Radiology: Cardiothoracic Imaging

Serial Quantitative Chest CT Assessment of COVID-19: Deep-Learning Approach

Lu Huang, MD, PhD

Rui Han, MD

Tao Ai, MD, PhD

Pengxin Yu, MS

Han Kang, MS

Qian Tao, PhD

Liming Xia, MD, PhD

Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Jiefang Avenue 1095, 430030 Wuhan, China (L.H., T.A., L.X.); Department of Radiology, Wuhan NO.1 Hospital, Wuhan, China (R.H.); Institute of Advanced Research, Infervison, Beijing, China (P.Y., H.K.); Division of Imaging Processing, Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands (Q.T.)

Corresponding author: Liming Xia

Address: Jiefang Avenue 1095, 430030 Wuhan, China

E-mail: lmxia@tjh.tjmu.edu.cn

Funding: Wuhan Science and technology program (Grant No. 2018060401011326), Hubei Provincial Novel Pneumonia Emergency Science and Technology Project (2020FCA021) and Huazhong University of Science and Technology Novel Coronavirus Pneumonia Emergency Science and Technology Project (2020kfyXGYJ014)

Conflicts of interest: authors had nothing to disclose.

Key Results

- The quantitative CT parameter calculated by the deep learning method showed significant differences at baseline among four clinical types (all $P < 0.01$).
- Lung opacification percentage may be used to monitor disease progression and help understand the course of COVID-19.

Summary

The severity of the pulmonary manifestations of COVID-19 can be quantitatively evaluated from chest CT using a deep-learning method. There were significant differences in lung opacification percentage, as measured by the deep learning algorithm, among patients with different clinical severity. This automated tool for quantification of lung involvement may be used to monitor the disease progression and understand the temporal evolution of COVID-19.

Abbreviations

ARDS = acute respiratory distress syndrome, COVID-19 = coronavirus disease 19, GGO = ground glass opacity, HRCT = high resolution computed tomography, RT-PCR = reverse transcription-polymerase chain reaction, SARS-Cov-2 = severe acute respiratory syndrome coronavirus 2, SpO2 =pulse oxygen saturation

Abstract

Purpose: To quantitatively evaluate lung burden changes in patients with COVID-19 using serial CT scan by an automated deep learning method.

Materials and Methods: Patients with COVID-19 who underwent chest CT between 1st January 2020 and 3rd February 2020 were retrospectively evaluated. Patients were divided into mild, moderate, severe, and critical types, according to their baseline clinical, laboratory, and CT findings. CT lung opacification percentage of the whole lung and five lobes were automatically quantified by a commercial deep learning software, and compared over follow-ups CT scans. Longitudinal changes of the CT quantitative parameter were also compared among the four clinical types.

Results: A total of 126 patients with COVID-19 (age 52 years ± 15 years, 53.2% males) were evaluated, including 6 mild, 94 moderate, 20 severe and 6 critical cases. CT-derived opacification percentage was significantly different among clinical groups at baseline, gradually progressing from mild to critical type (all $P < 0.01$). Overall, the whole-lung opacification percentage significantly increased between baseline CT and 1st follow-up CT (median [interquartile range]; 3.6% [0.5%,12.1%] vs 8.7% [2.7%,21.2%], $P < 0.01$). No significant progression of the opacification percentages was noted between the 1st follow-up and 2nd follow-up CT (8.7% [2.7%,21.2%] vs 6.0% [1.9%,24.3%], P=0.655).

Conclusion: The quantification of lung opacification in COVID-19 measured on chest CT by a commercially available deep-learning-based tool was significantly different among different clinical severity groups. This approach could potentially eliminate the subjectivity in the initial assessment and follow up of pulmonary findings in COVID-19.

Introduction

SARS-CoV-2 is a novel coronavirus initially identified in Wuhan, China, which causes a respiratory pandemic disease named Coronavirus Disease 2019 (COVID-19) (1,2). Chest CT has played a pivotal diagnostic role in the assessment of patients with COVID-19 in China (3).

Recent studies reported that the possible pathological mechanism in COVID-19 is diffuse alveolar damage and inflammatory exudation, which is similar to histologic findings seen in SARS-CoV pneumonia (1,4). The pathological evolution during the course of infection in COVID-19 has not been clarified, and the disparity of such changes in patients with different clinical severities are largely unknown. Chest CT, especially high-resolution CT (HRCT), can detect small areas of ground glass opacity (GGO) (5), and, therefore, is a promising imaging tool for monitoring the disease, if radiation dose is balanced to comply with ALARA principles. It is common practice for radiologists to evaluate the pneumonia severity qualitatively or semi-quantitatively by visual scoring (6). Visual evaluation of changes between two CT scan is subjective and its validity may depend on the radiologists' experience. Quantitative analysis of the CT scans using artificial intelligence (AI) tool, in particular deep learning, could provide an automatic and objective estimation of the disease burden, facilitating and expediting imaging interpretation during the COVID-19 pandemic (7).

The purpose of the present study was to assess a quantitative CT image parameter, defined as the percentage of lung opacification (QCT-PLO), calculated automatically using a deep learning tool. We evaluated QCT-PLO in COVID-19 patients at baseline and on follow-up scans, focusing on cross-sectional and longitudinal differences in patients with different degrees of clinical severity.

Materials and Methods

The local ethical review board approved this retrospective study and waived the requirement to obtain individual informed consent.

Study Population

Patients with COVID-19 who underwent chest CT in our department from 1st January to 3rd February 2020 were enrolled in this retrospective study. Inclusion criteria were: (a) positive SARS-Cov-2 nucleic acid in double swab tests (within an interval of 2 days, real time RT-PCR), and (b) with at least 2 chest CT scans in our hospital, and (c) without confirmation of another viral infection. Exclusion criteria were: (a) patients who underwent initial chest CT in other hospitals, or (b) CT images with respiratory artifacts that could not meet the image analysis requirement, or (c) inadequate deep-learning segmentation by the segmentation algorithm based on radiologist review, as will be explained in detail later. **Figure 1** shows the enrollment flowchart.

At baseline, all patients were classified into four clinical types: mild, moderate, severe and critical type, based on the Diagnosis and Treatment Protocol of Novel Coronavirus (trial version 5th) (3) from the National Health Commission of the People's Republic of China. The classification criteria of clinical types are described in **Appendix E1**.

