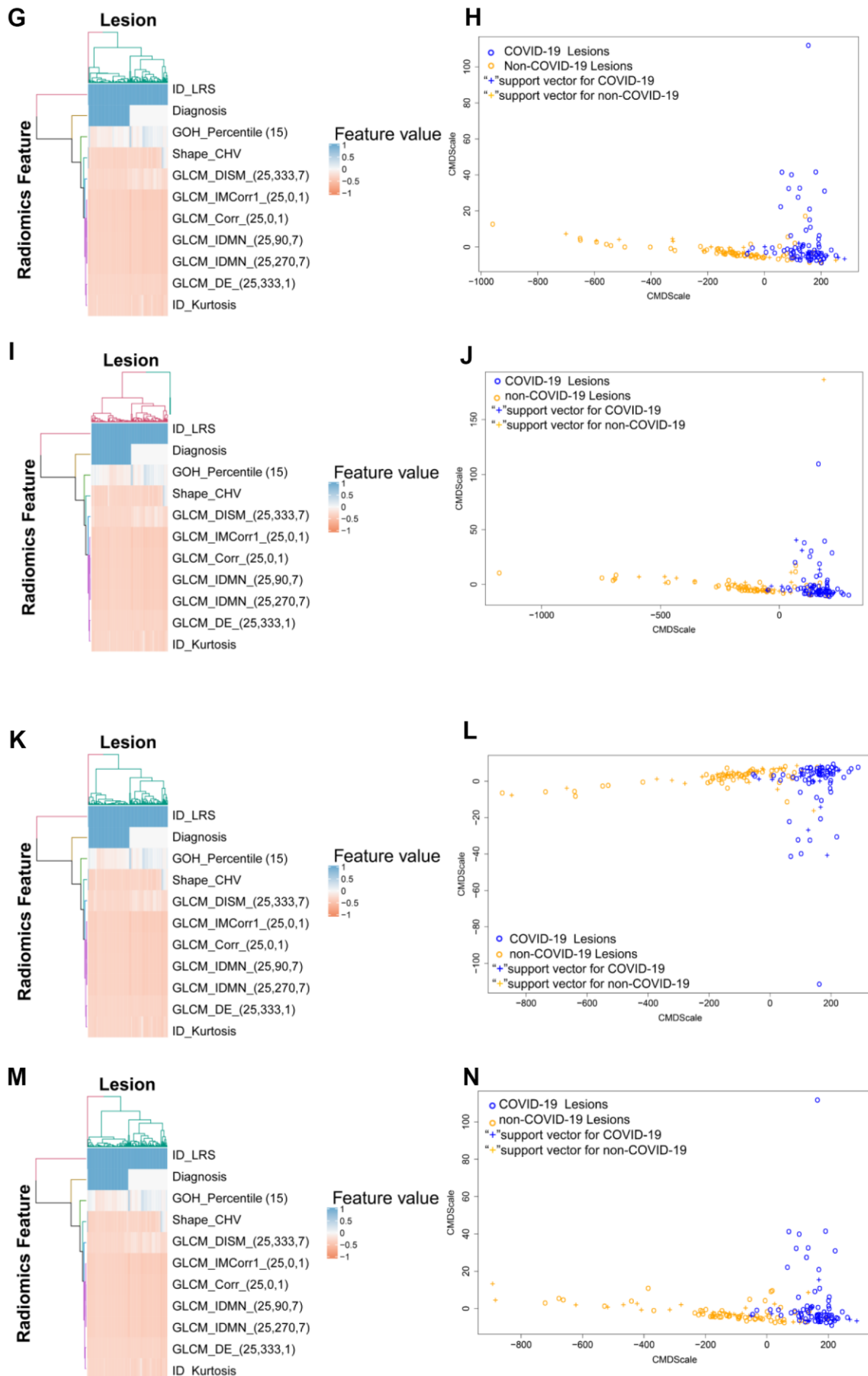


**Supplementary Figure 2. The LASSO process to select the radiomics features and clinical variables for the construction of the patient-based risk score combining radiomic features and clinical variables.** (A) The radiomic features and clinical variables selection procedure using LASSO regression method. To determine the best features combination for building the risk score, the control parameter  $\lambda$  value in the LASSO model was selected via 4-fold cross-validation with minimum criteria. The x-axis is the value of  $\log(\lambda)$  and the y-axis is the binomial deviance in the 4-fold cross validation method 100 times. The upper x-axis is the number of non-zero-coefficient features with a given  $\lambda$ . The red curve indicated the average binomial deviance value with the vertical bars showing the upper and lower boundaries. The left vertical dotted line defined the  $\lambda$  with the least binomial deviance. The right vertical dotted line indicates the largest value of  $\lambda$  such that the binomial deviance is within one standard error of the minimum binomial deviance. (B) The LASSO coefficient profiles of the 17 radiomic features and clinical variables. The figure shows the feature coefficient change with the tuning of  $\lambda$  value. The dotted line was plotted at the  $\lambda$  value determined in (A) resulting 5 non-zero-coefficient radiomic features and clinical variables.

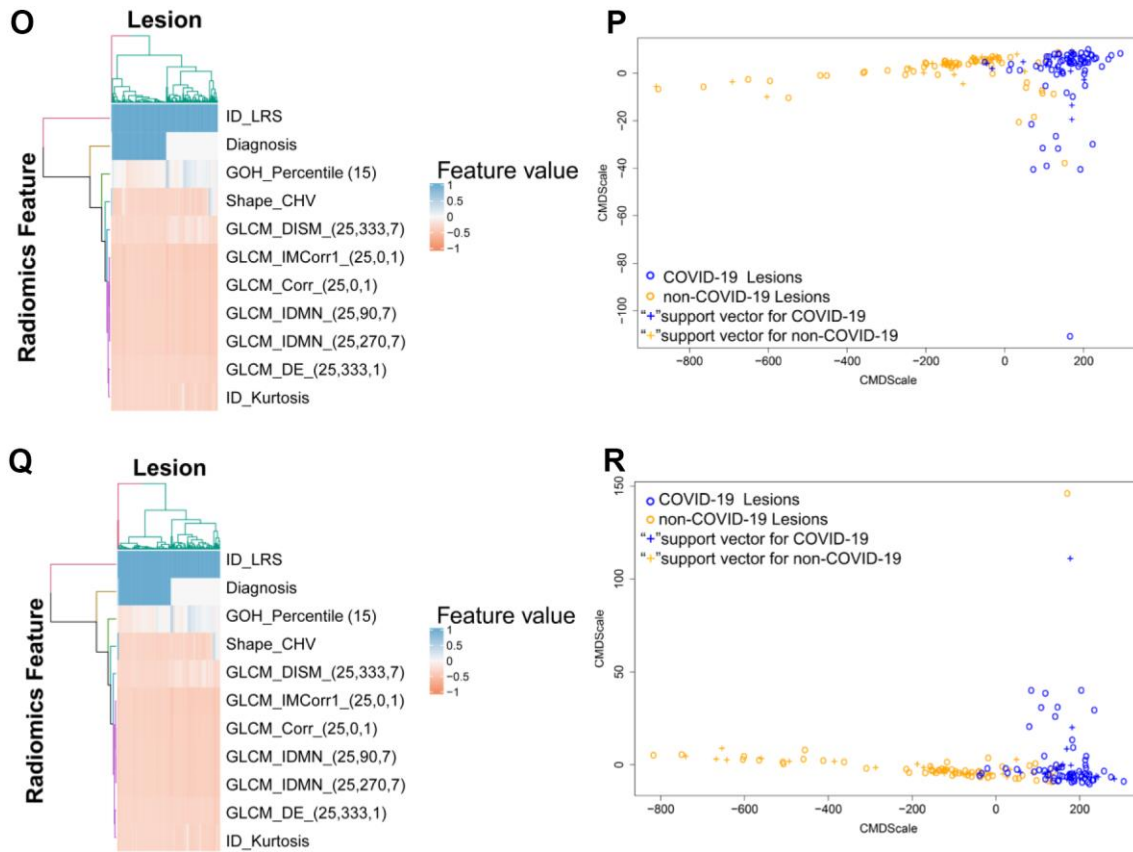


**Supplementary Figure 3. The LASSO process to select the radiomic features for the construction of lesion-based risk score using radiomic features alone.** (A) The radiomic features selection procedure using LASSO regression method. To determine the best features combination for building the risk score, the control parameter  $\lambda$  value in the LASSO model was selected via 4-fold cross-validation with minimum criteria. The x-axis is the value of  $\log(\lambda)$  and the y-axis is the binomial deviance in the 4-fold cross validation method 100 times. The upper x-axis is the number of non-zero-coefficient features with a given  $\lambda$ . The red curve indicated the average binomial deviance value with the vertical bars showing the upper and lower boundaries. The left vertical dotted line defined the  $\lambda$  with the least binomial deviance. The right vertical dotted line indicates the largest value of  $\lambda$  such that the binomial deviance is within one standard error of the minimum binomial deviance. (B) The LASSO coefficient profiles of the radiomic features. The figure shows the feature coefficient change with the tuning of  $\lambda$  value. The dotted line was plotted at the  $\lambda$  value determined in (A) resulting the 10 non-zero-coefficient radiomic features.

