Predicting COVID-19 Disease Progression with Chest CT Images

Hongqin Liang
First affiliated hospital to Army Medical University
Xiaoming Qiu
Huangshi Central hospital
Liqiang Zhu
The central theart command air force hospital of PLA
Lihua Chen
PLA 904 hospital
Xiaofei Hu
Southwest Hospital
Chen Liu
Southwest Hospital
Wei Chen
Southwest Hospital
Jian Wang (✉ wangjian_811@yahoo.com)
Southwest Hospital

Research article

Keywords: Coronavirus disease 2019 (COVID-19), lung, computed tomography (CT), Machine Learning, SARS-CoV-2

Posted Date: October 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-80956/v1

Abstract

Background: Some mild patients can deteriorate to moderate or severe within a week with the natural progression of COVID-19.it has been crucial to early identify those mild cases and give timely treatment . The chest computed tomography (CT) has shown to be useful to assist clinical diagnosis of COVID-19.In this study, machine learning was used to develop an early-warning CT feature model for predicting mild patients with potential malignant progression.

Methods: The total of 140 COVID-19 mild patients were collected. All patients at admission were divided into groups (alleviation group and exacerbation group) with or without malignant progression.The clinical and laboratory data at admission, the first CT, and the follow-up CT at critical stage of the two groups were compared with Chi-square test,.The CT features data (distribution, morphology,etc) were used to establish the prediction model by Fisher's linear discriminant method and Unconditional logistic regression algorithm. And the model was validated with 40 exception data.and the Area Under ROC curve (AUC) was used to evaluate the models.

Results: The model filtered out three variables of CT features including distal air bronchogram, fibrosis,and reversed halo sign. Notably, the distal air bronchograms was less common in alleviation group, while the fibrosis and reversed halo sign were more common.The sensitivity, specificity and Youden index of unconditional logistic regression were 86.1%, 92.6% and 78.7%, For the analysis of Fisher's linear discriminant, the sensitivity, specificity and Youden index were 83.3%, 94.1% and 77.4%. The generalization ability of both models were consistent with sensitivity of 95.89%, specificity of 100%, and Youden index of 83.33%.

Conclusions: The CT imaging features-based machine learning model has a high sensitivity for finding out the mild patients who are easy to deteriorate into severe/critical cases efficiently so that timely treatments came true for those patients,while largely help to relieve the medical pressure.

Background

In the middle of December 2019, an outbreak of a novel coronavirus has spread in mainland China and beyond. World Health Organization (WHO) officially named the disease as Coronavirus Disease 2019 (COVID-19) on February 12, 2020. At the same time, International committee on the Taxonomy of viruses (ICTV) named the disease as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). One of the primary manifestations for COVID-19 is pneumonia which greatly challenges the public medical system because of its high infection and mortality[1–4]. According to the Chinese epidemic data, the majority of the patients was the mild(81%)[5]. Some mild patients can deteriorate to moderate or severe within a week with the natural progression of COVID-19. It is important to identify those mild patients in order to prevent malignant progression and reduce the mortality of COVID-19. However, Most studies focused on cross-sectional descriptions, comparisons of clinical, laboratory and CT imaging findings, or analysis of

risk factors of death outcome[6–8], and less methods focused on predicting the mild COVID-19 patients with potential malignant progression.

Chest CT is a non-invasive imaging modality of COVID-19 with high speed and accuracy [9]. And the course of lesion evolution can be assessed by the distribution (subpleural, central), morphology (Ground glass opacification (GGO), mixed GGO, consolidation), distal vascular dilatation, etal [10–11]. Machine learning (ML) is largely used to develop automatic predictors in pneumonia classification but automatic predictors for COVID-19 disease progression are still in their infancy. Thus, to understand the benefits of ML in COVID-19 disease progression prediction, a study was designed to analyze multivariate heterogeneous data (clinical data and serial chest CT imaging) and to further develop an accurate and effective prediction model that combines Unconditional logistic regression and Fisher's linear discriminant analysis .Since the two learning-based methods had been widely adopted and had achieved great performance in disease classification and prediction.

Therefore, the purposes of our study are to develop CT features models to identify the mild patients who are easy to deteriorate into severe cases using Unconditional logistic regression and Fisher's linear discriminant analysis respectively.

Methods

Ethics statement

This study was conducted in accordance with the Declaration of Helsinki. Informed consent was waived due to the retrospective nature of the study and the analysis of anonymous clinical data.Institutional Review Board approval was obtained.

