Automatic CT Quantification of Coronavirus Disease 2019 pneumonia: An international collaborative development, validation, and clinical implication

Seung-Jin Yoo

Department of radiology, Hanyang University Medical Center, Hanyang University College of Medicine, 222-1, Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

Xiaolong Qi

CHESS Center, The First Hospital of Lanzhou University, Lanzhou 730000, China

Shohei Inui

Department of radiology, Japan Self-Defense Forces Central Hospital, 1-2-24, Ikejiri, Setagaya-ku, Tokyo, 154-0001, Japan / Department of Radiology, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

Sang Joon Park

Department of radiology, Seoul National University Hospital, Seoul National College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea / MEDICALIP Co. Ltd., Changgyeong Building, 174, Yulgok-ro, Jongno-gu, Seoul 03127, Korea

Hyungjin Kim

Department of radiology, Seoul National University Hospital, Seoul National College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

Yeon Joo Jeong

Department of Radiology, Pusan National University Hospital, Pusan National University School of Medicine and Biomedical Research Institute, 179 Gudeok-ro, Seo-gu, Busan 49241, Korea

Kyung Hee Lee

Department of Radiology, Seoul National University Bundang Hospital, 82 Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea

Young Kyung Lee

Department of Radiology, Seoul Medical Center, 156, Sinnae-ro, Jungnang-gu, Seoul 02053, Korea

Bae Young Lee

Departement of Radiology, Eunpyeong St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 1021, Tongil-ro, E·unpyeong-gu, Seoul 03341, Korea

Jin Yong Kim

Division of Infectious Diseases, Department of Internal Medicine, Incheon Medical Center, 217 Bangchuk-ro, Dong-gu, Incheon, 22532, Korea.

Kwang Nam Jin

Department of Radiology, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, 20 Boramae-ro 5-gil, Dongjak-gu, Seoul 07061, Korea.

Jae-Kwang Lim

Department of Radiology, School of Medicine, Kyungpook National University, 680 Gukchaebosang-ro, Jung-gu, Daegu 41944, Korea.

Yun-Hyeon Kim

Department of Radiology, Chonnam National University Medical School, 42 Jebong-ro, Dong-gu, Gwangju 61469, Korea

Ki Beom Kim

Department of Radiology, Daegu Fatima Hospital, 99 Ayang-ro, Dong-gu, Daegu 701-724, Korea

Zicheng Jiang

Department of Infectious Diseases, Ankang Central Hospital, An·kang, China

Chuxiao Shao

CHESS-COVID-19 Group, Lishui Central Hospital, Lishui, China

Junqiang Lei

Department of Radiology, The First Hospital of Lanzhou University, Lanzhou 730000, China

Shengqiang Zou

Department of Infectious Diseases, The Affiliated Third Hospital of Jiangsu University, Zhenjiang, China

Hongqiu Pan

Department of Infectious Diseases, The Affiliated Third Hospital of Jiangsu University, Zhenjiang, China

Ye Gu

CHESS-COVID-19 Group, The Sixth People's Hospital of Shenyang, Shenyang, China

Guo Zhang

CHESS-COVID-19 Group, The People's Hospital of Guangxi Zhuang Autonomous Region, Nanning, China

Jin Mo Goo

Department of radiology, Seoul National University Hospital, Seoul National College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

Soon Ho Yoon (yshoka@gmail.com)

Department of radiology, Seoul National University Hospital, Seoul National College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

---

Research Article

Keywords: COVID-19, Pneumonia, Deep learning, Computed Tomography

Posted Date: July 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-48290/v1

Version of Record: A version of this preprint was published at Journal of Computer Assisted Tomography on April 8th, 2022. See the published version at https://doi.org/10.1097/RCT.0000000000001303.

Abstract

Objectives

We aimed to develop and validate the automatic quantification of COVID-19 pneumonia on CT images.

Methods

This retrospective study included 176 chest CT scans of 131 COVID-19 patients from 13 Korean and Chinese institutions. Two experienced radiologists semi-automatically drew pneumonia, preparing 49,830 positive and negative CT slices to develop the 2D U-Net for segmenting pneumonia. The 2D U-Net was distributed as downloadable software. External validation for quantifications' accuracy was performed using Japanese, Italian, Radiopaedia, Chinese datasets. Primary measures for the accuracy of the network were correlation coefficients for extent (%) and weight (g) of pneumonia. Logistic regression analyses were performed to evaluate the clinical implication of the extent and weight regarding the presence of symptoms in the Japanese dataset and the occurrence of composite outcome in the Spanish dataset.

Results

In the internal validation dataset, the intraclass correlation coefficients between the 2D U-Net and reference values for the extent and weight were 0.990 and 0.993, respectively. In the Japanese dataset, the Pearson correlation coefficients between the U-Net outcomes and visual CT severity scores were 0.908 and 0.899, respectively. In the other external validation datasets, the intraclass correlation coefficients between the U-Net and reference values were between 0.951-0.970 (extent) and between 0.970-0.995 (weight), respectively. In multivariate logistic regression analyses, the extent and weight of pneumonia were independently associated with symptoms (OR, 4.142 and 4.434; p=.013 and .009, respectively), and poor prognosis (OR, 7.446 and 4.677; p=.004 and .029, respectively).

Conclusions

CT extent and weight of COVID-19 pneumonia were automatically quantifiable and independently associated with symptoms and prognosis.

Introduction

Coronavirus disease 2019 (COVID-19) is rapidly spreading worldwide, causing substantial morbidity and mortality on a global scale, and was declared a pandemic as of 12 March 2020. There are 8,242,999 confirmed cases and 445,535 deaths in 216 countries as of 18 June 2020 [1]. The clinical manifestation of COVID-19 varies from an asymptomatic form to a critical form that causes respiratory and multiorgan failure, requiring mechanical ventilation and support in an intensive care unit [2]. Four-fifths of COVID-19

patients experience mild disease, whereas one-fifth of COVID-19 patients have severe to critical illness [3]. Elderly patients and those with comorbidities are at a higher risk of severe disease, developing acute respiratory distress syndrome, and death [2-5].

