

Received:  
18 August 2020

Revised:  
03 December 2020

Accepted:  
26 January 2021

© 2021 The Authors. Published by the British Institute of Radiology under the terms of the Creative Commons Attribution 4.0 Unported License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Cite this article as:

Zhang B, Ni-jia-Ti M, Yan R, An N, Chen L, Liu S, et al. CT-based radiomics for predicting the rapid progression of coronavirus disease 2019 (COVID-19) pneumonia lesions. *Br J Radiol* 2021; **94**: 20201007.

## SHORT COMMUNICATION

# CT-based radiomics for predicting the rapid progression of coronavirus disease 2019 (COVID-19) pneumonia lesions

**<sup>1</sup>BIN ZHANG, <sup>2</sup>MA-YI-DI-LI NI-JIA-TI, <sup>1</sup>RUIKE YAN, <sup>3</sup>NAN AN, <sup>1</sup>LV CHEN, <sup>1</sup>SHUYI LIU, <sup>1</sup>LUYAN CHEN, <sup>1</sup>QIUYING CHEN, <sup>1</sup>MINMIN LI, <sup>1</sup>ZHUOZHI CHEN, <sup>1</sup>JINGJING YOU, <sup>4</sup>YUHAO DONG, <sup>5,6</sup>ZHIYUAN XIONG and <sup>1</sup>SHUIXING ZHANG**

<sup>1</sup>Department of Radiology, The First Affiliated Hospital of Jinan University, Guangzhou, China

<sup>2</sup>Department of Radiology, The First People's Hospital of Kashi area, Kashgar, China

<sup>3</sup>Yizhun Medical AI Co., Ltd, Beijing, China

<sup>4</sup>Department of Catheterization Lab, Guangdong Cardiovascular Institute, Guangdong Provincial Key Laboratory of South China Structural Heart Disease, Guangdong Provincial People's Hospital/Guangdong Academy of Medical Sciences, Guangzhou, China

<sup>5</sup>Jinan University, Guangzhou, China

<sup>6</sup>Department of Chemical and Bio-molecular Engineering, The University of Melbourne, Melbourne, Australia

Address correspondence to: Pro Shuixing Zhang

E-mail: [shui7515@126.com](mailto:shui7515@126.com)

**Objectives:** To develop and validate a radiomic model to predict the rapid progression (defined as volume growth of pneumonia lesions > 50% within seven days) in patients with coronavirus disease 2019 (COVID-19).

**Methods:** Patients with laboratory-confirmed COVID-19 who underwent longitudinal chest CT between January 01 and February 18, 2020 were included. A total of 1316 radiomic features were extracted from the lung parenchyma window for each CT. The least absolute shrinkage and selection operator (LASSO), Relief, Las Vegas Wrapper (LVW), L1-norm-Support Vector Machine (L1-norm-SVM), and recursive feature elimination (RFE) were applied to select the features that associated with rapid progression. Four machine learning classifiers were used for modeling, including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Decision Tree (DT). Accordingly, 20 radiomic models were developed on the basis of 296 CT scans and validated in 74 CT scans. Model performance was determined by the receiver operating characteristic curve.

**Results:** A total of 107 patients (median age, 49.0 years, interquartile range, 35–54) were evaluated. The patients underwent a total of 370 chest CT scans with a median interval of 4 days (interquartile range, 3–5 days). The combination methods of L1-norm SVM and SVM with 17 radiomic features yielded the highest performance in predicting the likelihood of rapid progression of pneumonia lesions on next CT scan, with an AUC of 0.857 (95% CI: 0.766–0.947), sensitivity of 87.5%, and specificity of 70.7%.

**Conclusions:** Our radiomic model based on longitudinal chest CT data could predict the rapid progression of pneumonia lesions, which may facilitate the CT follow-up intervals and reduce the radiation.

