Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction

Fei Shan*
Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China

Yaozong Gao*
Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200232, China

Jun Wang*
Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai Institute for Advanced Communication and Data Science, School of Communication & Information Engineering, Shanghai University, Shanghai 200444, China

Weiya Shi and Nannan Shi
Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China

Miaofei Han and Zhong Xue
Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200232, China

Dinggang Shen[a]
School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200232, China
Department of Artificial Intelligence, Korea University, Seoul 02841, Republic of Korea

Yuxin Shi[a]
Department of Radiology, Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China

Objective: Computed tomography (CT) provides rich diagnosis and severity information of COVID-19 in clinical practice. However, there is no computerized tool to automatically delineate COVID-19 infection regions in chest CT scans for quantitative assessment in advanced applications such as severity prediction. The aim of this study was to develop a deep learning (DL)-based method for automatic segmentation and quantification of infection regions as well as the entire lungs from chest CT scans.

Methods: The DL-based segmentation method employs the "VB-Net" neural network to segment COVID-19 infection regions in CT scans. The developed DL-based segmentation system is trained by CT scans from 249 COVID-19 patients, and further validated by CT scans from other 300 COVID-19 patients. To accelerate the manual delineation of CT scans for training, a human-in-volved-model-iterations (HIMI) strategy is also adopted to assist radiologists to refine automatic annotation of each training case. To evaluate the performance of the DL-based segmentation system, three metrics, that is, Dice similarity coefficient, the differences of volume, and percentage of infection (POI), are calculated between automatic and manual segmentations on the validation set. Then, a clinical study on severity prediction is reported based on the quantitative infection assessment.

Results: The proposed DL-based segmentation system yielded Dice similarity coefficients of $91.6\% \pm 10.0\%$ between automatic and manual segmentations, and a mean POI estimation error of 0.3% for the whole lung on the validation dataset. Moreover, compared with the cases with fully manual delineation that often takes hours, the proposed HIMI training strategy can dramatically reduce the delineation time to 4 min after three iterations of model updating. Besides, the best accuracy of severity prediction was $73.4\% \pm 1.3\%$ when the mass of infection (MOI) of multiple lung lobes and bronchopulmonary segments were used as features for severity prediction, indicating the potential clinical application of our quantification technique on severity prediction.

Conclusions: A DL-based segmentation system has been developed to automatically segment and quantify infection regions in CT scans of COVID-19 patients. Quantitative evaluation indicated high accuracy in automatic infection delineation and severity prediction. © *2020 American Association of Physicists in Medicine* [https://doi.org/10.1002/mp.14609]

1. INTRODUCTION

The outbreak of 2019 novel coronavirus in Wuhan, China has rapidly spread to other countries since Dec 2019.[1–7] The infectious disease caused by this virus was named as COVID-19 by the World Health Organization (WHO) on Feb 11, 2020.[8] To date (July 23rd, 2020), there have been 14 765 256 confirmed cases reported all around the world.[9] Each suspected case needs to be confirmed by real-time polymerase chain reaction (RT-PCR) assay of the sputum.[10] Although it is the gold standard for diagnosis, confirming COVID-19 patients using RT-PCR is time-consuming in many countries and has been reported to suffer from high false negative rates. On the other hand, because chest computed tomography (CT) scans collected from COVID-19 patients often show typical features such as bilateral multifocal patchy consolidation or ground glass opacities (GGO) in the lung,[11,12] it has been used as an important complementary indicator in COVID-19 screening due to high sensitivity.[13,14]

Due to fast progression of COVID-19, a considerable proportion of COVID-19 patients will progress to severe or even critically ill stage. According to a retrospective study on 138 hospitalized COVID-19 patients at Zhongnan Hospital of Wuhan University,[12] 26.1% of the patients were transferred to the intensive care unit (ICU) after enrollment. The median time from first symptom to dyspnea was only 5.0 days, and to acute respiratory distress syndrome (ARDS) was only 8.0 days. Similar observations were also reported by Chen et al.[15] and Huang et al.[11] Thus, timely identification of patients who may progress to the severity stage at the early stage is pivotal for subsequent active intervention.[16] Although CT provides rich imaging information, it only provides qualitative evaluation in the radiological reports owing to the lack of computerized tools to accurately quantify the infection regions and their longitudinal changes. Besides, contouring infection regions in the chest CT is necessary for quantitative assessment; however, manual contouring of lung lesions is a tedious and time-consuming work, and inconsistent delineation could also lead to subsequent assessment discrepancies. Thus, a fast auto-contouring tool for COVID-19 infection is urgently needed in the onsite applications for quantitative disease assessment.