CT Scanning

Non-contrast enhanced chest CT examinations were performed with three CT scanners (United Imaging uCT, United Imaging Healthcare, Shanghai, China; GE Optima 660, GE Healthcare, USA; Siemens SOMATOM Definition AS+, Siemens Healthineers, Germany). The patients were scanned in supine position during inspiratory breathhold. The scanning range was from apex to the base of lungs. Scanning parameters were as follows: tube voltage

80-120 kV, tube current 50-350 mAs, pitch 0.99~1.22 mm, matrix 512×512, slice thickness 10 mm, field of view 350 mm×350 mm. Reconstruction was performed with slice thickness of 0.625~1.250 mm, a lung window with a width of 1200HU and a level of -600HU, and a mediastinal window with a width of 350 HU and a level of 40HU.

CT Image Analysis

Quantitative analysis of lung opacification was performed by a deep-learning algorithm. This algorithm consists of three modules: (a) lung and lobes segmentation module; (b) lung opacity segmentation module; and (c) quantitative analysis module. The algorithms used in (a) and (b) were based on a deep-learning framework to learn the complex relationship between diverse features extracted from chest CT scans and regions of interest (lungs, lobes, and opacities). The deep-learning algorithm in module (b) employed a well-established fully convolutional neural network architecture (8) trained on annotated datasets of COVID-19. We describe the deep learning algorithms in detail in **Appendix E2**. Based on the segmentation results of lungs and lesions, the workstation provided a quantitative measure of lung opacification percentage (Figure 2).

Accurate segmentation of the lung opacities was the basis for quantitative analysis. Hence, all segmentation results derived from this deep-learning algorithm were visually evaluated by two radiologists (one with 7 years of experience in cardiopulmonary imaging and another with 8 years of experience in pulmonary imaging), who viewed the segmentation independently. Both radiologists were blinded to the patient's clinical status. The scoring procedure was as follows: both radiologists reviewed the segmentation results displayed as regions of interest overlaid on the CT images slice-by-slice. The readers did not adjust the automatic segmentation. The readers used a scoring criteria based on the adequacy of the segmentation task versus actual lung opacification. Specifically, the degree of matching was quantified using a Likert score from 0 to 5. The scoring criteria is described in detail in **Appendix E3**. To reduce the subjectivity of the radiologist's evaluation, the final score of was the average of two scores for each scan. A final score ≥ 3 was considered as sufficient to meet the quantitative analysis requirement.

Statistical Analysis

Statistical analysis was performed using SPSS software (version 23.0, IBM statistics, Armonk, NY, USA). Categorical variables were expressed as counts (percentage), and continuous variable as mean ± SD or median (interquartile range). Normality of distribution was tested using the Kolmogorov-Smirnov test. The difference between two paired groups were assessed by paired t-test or Wilcoxon tests. Moreover, Comparisons among different clinical types were performed by the analysis of variance (ANOVA) or Kruskal-Wallis test. Comparison between any of the two clinical types were performed by t-test or Mann-Whitney U test with continuous variable, or χ^2 test with categorical variable. Low frequency variables were compared with Fisher exact test. Two-side $P < 0.05$ was considered statistically significant.

Results

Clinical characteristics

One hundred and forty-eight patients with COVID-19 were initially enrolled, with 9 (6.1%) patients excluded due to respiratory motion artifacts and 13 (8.7%) excluded due to insufficient segmentation quality as determined by the scoring from the two radiologists (i.e., mean score < 3). Finally, a total of 126 patients (mean age, 52 years ± 15 years; age range, 14-86 years; 53.2% males) with COVID-19 were included. Baseline characteristics of COVID-19 patients are summarized in **Table 1**. All patients were classified into four clinical types, including 6 mild cases (4.8%), 94 moderate cases (74.6%), 20 severe cases (15.8%) and 6 critical cases (4.8%). The median of interval between baseline and 1st follow-up was 4 days (interquartile range 3-6 days), and the median of interval between the 1st and 2nd follow-up was 5 days (interquartile range 3-7 days).

Age and gender had no significant difference among the different clinical types of COVID-19 ($P > 0.05$). Duration between onset symptoms and initial CT scanning of mild and moderate type patients were shorter than those of severe and critical type (all $P<0.01$). In 117 patients of 126 (92.9%), fever was the initial symptom, while dyspnea was only observed in severe and critical types. Of the laboratory findings, WBC count, lymphocyte count, high-sensitivity C-reactive protein (hs-CRP), and pulse oxygen saturation (SpO_2) showed significant differences among the four clinical types of patients (all $P<0.05$). Compared to critical type patients, WBC count and hs-CRP were significantly lower in moderate type cases (both $P<0.001$), but lymphocyte count was higher in the moderate type ($P=0.004$).

Quantitative CT Parameters at baseline, 1st and 2nd follow-up CT scans

All 126 patients had two CT scans as per inclusion criteria, and 48 0f 126 (38.1%) patients had three CT scans. 236 of all 300 CT scans (78.6%) has a segmentation quality score in the range of 3~4, and 64 (21.4%) CT scans were in the range of 4~5.

The distribution of lung opacification percentage of all patients according to days since onset of symptoms is shown in **Figure 2a**, and peak lung opacification percentage of whole lung occurred at day 13. Overall, the whole-lung QCT-PLO significantly increased between baseline CT and 1st follow-up CT (median [interquartile range]; 3.6%[0.5%,12.1%] vs 8.7%[2.7%,21.2%], $P < 0.01$). No significant progression of whole-lung QCT-PLO was noted between the 1st follow-up and 2nd follow-up CT (8.7% [2.7%,21.2%] vs 6.0%

[1.9%,24.3%], *P*=0.655). Percentage changes in the CT derived opacification parameters of the 1^{st} and 2^{nd} follow-up are shown in **Table 2**.

Quantitative CT opacification parameters in different clinical type of COVID-19 patients

Differences in whole-lung QCT-PLO according to clinical severity subtype and days since onset of symptoms at the baseline CT is showed in **Figure 2b**.