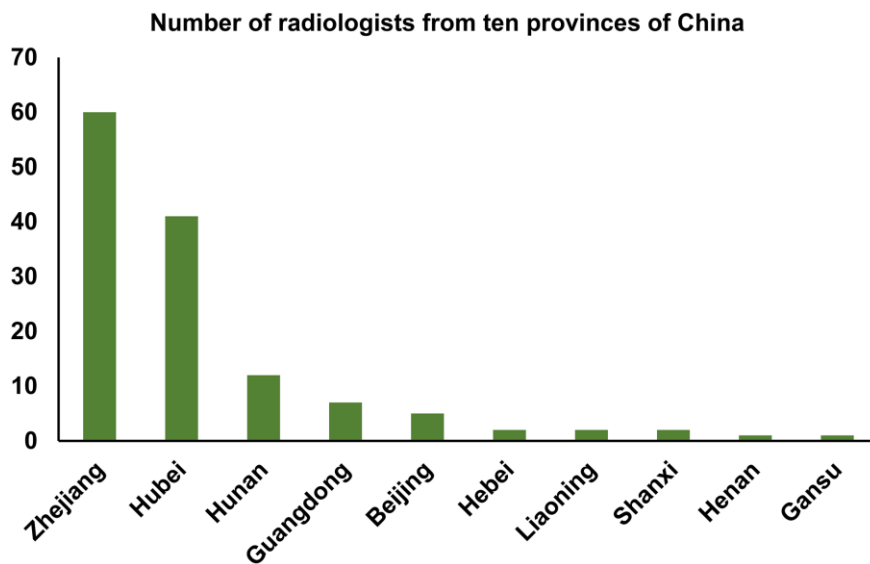




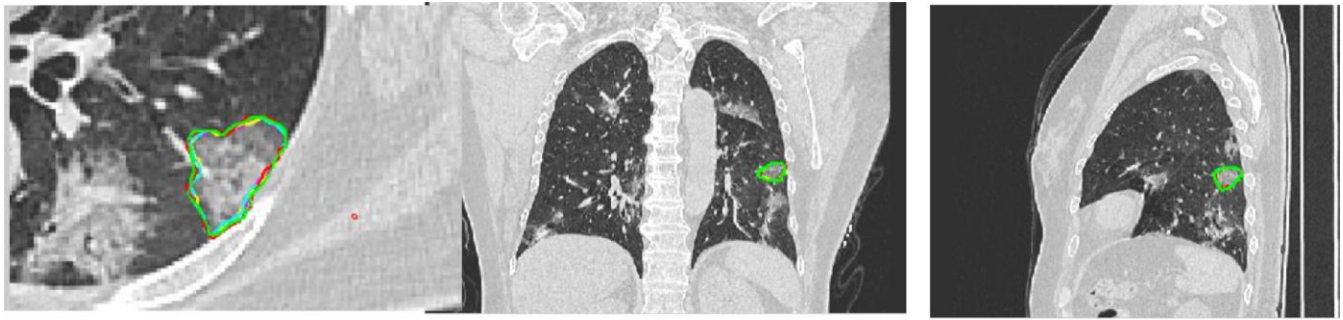




**Supplementary Figure 4. The correlation of radiomics features and COVID-19 for each constituent SVM. (A, C, E, G, I, K, M, O, Q):** The heat maps of correlation of radiomics features and COVID-19 for 9 individual constituent SVM. **(B, D, F, H, J, L, N, P, R):** The classical (metric) multidimensional scaling matrix to demonstrate the discriminative abilities of the individual constituent SVM to classify COVID-19 and non-COVID-19 using 9 constituent SVM individually. Abbreviations: SVM, support vector machine.



**Supplementary Figure 5. The geographic distribution of 130 radiologists from 10 provinces in China including Hubei province, the epicenter of COVID-19 outbreak in China.**



Supplementary Figure 6. Segmentations from 5 radiologists to delineate the volumes of interest in the same COVID-19 patient.

## Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 15, 16.

**Supplementary Table 1. The 32 radiomic features selected by principal component analysis (PCA) for dimensionality reduction to construct the patient-based risk score using radiomic features only.**

Radiomic features	Coefficients	Cumulative%	Explainable variance ratio%
Shape_Convex	0.913	37.116	37.116
Shape_Convex_Hull_Volume	0.926	54.347	17.231
GOH_Percentile_(15)	0.933	66.221	11.874
GOH_Percentile_(50)	0.979	75.455	9.234
GLCM_Cluster_Prominence_(25,333,4)	0.901	80.369	4.914
GLCM_Cluster_Prominence_(25,315,4)	0.972	84.221	3.852
GLCM_Cluster_Shade_(25,333,4)	0.754	86.194	1.973
GLCM_Cluster_Shade_(25,315,4)	0.829	87.825	1.631
GLCM_Cluster_Shade_(25,315,7)	0.936	89.214	1.389
GLCM_Contrast_(25,0,10)	0.689	90.605	1.391
GLCM_Contrast_(25,90,7)	0.757	91.711	1.106
GLCM_Contrast_(25,270,4)	0.837	92.582	0.871
GLCM_Correlation_(25,0,1)	0.306	93.401	0.819
GLCM_Correlation_(25,0,4)	0.954	94.124	0.723
GLCM_Correlation_(25,90,4)	0.661	94.813	0.689
GLCM_Difference_Entropy_(25,333,1)	0.396	95.446	0.633
GLCM_Dissimilarity_(25,333,7)	0.925	95.967	0.521
GLCM_Energy_(25,45,4)	0.931	96.548	0.581
GLCM_Energy_(25,45,7)	0.486	96.985	0.437
GLCM_Energy_(25,135,4)	0.654	97.386	0.401
GLCM_Information_Measure_Corr1_(25,0,1)	0.671	97.782	0.396
GLCM_Information_Measure_Corr2_(25,315,1)	0.432	98.097	0.315
GLCM_Inverse_Diff_Moment_Norm_(25,135,4)	0.779	98.395	0.298
GLCM_Inverse_Diff_Moment_Norm_(25,225,4)	0.828	98.672	0.277
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	0.669	98.884	0.212
GLCM_Inverse_Variance_(25,225,4)	0.534	99.093	0.209
GLCM_Inverse_Variance_(25,315,4)	0.289	99.287	0.194
GLCM_Max_Probability_(25,180,1)	0.275	99.454	0.167
GLCM_Sum_Entropy_(25,180,7)	0.439	99.602	0.148
ID_Kurtosis	0.354	99.74	0.138
ID_Local_Range_Std	0.119	99.879	0.139
IH_Percentile_Area_(30)	0.387	100	0.121

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 2. The 20 radiomic features plus 6 clinical variables selected by principal component analysis (PCA) for dimensionality reduction to construct the patient-based risk score combining radiomic features and clinical variables.**

<b>Radiomic features and clinical variables</b>	<b>Coefficients</b>	<b>Cumulative%</b>	<b>Explainable variance ratio%</b>
Lesion number	0.866	33.623	33.623
Lactate dehydrogenase	0.829	44.588	10.965
C-reactive protein	0.756	53.795	9.207
Gender	0.932	61.926	8.131
White blood cell	0.689	69.029	7.103
D-dimers	0.832	75.902	6.873
Shape_Compactness_(2)	0.911	81.423	5.521
Shape_Convex_Hull_Volume	0.874	85.786	4.363
Shape_Orientation	0.829	88.938	3.152
Shape_Roundness	0.543	91.042	2.104
GOH_Percentile_Area_(15)	0.886	92.476	1.434
GOH_Percentile_Area_(70)	0.851	93.494	1.018
GOH_Range	0.944	94.497	1.003
GLCM_Cluster_Prominence_(25,333,4)	0.756	95.499	1.002
GLCM_Cluster_Shade_(25,333,4)	0.813	96.400	0.901
GLCM_Correlation_(25,333,1)	0.925	97.143	0.743
GLCM_Correlation_(25180,7)	0.866	97.620	0.477
GLCM_Difference_Entropy_(25,333,1)	0.862	98.081	0.461
GLCM_Information_Measure_Corr_(2,25,333,7)	0.642	98.533	0.452
GLCM_Max_Probability_(25,333,1)	0.473	98.845	0.312
GLCM_Max_Probability_(250,7)	0.455	99.078	0.233
GLRLM_Long_Run_Low_Gray_Level_Empha_(2590)	0.856	99.289	0.211
ID_Global_Max	0.865	99.498	0.209
ID_Local_Entropy_Std	0.674	99.699	0.201
ID_Percentile_(30)	0.498	99.863	0.164
IH_Percentile_(30)	0.563	100.000	0.137