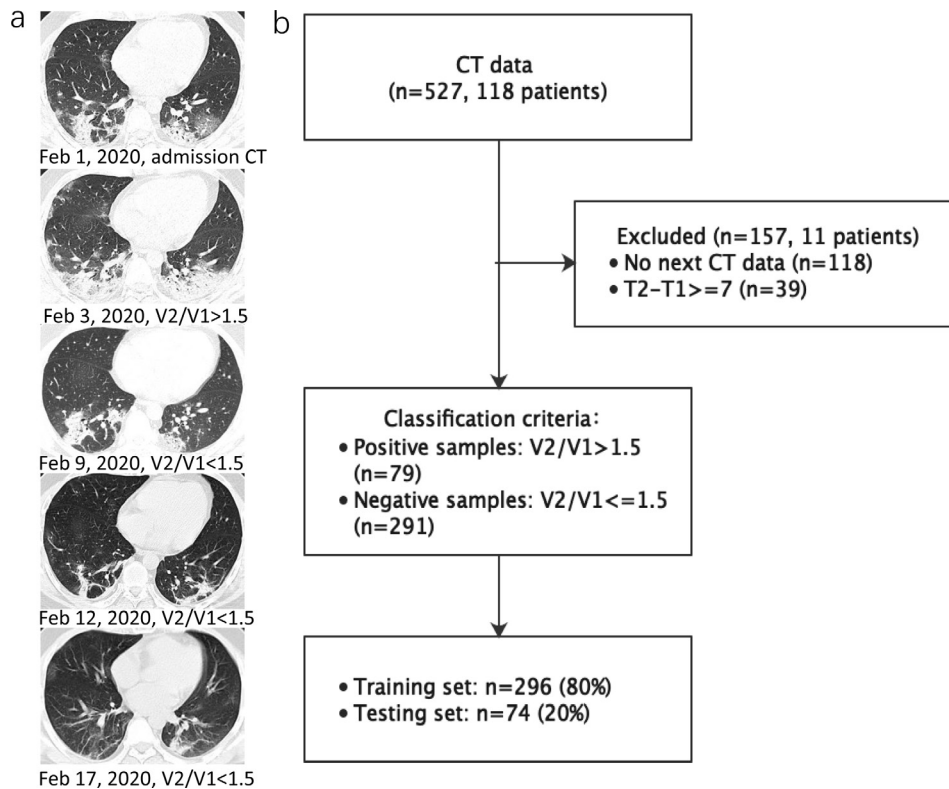
**Advances in knowledge:** Radiomic features extracted from the current chest CT have potential in predicting the likelihood of rapid progression of pneumonia lesions on the next chest CT, which would improve clinical decision-making regarding timely treatment.

## INTRODUCTION

The rapid spread of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as a potentially fatal disease is a major and urgent threat to global health.<sup>1</sup> As of August 14, 2020, there are more than 21.05 million confirmed cases by the World Health Organization (WHO) with 752,378 deaths.<sup>2</sup> Since the outbreak of COVID-19, chest CT plays an indispensable role in the detection, diagnosis, and follow-up of COVID-19 pneumonia.<sup>3</sup> Chest CT not only

presents the clinical course of COVID-19 infection and the disease severity but also predicts the poor outcomes of patients.<sup>4–6</sup> However, multiple CT scans in short time during the COVID-19 pandemic arouses great concern about the radiation burden of the patients and health-care workers. It is widely accepted that ionizing radiation increases the lifetime likelihood of developing cancer.<sup>7</sup> Some previous studies have tried a low-dose chest CT scan in the diagnosis of COVID-19 pneumonia to reduce radiation dose.<sup>8–12</sup> However, low-dose CT may miss some key

Figure 1. (a) Representative follow-up chest CT images in a male patient (29-year-old) with COVID-19 pneumonia. (b) Flowchart of study inclusion. Note: T2 = date of next CT examination, T1 = date of this CT examination, V2 = the pneumonia volume on the next CT examination, V1 = the pneumonia volume on this CT examination.



signs of COVID-19 pneumonia compared with standard-dose CT. In this present study, for the first time, we aimed to develop a CT-based radiomic model to predict the probability of rapid progression of COVID-19 pneumonia to guide the follow-up interval of chest CT scan, which may reduce ionizing radiation dose and estimated cancer risk.

## METHODS

### Patient data

This study was approved by the institutional review board and the need for written informed consent was waived. A total of 118 COVID-19 patients from two designated hospitals were retrospectively included between January 8, 2020 and February 25, 2020. Adult patients had a laboratory-confirmed COVID-19, which was achieved by real-time reverse transcription-polymerase chain reaction (RT-PCR) assay of throat swab samples (at least two samples were taken, at least 24 h apart) for COVID-19 according to the protocol established by the WHO. The 118 CT scans from 11 patients were excluded due to no follow-up CT scans or the interval time of two adjacent CT scans  $>7$  days. Finally, a total of 370 CT scans from 107 patients were analyzed in this study. The whole dataset was randomly divided into two subsets, 80% for training and the remaining 20% for validation using 10-fold cross-validation. A representative case and the flowchart of patients and CT scans inclusion are shown in Figure 1. The distribution of number of CT scans and patients under different time interval between two adjacent CT scans is illustrated in Figure 2.

### CT image acquisitions

Patients underwent chest CT scans by CT 64 scanner (GE Medical System), Siemens Emotion 16 scanner (Siemens Healthineers; Erlangen, Germany), or ICT 128 scanner (Philips Healthcare, Netherlands). No contrast agent was administered. CT acquisition parameters of the three CT scanners were shown in Table 1.

Figure 2. Histogram of the distribution of number of CT scans and patients under different time interval between two adjacent CT scans.