Significant differences of QCT-PLO were found among the four different clinical types at the baseline and at the 1^{st} follow-up (all $P<0.05$, **Table 3**). All of the 6 mild COVID-19 patients had negative CTs at the baseline, and were found positive at the 1^{st} follow-up CT scan **(Figure 3**). QCT-PLO of right and left lower lobes were elevated in the 2^{nd} follow-up CT scan (both $P < 0.05$, **Table E1**). Compared to baseline CT scan, whole-lung and per lobe QCT-PLO increased significantly in moderate type patients (all $P < 0.05$, supplement 3) (**Figure 4**), while no remarkable difference was found between the 1^{st} and 2^{nd} follow-up scans (all $P > 0.05$, **Table E2**). In severe and critical type patients, the whole-lung and per lobe QCT-PLO showed no significant differences between baseline, 1^{st}, or 2^{nd} follow-up CTs (**Figures 5 and 6**, respectively).

Discussion

In this study, we evaluated the longitudinal changes of pneumonia severity in different clinical types COVID-19 at baseline and follow-up imaging using a quantitative image parameter (QCT-PLO), which was automatically generated by a deep-learning tool from chest CT scans. Our major findings were: (a) This quantitative parameter based on deep learning could identify differences in the lung opacity burden on CTs from COVID-19 patients of different clinical severities; (b) Overall, the whole lung and per lobe QCT-PLO at the 1st follow-up CT increased in comparison with the baseline scans (median interval 4 days), while no remarkable progress was found at the 2nd follow-up (median interval 5 days).

Mild and moderate COVID-19 patients had shorter duration between onset symptoms and initial CT scan, which indicates that these patients could have presented at a relative early stage of disease. This was confirmed by the lower whole-lung and per lobe QCT-PLO at baseline CT. SpO_2 of all severe and critical type patients were less than 90% and more than half had dyspnea, which concords to the higher lung opacification percentage assessed by the deep-learning tool. According to prior studies (9,10), severe and critical type patients had multiple GGO with consolidation, which can lead to ventilatory dysfunction and even respiratory failure. Moreover, hs-CRP was significantly elevated in severe and critical type patients, which indicates an inflammatory type of response.

We observed in our data that whole-lung and per lobe QCT-PLO were higher at the 1st follow up than at baseline, suggesting a sustained progression of imaging findings from presentation, plateauing on the 2nd follow-up CT. Such pattern could be attributed to many factors, including characteristics of our cohort, clinical severity at admission, treatment effect, and the natural history of disease. Depending on the initial clinical type and time of scan, patients could present at any of the stages described here. A combined analysis of our quantitative results suggests that pulmonary involvement in COVID-19 ramps up after the beginning of symptoms, peaking at 13 days, which is in keeping with prior a prior observation (11).

There are several limitations of the present study. First, not all patients had a serial of three CT scans, therefore we cannot systemically evaluate the changes for all patients at the 1st and 2nd follow up. Second, there was no systematic confirmation of the pulmonary opacities as being directly caused by the pathological effects of the coronavirus. Last, although the commercial software can quantitatively evaluate lung opacification percentage, the current version still needs radiologists' supervision. Noticeably, 8.7% (13/148) of the cases had insufficient segmentation quality to ensure appropriate quantification.

In conclusion, the pulmonary involvement of COVID-19 could be objectively assessed by deep-learning-based quantitative CT. This automated tool may be used for quantifying the disease burden and monitoring disease progression or response to treatment.

References

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;6736(20):1–10.

2. Naming the 2019 Coronavirus. The International Committee on Taxonomy of Viruses (ICTV) website. https://talk.ictvonline.org. Published February 11, 2020.

3. Diagnosis and Treatment Protocol of Novel Coronavirus (trial version 5th). National Health Commission of the People's Republic of China website. http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml. Published February 4,2020.

4. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respiratory Medicine. February 18, 2020. DOI: https://doi.org/10.1016/S2213-2600(20)30076-x.

5. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. Radiology. 2017;284(1):228–243.

6. F. Pan, T. Ye, P. Sun, et al. Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. Radiology 0(0): 200370. Published Online: Feb 13 2020.https://doi.org/10.1148/radiol.2020200370.

7. Science and technology anti-epidemic: inferVISION company first launched pneumonia AI system into clinical use. cn-healthcare website. https://www.cn-healthcare.com/article/20200131/wap-content-529622.html. Published January 31, 2020.

8. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[M]. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Springer International Publishing, 2015:234-241.

9. Huang L, Rui Han, Yu X, et al. A correlation study of CT and clinical features of different clinical types of 2019 novel coronavirus pneumonia. Chin J of Radiol. 2020,54(00): E003-E003. Published Feb 8, 2020. DOI: 10.3760/cma.j.issn.1005-1201.2020.0003.

10. Song F, Shi N, Shan F, et al. Emerging Coronavirus 2019-nCoV Pneumonia. Radiology. 2002742020 Feb 6 [Online ahead of print]. DOI: 10.1148/radiol.2020200274.

11. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng D, Wang J, Hesketh RL, Yang L, Zheng C. Time Course of Lung Changes On Chest CT During Recovery From 2019

Novel Coronavirus (COVID-19) Pneumonia. Radiology. 2020 Feb 13:200370. doi: 10.1148/radiol.2020200370. [Epub ahead of print] PubMed PMID: 32053470.

Table 1. Baseline characteristics of patients with COVID-19, according to clinical severity