Another prognostic indicator in patients with COVID-19 may be the radiologic extent of pneumonia, in accordance with experiences from the earlier SARS and MERS outbreaks [6]. Chest radiography is an easily-accessible imaging modality, but its sensitivity is only 30-70% in detecting COVID-19 pneumonia [7; 8]. Chest CT provides a comprehensive evaluation of the pulmonary manifestations of COVID-19 [9], and typical CT findings are bilateral predominant ground-glass opacities (GGO) with or without consolidation in the peripheral lungs [10; 11]. COVID-19 patients with severe to critical illness were found to have a larger extent of pulmonary disease based on a visual assessment, along with lymphadenopathy, pleural effusion, and traction bronchiectasis [10; 12-14]. Chest CT is indicated to assess COVID-19 patients with a moderate-to-severe disease or at risk for progression [15]. However, the visual evaluation of disease extent requires considerable reader experience and is prone to intra- and inter-reader variability.

The purpose of our study was to develop and validate the automatic quantification of the extent (%) and weight (g) of COVID-19 pneumonia on CT images.

Materials And Methods

The institutional review board of the participating hospitals approved this retrospective study, and the requirement for patient consent was waived.

Study population

We retrospectively collected anonymized 176 chest CT scans of 131 reverse transcription-polymerase chain reaction (RT-PCR)-proven COVID-19 patients (mean age, 47.2±18.1 years; male to female ratio, 59%:41%) that were obtained at 13 Korean and Chinese institutions from 23 January to 15 March 2020. Among them, 12, five, and three patients had additional follow-up chest CT scans once, twice, and three times respectively. All 176 chest CT scans were performed using one of the 17 multi-detector CT scanners (Supplemental text).

Preparation of training CT data

The CT images were uploaded to a commercially available software program for semi-automatic segmentation (MEDIP PRO v2.0.0.0, MEDICALIP Co. Ltd., Seoul, Korea). The lung parenchyma was segmented by a previously developed deep neural network (DeepCatch v1.0.0.0, MEDICALIP Co. Ltd., Seoul, Korea; submitted), which automatically extracts lung parenchyma with an accuracy higher than 99% in CT images containing extensive lung disease. All parenchymal abnormalities of COVID-19 were initially segmented by two technicians. After reviewing the tentative lung and lesion masks, one of the two thoracic radiologists (S.H.Y. and S.J.Y. with 15 and 5 years of experience with chest CT interpretation, respectively) determined the presence of COVID-19 lesions and adjusted the masks in every axial CT

image slice. The adjustment was further supplemented by modification of the mask on coronal and sagittal images. The radiologists excluded parenchymal lesions other than COVID-19, such as peripheral reticulations and honeycombing, tuberculous sequelae, calcified nodules, dependent densities, pleural effusions, and areas of motion artifacts.

Development of the deep neural network

The 176 CT scans were randomly assigned into one of the three following data sets: 146 cases for the training set, 10 cases for the tuning set, and 20 cases for the internal validation set. A majority of the CT scans consisted of 1-mm-section CT images with standard- to low-dose CT protocols. Data were preprocessed by changing the Hounsfield unit (HU) values of the area outside the lung to -3024, and axial slices without the lung were not included in the training set. In total, 24,915 slices of axial data with areas of pneumonia and 30,711 slices of axial data without areas of pneumonia were available in the dataset. To minimize the possibility of reduced performance due to an unbalanced dataset, for every epoch, all positive slices were included, whereas all 24,915 negative slices were randomly selected. The training data were normalized by using the lung window setting.

Our 2D U-Net received an input with a size of 512×512×1 and consisted of initial convolutions, four encoders, four decoders, and a final convolution. Except for the final convolution, which was a 1×1 convolution, every convolutional layer consisted of a 3×3 convolution followed by batch normalization [16] and the rectified linear unit (ReLu) activation function [17]. For decoders, upsampling with bilinear interpolation was used, followed by concatenation to conserve information before down-sampling (Figure 1).

The Kaiming He initialization method [18] was used for weight initialization. A sigmoid function was used in the final layer, and the model was trained using the stochastic gradient descent algorithm and the binary cross entropy loss function. After the training was completed, the tuning dataset was used to choose the best weight, which was saved after each epoch.

The 2D U-Net was distributed as free standalone software (MEDIP COVID19) on 18 March 2020 and updated with the current version of v1.2.0.0 on 27 April. The software conducted a quantification of pneumonia in 1 minute with the recommended speculations (supplemental text). Based on the area of COVID-19 pneumonia segmented by the network, the software measured the volume of the pneumonia and lung parenchyma and automatically calculated the extent (%) and weight (g) of pneumonia (supplemental text).

External validation

We used five datasets for external validation (supplemental text). The first dataset included 103 non-enhanced chest CT scans of RT-PCR–proven Japanese COVID-19 cases with information about the presence of symptoms in the Diamond Princess cruise ship [12]. The second and third dataset was a public CT dataset that comprised 99 single-slice CT images of Italian COVID-19 patients [19] and nine

volumetric CT scans from website ‘Radiopaedia’ [20]. The fourth dataset was a public dataset comprised of 10 COVID-19 CT scans form China [21]. The fifth dataset was obtained from de-identified ‘COVID data save lives’ volumetric CT scan in 97 patients from HM hospitales, Spain [22].

Statistical analysis

The primary measures for the performance of the network were correlation coefficients for the extent and weight of COVID-19 pneumonia between the reference datasets and 2D U-Net. The following served as reference values: in the internal dataset, radiologists’ mask-driven extent and weight; in the first external dataset, radiologists’ visual CT severity score; and in the second, third and fourth external dataset, the provided mask-driven extent and weight. Intraclass correlation coefficients (ICCs) were used for the internal, the second, third and fourth external validation dataset, whereas Pearson correlation coefficients were used in the first external validation dataset.