To this end, we developed a deep learning (DL)-based segmentation system for quantitative infection assessment. The system not only performs auto-contouring of infection regions, but also accurately estimates their volumes and percentage of infection (POI) in CT scans of COVID-19 patients. In order to provide delineation for hundreds of the training COVID-19 CT scans, which is a tedious and time-consuming work, we proposed a human-involved-model-iterations (HIMI) strategy to iteratively generate the training samples. This method involves radiologists to efficiently intervene DL-based segmentation results and iteratively add more training samples to update the model, and thus greatly accelerates the algorithm development cycle. To the best of our knowledge, there are no literature that have reported the utilization of HIMI strategy in delineating COVID-19 infections in chest CT scans.

2. MATERIALS AND METHODS

2.A. Datasets

The protocol of this retrospective study was approved by the Ethics Committees of Shanghai Public Health Clinical Center and other centers outside Shanghai. Informed consent was waived because of the retrospective nature of the study, and all the private information of patients was anonymized by the investigators after data collection. Totally 305 CT scans from 305 COVID-19 patients (from Shanghai) were collected for validation. About 249 CT scans of 249 COVID-19 patients were collected from other centers (outside Shanghai) for training the segmentation network. Among 305 validation patients, 97 patients were considered as severe cases, which follows the clinical classification criterion in *The Handbook of COVID-19 Prevention and Treatment*.[17] Seven patients died eventually. All COVID-19 patients were confirmed with a laboratory examination through real-time PCR (RT-PCR) detection by the local Center for Disease Control (CDC), and rechecked by national CDC. According to the Hospitalization Information System, the average hospital stay for COVID-19 patients was about 9 days. The inclusion criteria are listed as follows:

a Patients with a positive new coronavirus nucleic acid and confirmed by the CDC;
b Age >= 18;
c Chest CT showed pneumonia.

Since the 249 CT scans were only used in training the segmentation network and not used for severity prediction, the severity information is not provided by the hospitals. Five COVID-19 Patients (from Shanghai) with CT scans showing obvious motion artifacts or pre-existing lung cancer conditions were excluded in this study. Therefore, the total number of patients used for validation is 300.

2.B. Image acquisition parameters

All COVID-19 patients underwent thin-section CT scan. The CT scanners used in our study included uCT780 from UIH, Optima CT520, LightSpeed 16 from GE, Aquilion ONE from Toshiba, SOMATOM Force from Siemens, and Scenaria 64 CT from Hitachi. The median duration from illness onset to CT scan was 4 days, ranging from 1 to 14 days. The CT protocol was as follows: 120 kV; automatic tube current (180–400 mA); iterative reconstruction; 64 mm detector; rotation time, 0.35 s; slice thickness, 1 mm; collimation, 0.625 mm; pitch, 1.5; matrix, 512 × 512; and breath hold at full inspiration. The reconstruction kernel used is set as "lung smooth with a thickness of 1 mm and an interval of 0.8 mm". During reading, the lung windows (with window width 1200 HU and window level-600 HU) were used.