Variables	COVID-19 (N=126)	Mild type (N=6)	Moderate type (N=94)	Severe type (N=20)	Critical type (N=6)	*P* value*
Age (year) †	52±15	47±15	51±16	56±13	66±8	0.074
Sex (male)	67 (53.2)	3 (50.0)	49 (52.1)	11 (55.0)	4 (66.7)	0.946
Duration between onset symptoms and the initial CT scanning †	2.5 (1,5)	1 (1,1)	2 (1,3)	6.5 (5,7.3)	6 (5,7.8)	<0.001
Comorbidity						
Hyertension	10 (7.9)	1 (16.7)	4 (5.3)	2 (10.0)	3 (50)	0.018
Diabetes	7 (5.6)	0	3 (3.2)	2 (10.0)	2 (33.3)	0.036
COPD	2 (1.6)	1 (16.7)	0	0	1 (16.7)	0.008
CAD	7 (5.6)	0	4 (4.3)	1 (5.0)	2 (33.3)	0.085
Symptoms						
Fever	117 (92.9)	6 (100)	85 (90.4)	20 (100)	6 (100)	0.598
Normal	9 (7.1)	0	9 (9.6)	0	0	<0.001
37.3-38.0°C	8 (6.3)	1 (16.7)	7 (8.2)	0	0	0.344
38.1-39.0°C	94 (74.6)	4 (66.7)	78 (91.8)	11 (55.0)	1 (16.7)	<0.001
≥39.1°C	15 (11.9)	1 (16.7)	0	9 (45.0)	5 (83.3)	<0.001
Cough	35 (27.8)	3 (50)	22 (23.4)	9 (45.0)	1 (16.7)	0.202
Fatigue	19 (15.1)	0	14 (14.9)	5 (25.0)	0	0.416
Dyspnea	14 (11.1)	0	0	8 (40.0)	6 (100)	<0.001
Chest distress	9 (7.1)	0	7 (7.4)	2 (10.0)	0	0.865
Headache	5 (4.0)	0	3 (3.2)	2 (10.0)	0	0.526
Diarrhea	4 (3.2)	0	4 (4.3)	0	0	1.000
Sore throat	2 (1.6)	0	1 (1.2)	1 (5.0)	0	0.445
Laboratory findings						
WBC count (*10^9/L) †	4.8 (3.8,6.1)	3.2 (3.1,5.7)	4.6 (3.8,5.8)	6.3 (5.0,11.7)	8.1 (6.7,9.3)	0.014
Lymphocyte count (*10^9/L) †	0.9 (0.7,1.3)	0.7 (0.7,1.0)	1.1 (0.9,1.3)	0.8 (0.7,0.9)	0.6 (0.5,0.7)	0.016
Hs CRP (mg/L) †	18.9 (10.2,45.7)	11.7 (10.85,14.65)	16.1 (10.1,25.7)	97.4 (34.6,122.5)	123.9 (114.6,136.3)	<0.001
SpO_2 < 90	26 (20.6)	0	0	20 (100)	6 (100)	<0.001

Note.—Unless otherwise specified, data are numbers, with percentages in parentheses. COPD= chronic obstructive pulmonary disease, CAD=coronary artery disease, COVID-19 = coronavirus disease 19; WBC= white blood cell; hs-CRP= high-sensitivity C-reactive protein, SpO_2=pulse oxygen saturation.

**P* value is for four clinical types. $P < 0.05$ is considered to indicate statistical significance.

†Data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution.

Table 2. Percent changes in QCT-PLO at 1st and 2nd follow-up

Percent Changes	1st follow-up	2nd follow-up
Total opacification percentage of whole lung	69.3 (-14.5,605.8)	3.0 (-59.1,223.0)
Opacification percentage of right upper lobe	0 (-12.0,170.9)	0 (-43.2,93.4)
Opacification percentage of right middle lobe	0 (-30.9,47.9)	0 (-86.3,54.0)
Opacification percentage of right lower lobe	14.1 (-14.5,431.9)	0 (-68.0,155.8)
Opacification percentage of left upper lobe	0 (-10.5,163.4)	0 (-52.7,135.1)
Opacification percentage of left lower lobe	0.7 (9.1,370.6)	0 (-74.9,453.5)

Note.—Unless otherwise specified, data are median (interquartile range). QCT-PLO, quantitative CT – percentage of lung opacification.

Table 3. QCT-PLO according to clinical severity in COVID-19, baseline CT

Parameters	Mild type (n=6)	Moderate type (n=94)	Severer type (n=20)	Critical type (n=6)	*P* value*
Total opacification percentage of whole lung	0	2.2 (0.4,7.1)	28.9±19.2†	49.6±14.8†	<0.001
Opacification percentage of right upper lobe	0	0.4 (0,2.7)	28.1±21.0†	56.2±21.9†	<0.001
Opacification percentage of right middle lobe	0	0.2 (0,1.8)	24.5±20.4†	42.3±25.9†	<0.001
Opacification percentage of right lower lobe	0	2.9 (0.2,13.6)	43.3±30.7†	61.1±17.7†	<0.001
Opacification percentage of left upper lobe	0	0.3 (0,3.0)	12.3 (4.4,22.6) †	44.8±24.8†‡	<0.001
Opacification percentage of left lower lobe	0	1.3 (0,7.0)	33.3±21.8†	42.8±34.0†	<0.001

Note.—Unless otherwise specified, data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution. COVID-19 = coronavirus disease 19. QCT-PLO, quantitative CT – percentage of lung opacification.

**P* value is for four clinical types. $P < 0.05$ is considered to indicate statistical significance.

†*P* value < 0.05/6 compared to moderate type.

‡*P* value < 0.05/6 compared to severe type.

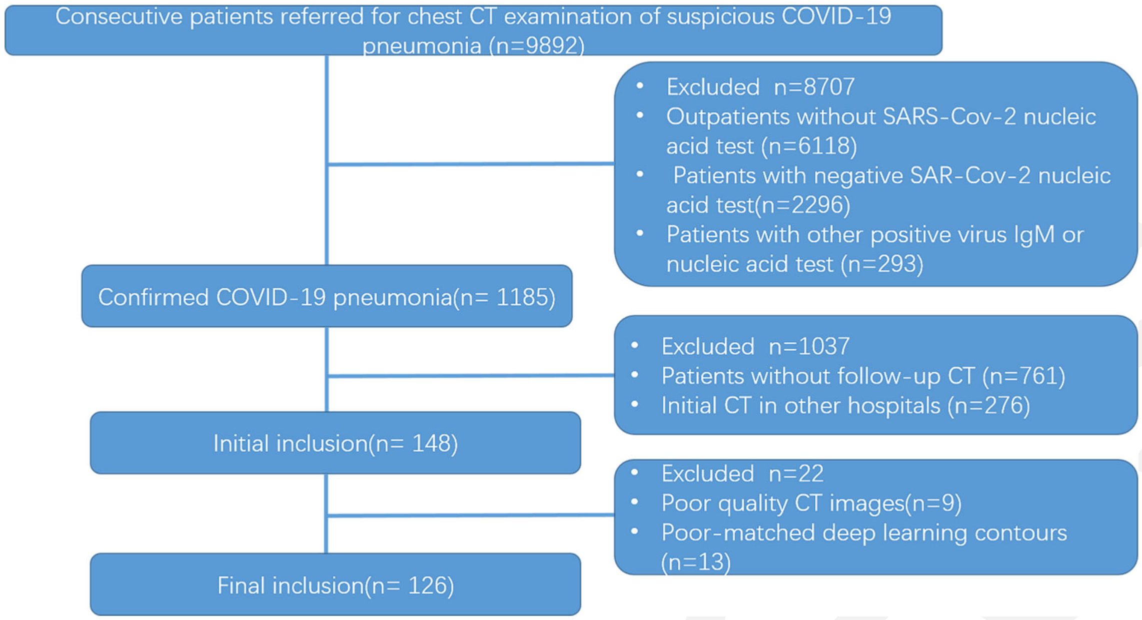

Figure 1: Flowchart shows the patient selection process. COVID-19=coronavirus disease 19.