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 3. The 20 radiomic features selected by the Mann-Whitney U test for dimensionality reduction to construct the patient-based risk score using radiomic features only.**

<b>Radiomic feature</b>	<b>AUC</b>	<b>95% CI</b>	<b>P</b>
Shape_Convex_Hull_Volume	0.704	0.664-0.808	0.000
GOH_Percentile_(15)	0.890	0.828-0.921	0.000
GOH_Percentile_(50)	0.688	0.618-0.778	0.321
GLCM_Cluster_Prominence_(25,333,4)	0.544	0.464-0.632	0.467
GLCM_Cluster_Prominence_(25,315,4)	0.509	0.458-0.613	0.218
GLCM_Cluster_Shade_(25,315,7)	0.438	0.367-0.557	0.000
GLCM_Correlation_(25,0,1)	0.853	0.803-0.903	0.000
GLCM_Contrast_(25,0,1)	0.972	0.912-0.998	0.000
GLCM_Contrast_(25,270,4)	0.781	0.713-0.872	0.000
GLCM_Difference_Entropy_(25,333,1)	0.940	0.876-0.984	0.000
GLCM_Dissimilarity_(25,333,7)	0.532	0.429-0.613	0.311
GLCM_Energy_(25,135,4)	0.557	0.465-0.604	0.231
GLCM_Energy_(25,225,4)	0.532	0.487-0.612	0.367
GLCM_Information_Measure_Corr1_(25,0,1)	0.623	0.539-0.719	0.000
GLCM_Information_Measure_Corr2_(25,315,1)	0.567	0.509-0.633	0.000
GLCM_Inverse_Diff_Moment_Norm_(25,90,7)	0.540	0.512-0.634	0.000
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	0.677	0.621-0.759	0.000
GLCM_Sum_Entropy_(25,180,7)	0.684	0.589-0.704	0.000
ID_Kurtosis	0.734	0.633-0.811	0.000
ID_Local_Range_Std	0.976	0.909-0.994	0.216

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 4. The 11 radiomic features and 6 clinical variables selected by the Mann-Whitney U test for dimensionality reduction to construct COVID-19 risk score using radiomic features and clinical variables.**

<b>Radiomic features and clinical variables</b>	<b>AUC</b>	<b>95% CI</b>	<b>P Value</b>
Lesion number	0.956	0.909-0.994	0.000
Gender	0.463	0.274-0.554	0.673
White blood cell	0.394	0.369-0.551	0.421
C-reactive protein	0.674	0.593-0.751	0.000
Lactate dehydrogenase	0.887	0.804-0.915	0.000
Creatine kinase isoenzyme	0.069	0.036-0.081	0.000
Shape_Compactness_(2)	0.365	0.278-0.434	0.003
GOH_Range	0.678	0.576-0.717	0.000
GLCM_Cluster_Prominence_(25,333,4)	0.604	0.512-0.723	0.004
GLCM_Correlation_(25,333,1)	0.287	0.209-0.353	0.000
GLCM_Difference_Entropy_(25,333,1)	0.724	0.667-0.771	0.000
GLCM_MaxProbability_(250,7)	0.365	0.312-0.423	0.009
GLRLM_LRLGE_(25,90)	0.767	0.676-0.812	0.000
ID_GlobalMax	0.791	0.695-0.853	0.000
ID_Percentile_(30)	0.368	0.311-0.467	0.001
IH_Percentile_(30)	0.303	0.286-0.431	0.003
IH_Percentile_Area	0.387	0.271-0.463	0.000

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 5. The 16 radiomic feature selected by the LASSO for dimensionality reduction to construct the patient-based risk score using radiomic features only.**