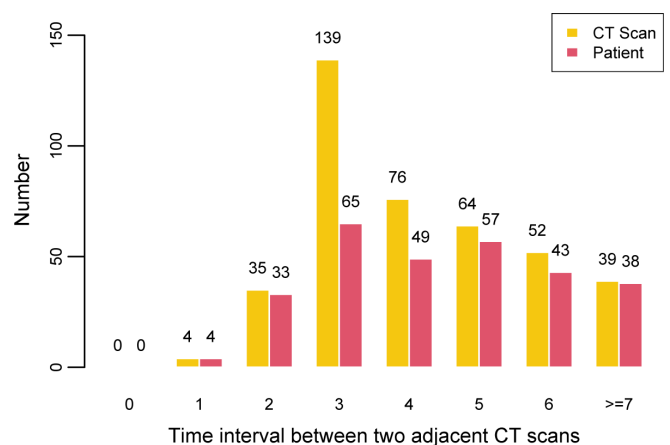


Table 1. The parameters of CT image acquisitions

| CT scanner  | Manufacturer | Tube voltage | Tube current | Pitch | Slice Thickness |
|-------------|--------------|--------------|--------------|-------|-----------------|
| 64 scanner  | GE           | 120 kV       | 260 mA       | 0.984 | 0.625 mm        |
| 16 scanner  | Siemens      | 130 kV       | automatic    | 1.5   | 1.0 or 0.6 mm   |
| 128 scanner | Philips      | 120 kV       | automatic    | 0.7   | 1.0 or 0.67 mm  |

### CT image segmentation

We used a previously trained 3D U-net ++that based on 2000 COVID-19 pneumonia cases to pre-segment the COVID-19 pneumonia lesions in this present study. Several days later, an experienced radiologist (with more than 15 years' experience in chest imaging) edited and verified the pre-segmentation results, removed false positives and segmented the missed lesions.

### Radiomic feature extraction

All raw CT images were preprocessed by 1 mm\*1 mm\*1 mm resampling. Radiomic features were extracted from the lung window (window width: 1500 Hounsfield Unit [HU], window level: 600 HU) in the Python (v.3.7.0, Beaverton, Ore; <https://www.python.org/>) by using the Pyradiomics package (v.3.0; <https://github.com/Radiomics/pyradiomics>). The parameters used in different transforms in c were presented in Table 2. A total of 1316 radiomic features were extracted under seven image types including Original, Wavelet, LoG, Square, SquareRoot, Exponential and Logarithm. The class and corresponding number of radiomics features are presented in Table 3. All radiomic features were normalized by min\_max.

### Radiomic feature selection

Considering the high-dimensional radiomic features may contain redundant information, five feature selectors including the least absolute shrinkage and selection operator (LASSO),<sup>13</sup> Relief,<sup>14</sup> Las Vegas Wrapper (LVW),<sup>15</sup> L1-norm-Support Vector Machine (L1-norm-SVM),<sup>16</sup> and recursive feature elimination (RFE)<sup>17</sup> were used to reduce the dimensions of the features before the machine learning was applied to train the models.

### Machine-learning-based radiomic model construction

Rapid progression of pneumonia lesions meant volume growth >50% within seven days, which was calculated as the ratio of the pneumonia volume on the next CT scan (V2) to the pneumonia volume on the current CT scan (V1). The threshold of 50% was identified according to the COVID-19 guidelines (trial version 6) released by the National Health Commission of China.<sup>18</sup> V2/V1 >1.5 indicates that the current CT scan is a positive sample, otherwise, a negative sample. For unbiased estimates of diagnostic accuracy, our dataset was randomly split into training and validation datasets with a ratio of 4:1. The proportions of positive and negative samples in training and testing datasets were the

same when splitting the dataset. Four common machine learning algorithms including Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR) and Decision Tree (DT) were applied to predict the occurrence of rapid progression. To select the optimal model and hyperparameters for each model, we conducted 10-fold cross-validation on each training dataset. The hyperparameters of 20 combination models were showed in Supplementary Material 1. The model with the highest area under the curve (AUC) is considered to be the optimal model. All models building was performed in the Python environment (v.3.7.0, <https://www.python.org/>) by using the Scikit-learn package (v.0.23.1; <https://scikit-learn.org/>).

### Statistical analysis

Categorical variables were expressed as counts and percentage, while continuous variables are shown as median and interquartile range. All the statistical analyses were performed using Python, v.3.7.0 (Beaverton, Ore; <https://www.python.org/>). The packages were used as follows: "mlr" for LR, "randomForest" for RF, and "e1071" for SVM. Receiver operating characteristic curve (ROC) analyses were performed to evaluate the performance of different models to predict the occurrence of rapid progression. The AUC comparison of different models used Delong test.<sup>19</sup> A  $p < 0.05$  was considered significant.

## RESULTS

A total of 107 patients with COVID-19 were included in this study, with median age of 49.0 years (interquartile range, 35–54) and 60 (56.6%) were males. 11 (10.3%) had hypertension, 5 (4.7%) had diabetes, 3 (2.8%) had coronary heart disease, 2 (1.9%) had chronic liver disease, 2 (1.9%) had chronic lung disease, and 5 (4.7%) had previous surgery. The patients underwent a total of 370 chest CT scans with a median interval of 4 days (interquartile range, 3–5 days). The median time from the initial to the last CT scan was 15 days (interquartile range, 12–22).