Secondary measures included differences in the extent and weight of pneumonia and the morphological similarity of lung opacities between human experts and 2D U-Net. The degree of these differences was evaluated using Bland-Altman plots. The limits of agreement in the difference between the masks were explored using the original values and square root-transformation of the extent and weight of pneumonia to check the whether the degree of the difference depended on the extent and the weight [23]. Morphological similarity was assessed using the Dice similarity coefficient (DSC), sensitivity, and positive predictive value (PPV) between the network-driven and reference masks. The DSC considers correctly (true positive) and incorrectly segmented (false positive) areas, as well as missing target areas (false negative), to measure the performance of a classifier [24; 25]. The PPV, also known as precision, is an accuracy measurement that reflects the proportion of correctly predicted areas compared to all predicted areas [24; 25].

Univariate and multivariate logistic regression were conducted to evaluate risk factors for presence of symptom in Japanese dataset and composite outcome including respiratory failure, intensive care unit admission, and mortality in the Spanish dataset. The optimal cutoff value for pneumia extent and weight were identified in the ROC curve analysis and used in the logistic regression analysis. Multivariate logistic regression analyses with a backward selection were done including variables with a p-value of 0.200 or smaller in univariate analyses.

SPSS version 25 (IBM Corp., Armonk, NY, USA) was used for all statistical analyses.

Results

The ICCs for pneumonia extent and weight between the 2D U-Net and reference were 0.990 and 0.993 in the internal validation dataset (Table 1). The mean differences in pneumonia extent and weight between the reference and 2D U-Net were 0.7% and 7.04 g, respectively (Table 1). The magnitude of the measurement difference between the reference sources and 2D U-Net depended on the extent and weight (Figure 2): 10% extent: 95% limits of agreement (LOA), -4.0% to 5.4%; 50% extent: (95% LOA, -9.7% to

11.1%); 100 g: 95% LOA, -41.4 g to 55.5 g; 500 g: -101.3 g to 115.4 g. The DSC, sensitivity, and PPV for the 2D U-Net masks relative to the reference masks were 77.8±17.1%, 81.4±10.2%, and 80.3±20.6%, respectively.

In the first external validation (Japanese dataset), the Pearson correlation coefficients were 0.908 and 0.899 between the visual CT severity score and extent and pneumonia weight, respectively (Table 1). In the other external validation datasets, the intraclass correlation coefficients between the U-Net and reference values were between 0.951-0.970 (extent) and between 0.970-0.995 (weight), respectively (Table 1). The limit of agreements of pneumonia extent and weight for these external dataset are described in supplementary Figure 1 and Table 1.

Univariate logistic regression analysis for identifying the risk factor for presence of symptom in the Japanese dataset revealed age less than 60 (p=0.048), lymphopenia less than 1000 cells/μl (p=0.015), lactate dehydrogenase more than 245 U/L (p=0.025), pneumonia extent greater than 0.5% (p=0.015), and pneumonia weight heavier than 10g (p=0.008) were significant risk factors (Table 2). Two models of multivariate logistic regression analyses was done subsequently. Each model included age less than 60, lymphopenia, LDH over 245 U/L, CRP over 1 mg/dL and one of either pneumonia extent or weight. Age less than 60 (pneumonia extent model p=0.005, pneumonia weight model p=0.005), lymphopenia (pneumonia extent model p=0.038, pneumonia weight model p=0.048), pneumonia extent more than 0.5% (p=0.013), and pneumonia weight more than 10g (p=0.009) were independent risk factors for presence of symptom in COVID-19 patients (Table 2).

With regards to the composite outcome, univariate logistic regression analysis revealed that significant risk factors were age equal or more than 60 (p=0.007), LDH equal or more than cut off value of 375 U/L or more (p=0.027), lymphocyte less than 20% (p=0.029), neutrophil-to-lymphocyte ratio more than cut off value of 5.8 (p=0.005), D-dimer equal or more than 1000 ng/ml (p=0.005), pneumonia extent more than 5% (p=0.000) and pneumonia weight more than 75g (p=0.005) (Table 3). Two models of multivariate logistic regression analyses included age equal or more than 60, comorbidity, percentage of lymphocyte less than 20%, neutrophil-to-lymphocyte ration more than 5.8, LDH more than 375, D-dimer equal or more than 1000 and one of either pneumonia extent more than 5% or pneumonia weight more than 75g. The analyses identified that independent risk factors were age equal or more than 60 (p=0.0424), pneumonia extent more than 5% (p=0.004) in pneumonia extent model, and age equal or more than (p=0.011), LDH over 375 (p=0.048), pneumonia weight more than 75g (p=0.029)(Table 3).

Discussion And Conclusion

The 2D U-Net developed in this study was trained with CT images with COVID-19 pneumonia that were obtained from 17 CT scanners with varying CT parameters, including devices from the major five CT vendors that accounted for approximately 90% of the global market [26] in 2018. The correlations observed between the 2D U-Net and reference in the Korean-Chinese internal validation dataset were relatively well reproduced in the external validation datasets. The Japanese dataset used a visual CT

severity score that differed in units from pneumonia extent and weight, but a linear correlation was observed between those scores and the two outcomes of the software. The degree of correlation was slightly lower in the external datasets than in the internal validation dataset. This may partly originate from differences in the individual standard of each radiologist in discriminating subtle GGO and normal lung parenchyma (Supplementary Figure 2).

Recent publications have indicated a close relationship between clinical and radiologic severity in COVID-19 [10; 27]. When clinical severity was categorized into mild, severe, and critical cases, severe to critical COVID-19 cases had more frequent bilateral disease [3] and a greater extent of COVID-19 than mild COVID-19 cases [5]. These results imply that CT severity can be a surrogate parameter for estimating the pneumonia burden of COVID-19. The extent of COVID-19 pneumonia is a straightforward outcome indicating radiologic severity, but does not reflect the CT density of COVID-19 lesions. As the CT density of COVID-19 pneumonia varies from pure faint GGO to dense consolidation, reflecting the density and extent together by calculating the weight of pneumonia would provide a more accurate assessment of pneumonia burden than calculating the extent alone. To validate the usefulness of CT severity for severity stratification or prognostication, CT severity should be assessed using a uniform measuring tool in multiple cohorts with a sufficient number of cases. Such cohorts should be collected at multiple centers, requiring an accurate, reproducible, and easily-accessible measuring tool. A deep neural network is a potential candidate for this purpose.