Data collection

A total of 151 patients were hospitalized and had confirmed COVID-19 infection in Chongqing city and Hubei province from December 21, 2019 to February 21, 2020. Patients tested positive for the nucleic acids of COVID-19 were identified as confirmed cases.All the clinical data on signs and symptoms, epidemiology (including the history of recent exposure) and underlying comorbidities as well as laboratory results were retrospectively collected from electronic medical rebands. After excluding invalid information,140 patients were in this condition,as showed in the flowchart (Fig. 1).The degree of severity of COVID-19 patients (severe vs. non-severe) at the time of admission were defined according to the American Thoracic Society guidelines for community-acquired pneumonia [12]. Only non-severe cases (blood oxygen saturation≥90%) were included in the analysis. They received treatment focused on symptomatic and respiratory support, while systemic corticosteroid therapy was not applied. Of these patients, 62 patients had two follow-up CT scans, and 78 patients had more than two. The follow-up interval was ranged from 1 day to 21 days. According to the results of clinical follow-up, the patients were divided into two groups, the Alleviation group (n = 72) and the Exacerbation group((n = 68).

CT image acquisition and analysis

All patients underwent chest non-contrast enhanced CT scans. The obtained images were reconstructed with a slice thickness of 1.5 mm. Lesions and imaging features were assessed in each lung segment of each patient.All imaging features were reviewed and evaluated by two experienced radiologists (9 and 15 years of experience in chest CT) independently blinded to the clinical information, and the discrepancy was resolved by consulting another radiologist (18 years' experience in chest CT).

The CT features of lung lesions, including the distribution characteristics (peripheral, central, single, multiple, diffuse), morphology (GGO and mixed GGO, consolidation), and associated manifestations (distal air bronchogram sign, fibrosis, reversed halo sign, etc.), before and during follow-up were evaluated.Meanwhile, the location of the lesion was considered as peripheral if it was in the outer one-third of the lung; otherwise, it was deemed as central [13].

Statistical analysis

All the statistical analysis was performed using SPSS (Version 18.0) with statistical significance set at 0.05. The continuous variables were analyzed by the t-test of two independent samples, while the categorical variables were analyzed by the chi-squared test. The discriminant model employed Unconditional logistic regression and Fisher's linear discriminant method.The discriminant equation was obtained, and the discriminant function was used to determine the machine learning. The machine learning results of stepwise logistic regression and Fisher's linear discriminant were then inputted into MedCalc 19.2, and the Area Under the Curve(AUC) of the two models were analyzed. Finally, Z test was used to compare the prognostic efficacy of the two models.

Results

Demographic, clinical and CT characteristics

140 patients with mild COVID-19 pneumonia at admission included 92 male and 48 female, age ranged from 18 to 82 (45 ± 15) years, 68 patients (68/140,48.5%) malignantly progressed to severe periods during the hospitalization, while the remaining 72 patients (72/140, 51.4%) did not. The demographic characteristics of 140 COVID-19 patients are shown in Table 1. The laboratory data and clinical manifestations including leucocytes,lymphocytes,elevated C-reactive protein level,fever,cough,etal are shown in Table 2. At the time of admission, the levels of C-reactive protein were obvious increased, the level of lymphocytes was decreased ,and the level of leucocytes was normal in most patients. Fever and cough were the most common symptoms, whereas headache and no obvious symptoms were rarely observed. Serial CT imaging features of patients with and without severe progression were summarized in Tables 3. In brief, comparing to the patients with exacerbation progression, the disease distribution of alleviation group were single and multiple mainly. fibers and reversed halo sign were common in the alleviation group while the distal air bronchogram sign was common in the exacerbation group.

Table 1
Demographic and clinical characteristics of the 140 patients.

Items	n	Percentage, %
Age	45 ± 15 (18–82)	-
Gender (Male/Female)	92/48	65.7/34.2
Exposure to Wuhan or infected patient	129	92.1
Chronic obstructive pulmonary disease	11	7.8
Hypertension	12	8.5
Diabetes	6	4.2
Cardiovascular disease	12	8.5
Digestive system diseases	5	3.5

Table 2
Laboratory data and clinical manifestations of the 140 patients.

Items	n	Percentage, %
Positive for (SARS-CoV-2) nucleic acid test byReal-time PCR	130/140	92.8
Leucocytes (× 10^9/L)		
<4	39	27.8
4–10	89	63.6
>4	12	8.6
Lymphocytes (× 10^9/L)		
>1	31	22.1
<1	109	77.9
Elevated C-reactive protein level (mg/L)	119	85.0
Fever	102	72.8
Cough	86	61.4
Sputum production	42	30.0
Fatigue weakness	39	27.8
Headache	3	2.1
Diarrhea	12	8.6
No obvious symptoms	2	1.4
Temperature (℃)		
<37.3	38	27.1
37.3–38	41	29.2
38.1–39	42	30.0
>39	19	13.6
Blood oxygen saturation, % ↓	11	7.8
Heart rate (beats per minute)	81 ± 8(64–102)	-

Table 3
Imaging findings on chest CT of patients with SARS-CoV-2.