After selection by LASSO, Relief, LVW, L1-norm SVM, and RFE, 11, 18, 2–24, 17, and 44 features were identified respectively. Among the 20 combined models, L1-norm SVM +SVM with 17 radiomic features achieved the optimal predictive performance, with an AUC of 0.885 (95% CI: 0.845–0.924), sensitivity of 93.7%, specificity of 63.5%, and accuracy of 69.9% in the training dataset and an AUC of 0.857 (95% CI: 0.766–0.947), sensitivity of 87.5%, specificity of 70.7%, and accuracy of 74.3%

Table 2. The parameters used in different transforms in Pyradiomics

| Transform  | Original | Wavelet | LoG         | Square  | SquareRoot | Exponential | Logarithm |
|------------|----------|---------|-------------|---------|------------|-------------|-----------|
| Parameters | default  | default | sigma = 3.0 | default | default    | default     | default   |

Table 3. The number of radiomics features in different categories under different image types

| Image type Feature type | Original | Wavelet | LoG | Square | SquareRoot | Exponential | Logarithm |
|-------------------------|----------|---------|-----|--------|------------|-------------|-----------|
| Shape-based features    | 14       | –       | –   | –      | –          | –           | –         |
| First-order statistics  | 18       | 144     | 18  | 18     | 18         | 18          | 18        |
| GLCM                    | 24       | 192     | 24  | 24     | 24         | 24          | 24        |
| GLDM                    | 14       | 112     | 14  | 14     | 14         | 14          | 14        |
| GLRLM                   | 16       | 128     | 16  | 16     | 16         | 16          | 16        |
| GLSZM                   | 16       | 128     | 16  | 16     | 16         | 16          | 16        |
| NGTDM                   | 5        | 40      | 5   | 5      | 5          | 5           | 5         |

GLCM, Gray-level co-occurrence matrix; GLDM, Gray-level dependence matrix; GLRLM, Gray-level run length matrix; GLSZM, Gray-level size zone matrix; NGTDM, Neighboring gray tone difference matrix.

in the internal-validation dataset (Figure 3, Table 4) The AUC of L1-norm SVM +SVM was higher than that of all classifiers with the exception of LASSO + SVM (0.844), Relief + LR (0.842), L1-norm SVM +LR (0.851), and RFE + LR (0.834) (Delong test:  $p = 0.402, 0.218, 0.725,$  and  $0.271,$  respectively (Supplementary Table 1). The radiomic risk signature based on three first-order statistics and 14 texture features is illustrated as follows:

Radiomic risk signature =

$4.35 * \text{original\_glcm\_Imc2}$

$+1.56 * \text{wavelet\_LHL\_firstorder\_TotalEnergy}$

$+1.58 * \text{wavelet\_LHL\_gldm\_LargeDependenceLowGrayLevelEmphasis}$

$+8.42 * \text{wavelet\_LHL\_glrlm\_GrayLevelNonUniformityNormalized}$

$-1.25 * \text{wavelet\_LHL\_glrlm\_RunEntropy}$

$+1.81 * \text{wavelet\_LHL\_glszm\_GrayLevelNonUniformityNormalized}$

$-8.60 * \text{wavelet\_LHH\_glszm\_ZoneVariance}$

$-1.30 * \text{wavelet\_HHL\_glrlm\_LongRunHighGrayLevelEmphasis}$

$+2.46 * \text{wavelet\_HHH\_firstorder\_Kurtosis}$

$+0.82 * \text{wavelet\_HHH\_gldm\_DependenceEntropy}$

$-2.01 * \text{wavelet\_HHH\_glszm\_GrayLevelNonUniformityNormalized}$

$+3.54 * \text{wavelet\_LLL\_glszm\_GrayLevelVariance}$

$-8.49 * \text{square\_glcm\_JointEnergy}$

$-6.82 * \text{squareroot\_glcm\_Imc2}$

$-1.07 * \text{exponential\_glrlm\_RunVariance}$

Figure 3. The ROC curves of the training dataset and testing dataset with the optimal model. The left panel shows the mean ROC curve and the 95% CI for the training dataset (a). The right panel shows the mean ROC curve and the 95% CI for the validation dataset (b).

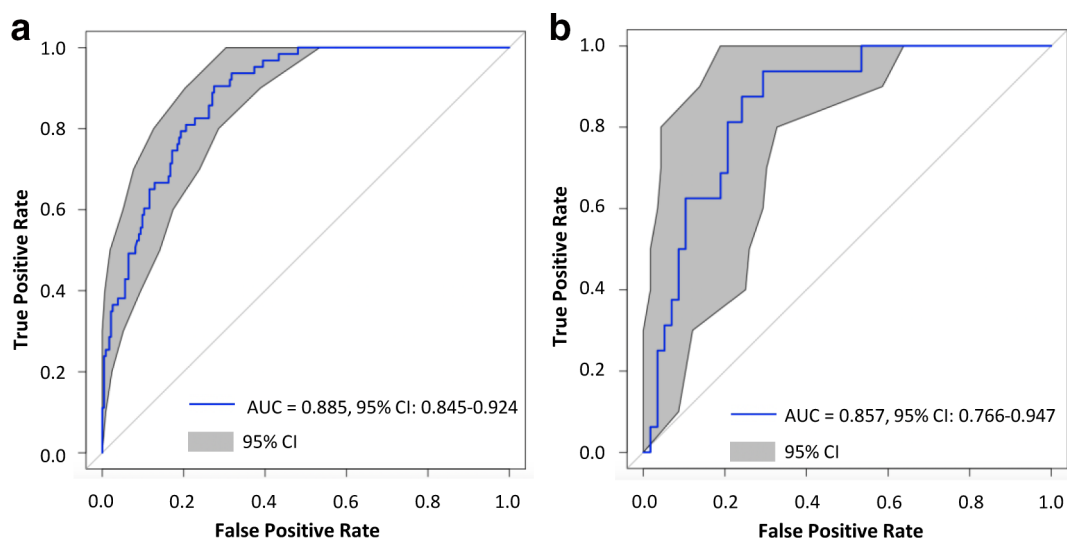


Table 4. The performance of 20 combinations of machine learning methods in predicting the occurrence of rapid progression in patients with COVID-19