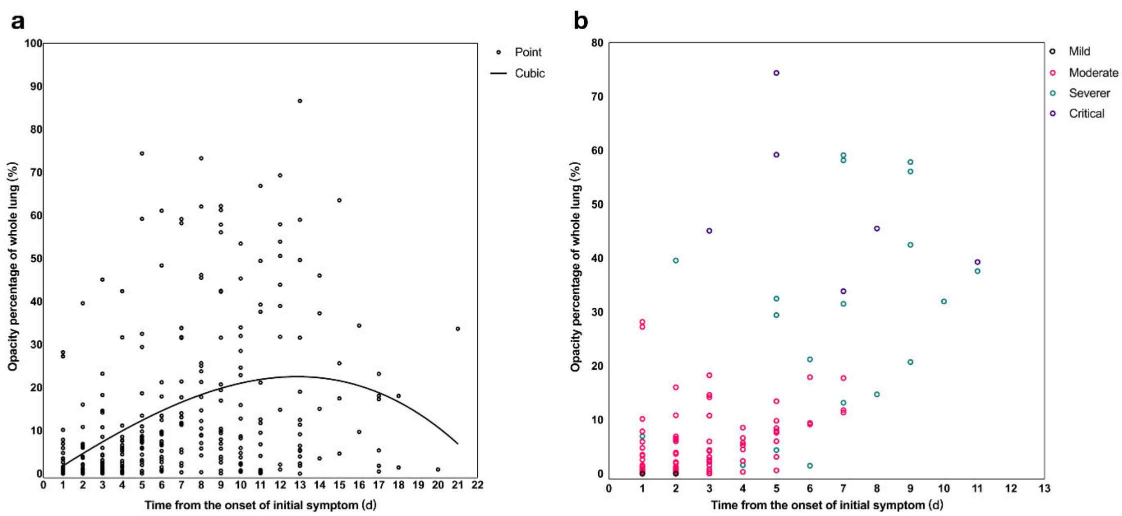

Figure 2: Scatter plots with the distribution of lung opacification percentage according to days since initial symptoms. (a) The dynamic change in lung opacification percentage of whole lung (curve fitting equation: $y=2.956*x^3-0.03065*x^2-0.004374x-1.106$, in which x=time from the onset of initial symptoms, y=lung opacification percentage of whole lung; R2=0.161, $p < 0.001$), (b) The distribution of percentage of lung opacification on quantitative CT in different clinical types according to days since initial symptoms at the baseline CT.

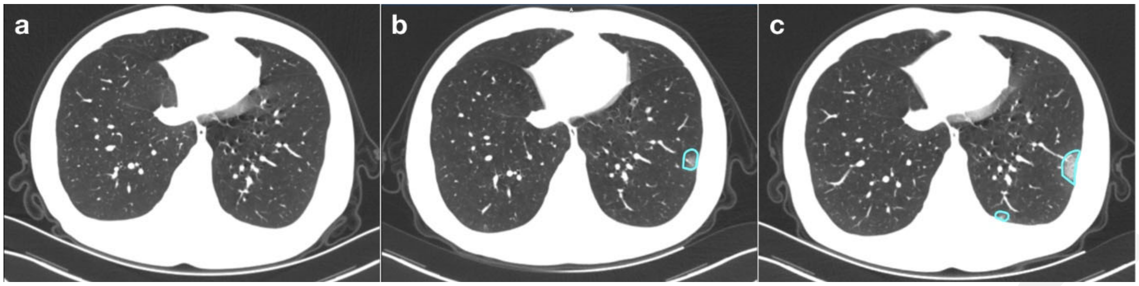

Figure 3: A 29-year-old male patient with mild COVID-19, axial chest CT images at baseline and follow-up. (a) baseline: negative CT; (b) 1st follow-up: ground-glass opacity (GGO) is observed in the left lower lobe (opacification percentage of the left lower lobe: 0.24%); (c) 2nd follow-up: increased size and new GGO (opacification percentage of the left lower lobe: 2.55%).

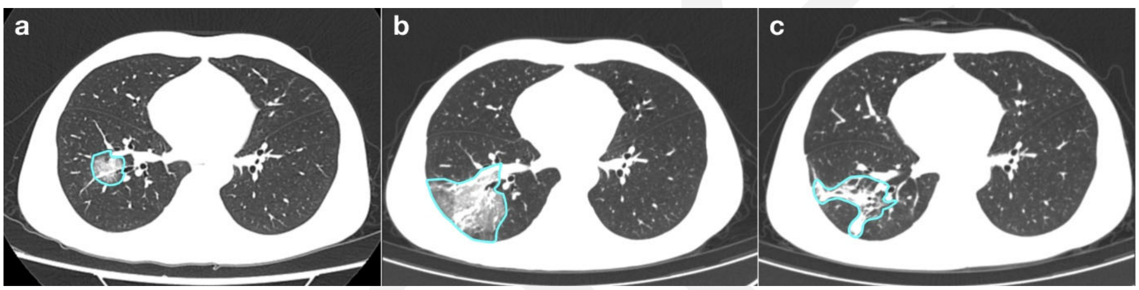

Figure 4: A 41-year-old male with moderate COVID-19, axial chest CT images at baseline and follow-up. (a) baseline: ground-glass opacity (GGO) is found in the right lower lobe (opacification percentage of the right lower lobe: 1.33%); (b) 1st follow-up: increased patchy GGO with new consolidation in the right lower lobe (opacification percentage of the right lower lobe: 12.56%); (c) 2nd follow-up: GGO is partially absorbed and development of perilobular pattern (opacification percentage of the right lower lobe: 9.28%).