Items	Alleviation (n = 72)	Exacerbation(n = 68)(n = 68)	χ2	P-value
Distribution				
Peripheral	71 (98.6)	56 (82.4)	10.974	0.001
Central	1 (1.4)	12 (17.6)		
Single, multiple or diffuse				
Single	3 (4.2)	10 (14.7)	5.723	0.057
Multiple	67 (93.1)	54 (79.4)		
Diffuse	2 (2.8)	4 (5.9)		
Morphology				
Mixed GGO	52 (72.2)	39 (57.4)	11.207	0.004
Consolidation	5 (6.9)	0 (0)		
GGO	15 (20.8)	29 (42.6)		
Bronchial wall thickening				
N	53 (73.6)	40 (58.8)	3.429	0.064
Y	19 (26.4)	28 (41.2)		
Distal air bronchograms				
N	41 (56.9)	26 (38.2)	4.905	0.027
Y	31 (43.1)	42 (61.8)		
Fibers				
N	12 (16.7)	64 (94.1)	84.534	0.000
Y	60 (83.3)	4 (5.9)		
Crazy paving pattern				
N	70 (97.2)	59 (86.8)	5.283	0.022
Y	2 (2.8)	9 (13.2)		
Pericardial effusion				
N	55 (76.4)	66 (97.1)	12.738	0.000

N: without manifestation; Y: with manifestation.

Items	Alleviation (n = 72)	Exacerbation(n = 68)(n = 68)	χ2	P-value
Y	17 (23.6)	2 (2.9)		
Reversed halo sign				
N	26 (36.1)	67 (98.5)	26.959	0.000
Y	46 (63.9)	1 (1.5)		
N: without manifestation; Y: with manifestation.				

Establishment of prediction model

The machine learning models were established with the significant influencing factors such as distribution, morphology, distal air bronchogram sign, etc. Finally, the model filters out three variables including distal air bronchogram sign, fibrosis, and reversed halo sign.Notably, during patient follow-up, the CT feature of distal air bronchograms was less common in alleviation group, while the fibrosis and reversed halo sign were more common(Fig. 2–4).

Unconditional logistic regression

The sensitivity, specificity and Youden index of the regression model were 86.1%, 92.6% and 78.7%, respectively. The model was constructed as follows: LogitP = 1.241 + 1.402 × distal air bronchograms − 4.220 × whether the fibrous bands formed − 3.488 × reversed halo.(Table 4)

Table 4
Unconditional logistic regression results.

Disease outcome	Model regression results [n, (%)]		Total
	Alleviation	Exacerbation	
Alleviation	62 (86.1)	10 (13.9)	72 (100)
Exacerbation	5 (7.4)	63 (92.6)	68 (100)

Fisher's linear discriminant analysis

The sensitivity, specificity and Youden index of the machine model were 83.3% (60/72), 94.1% (64/68) and 77.4%, respectively. The machine functions were as follows: P0= -4.612 + 1.146 × distal air bronchograms + 8.043 × whether the fibrous bands formed + 1.456 × reversed halo; P1= -1.504 + 2.588 × distal air bronchograms + 0.553 × whether the fibrous bands formed − 0.572 × reversed halo (Table 5).

Table 5
Fisher's linear discriminant results.

Disease outcome	Model classification results [n, (%)]		Total
	Alleviation	Exacerbation	
Alleviation	60 (83.3)	12 (16.1)	72 (100)
Exacerbation	4 (5.9)	64 (94.1)	68 (100)

Comparison of model machine ability

The machine results of stepwise logistic regression model and Fisher's linear machine were inputted into MedCalc 19.2, and the area under the curve (AUC) of the two models were 0.894 (95%CI: 0.835 ~ 0.953) and 0.887 (95%CI: 0.827 ~ 0.948), respectively(figure 5). Z test was used to compare the diagnostic efficacy of the area evaluation model under the ROC curve of the two models.The results showed that the Z value was 0.162, and the difference was not statistically significant (P = 0.871). The generalization ability of the discriminant model was consistent with that of the unconditional logistic regression, with sensitivity of 95.89%, specificity of 100% and Youden index = 83.33%.

Discussion

Assessing the progression of COVID-19 is crucial for the disease treatment and control. In the early stage, the lung stroma of COVID-19 patient was mostly invaded, which could be manifested by the thickening of interlobular septa, angioedema dilatation and GGO appearance. As the disease progresses, the alveolar structure was gradually affected by inflammation, while alveolar edema, exudation and bleeding might occur. On the CT image, lung consolidation and mixed GGO can be manifested. Parts of the lesions exhibited distal air bronchograms sign and thickening of the bronchial wall, while the remaining parts displayed other signs. Based on the clinical data published in recent literature, almost all patients with COVID-19 had characteristic CT features during the course of the disease, including angioedema dilatation sign, paving stone sign, etc[14].