Recent studies of severe COVID-19 patients reported that the elderly, and several laboratory findings such as lymphocytopenia, elevated neutrophil-to-lymphocyte ratio, elevated lactate dehydrogenase, elevated C-reactive protein, elevated procalcitonin, elevated d-dimers, elevated serum ferritin, and cardiac troponin were predictive parameters for risk stratification of COVID-19 patients [28-31]. Our univariate and multivariate analyses showed consistent results with those studies regarding clinical and laboratory parameters. In addition, automatically quantified pneumonia extent and weight were also predictive factors for the symptom presence and composite outcome of patients with COVID-19 pneumonia. These results suggest that the extent and weight of pneumonia measured in CT scans could be useful for risk stratification in COVID-19 pneumonia patients.

Measurement differences between the reference and 2D U-Net values inevitably occurred and depended on the extent and weight. The DSC, sensitivity, and PPV between the reference masks and the 2D U-Net masks were slightly lower than 80% (Table 1). Differences mainly occurred under the following circumstances: the current version of 2D U-Net sometimes misrecognized partial volumes of respiratory and cardiac motion artifacts or pulmonary vessels in the basal lungs as pneumonia (Figure 3 and 4), whereas minute lesions in the apical end of lung parenchyma tended to be missed. Those errors can be readily adjusted manually in the software.

Several limitations exist in this study. First, we retrospectively collected data from 131 patients, and their CT images may not cover a diverse range of radiologic manifestations of COVID-19. Second, we did not include sufficient numbers of COVID-19 patients who had underlying parenchymal disease (i.e., extensive

metastasis), although some CT images containing pulmonary lesions other than COVID-19 pneumonia were included in the 2D U-Net training set. Third, we tentatively validated correlations of pneumonia extent and weight with clinical parameters or outcomes. The 3D U-Net performed better in terms of minimizing misperceptions (unpublished data), but required higher computer specifications than 2D U-Net. Software based on 2D U-Net also requires high computer specifications with a particular GPU version, limiting its accessibility. Implementing this program in CT scanner consoles may be considered as a way of expanding its accessibility in locations where CT scanners exist.

In conclusion, **the extent and weight of COVID-19 pneumonia on CT images were automatically quantifiable and independently associated with symptoms and prognosis.** The quantification of COVID-19 pneumonia at multiple sites using a uniform measuring method can facilitate research toward better severity stratification and prognostication of COVID-19 patients at risk for morbidity and mortality.

Abbreviations

CT = computed tomography

COVID-19 = coronavirys disease 2019

GGO = ground glass opacity

RT-PCR = reverse transcription-polymerase chain reaction

HU = Hounsfield unit

ICCs = Intraclass correlation coefficients

DSC = Dice similarity coefficient

PPV = positive predictive value

LOA = limits of agreement

Declarations

Funding

None

Declaration of interests

Sang Joon Park is the CEO of Medical IP.

Gin Mo Goo has grants from Infinitt Healthcare, grants from Dongkook Lifescience, outside the submitted work.

All the other authors have no potential conflicts of interest to disclose.

Acknowledgement

The authors gratefully acknowledge the provision of de-identified ‘COVID data save lives’ by HM hospitales and Andrew Dombrowski, PhD (Compecs, Inc.) for his assistance in improving the use of English in this manuscript.

References

1 WHO (2020) Coronavirus disease (COVID-19) outbreak situation. World Health Organization. Available via https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed May 17 2020

2 Wu Z, McGoogan JM (2020) 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. 10.1001/jama.2020.2648

3 Guan WJ, Ni ZY, Hu Y et al (2020) Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 10.1056/NEJMoa2002032

4 Yoon SH, Lee KH, Kim JY et al (2020) Chest Radiographic and CT Findings of the 2019 Novel Coronavirus Disease (COVID-19): Analysis of Nine Patients Treated in Korea. Korean J Radiol 21:494-500

5 Yang X, Yu Y, Xu J et al (2020) Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 10.1016/S2213-2600(20)30079-5

6 Hosseiny M, Kooraki S, Gholamrezanezhad A, Reddy S, Myers L (2020) Radiology Perspective of Coronavirus Disease 2019 (COVID-19): Lessons From Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. AJR Am J Roentgenol. 10.2214/AJR.20.22969:1-5

7 Choi H, Qi X, Yoon SH et al (2020) Extension of Coronavirus Disease 2019 (COVID-19) on Chest CT and Implications for Chest Radiograph Interpretation. Radiology: Cardiothoracic Imaging 2:e200107

8 Wong HYF, Lam HYS, Fong AH-T et al (2020) Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. Radiology:201160

9 Fang Y, Zhang H, Xie J et al (2020) Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology. 10.1148/radiol.2020200432:200432

10 Zhao W, Zhong Z, Xie X, Yu Q, Liu J (2020) Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. AJR Am J Roentgenol. 10.2214/AJR.20.22976:1-6

11 Chung M, Bernheim A, Mei X et al (2020) CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). Radiology 295:202-207

12 Inui S, Fujikawa A, Jitsu M et al (2020) Chest CT Findings in Cases from the Cruise Ship “Diamond Princess” with Coronavirus Disease 2019 (COVID-19). Radiology: Cardiothoracic Imaging 2:e200110

13 Li K, Wu J, Wu F et al (2020) The Clinical and Chest CT Features Associated with Severe and Critical COVID-19 Pneumonia. Invest Radiol. 10.1097/RLI.0000000000000672

14 Lyu P, Liu X, Zhang R, Shi L, Gao J (2020) The performance of chest CT in evaluating the clinical severity of COVID-19 pneumonia: identifying critical cases based on CT characteristics. Invest Radiol. 10.1097/RLI.0000000000000689

15 Bernheim A, Mei X, Huang M et al (2020) Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. Radiology 295:200463

16 Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:150203167

17 Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networksAdvances in neural information processing systems, pp 1097-1105