<b>Radiomic feature</b>	<b>Lambda</b>
Intercept	1.04E-02
Shape_Convex_Hull_Volume	2.07E-01
GOH_Percentile_(15)	-8.58E-02
GOH_Percentile_(50)	-4.07E-05
GLCM_AutoCorrelation(250,1)	-3.58E-06
GLCM_ClusterProminence(25,315,4)	-2.02E-05
GLCM_ClusterShade(25,315,4)	4.63E-02
GLCM_Correlation_(25,0,1)	4.04E-01
GLCM_Difference_Entropy_(25,333,1)	-4.26E-01
GLCM_Dissimilarity_(25,333,7)	6.11E+02
GLCM_Information_Measure_Corr1_(25,0,1)	1.49E+01
GLCM_Inverse_Diff_Moment_Norm_(25,90,7)	7.88E+01
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	7.91E-13
GLCM_InverseDiffMomentNorm(25,270,4)	3.54E+00
GLCM25180_7SumEntropy	-1.83E-01
ID_Kurtosis	3.45E-02
ID_Local_Range_Std	1.04E-02

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 6. The 2 radiomic features and 3 clinical variables selected by the LASSO for dimensionality reduction to construct COVID-19 risk score using radiomic features and clinical variables.**

<b>Radiomic features and clinical variables</b>	<b>Lambda</b>
Intercept	-1.088064e+00
Lesion number	3.450725e-01
Lactate dehydrogenase	4.697555e-01
Creatine kinase isoenzyme	3.636614e-05
GLRLM_LRLGE_(25,90)	5.842366e-03
ID_GlobalMax	-1.605967e-01

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 7. The 2 radiomic features selected by the multivariable logistic regression to construct the patient-based risk score using radiomic features only in the training set.**

	<b>Coef</b>	<b>S.E.</b>	<b>Wald Z</b>	<b>Pr(&gt; Z )</b>	<b>AUC</b>	<b>95% CI</b>
Intercept	-3.785	27.834	-4.25	0.898	..	..
GLRLM_LRLGE_(25, 90)	19.563	175.031	7.27	<0.0001	0.813	0.698-0.876
ID_Global_Max	0.002	0.017	0.14	0.002	0.853	0.668-0.814

The values in the brackets represent the parameters used to calculate that particular radiomic feature.



**Supplementary Table 8. The 2 radiomic features and 3 clinical variables used by the multivariable logistic regression to construct the COVID-19 risk score using radiomic features and clinical variables in training set.**

	<b>Coef</b>	<b>S.E.</b>	<b>Wald Z</b>	<b>Pr(&gt; Z )</b>	<b>AUC</b>	<b>95% CI</b>
Intercept	-114.053	559.175	-0.11	0.765	..	..
Lesion number	9.311	31.132	9.59	<0.0001	0.823	0.793-0.906
GLRLM_LRLGE_(25, 90)	122.045	687.176	2.14	<0.0001	0.718	0.624-0.845
ID_Global_Max	0.0196	0.045	3.08	0.001	0.786	0.667-0.822
Lactate dehydrogenase	0.334	2.063	7.21	<0.0001	0.861	0.779-0.916
creatinine kinase isoenzymes	-7.593	38.179	-6.23	<0.0001	0.856	0.793-0.991

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 9. The 3 clinical variables used by the multivariable logistic regression to construct the patient-based risk score using clinical variables only in the training set.**

	<b>Coef</b>	<b>S.E.</b>	<b>Wald Z</b>	<b>Pr(&gt; Z )</b>	<b>AUC</b>	<b>95% CI</b>
Intercept	-15.680	134.812	-8.13	0.898	..	..
Lesion number	2.833	17.532	2.31	<0.0001	0.812	0.749-0.917
Lactate dehydrogenase	0.104	1.337	1.19	<0.0001	0.857	0.813-0.906
creatinine kinase isoenzymes	-1.674	11.294	-1.76	<0.0001	0.844	0.794-0.974

**Supplementary Table 10. The 32 radiomics features selected by principal component analysis (PCA) for dimensionality reduction to construct lesion-based risk score using radiomic features only.**