The imaging manifestations of the influencing factors were successfully extracted by machine learning model during the course of COVID-19. Finally, the three objective variables, namely, fibrosis formation, distal air bronchogram sign and reversed halo sign, were incorporated into the model, which could serve as potential indicators to predict the disease outcomes. In the follow-up CT images, we found that 61.8% (42/68) of patients with exacerbation had distal air bronchogram sign. Among the alleviated cases, the CT features of fibrosis formation, reversed halo sign were observed to be 83.3% (60/72), 63.9% (46/72), respectively. The pathological mechanism of fibrosis formation is that the immune response of human body is intense or when the wall of small blood vessel is damaged by edema, the permeability of blood vessel wall is increased, the plasma and fibrin exudate, which can be interwoven into a net to limit the spread of pathogens and attenuate the lesion[15]. As for the occurrence of reversed halo, it represents a rare sign of a focal ground glass area surrounded by a complete ring of consolidation. Surgical pathology

confirmed that the central GGO was actually alveolar septal inflammation and cellular debris, and the lesions surrounding alveoli tended to be mechanical inflammation, Some literature has suggested that the lesions turn out to be benign when their center part began to be absorbed [16–18]. The bright bronchogram seen in the area of diseased lung tissue is known as air bronchogram sign, which can be considered as strong evidence of inflammatory lesions. It has been reported that distal air bronchogram sign is helpful to distinguish the lung and pleura or mediastinal lesions,Alveolar lesions can be detected by air bronchogram sign, whereas thoracic reef and mediastinal lesions display no such signs[19–20]. So from this study, we observed the presence of more distal air bronchograms in the follow-up CT images of patients with exacerbation, suggesting that the lesion is further aggravated by expanding from the septal injury to the alveoli. fibrosis and reverse - halo signs are the prediction of benign outcome.

Unconditional logistic regression and Fisher's linear discriminant analysis are very important tasks in machine learning, which can be used to automatically derive the generalized description of a given dataset from known historical data, in order to predict future events [21–22].The results of the two models are relatively satisfactory, and consequently afford greater confidence in the assessment of COVID-19. In addition, the above CT features indicate that the lesion is in the critical period, and this trend change is helpful for clinicians to judge the therapeutic effect and predict the outcome of the disease. In the single factor analysis, the model variable, such as crazy paving pattern, is associated with the evolution of COVID-19. However, in the multiple factor analysis, it can be influenced by other factors with "false association" claims, and hence the "false association" should been adjusted in the analysis.

This study has several limitations. First, there was no long-term clinical follow-up and the CT examination data of discharged patients was lacking. Hence, the severity of pulmonary fibrosis at the time of its formation and later changes needs to be further observed. Second, severe cases were not included. The prognosis of severe patients can be affected by many factors.

Conclusions

In summary, the CT imaging features-based machine learning model provides a non-invasive and easy-to-use method for the outcome prediction of COVID-19 patients. Our future work will focus on mining richer spatial information using deep learning technologies to screen the risk factors of CT and clinic.

Abbreviations

GGO Ground glass opacification

RT-PCR Real-time fluorescence reverse transcription-polymerase chain reaction

AUC Area under the curve

WHO World Health Organization

COVID-19 Coronavirus Disease 2019

ICTV International committee on the Taxonomy of viruses

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus 2

MERS-CoV Middle East respiratory syndrome coronavirus

ML Machine Learning

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee on First Affiliated Hospital to Army Medical University,PLA (No.KY2020036). Written informed consent was waived due to the retrospective nature of the study and the analysis of anonymous clinical data.

Consent for publication

Not applicable.

Availability of data and materials

The datasets included in the present study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the technology innovation and application development of Chongqing city (No.cstc2019jscx-msxmX0126 and cstc2017shmsA130037) ; and the Chongqing Key technology and application demonstration of medical imaging depth intelligent diagnostic platform (cstc2018jszx-cyztzxX0017).The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Authors' contributions

Prof.CW and WJ take responsibility of conception and design of the study as the corresponding author.

LQ designed the study, and wrote the manuscript. QM collected and analyzed datasets

from study patients.They made equal contributions as the first author.

ZQ was in charge of statistics. CH, HF and LC revised the initial manuscript draft.

All authors read and approved the final manuscript.

Acknowledgement

The authors would like to express their gratitude to Fang Wang (Huiyihuiying Medical Technology Co. Ltd) for help. We are grateful to the study participants for their involvementin the study.