18 He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classificationProceedings of the IEEE international conference on computer vision, pp 1026-1034

19 Holshue ML, DeBolt C, Lindquist S et al (2020) First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 382:929-936

20 Bell DJ (2020) COVID-19. Available via https://radiopaedia.org/articles/covid-19-3. Accessed May 17 2020

21 Jun M, Cheng G, Yixin W et al (2020) COVID-19 CT Lung and Infection Segmentation Dataset. Available via https://zenodo.org/record/3757476#.Xp_VDMgzbPt Accessed May 17 2020

22 HMhospitales (2020) Covid Data Save Lives. Available via https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version. Accessed May 17 2020

23 Yoon J-H, Yoon SH, Hahn S (2019) Development of an algorithm for evaluating the impact of measurement variability on response categorization in oncology trials. BMC Medical Research Methodology 19:90

24 Mansoor A, Bagci U, Foster B et al (2015) Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. Radiographics 35:1056-1076

25 Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15:29

26 Davidson A (2018) CT Equipment Market: Made in China...and the USA. Available via https://technology.informa.com/606142/ct-equipment-market-made-in-chinaand-the-usa. Accessed April 4th 2020

27 Yang R, Li X, Liu H et al (2020) Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiology: Cardiothoracic Imaging 2:e200047

28 Qin C, Zhou L, Hu Z et al (2020) Dysregulation of immune response in patients with COVID-19 in Wuhan, China. Clin Infect Dis. 10.1093/cid/ciaa248

29 Zhou F, Yu T, Du R et al (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395:1054-1062

30 Tan L, Wang Q, Zhang D et al (2020) Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther 5:33

31 Velavan TP, Meyer CG (2020) Mild versus severe COVID-19: Laboratory markers. Int J Infect Dis 95:304-307

Tables

Table 1. Performance of the network predicting COVID-19 pneumonia extent and weight in CT scans, evaluated with correlation coefficients, calculations of spatial overlaps and Bland-Altman analysis in internal and external validation datasets.

Datasets	Internal validation		External validation							
			Japan (1st)		Italy (2nd)		Radiopaedia (3rd)		China (4th)	
	Extent	Weight	Extent	Weight	Extent	Weight	Extent	Weight	Extent	Weight
Correlation coefficients										
ICC	0.990	0.993			0.951	0.995	0.964	0.976	0.970	0.994
Pearson correlation coefficient			0.908	0.899						
Calculation of spatial overlaps of AI-driven mask and reference mask										
Dice	77.8±17.1%				73.4±14.0%		71.9±25.9%		77.0±10.9%	
sensitivity	81.4±10.2%				70.6±18.3%		79.9±11.8%		70.7±15.4%	
PPV	80.3±20.6%				80.6±10.1%		71.9±27.6%.		87.7±5.4%	
Bland-Altman Analysis										
95% LOA										
10% extent	-4.0% to 5.4%				-10.1 to 15.1%		-4.0 to 7.4%		-4.6 to 8.4%	
50% extent	-9.7% to 11.1%				-25.6 to 30.6%		-11.0 to 14.4%		-12.6 to16.4%	
10g						-3.6 to 3.5g				
50g						-7.9 to 7.8g				
100g		-41.4 to 55.5g						-57.9 to 97.2g		-32.4 to 68.2g
500g		101.3 to 115.4g						-153.8 to 193.1g		-94.5 to 130.3g

ICC=intraclass correlation coefficient, LOA=Limit of agreement

Table 2. Univariate and Multivariate Logistic Regression Analysis of Risk Factors for presence of symptom in the Japanese dataset.

	Univariate analysis			Multivariate analysis 1 – pneumonia extent model			Multivariate analysis 2 – pneumonia weight model		
Variables	**Odd ratios**	**95% confidence interval**	**P-value**	**Odd ratios**	**95% confidence interval**	**P-value**	**Odd ratios**	**95% confidence interval**	**P-value**
Age <60 years (vs. ≥60 years)	2.482	1.009-6.002	**0.048***	5.007	1.632-15.363	**0.005***	4.917	1.610-15.023	**0.005***
Male (vs. female)	0.675	0.280-1.631	0.383						
Smoking (vs. non-smoker)	1.934	0.667-5.605	0.224						
Presence of comorbidity (vs. absence of comorbidity)	0.880	0.360-2.152	0.780						
Lymphocytopenia <1000 cells/µl (vs. ≥1000 cells/µl)	3.332	1.260-8.811	**0.015***	3.265	1.067-9.989	**0.038***	3.106	1.272-12.323	**0.048***
Thrombocytopenia <150,000 cells/µl (vs. ≥150,000 cells/µl)	1.775	0.394-7.989	0.455						
LDH >245 U/L (vs. ≤245 U/L)	3.300	1.166-9.341	**0.025***						
C-reactive protein (CRP) >1 mg/dL (vs. ≤1 mg/dL)	2.215	0.872-5.627	0.094						
Aspartate aminotransferase (AST) >40 U/L (vs. ≤ 40 U/L)	1.295	0.364-4.608	0.690						
Alanine aminotransferase (ALT) >40U/L (vs. ≤ 40 U/L)	0.852	0.279-2.603	0.779						
AI-driven pneumonia extent >0.5% (vs. ≤0.5%)	3.171	1.246-8.065	**0.015***	4.142	1.356-12.647	**0.013***			
AI-driven pneumonia weight >10g (vs. ≤10g)	3.560	1.394-9.091	**0.008***				4.434	1.455-13.513	**0.009***

*Statistical significane below 0.05.

Table 3. Univariate and Multivariate Logistic Regression Analysis of Risk Factors for composite outcome in the Spanish dataset.