| Model           | No. of features | Train_ ACC | Train_ SN | Train_ SP | Train_ PPV | Train_ NPV | Train_ AUC (95% CI)    | Test_ ACC | Test_ SN | Test_ SP | Test_ PPV | Test_ NPV | Test_ AUC (95% CI)     |
|-----------------|-----------------|------------|-----------|-----------|------------|------------|------------------------|-----------|----------|----------|-----------|-----------|------------------------|
| LASSO + SVM     | 11              | 69.6%      | 92.1%     | 63.5%     | 40.6%      | 96.7%      | 0.854<br>(0.809–0.898) | 71.6%     | 87.5%    | 67.2%    | 42.4%     | 95.1%     | 0.844<br>(0.750–0.938) |
| LASSO + LR      | 11              | 75.7%      | 81.0%     | 74.3%     | 46.0%      | 93.5%      | 0.860<br>(0.817–0.904) | 77.0%     | 62.5%    | 81.0%    | 47.6%     | 88.7%     | 0.825<br>(0.724–0.927) |
| LASSO + DT      | 11              | 98.7%      | 100.0%    | 98.3%     | 94.0%      | 100.0%     | 1.000<br>(0.999–1.000) | 75.7%     | 43.8%    | 84.5%    | 43.8%     | 84.5%     | 0.645<br>(0.510–0.780) |
| LASSO + RF      | 11              | 97.0%      | 87.3%     | 99.6%     | 98.2%      | 96.7%      | 0.998<br>(0.996–1.000) | 75.7%     | 0        | 96.6%    | 0         | 77.8%     | 0.736<br>(0.611–0.860) |
| Relief + SVM    | 18              | 65.2%      | 73.0%     | 63.1%     | 34.9%      | 89.6%      | 0.777<br>(0.717–0.837) | 71.6%     | 81.3%    | 69.0%    | 41.9%     | 93.0%     | 0.802<br>(0.701–0.903) |
| Relief + LR     | 18              | 73.0%      | 81.0%     | 70.8%     | 42.9%      | 93.2%      | 0.831<br>(0.779–0.883) | 77.0%     | 68.8%    | 79.3%    | 47.8%     | 90.2%     | 0.842<br>(0.749–0.935) |
| Relief + DT     | 18              | 100.0%     | 100.0%    | 100.0%    | 100.0%     | 100.0%     | 1.000<br>(1.000–1.000) | 73.0%     | 12.5%    | 89.7%    | 25.0%     | 78.8%     | 0.511<br>(0.418–0.603) |
| Relief + RF     | 18              | 90.9%      | 60.3%     | 99.1%     | 95.0%      | 90.2%      | 0.920<br>(0.878–0.962) | 81.1%     | 12.5%    | 100.0%   | 100.0%    | 80.6%     | 0.713<br>(0.577–0.850) |
| LVW + SVM       | 2               | 60.1%      | 74.6%     | 56.2%     | 31.5%      | 89.1%      | 0.758<br>(0.696–0.820) | 54.1%     | 50.0%    | 55.2%    | 23.5%     | 80.0%     | 0.642<br>(0.492–0.792) |
| LVW + LR        | 5               | 68.2%      | 77.8%     | 65.7%     | 38.0%      | 91.6%      | 0.789<br>(0.732–0.845) | 71.6%     | 75.0%    | 70.7%    | 41.4%     | 91.1%     | 0.780<br>(0.669–0.892) |
| LVW + DT        | 24              | 99.0%      | 100.0%    | 98.7%     | 95.5%      | 100.0%     | 0.999<br>(0.997–1.000) | 68.9%     | 31.3%    | 79.3%    | 29.4%     | 80.7%     | 0.554<br>(0.424–0.685) |
| LVW + RF        | 6               | 98.0%      | 90.5%     | 100.0%    | 100.0%     | 97.5%      | 0.999<br>(0.997–1.000) | 77.0%     | 12.5%    | 94.8%    | 40.0%     | 79.7%     | 0.671<br>(0.524–0.818) |
| L1-norm-SVM+SVM | 17              | 69.9%      | 93.7%     | 63.5%     | 41.0%      | 97.4%      | 0.885<br>(0.845–0.924) | 74.3%     | 87.5%    | 70.7%    | 45.2%     | 95.4%     | 0.857<br>(0.766–0.947) |
| L1-norm-SVM+LR  | 17              | 77.7%      | 84.1%     | 76.0%     | 48.6%      | 94.7%      | 0.889<br>(0.851–0.927) | 81.1%     | 75.0%    | 82.8%    | 54.6%     | 92.3%     | 0.851<br>(0.753–0.949) |
| L1-norm-SVM+DT  | 17              | 100.0%     | 100.0%    | 100.0%    | 100.0%     | 100.0%     | 1.000<br>(1.000–1.000) | 81.1%     | 31.3%    | 94.8%    | 62.5%     | 83.3%     | 0.630<br>(0.510–0.751) |
| L1-norm-SVM+RF  | 17              | 98.7%      | 93.7%     | 100.0%    | 100.0%     | 98.3%      | 1.000<br>(0.999–1.000) | 81.1%     | 31.3%    | 94.8%    | 62.5%     | 83.3%     | 0.728<br>(0.581–0.876) |
| RFE + SVM       | 44              | 88.2%      | 100.0%    | 85.0%     | 64.3%      | 100.0%     | 0.954<br>(0.932–0.976) | 73.0%     | 62.5%    | 75.9%    | 41.7%     | 88.0%     | 0.776<br>(0.666–0.875) |
| RFE + LR        | 44              | 83.1%      | 92.1%     | 80.7%     | 56.3%      | 97.4%      | 0.930<br>(0.901–0.960) | 77.0%     | 62.5%    | 81.0%    | 47.6%     | 88.7%     | 0.834<br>(0.738–0.931) |