References

1. Zhu N, Zhang D, Wang W. **et al**. A Novel Coronavirus from Patients with Pneumonia in China.N Engl. J Med. 2019;382(8):727–33.
2. Huang C, Wang Y, Li X. **et al**. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. China Lancet. 2020;395(10223):497–506.
3. Li Q, Guan X, Wu P. **et al**. Early Transmission Dynamics in Wuhan, **China, of Novel**. **Coronavirus-Infected Pneumonia**. N Engl J Med. 2020;382(13):1199–207.
4. **World Health Organization**.**Corona-virus disease** (**COVID-19**) **outbreak**.**Available at**:https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19.
5. Wu Z, McGoogan JM. **Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases**. **From the**. **Chinese Center for Disease Control and Prevention**. **Jama** 2020.
6. Chen N, Zhou M, Dong X. **et al**. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395:507 – 13.
7. Huang C, Wang Y, Li X. **et al**. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497–506.
8. Zhou F, Yu T, Du R. **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**(**10229**):**1054**–**1062**.
9. Guan WJ, Ni ZY, Zhong NS. **et al**.**Clinical Characteristics of Coronavirus Disease 2019 in China**. N Engl J Med.2020; **Feb 28**.
10. Xu X. **et al**. **Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2**. European journal of nuclear medicine and molecular imaging. 2020.

11. Shi H, **et al**. **Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan**, **China**: **a descriptive study**. **The Lancet Infectious Diseases**.2020.
12. Metlay JP, Waterer GW, Long AC. **et al**. Diagnosis and treatment of adults with community-acquired pneumonia: an official clinical practice guideline of the American Thoracic Society and Infectious Disease Society of America. Am J Respir Crit Care Med. 2019;200(7):45–67.
13. Pan Y, Guan H, Zhou S,**et al**. **Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan**, China.Eur Radiol.2020; **Feb 13**.
14. Shi H, Han X, Zheng C. Evolution of CT manifestations in a patient recovered from 2019 novel coronavirus (2019-nCoV) pneumonia in Wuhan, China. Radiology. 2020;295(1):20.
15. Duan YN, Jie Q. **Pre- and Posttreatment Chest CT Findings: 2019 Novel Coronavirus**. -nCoV). Pneumonia Radiology. 2020;295(1):21.
16. Churg A, Wright JL, Bilawich A. Cicatricial organising pneumonia mimicking a fibrosing interstitial pneumonia. Histopathology. 2018;72(5):846–54.
17. Artunduaga M, Rao D, Friedman J. **et al**. **Pediatric Chest Radiographic and CT Findings of Electronic Cigarette or Vaping Product Use-associated Lung Injury (EVALI)**. **Radiology.2020;Mar** 3:**192778**.
18. Sullivan T, Rana M.**The reversed halo sign and mucormycosis**. **Lancet Infect Dis**. 2019.;19(**12**):**1379**.
19. Lichtenstein D, Mezière G, Seitz J. The dynamic air bronchogram. A lung ultrasound sign of alveolar consolidation ruling out atelectasis. Chest. 2009;135(6):1421–5.
20. Qu H, Zhang W, Yang J. The value of the air bronchogram sign on CT image in the identification of different solitary pulmonary consolidation lesions. Medicine. 2018;97(35):e11985.
21. Zheng B, Zhou X, Chen J. **et al**. A modified model for preoperatively predicting malignancy of solitary pulmonary nodules: an Asia cohort study. Ann Thorac Surg. 2015;100:288–94.
22. Lu Y, Lai Z, Wong WK. **et al**.Low-rank discriminative regression learning for image machine. Neural Netw. 2020;125:245–57.