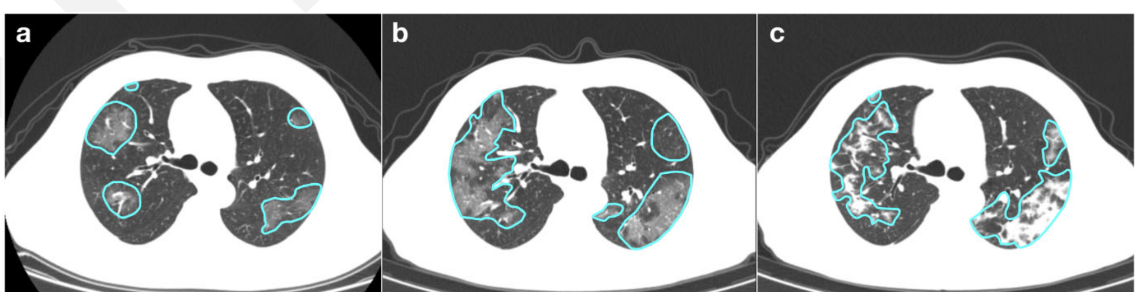

Figure 5: A 56-year-old male with severe COVID-19, axial chest CT images at baseline and

follow-up. (a) baseline: multiple ground-glass opacities (GGO) are observed in the right and left upper lobes (opacification percentages of right and left lobes: 19.78% and 17.79%, respectively); (b) 1^{st} follow-up: multiple patchy GGO are increased bilaterally (opacification percentages of right and left lobes: 30.39% and 29.72%, respectively); (c) 2^{nd} follow-up: GGO is absorbed, with development of consolidation and perilobular pattern (opacification percentages of right and left lobes: 24.21% and 19.73%, respectively).

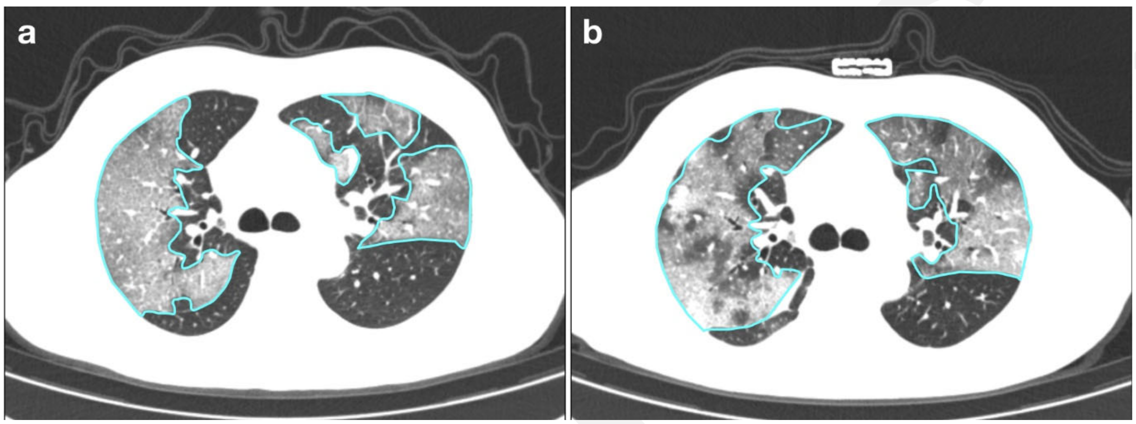

Figure 6: A 53-year-old male with critical COVID-19, axial chest CT images at baseline and follow-up. (a) baseline: multiple ground-glass opacities (GGO) are observed in the right and left upper lobes (opacification percentages of right and left lobes: 53.55% and 45.89%, respectively); (b) 1^{st} follow-up: multiple patchy GGO are increased bilaterally, with development of consolidation (opacification percentages of right and left lobes: 59.36% and 67.77%, respectively).

SUPPLEMENTAL MATERIALS

Appendix E1 Clinical classification of COVID-19

According to Diagnosis and Treatment Protocol of Novel Coronavirus (trial version 5^{th}) from National Health Commission of the People's Republic of China [1], COVID-19 pneumonia was classified into four types, including mild, moderate, severe and critical types. The detailed information is following:

(1) Mild type: Patients have mild clinical symptoms without CT findings of pneumonia.

(2) Moderate type: Patients have fever and respiratory symptoms with CT findings of pneumonia.

(3) Severe type: Patients are confirmed this type, if they had any of following criteria, (a) respiratory distress (respiratory rate≥30bpm), (b) $SpO_2 \leq 93\%$ at rest, (c) Pa $O_2/FiO_2 \leq 300$mmHg(1mmHg=0.133kPa)

(4) Critical type: Patients are confirmed this type, if they had any of following criteria, (a) respiratory failure with mechanical ventilation, (b) shock, (c) combined with other organs dysfunction with ICU therapy.

Appendix E2 Development of deep learning algorithms

InferReadTM CT Pneumonia consists of three modules: (a) lung and lobes region extraction module; (b) pneumonia segmentation module; (c) quantitative analysis module. The algorithms used in (a) and (b) were based on deep learning. In this study, the development process of the deep learning algorithm of pneumonia segmentation was mainly described. Specially, the deep learning algorithm employed a popular convolutional neural network architecture of U-Net [2] and was trained using a annotated dataset of COVID-19.

A total of 842 patients (all confirmed to have COVID-19) were collected retrospectively for lung opacity segmentation training and testing, who underwent chest CT scans between 10 January 2020 and 25 January 2020 in Tongji Hospital, Wuhan, China. This data set did not overlap with the dataset in the body of this study. Among them, 774 cases were randomly selected for training the U-Net, and other 68 patients were used for testing.

Manual segmentation for lung opacities was performed by two radiologists (not the same person as the two radiologists who evaluated the score in the body of the study) in a consensus reading using InferScholarTM Center (InfervisionTM, Beijing, China). To reduce time consumption of manual annotation, radiologists segmented lung opacities every five slices on the training dataset. In the testing dataset, manual annotation was performed in a slice-by-slice manner. In the end, 14,482 slices were annotated in the training set and 5,303 slices were annotated in the test set.