2.C. DL-based segmentation network: VB-Net

Although segmentation methods have been fully investigated in medical imaging applications,[18,19] segmenting infection regions from the chest CT scans is still very challenging due to the low contrast of the infection regions manifested as GGO in CT images and large variation of both shape and position across different patients. We therefore developed a DL-based network called VB-Net, for this purpose. It is a modified 3D convolutional neural network that combines V-Net[20] with a bottle-neck structure.[21] Similar to V-Net,[20] VB-Net consists of two paths (Fig. 1). The first path is a contracting path including down-sampling and convolution operations to extract global image features. The second path is an expansive path including up-sampling and convolution operations to integrate fine-grained image features. In the contracting path, the number of channels after the first convolution layer is 16. The number of channels doubles, and the spatial size decreases by half after each down-sampling layer. In the expansive path, the number of channels decreases by half, and the spatial size doubles after each up-sampling layer. In the VB-Net, the bottle-neck structure replaces the $5 \times 5 \times 5$ convolution operation in the V-Net by a sub-network including three convolutional layers. Specifically, the first convolutional layer reduces the channels of feature maps by applying $1 \times 1 \times 1$ convolution kernel. The second convolutional layer performs spatial convolution with $3 \times 3 \times 3$ kernel sizes. The last convolutional layer increases the channels of feature maps by applying $1 \times 1 \times 1$ convolution kernel. Compared with V-Net,[20] the speed of VB-Net is much faster because the bottle-neck structure is integrated in the VB-Net, as detailed in Fig. 1.[22,23]

2.D. Training VB-Net with human-involved-model-iterations (HIMI) strategy

Training samples with detailed delineation of each infection region are required for the proposed VB-Net. However, it is a labor-intensive work for radiologists to manually annotate hundreds of COVID-19 CT scans. We, therefore, adopted the human-involved-model-iterations (HIMI) strategy to iteratively update the DL model. Specifically, the training data are divided into several batches, with 30–50 CT scans in each batch. First, radiologists manually contour the CT data in the smallest batch (i.e., 36 CT scans). Then, the segmentation network is trained by this batch as an initial model. This initial model is applied to segment infection regions in the next batch, which will be manually corrected by radiologists. These corrected segmentation results together with the first batch contoured by the radiologists are then combined as new training data, and the model can be updated with increased training dataset. In this way, we iteratively increase the training dataset by adding the newly corrected batch to the previous ones. When the segmentation accuracy on the new batch becomes stable, the training process converges and the final segmentation network is constructed. In the testing stage, the trained segmentation network segments the infection regions
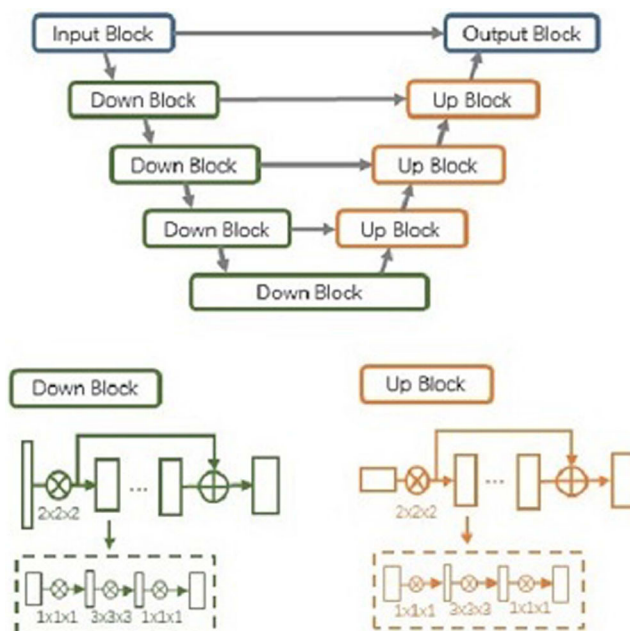

Fig. 1. The network structure for COVID-19 infection segmentation. The dashed boxes show the bottle-neck structures inside the V-shaped network. [Color figure can be viewed at wileyonlinelibrary.com]

on a new CT scan via a forward propagation of neural network. According to our experience, this HIMI training strategy converged after 3-4 batches were used in the training process. Figure 2 illustrates the process of the proposed HIMI training strategy.

2.E. Quantification and assessment of COVID-19 infection

Besides COVID-19 infection regions, the whole lung, lung lobes and bronchopulmonary segments of each subject were also segmented using our system. To acquire the ground-truth for bronchopulmonary segments, radiologists annotated bronchopulmonary segments on CT images, which we used as the ground-truth for bronchopulmonary segmentation. Specifically, radiologists first annotated lung vessels and bronchus using semi-automated tools on 3D Slicer.[24] Then, based on the 3D surface rendering of vessels and bronchus, radiologists divide the lung volume into multiple bronchopulmonary segments.