Figures

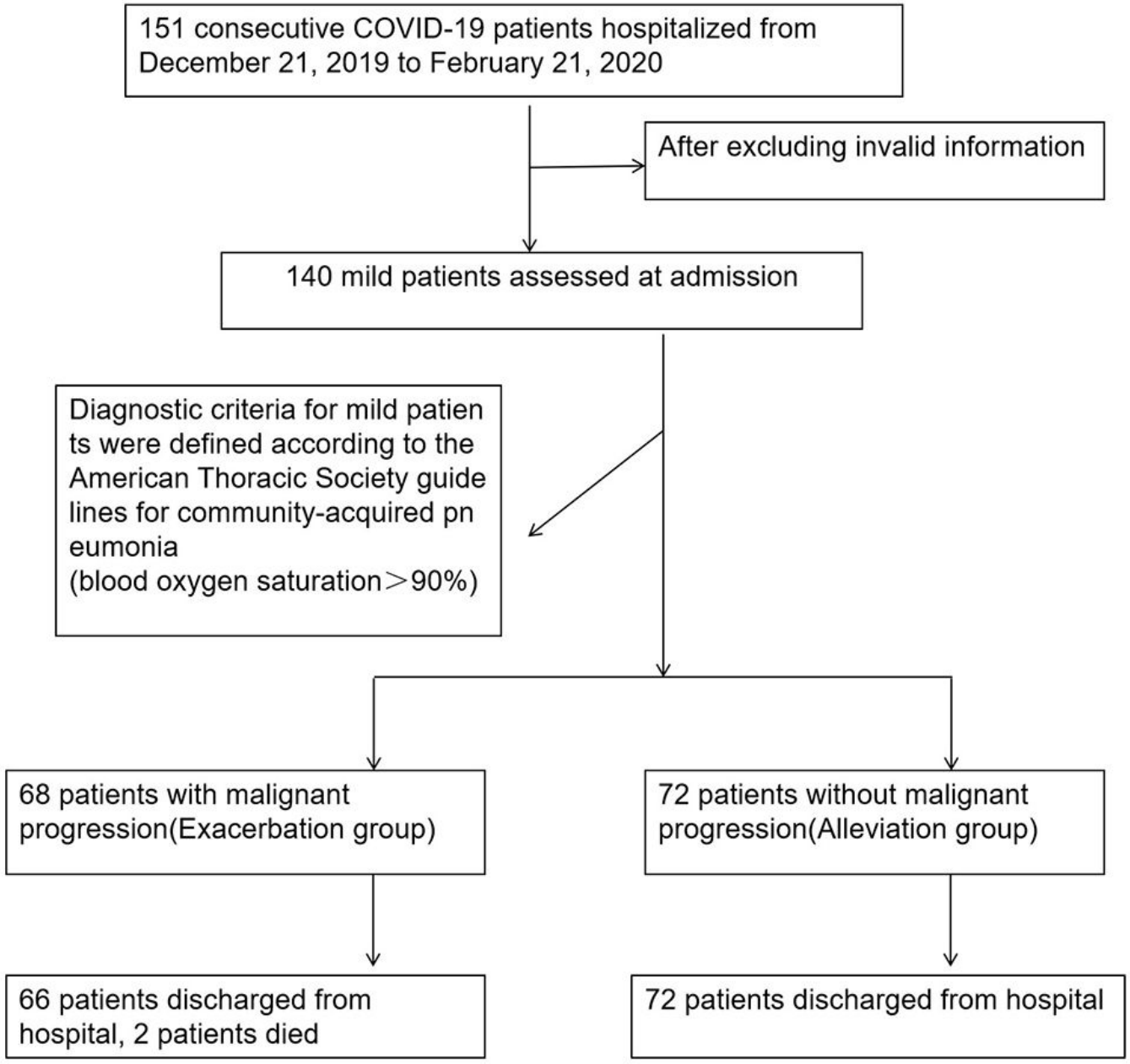

Figure 1

Flowchart of patient selection

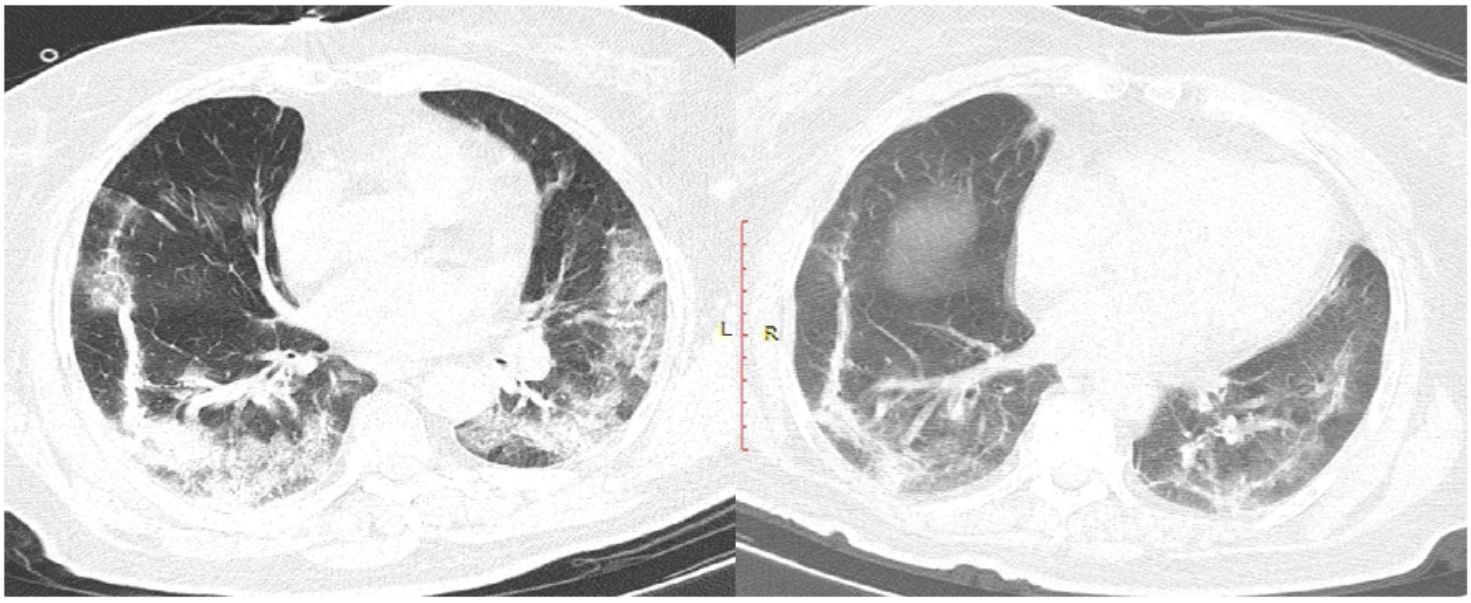

Figure 2

Chest CT of a male patient . He underwent a chest CT examination on February 1st(a), which showed subtle peripheral mixed ground glass opacity and some fibers. The CT images on February 3rd (b)demonstrated significant decreases in lesion number and density.