The U-Net architecture (shown in Figure E1) consisted of a downsampling path and an upsampling path, which reduced the 512×512 input image to a $16 \times 16 \times 256$ feature map to capture context and features, and then upsampled it into a $512 \times 512 \times 2$ output for precise lesion localization. In the downsampling path, each step consisted of a series of 3×3 convolutions, which followed by a Rectified Linear Unit (ReLU) activation function, and a 2×2 max-pooling layer. In the upsampling path, each of the first three steps combined an up-sampling operation, a merge layer, and blocks of 3×3 convolutions with ReLU. The final step in the upsampling path consisted of a upsampling layer and a convolution with a $1 \times 1 \times 2$ kernel followed by a softmax function, which output a score for each of the two classes (background and pneumonia lesion). The final segmentation was obtained by selecting the class with the highest softmax score for each pixel.

The dice similarity coefficient (DSC) was used to evaluate segmentation performance, and defined as $DSC=2TP / (FP + 2TP + FN)$, where TP, FP, and FN are the numbers of true positive, false positive, and false negative detections, respectively. The segmentation performance was evaluated on the testing dataset, and the median DSC and range was 0.8481 (0.6526 - 0.9094).

Appendix E3 Radiologists evaluation

Radiologists reviewed the segmentation results overlaid each CT image. The scoring criteria were based on the agreement between the results of the automatic segmentation task and the actual lung opacities. The degree of matching was described in a Likert score from 0 to 5 (Figure E2). A score of 0 was assigned in two cases (Figure E2a), corresponding to large areas of false positive or false negative contours in at least one slice or medium-sized area of false positive or false negative contours in at least three slices. Under the condition that score 0 was not met, if at least one slice had a medium-sized area of false positive or false negative contours, it was defined as score 1. Score 3 indicated that there were no obvious false positive or false negative contours on all sliced (Figure E2b). Score 5 indicated that segmentation results had a perfect fit to actual lung opacification on all slices (Figure E2c). Score 2 and score 4 were assigned to intermediate conditions that did not meet the predefined criteria above.

For the scans included in the analysis, we calculated the distribution of the scores and compared the consistency among the radiologists' scores. Radiologist 1 assigned a score of 3 in 32 (10.6%), score of 4 in 212 (70.6%), and a score 5 in 56 CTs (18.8%). Radiologist 2 assinged a score of 3 in 35 (11.6%), a score of 4 in 212 (70.6%), and score 5 in 56 CTs (18.8%). The kappa coefficient [3] between the two radiologists was 0.75.

References

1. Diagnosis and Treatment Protocol of Novel Coronavirus (trial version 5th). National Health Commission of the People's Republic of China website. http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml. Published February 4,2020.

2. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Springer International Publishing, 2015:234-241.

3. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960, 20. 37-46. 10.1177/001316446002000104.

Table E1. CT derived parameters of mild type COVID-19 patients at base, 1st and 2nd follow-up

Parameters	Baseline CT	1st follow-up CT	*P* value*	2nd follow-up CT	*P* value†
Opacification percentage of whole lung	0	0.4 (0.1,6.9)	<0.001	2.7 (0.6,14.7)	0.80
Opacification percentage of right upper lobe	0	0 (0,6.3)	<0.001	0 (0,13.4)	1.00
Opacification percentage of right middle lobe	0	0.3 (0,2.6)	0.007	1.3 (0.2,6.5)	0.043
Opacification percentage of right lower lobe	0	0.5 (0.1,11.5)	<0.001	2.3 (0.7,17.9)	0.5
Opacification percentage of left upper lobe	0	0.2 (0,5.7)	<0.001	1.5 (0,14.5)	0.068
Opacification percentage of left lower lobe	0	0.4 (0.2,10.9)	<0.001	2.6 (1.2,26.6)	0.043

Note.—Unless otherwise specified, data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution. COVID-19 = coronavirus disease 19.

**P* value is for baseline CT versus 1st follow-up CT.

†*P* value is for 1st follow-up CT versus 2nd follow-up. $P < 0.05$ is considered to indicate statistical significance.

Table E2. CT derived parameters of moderate type COVID-19 patients at base, 1st and 2nd follow-up

Parameters	Baseline CT	1st follow-up CT	*P* value*	2nd follow-up CT	*P* value†
Total opacification percentage of whole lung	2.2 (0.4,7.1)	6.9 (2.3,12.5)	<0.001	5.6 (1.8,23.9)	0.394
Opacification percentage of right upper lobe	0.4 (0,2.7)	1.2 (0.1,8.7)	<0.001	4.5 (0.2,28.3)	0.235
Opacification percentage of right middle lobe	0.2 (0,1.8)	0.5 (0,3.8)	0.029	0.4 (0,15.2)	0.249
Opacification percentage of right lower lobe	2.9 (0.2,13.6)	13.5 (2.3,27.0)	<0.001	7.9 (1.4,38.5)	0.886
Opacification percentage of left upper lobe	0.3 (0,3.0)	1.6 (0.1,9.2)	<0.001	2.0 (0.5,17.6)	0.191
Opacification percentage of left lower lobe	1.3 (0,7.0)	6.0 (0.4,16.4)	<0.001	3.9 (0.5,18.3)	0.922

Note.—Unless otherwise specified, data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution. COVID-19 = coronavirus disease 19.

**P* value is for baseline CT versus 1st follow-up CT.

†*P* value is for 1st follow-up CT versus 2nd follow-up CT. $P < 0.05$ is considered to indicate statistical significance.

Table E3. CT derived parameters of severe type COVID-19 patients at base, 1st and 2nd follow-up

Parameters	Baseline	1st follow-up	*P* value*	2nd follow-up	*P* value†
Total opacification percentage of whole lung	28.9±19.2	35.6±18.1	0.166	28.0±30.5	0.953
Opacification percentage of right upper lobe	28.1±21.0	33.5±21.3	0.261	25.1±27.9	0.891
Opacification percentage of right middle lobe	24.5±20.4	29.1±20.0	0.278	23.0±28.6	0.83
Opacification percentage of right lower lobe	43.3±30.7	49.4±26.0	0.436	38.8±30.7	0.898
Opacification percentage of left upper lobe	12.3 (4.4,22.6)	26.9±19.2	0.067	23.4 (11.9,41.9)	0.463
Opacification percentage of left lower lobe	33.3±21.8	44.6±23.9	0.122	36.4±32.8	0.852

Note.—Unless otherwise specified, data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution. COVID-19 = coronavirus disease 19.