(Continued)

Table 4. (Continued)

| Model    | No. of features | Train_ ACC | Train_ SN | Train_ SP | Train_ PPV | Train_ NPV | Train_ AUC (95% CI) | Test_ ACC | Test_ SN | Test_ SP | Test_ PPV | Test_ NPV | Test_ AUC (95% CI)  |
|----------|-----------------|------------|-----------|-----------|------------|------------|---------------------|-----------|----------|----------|-----------|-----------|---------------------|
| RFE + DT | 44              | 100.0%     | 100.0%    | 100.0%    | 100.0%     | 100.0%     | 1.000 (1.000–1.000) | 70.3%     | 18.8%    | 84.5%    | 25.0%     | 79.0%     | 0.516 (0.407–0.626) |
| RFE + RF | 44              | 97.3%      | 87.3%     | 100.0%    | 100.0%     | 96.7%      | 0.998 (0.996–1.000) | 78.4%     | 6.3%     | 98.3%    | 50.0%     | 79.2%     | 0.658 (0.506–0.809) |

ACC, Accuracy; AUC, Area under the curve; DT, Decision tree; LASSO, Least absolute shrinkage and selection operator; LR, Logistic regression; LVW, Las Vegas wrapper; NPV, Negative predictive value; PPV, Positive predictive value; RF, Random forest; SP, Specificity; SVM, Support vector machine.

+2.92 logarithm\_firstorder\_InterquartileRange

−0.65\*logarithm\_glszm\_SmallAreaLowGrayLevelEmphasis

−1.72

## DISCUSSION

In this present study, we developed and validated radiomic models based on this chest CT scan to predict the probability of rapid progression of COVID-19 pneumonia on next CT scan. We combined five feature selection methods and four classification methods and the results showed that the combination of L1-norm SVM +SVM outperformed other combinations, yielding an AUC of 0.857 (95% CI: 0.766–0.947), sensitivity of 87.5%, specificity of 70.7%, and accuracy of 74.3%.

Radiomics is a quantitative tool for medical imaging, which enhances the existing data available to clinicians by means of advanced mathematical analysis from the field of machine learning. This novel approach has been widely used in diagnosing, staging, predicting treatment response and prognosis of cancers. Since the outbreak of COVID-19, some studies have applied CT-based radiomics to copy with this emergent infectious diseases. By high-throughput extracting huge amounts of features from chest CT images of COVID-19 pneumonia, radiomics can reflect underlying information that associates with disease heterogeneity. Fang et al developed a radiomic nomogram with high performance in differentiating COVID-19 from other types of viral pneumonia.<sup>20</sup> Recent studies also used radiomics to diagnose and predict the outcomes of COVID-19. For example, researchers proposed a non-invasive and quantitative radiomic model using CT to predict poor outcomes in advance among COVID-19 patients.<sup>21–23</sup> Wei et al<sup>24</sup> and Tang et al<sup>25</sup> found that CT texture features could distinguish non-severe and severe COVID-19 patients. Fu et al<sup>26</sup> used radiomics and machine learning to predict the progressive from stable COVID-19 infection in the early stage. Yue et al<sup>27</sup> developed and tested machine learning-based CT radiomics models for predicting hospital stay in patients with COVID-19 pneumonia. Cai et al<sup>28</sup> built a model based on chest CT radiomic features and clinical characteristics to predict RT-PCR turning negative during clinical treatment. Current evidence demonstrated that radiomics has potential in the clinical management of COVID-19. Unlike previous studies, this study provided a radiomic tool to predict COVID-19 patients who had high-risk of rapid progression of pneumonia within seven days and the results showed promising.