<b>Radiomic Features</b>	<b>Coefficients</b>	<b>Cumulative%</b>	<b>Explainable variance ratio%</b>
Shape_Convex	0.919	36.680	36.680
Shape_Convex_Hull_Volume	0.977	19.417	56.098
GOH_Percentile_(15)	0.867	10.496	66.593
GOH_Percentile_(50)	0.867	9.244	75.837
GLCM_Cluster_Prominence_(25,333,4)	0.912	4.470	80.307
GLCM_Cluster_Prominence_(25,315,4)	0.858	2.653	82.961
GLCM_Cluster_Shade_(25,333,4)	0.673	1.839	84.800
GLCM_Cluster_Shade_(25,315,4)	0.895	1.568	86.368
GLCM_Cluster_Shade_(25,315,7)	0.920	1.319	87.687
GLCM_Contrast_(25,0,10)	0.494	1.255	88.942
GLCM_Contrast_(25,90,7)	0.695	1.161	90.103
GLCM_Contrast_(25,270,4)	0.817	1.012	91.116
GLCM_Correlation_(25,0,1)	0.208	0.922	92.038
GLCM_Correlation_(25,0,4)	0.850	0.844	92.882
GLCM_Correlation_(25,90,4)	0.659	0.785	93.667
GLCM_Difference_Entropy_(25,333,1)	0.569	0.734	94.401
GLCM_Dissimilarity_(25,333,7)	0.819	0.686	95.087
GLCM_Energy_(25,45,4)	0.878	0.658	95.745
GLCM_Energy_(25,45,7)	0.401	0.647	96.392
GLCM_Energy_(25,135,4)	0.418	0.634	97.026
GLCM_Information_Measure_Corr1_(25,0,1)	0.618	0.595	97.621
GLCM_Information_Measure_Corr2_(25,315,1)	0.412	0.536	98.157
GLCM_Inverse_Diff_Moment_Norm_(25,135,4)	0.468	0.515	98.673
GLCM_Inverse_Diff_Moment_Norm_(25,225,4)	0.781	0.383	99.056
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	0.676	0.132	99.189
GLCM_Inverse_Variance_(25,225,4)	0.476	0.107	99.295
GLCM_Inverse_Variance_(25,315,4)	0.178	0.093	99.388
GLCM_Max_Probability_(25,180,1)	0.211	0.086	99.474
GLCM_Sum_Entropy_(25,180,7)	0.365	0.081	99.555
ID_Kurtosis	0.313	0.064	99.620
ID_Local_Range_Std	0.121	0.045	99.665
IH_Percentile_Area_(30)	0.371	0.043	99.708

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 11. The 20 radiomics features selected by the Mann-Whitney U test for dimensionality reduction to construct the lesion-based risk score using radiomic features only.**

<b>Radiomic feature</b>	<b>AUC</b>	<b>95% CI</b>	<b>P</b>
Shape_Convex_Hull_Volume	0.715	0.618-0.757	0.000
GOH_Percentile_(15)	0.837	0.768-0.954	0.000
GOH_Percentile_(50)	0.669	0.607-0.779	0.172
GLCM_Cluster_Prominence_(25,333,4)	0.499	0.401-0.681	0.305
GLCM_Cluster_Prominence_(25,315,4)	0.480	0.396-0.557	0.172
GLCM_Cluster_Shade_(25,315,7)	0.441	0.348-0.567	0.000
GLCM_Correlation_(25,0,1)	0.809	0.731-0.898	0.001
GLCM_Contrast_(25,0,1)	0.884	0.715-0.918	0.000
GLCM_Contrast_(25,270,4)	0.713	0.647-0.862	0.003
GLCM_Difference_Entropy_(25,333,1)	0.837	0.778-0.925	0.000
GLCM_Dissimilarity_(25,333,7)	0.493	0.418-0.553	0.551
GLCM_Energy_(25,135,4)	0.511	0.449-0.654	0.106
GLCM_Energy_(25,225,4)	0.474	0.388-0.591	0.493
GLCM_Information_Measure_Corr1_(25,0,1)	0.570	0.469-0.670	0.001
GLCM_Information_Measure_Corr2_(25,315,1)	0.529	0.844-0.635	0.000
GLCM_Inverse_Diff_Moment_Norm_(25,90,7)	0.542	0.684-0.627	0.000
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	0.712	0.667-0.749	0.001
GLCM_Sum_Entropy_(25,180,7)	0.732	0.675-0.138	0.000
ID_Kurtosis	0.714	0.677-0.845	0.000
ID_Local_Range_Std	0.932	0.889-0.989	0.305

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 12. The 10 radiomic features selected by the LASSO for dimensionality reduction to construct the lesion-based risk score using radiomic features only.**

<b>Radiomic feature</b>	<b>Lambda</b>
Intercept	-5.023
Shape_Convex_Hull_Volume	0.006
GOH_Percentile_(15)	0.011
GLCM_Correlation_(25,0,1)	-3.858
GLCM_Difference_Entropy_(25,333,1)	1.916
GLCM_Dissimilarity_(25,333,7)	-0.080
GLCM_Information_Measure_Corr1_(25,0,1)	0.595
GLCM_Inverse_Diff_Moment_Norm_(25,90,7)	0.498
GLCM_Inverse_Diff_Moment_Norm_(25,270,7)	0.001
ID_Kurtosis	0.035
ID_Local_Range_Std	0.004

The values in the brackets represent the parameters used to calculate that particular radiomic feature.