	Univariate analysis			Multivariate analysis 1 – pneumonia extent model			Multivariate analysis 2 – pneumonia weight model		
Variables	Odd ratios	95% confidence interval	P-value	Odd ratios	95% confidence interval	P-value	Odd ratios	95% confidence interval	P-value
Age≥ 60 years (vs. <60 years)	4.895	1.535-15.613	**0.007***	4.366	1.215-15.686	**0.024***	5.083	14.459-17.712	**0.011***
Male (vs. female)	0.698	0.287-1.696	0.427						
Comorbity†	1.837	0.739-4.571	0.191						
Lymphopenia <1000 cells/µl (vs. ≥1000 cells/µl)	1.034	0.426-2.514	0.940						
Percentage of lymphocyte <20% (vs. ≥20%)	3.335	1.134-9.809	**0.029***						
Neutrophil to lymphocyte ratio >5.8 (vs <=5.8)	3.779	1.510-9.497	**0.005***						
LDH >375 U/L (vs. ≤375 U/L)	10.260	1.302-80.874	**0.027***	6.527	0.684-62.260	0.103	9.062	1.016-80.869	**0.048***
D-dimer ≥1000 ng/ml (vs. <1000 ng/ml)	4.240	1.530-11.748	0.005	2.988	0.934-9.565	0.065	2.848	0.916-8.853	0.070
AI-driven pneumonia extent >5% (vs. ≤5%)	10.833	2.986-39.303	**0.000***	7.446	1.899-29.193	**0.004***			
AI-driven pneumonia weight >75g (vs. ≤75g)	6.410	1.767-23.256	**0.005***				4.677	1.167-18.733	**0.029***

*Statistical significane below 0.05.

†Comorbidity includes hypertension, diabetes mellitus type 2, coronary artery disease, heart failure, cerebrovascular disease, chronic kidney disease

Figures

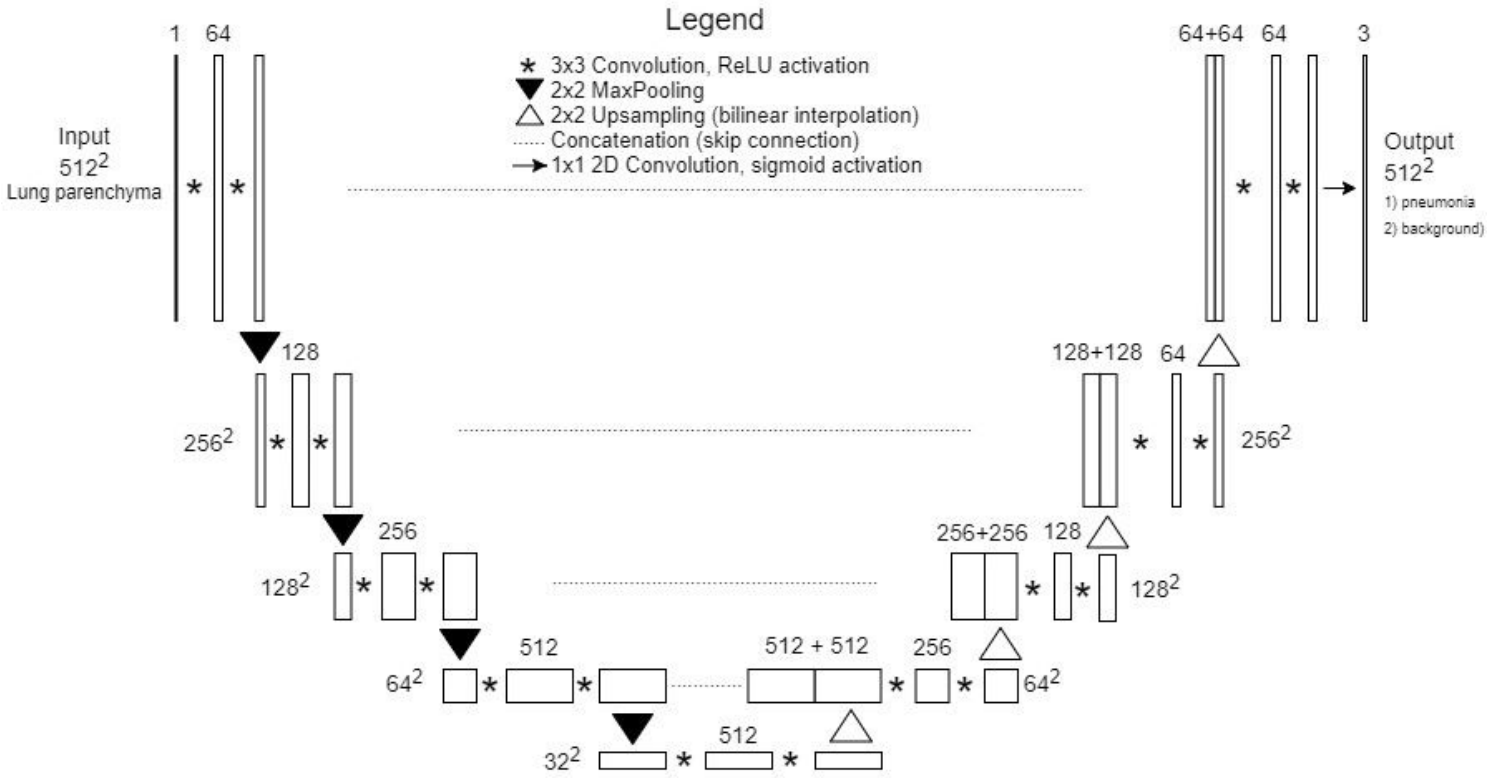

Figure 1

2D U-Net architecture for COVID-19 pneumonia segmentation.