**P* value is for baseline CT versus 1st follow-up CT.

†*P* value is for 1st follow-up CT versus 2nd follow-up CT. $P < 0.05$ is considered to indicate statistical significance.

Table E4. CT derived parameters of critical type COVID-19 patients at base and 1st follow-up

Parameters	Baseline CT	1st follow-up CT	*P* value*
Total opacification percentage of whole lung	49.6±14.8	48.5±23.6	0.907
Opacification percentage of right upper lobe	56.2±21.9	52.9±12.1	0.759
Opacification percentage of right middle lobe	42.3±25.9	50.2±30.5	0.468
Opacification percentage of right lower lobe	61.1±17.7	51.3±30.9	0.473
Opacification percentage of left upper lobe	44.8±24.8	43.6±24.2	0.889
Opacification percentage of left lower lobe	42.8±34.0	44.2±34.4	0.91

Note.—Unless otherwise specified, data are means ± standard deviation with normal distribution or median (interquartile range) with non-normal distribution. COVID-19 = coronavirus disease 19.

*P value is for baseline CT versus 1st follow-up CT. $P < 0.05$ is considered to indicate statistical significance.

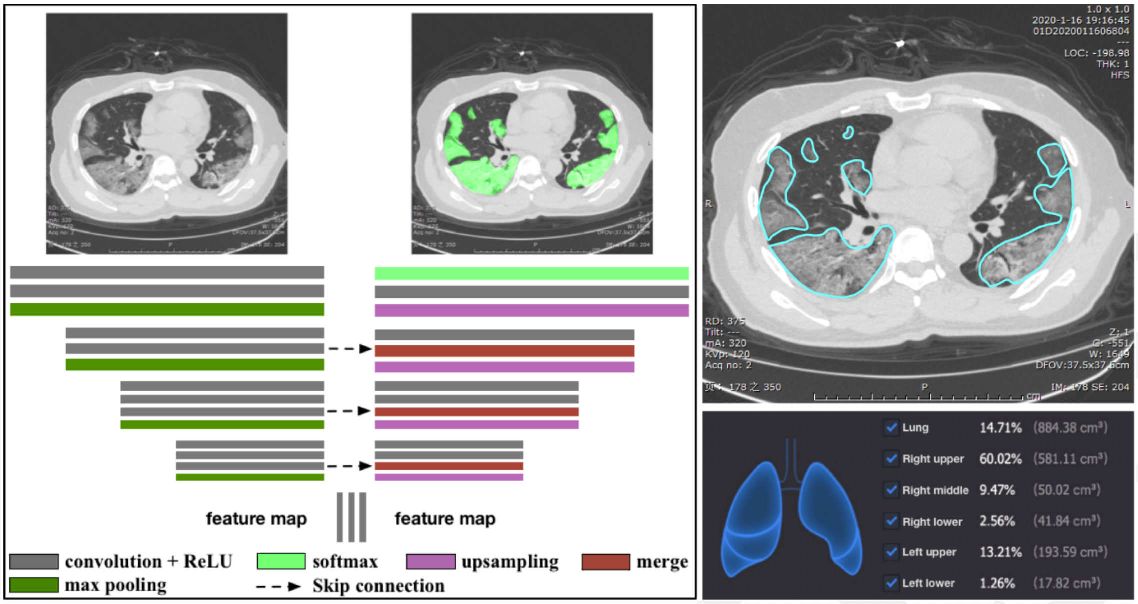

Figure E1: The U-Net architecture used in InferReadTM CT Pneumonia and an example on quantitative analysis. The architecture consisted of a downsampling path and an upsampling path, which reduced the input image to map for capturing context and features, and then upsampled for opacity segmentation. In the example, a 55-year-old male who underwent chest CT scan for COVID-19, confirmed by RT-PCR. InferReadTM CT Pneumonia outlined the regions that were considered as lung opacities with blue lines and showed the volume and proportion of lung opacities in the lung and the each of lobes.

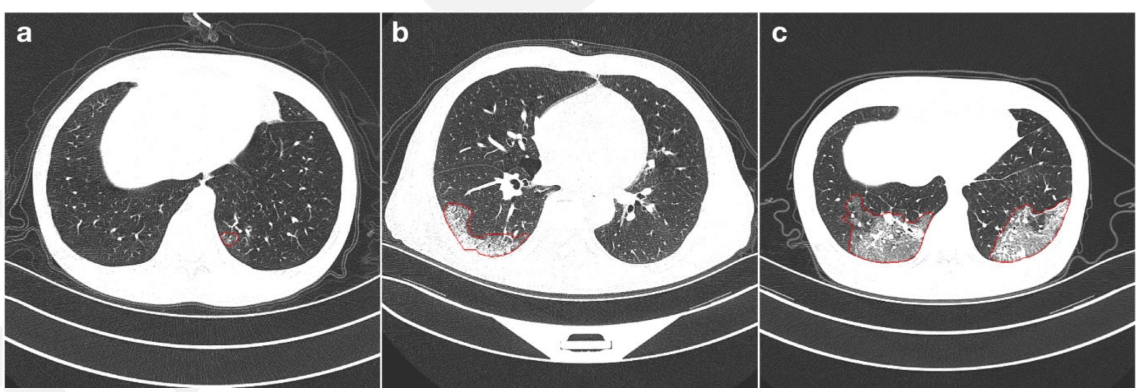

Figure E2: Examples of different scores. a. Example of a score of 0. The deep-learning algorithm did not correctly segmented the ground glass opacity in the left lower lobe. The occurrence of such pattern in at least three slices results in a score of 0. b. Example of a score of 3. Deep-learning algorithm satisfactory segmentation of the lung opacity in the right lower lobe, with minimal imperfections. The occurrence of such pattern in all slices, the scan score is 3. c. Example of score 5. Deep learning algorithms perfectly segment opacities of the lung. When the segmentation results of most slices in a scan meet this situation, the scan score is 5.