The performance of a radiomic model would be affected by the each step of radiomic workflow. The differences in noise and resolution of CT images from the different CT systems may impact the reproducibility of radiomics, for instance, different values of radiomic features.<sup>29</sup> We used image pre-processing to reduce the bias caused by different scanners and imaging protocols. According to previous studies, we extracted three common classes of radiomic features (first-order statistics, shape-based features and texture features) by using the Pyradiomics package. Regarding the various choices of feature selection and modelling methodologies, the identification of optimal machine learning

methods for radiomic applications is a crucial step towards stable and clinically relevant clinical-decision support systems; thus, multiple machine-learning methods should be employed and compared. In this study, we chose five feature selectors and four modeling methods to identify the best combination and found that L1-norm SVM +SVM achieved the highest performance in the specific task of predicting rapid progression of COVID-19. Zhang et al<sup>30</sup> evaluated six feature selection methods and nine classifiers to predict the recurrence and distant metastasis in patients with advanced nasopharyngeal carcinoma and found that the combination methods Random Forest (RF) +RF performed the best. In this study, we identified 17 radiomic features that were most strongly related to the prediction outcome, consisting of 3 first-order statistics and 14 texture features. All are associated with image uniformity and heterogeneity. COVID-19 pneumonia lesions with high-risk of rapid progression were more heterogeneous (e.g. mixed ground-glass opacity and consolidation) than those with low-risk of rapid progression. Previous studies have showed that radiomics or texture analysis can characterize tumor phenotypes and reflect the tumor heterogeneity.<sup>31,32</sup> This study also demonstrated that radiomics features can serve as an effective biomarker of COVID-19 pneumonia by reflecting the heterogeneity of lesions.

This study also has some limitations. First, the retrospective nature of this study. Second, the clinical and laboratory variables did not be integrated into the prediction model because they were not matched with the each chest CT examinations. Third, the effect of treatment on the COVID-19 pneumonia did not be considered because there were no specific treatment of COVID-19. The drugs of COVID-19 used in clinical setting were mixed although there were recommends of guidelines. Finally, this model lacks of external validation, whose generalization needs to be tested in other institutions.

In conclusion, we proposed a CT-based radiomic model to predict the rapid progression of COVID-19 pneumonia, which may rationalize the chest CT follow-up intervals of COVID-19 patients and would benefit the clinical management of COVID-19 patients.

#### ACKNOWLEDGEMENTS

Thanks to all the medical workers for their fighting against the COVID-19, and to the people of the country and the world for their contributions to this campaign.

#### REFERENCES

- Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC, et al. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19). *JAMA* 2020; **324**: 782. doi: <https://doi.org/10.1001/jama.2020.12839>
- World Health Organization. Coronavirus disease 2019 (COVID-19) situation report –207. 2020. Available from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- Liu J, Yu H, Zhang S. The indispensable role of chest CT in the detection of coronavirus disease 2019 (COVID-19). *Eur J Nucl Med Mol Imaging* 2020; **47**: 1638–9. doi: <https://doi.org/10.1007/s00259-020-04795-x>
- Akl EA, Blažić I, Yaacoub S, Frija G, Chou R, Appiah JA, et al. Use of chest imaging in the diagnosis and management of COVID-19: a who rapid advice guide. *Radiology* 2021; **298**: E63–9. doi: <https://doi.org/10.1148/radiol.2020203173>
- Leonardi A, Scipione R, Alfieri G, Petrillo R, Dolciami M, Ciccarelli F, et al. Role of computed tomography in predicting critical disease in patients with covid-19 pneumonia: a retrospective study using a semiautomatic quantitative method. *Eur J Radiol* 2020; **130**: 109202. doi: <https://doi.org/10.1016/j.ejrad.2020.109202>
- Mirza-Aghazadeh-Attari M, Zarrintan A, Nezami N, Mohammadi A, Zarrintan A, Mohebbi I, et al. Predictors of coronavirus disease 19 (COVID-19) pneumonitis outcome based on computed tomography (CT) imaging obtained prior to hospitalization: a retrospective study. *Emerg Radiol* 2020; **27**: 653–61. doi: <https://doi.org/10.1007/s10140-020-01833-x>
- Power SP, Moloney F, Twomey M, James K, O'Connor OJ, Maher MM. Computed tomography and patient risk: facts, perceptions and uncertainties. *World J Radiol* 2016; **8**: 902–15. doi: <https://doi.org/10.4329/wjr.v8.i12.902>
- Tabatabaei SMH, Talari H, Gholamrezanezhad A, Farhood B, Rahimi H, Razzaghi R, et al. A low-dose chest CT protocol for the diagnosis of COVID-19 pneumonia: a prospective study. *Emerg Radiol* 2020; **27**: 607–15. doi: <https://doi.org/10.1007/s10140-020-01838-6>
- Schulze-Hagen M, Hübel C, Meier-Schroers M, Yüksel C, Sander A, Sähn M, et al. Low-Dose chest CT for the diagnosis of COVID-19—A systematic, prospective comparison with PCR. *Dtsch Arztebl Int* 2020; **117**(22-23): 389–95. doi: <https://doi.org/10.3238/arztebl.2020.0389>
- Marando M, Tamburello A, Gianella P. False-Negative nasopharyngeal swab RT-PCR assays in typical COVID-19: role of Ultra-low-dose chest CT and bronchoscopy in diagnosis. *Eur J Case Rep Intern Med* 2020; **7**: 001680. doi: [https://doi.org/10.12890/2020\\_001680](https://doi.org/10.12890/2020_001680)
- Steuwe A, Rademacher C, Valentin B, Köhler M-H, Appel E, Keitel V, et al. Dose-optimised chest computed tomography for diagnosis of coronavirus disease 2019 (COVID-19) – evaluation of image quality and diagnostic impact. *J Radiol. Prot.* 2020; **40**: 877–91. doi: <https://doi.org/10.1088/1361-6498/aba16a>
- Li J, Wang X, Huang X, Chen F, Zhang X, Liu Y, et al. Application of CareDose 4D combined with Karl 3D technology in the low dose computed tomography for the follow-up of COVID-19. *BMC Med Imaging* 2020; **20**: 56. doi: <https://doi.org/10.1186/s12880-020-00456-5>
- Tibshirani R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B* 1996; **58**: 267–88.
- Kira K, Rendell LA. The feature selection problem: traditional methods and a new algorithm. *In Aaai* 1992; **2**: 129–34.
- Liu H, Setiono R. *Feature selection and Classification-A probabilistic wrapper approach*; 1996. pp. 419–24.
- Haq AU, Li JP, Memon MH, Khan J, Malik A, Ahmad T, et al. Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access* 2019; **7**: 37718–34. doi: <https://doi.org/10.1109/ACCESS.2019.2906350>