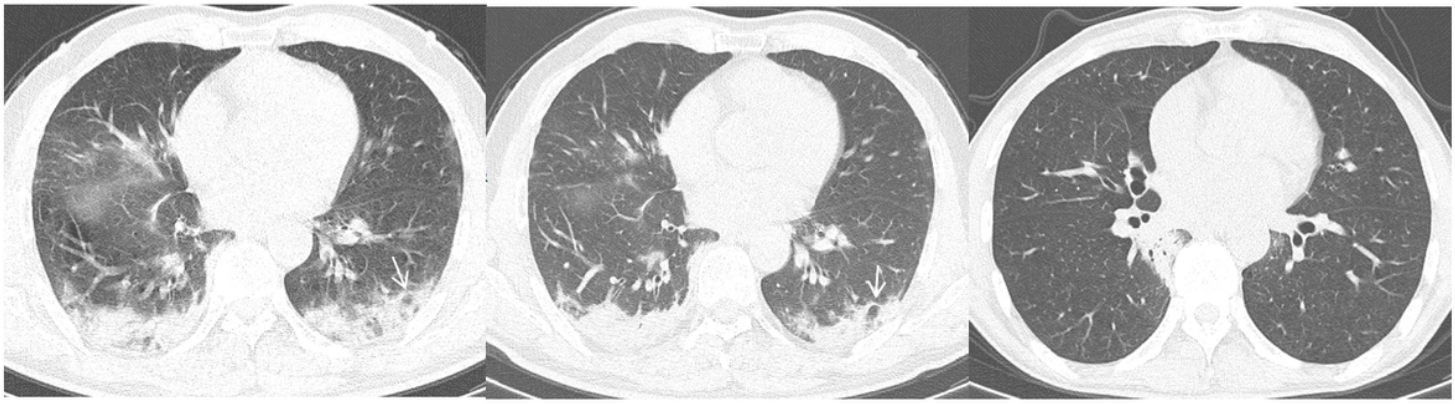

Figure 3

Chest CT of a male patient. His onset chest CT examination on January 21thshowed subtle peripheral consolidation, mix ground glass opacity and reversed halo sign(white arrow). The CT images on January 23thshowed a decrease in lesion size, especially at the left lower lobes, as well as the adsorption of the central part (areas of reversed halo sign). The results of the follow-up CT examination on February 1st indicated that the lesions were obviously absorbed.