**Supplementary Table 13. The 3 radiomic features used by the multivariable logistic regression to construct the lesion-based risk score using radiomic features only.**

	Coefficient	Standard error	Wald Z	Pr(> Z )	AUC	95% CI
Intercept	-55.389	62.895	-0.976	0.413	-	-
GLCM_Correlation_(25, 0, 1)	-6.769	5.811	-5.023	<0.0001	0.718	0.665-0.874
ID_Local_Range_Std	0.033	0.002	3112	<0.0001	0.753	0.633-0.819
GOH_Percentile_(15)	0.136	0.052	4.153	0.002	0.798	0.771-0.804

**Supplementary Table 14. The diagnostic performance of individual 9 constituent SVMs.**

9 Individual constituent SVM	Validation dataset	Precision	Recall	F1	AUC	95% CI	P value
SUB-SVM1	G1	1.000	0.846	0.917	0.909	0.895-0.923	
SUB-SVM2	G2	0.952	1.000	0.976	0.963	0.934-0.978	
SUB-SVM3	G3	1.000	0.833	0.909	0.9	0.892-0.912	
SUB-SVM4	G4	1.000	0.871	0.931	0.889	0.878-0.892	
SUB-SVM5	G5	1.000	0.839	0.912	0.96	0.943-0.965	
SUB-SVM6	G6	1.000	0.867	0.929	0.937	0.923-0.945	
SUB-SVM7	G7	1.000	0.893	0.943	0.944	0.954-0.962	
SUB-SVM8	G8	1.000	0.906	0.951	0.933	0.925-0.943	
SUB-SVM9	G9	1.000	0.879	0.935	0.929	0.904-0.937	
SUB-SVM-average	AVERAGE	0.995	0.882	0.934	0.929	0.912-0.934	0.231
SVM	TOTAL	0.962	0.979	0.970	0.882	0.823-0.915	
Model-G1	TOTAL	1.000	0.862	0.926	0.920	0.907-0.933	
Model-G2	TOTAL	1.000	0.877	0.935	0.930	0.918-0.943	
Model-G3	TOTAL	1.000	0.867	0.929	0.924	0.913-0.938	
Model-G4	TOTAL	1.000	0.868	0.930	0.924	0.912-0.937	
Model-G5	TOTAL	1.000	0.872	0.932	0.927	0.914-0.939	
Model-G6	TOTAL	1.000	0.873	0.932	0.928	0.915-0.940	
Model-G7	TOTAL	1.000	0.873	0.932	0.928	0.915-0.940	
Model-G8	TOTAL	1.000	0.869	0.930	0.925	0.912-0.938	
Model-G9	TOTAL	1.000	0.875	0.933	0.923	0.906-0.940	
Model-G-average	AVERAGE	1.000	0.871	0.931	0.925	0.912-0.939	0.226
WSVM	TOTAL	0.999	0.884	0.968	0.958	0.943-0.967	0.000

The COVID-19 lesion groups was randomly decomposed into 9 partitions, and all the non-COVID-19 lesions were combined with each partition of COVID-19 lesions to form an individual subset (i.e., group1 (G1) –group9 (G9)).

P values: P = 0.231 represents that there are no significant differences between SUB-SVM<sub>i</sub> (i=1-9); P = 0.226 represents that there are no significant differences between Model-G<sub>i</sub> (i=1-9); P = 0.000 represents that there are significant differences between SUB-SVM-average, Model-G-average and WSVM.

Abbreviations: WSVM, weighted support vector machine.

**Supplementary Table 15. The comparison of the variation of contours by different radiologists and its impact on the calculation of COVID-19 risk score using radiomic feature only (patient based analysis).**

**Supplementary Table 16. The comparison of the variation of contours by different radiologists and its impact on the calculation of COVID-19 risk score using radiomic feature only (lesion based analysis).**