17. Guyon I, Weston J, Barnhill S, Vapnik V, et al. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002; **46**(1-3): 389–422. doi: <https://doi.org/10.1023/A:1012487302797>
18. National Health Commission of the People's Republic of China. Guidelines for the diagnosis and treatment of novel coronavirus infection (trial version 6) (in Chinese). *Chinese Journal of Infection Control* 2020; **19**: 192–5.
19. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett* 2014; **21**: 1389–93. doi: <https://doi.org/10.1109/LSP.2014.2337313>
20. Fang X, Li X, Bian Y, Ji X, Lu J. Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2. *Eur Radiol* 2020; **30**: 6888–901. doi: <https://doi.org/10.1007/s00330-020-07032-z>
21. Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, et al. Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. *Theranostics* 2020; **10**: 7231–44. doi: <https://doi.org/10.7150/thno.46428>
22. Homayounieh F, Babaei R, Karimi Mobin H, Arru CD, Sharifian M, Mohseni I, et al. Computed tomography Radiomics can predict disease severity and outcome in coronavirus disease 2019 pneumonia. *J Comput Assist Tomogr* 2020; **44**: 640–6. doi: <https://doi.org/10.1097/RCT.0000000000001094>
23. Cai W, Liu T, Xue X, Luo G, Wang X, Shen Y, et al. CT quantification and Machine-learning models for assessment of disease severity and prognosis of COVID-19 patients. *Acad Radiol* 2020; **27**: 1665–78. doi: <https://doi.org/10.1016/j.acra.2020.09.004>
24. Wei W, Hu X-W, Cheng Q, Zhao Y-M, Ge Y-Q. Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics. *Eur Radiol* 2020; **30**: 6788–96. doi: <https://doi.org/10.1007/s00330-020-07012-3>
25. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu J, et al. Severity assessment of COVID-19 using CT image features and laboratory indices. *Phys Med Biol* 2020;08 Oct 2020. doi: <https://doi.org/10.1088/1361-6560/abbf9e>
26. Fu L, Li Y, Cheng A, Pang P, Shu Z. A novel machine Learning-derived radiomic signature of the whole lung differentiates stable from progressive COVID-19 infection: a retrospective cohort study. *J Thorac Imaging* 2020;16 Jun 2020. doi: <https://doi.org/10.1097/RTI.0000000000000544>
27. Yue H, Yu Q, Liu C, Huang Y, Jiang Z, Shao C, et al. Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. *Ann Transl Med* 2020; **8**: 859. doi: <https://doi.org/10.21037/atm-20-3026>
28. Cai Q, Du S-Y, Gao S, Huang G-L, Zhang Z, Li S, et al. A model based on CT radiomic features for predicting RT-PCR becoming negative in coronavirus disease 2019 (COVID-19) patients. *BMC Med Imaging* 2020; **20**: 118. doi: <https://doi.org/10.1186/s12880-020-00521-z>
29. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; **14**: 749–62. doi: <https://doi.org/10.1038/nrclinonc.2017.141>
30. Zhang B, He X, Ouyang F, Gu D, Dong Y, Zhang L, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett* 2017; **403**: 21–7. doi: <https://doi.org/10.1016/j.canlet.2017.06.004>
31. Gillies RJ, Kinahan PE, Hricak H, Gillies Robert J, Hedvig H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; **278**: 563–77. doi: <https://doi.org/10.1148/radiol.2015151169>
32. Limkin EJ, Sun R, Derclé L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 2017; **28**: 1191–206. doi: <https://doi.org/10.1093/annonc/mdx034>