After segmentation, various metrics were computed to quantify the COVID-19 infection, including volumes of infection in the whole lung, and volumes of infection in each lobe and each bronchopulmonary segment. In addition, the POIs in the whole lung, each lobe and each bronchopulmonary segment were also computed, respectively, to measure the severity of COVID-19 and the distribution of infection within the lung. The Hounsfield unit (HU) histogram within the infection region can also be visualized for evaluation of GGO (-750 to -300 HU) and consolidation components (-300 to 50 HU) inside the infection region delineated by our system.[25] Besides, a small number of voxels with HU values not
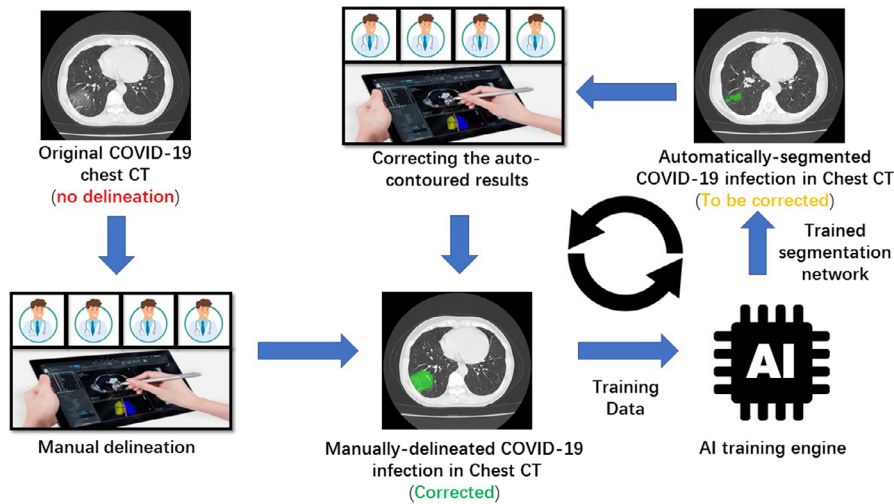
FIG. 2. The human-involved-model-iterations (HIMI) workflow. [Color figure can be viewed at wileyonlinelibrary.com]

falling within the interval $[-750, 50]$ are surrounded by GGOs or consolidation area. They are also delineated as the infection area by the system.

Figure 3 shows the entire pipeline for quantitative COVID-19 assessment. A chest CT scan is first fed to the DL-based segmentation system, which generates infection areas, the whole lung, lung lobes, and all the bronchopulmonary segments, respectively. Then, the aforementioned quantitative metrics are calculated to quantify infection regions of the patient. The quantification provides the basis for measuring the severity of COVID-19 from the CT perspective and for tracking longitudinal changes during the treatment course.

2.F. Quantitative evaluation on segmentation and measurement accuracy

To quantitatively evaluate the accuracy of segmentation and measurement, infection regions on 300 CT scans of 300 COVID-19 patients were first manually contoured by two radiologists (W.S. and F.S., with 12 and 19 yr of experience in chest radiology, respectively). Each case was manually contoured by one radiologist and then reviewed by the other. In case of disagreement, the final results were determined by consensus between the two radiologists. After that, the infection regions were automatically segmented by the system. Finally, the automatically segmented infection regions were compared to the reference standard in terms of overlap ratio (measured by Dice similarity coefficient), volume, the percentage of infection (POI) in the whole lung, POI in each lung lobe, and POI in each bronchopulmonary segment.

Due to the time-consuming manual labeling process, the inter-rater variability was assessed by randomly sampling 50 CT scans of COVID-19 patients from the entire validation set. The two radiologists first independently contoured the infection regions in these CT scans. Their manual segmentation was then compared using the same metrics as mentioned above.