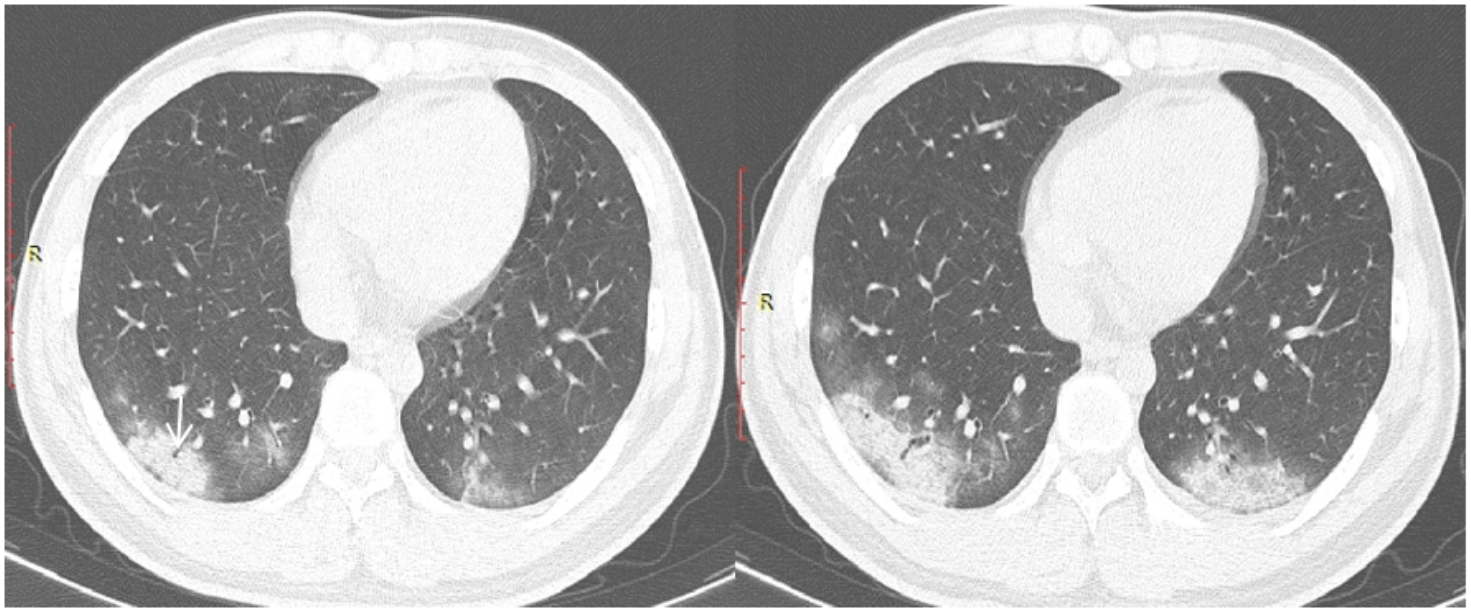

Figure 4

Chest CT of a male patient . He underwent a chest CT examination on January23th,which showed subtle peripheral mixed ground glass opacity and distal air bronchogram sign(white arrow).The CT images on January 24th demonstrated significant increases in lesion size and density.

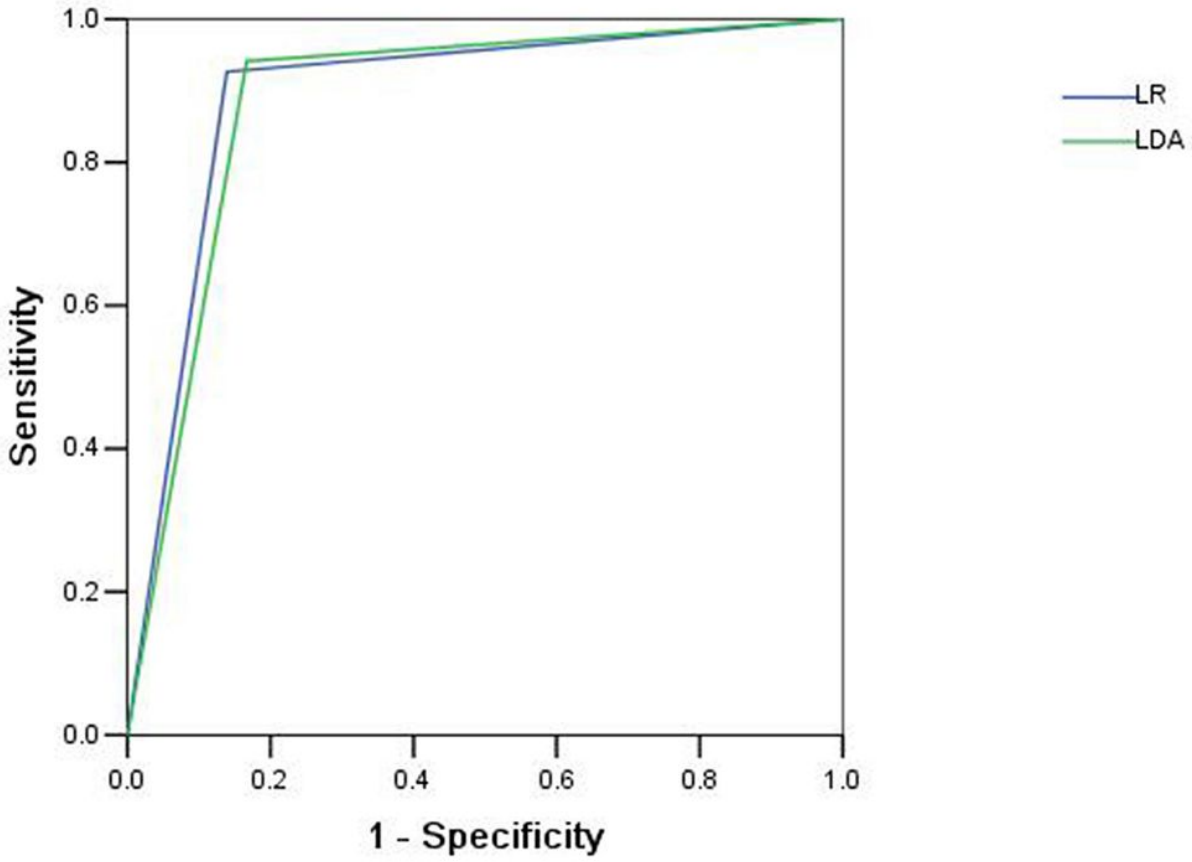

Figure 5

ROC curves of different methods on two models.