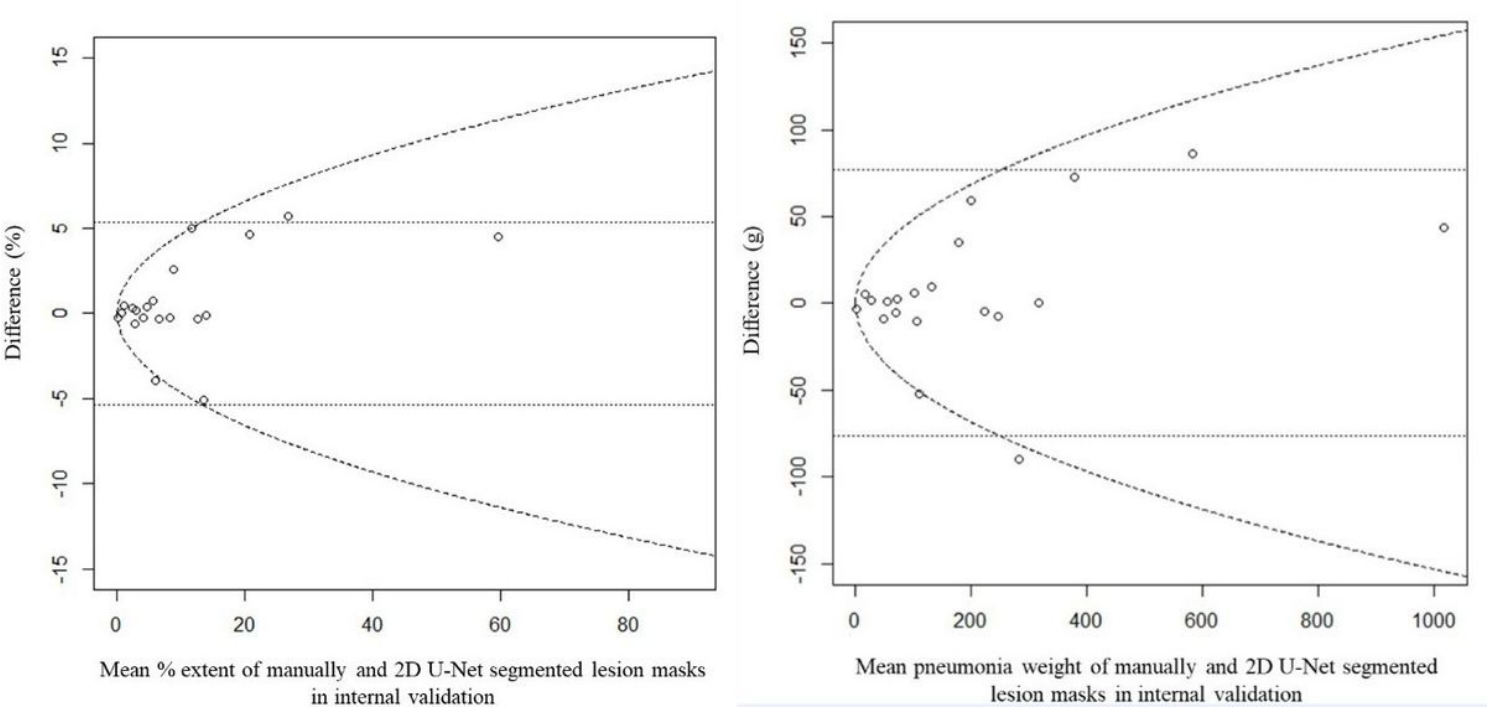

Figure 2

Bland-Altman plot and limits of agreement for pneumonia extent. Left(A) and weight Right (B) between the reference and 2D U-Net masks in the internal validation dataset. Horizontal lines: limits of agreement

(LOA) from the model using the original values of extent and weight of pneumonia; curved lines: LOA from the model using the square root-transformation of the extent and weight of pneumonia.

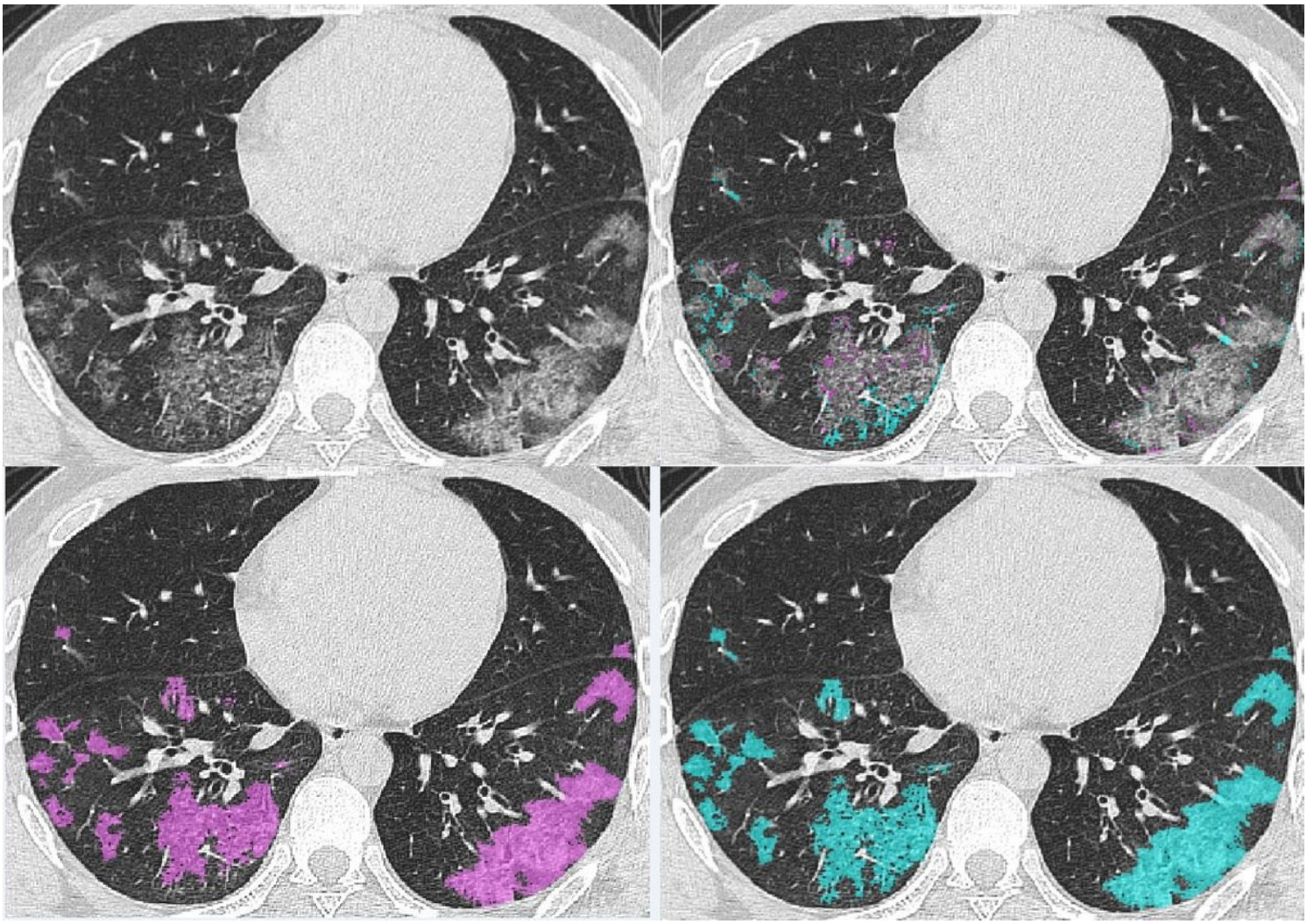

Figure 3

Representative images of a 32-year-old man with COVID-19 pneumonia in the internal validation dataset. A chest CT image shows peripheral and peribronchial ground-glass opacities in both lungs.Top Left (A). The reference mask Bottom Left: (C) and 2D U-Net mask. Bottom right:(D) match in most areas of the lesions except the blurry peripheral margin of the ground-glass opacity visible in the subtracted mask. Top right (B). The Dice similarity coefficient, sensitivity, and positive predictive value were 90.2%, 91.4%, and 88.9%, respectively.

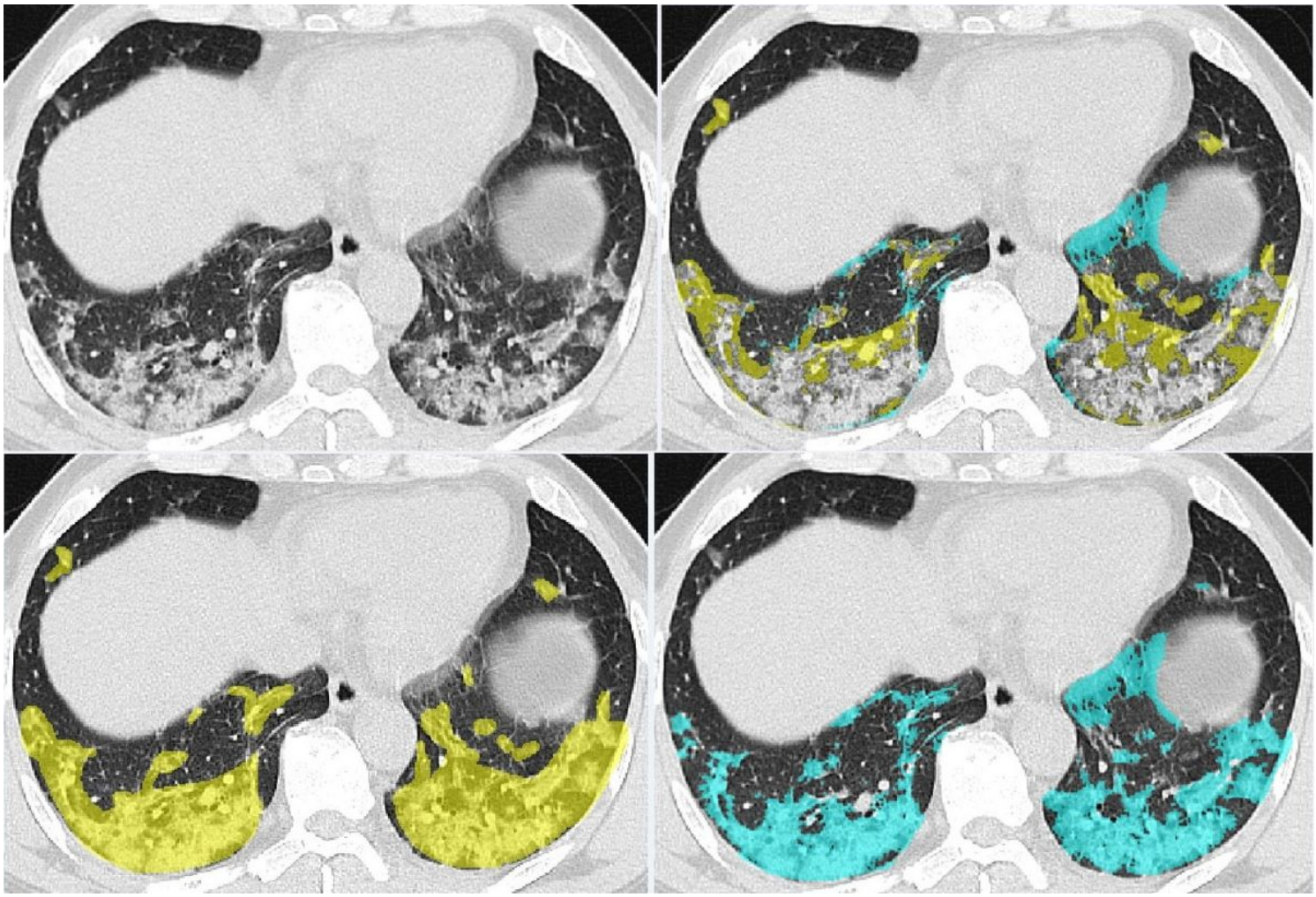

Figure 4

Representative images of a COVID-19 pneumonia in the fourth external validation dataset. A chest CT image shows peribronchial and subpleural ground-glass opacities in the right and left lower lobes Top Left (A). Subtraction Top Right(B) of the reference mask Bottom Left(C) and 2D U-Net mask Bottom Right(D) shows inaccurate lesion segmentation by 2D U-Net in the left retrocardiac and basal lung as pneumonia, whereas it was actually an artifact due to cardiac and diaphragm motion. However, other areas of pneumonia were segmented precisely. The Dice similarity coefficient, sensitivity, and positive predictive value were 84.9%, 77.0%, and 94.4%, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.png
- FigureS1final.jpg
- FigureS2final.jpg