In our study, both VB-Net and V-Net[20] had identical hyper-parameter settings including the number of channels

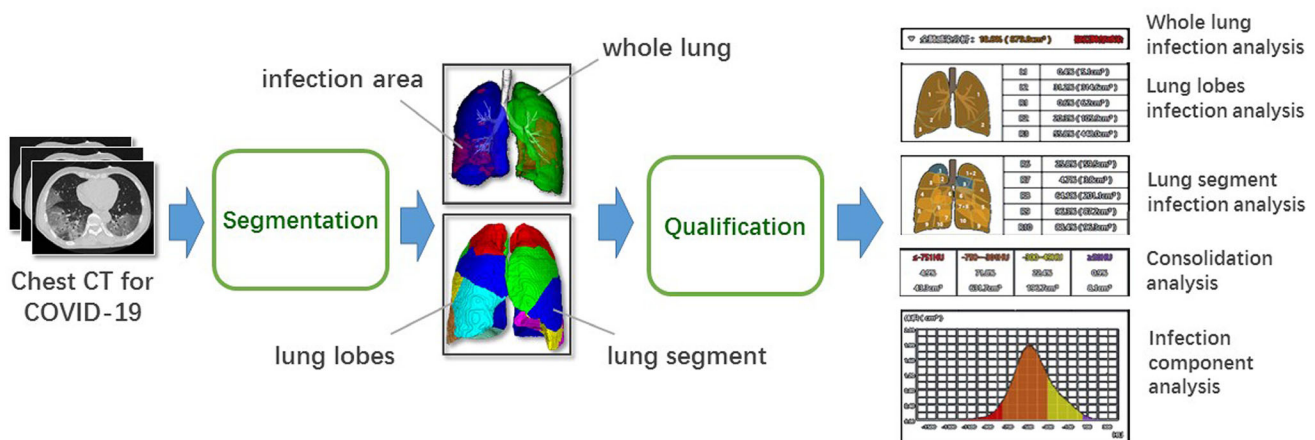FIG. 3. Pipeline for quantifying COVID-19 infection. A chest computed tomography (CT) scan is first fed into the DL-based segmentation system. Then, quantitative metrics are calculated to characterize infection regions in the CT scan, including (but not limited to) infection volumes and POIs in the whole lung, lung lobes and bronchopulmonary segments. [Color figure can be viewed at wileyonlinelibrary.com]

and the filter sizes in each layer. In training the network, the Adam optimizer was used. We used the default setting of Pytorch implementation. The initial learning rate was set as $\alpha = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$. The training was stopped when loss values were stabilized.

To show the advantage of VB-Net over classical deep learning methods, the segmentation results obtained by VB-Net were also compared with those obtained by U-Net,[26–28] which was popularly used in medical image segmentation. Notice that our experiments were carried out in the same experimental condition, and the U-Net segmentation experiments also used the HIMI training strategy.

2.G. Clinical use of quantification

One clinical application of the above quantification technique is performing severity prediction for the COVID-19 patients. Based on the segmentation results obtained from the CT scans taken from COVID-19 patients on their first-day enrollment, the quantified radiological features, including the percentage of consolidation (POC), the percentage of infection (POI), and the mass of infection, were calculated. Here, POC was calculated as the ratio between the volume of consolidation and the whole infection volume of the lung lobes or bronchopulmonary segments, and POI was calculated as the ratio between the infection volume and the volume of the lung lobes or bronchopulmonary segments. The mass of infection region was calculated as Ref. [29]

$$M = V \cdot [(A_{mean} + 1000) \times 0.001] \qquad (1)$$

where M is the mass in milligrams per millimeter, V is the infection volume, and A_{mean} is the mean attenuation in Hounsfield units. MOI, POC and POI of 5 lung lobes and 18 bronchopulmonary segments were calculated for each patient. Table I lists the lung regions, from which the radiological features are extracted.

The pneumonia severity index (PSI) is a clinical prediction rule that is often used to calculate the probability of morbidity and mortality among patients with community acquired pneumonia.[30,31] It is calculated based on demographics, coexistence of comorbidity illnesses, and abnormal physical and laboratory examinations. To verify whether PSI was suitable for severity prediction for COVID-19 patients, we also did severity prediction experiments using only PSI. In our experiments, PSI was estimated for each patient on his/her first-day enrollment.

In our experiments of severity prediction, the 300 validation scans were used, where 97 cases were severe cases. We performed 10-folded validation and repeated the experiments 20 times.

Since both severe and nonsevere cases are highly mixed in the feature space, the classical linear classifiers cannot obtain satisfactory results. We thus used the support vector machine (SVM) as the classifier for severity prediction. To handle data imbalance between the severe and nonsevere groups, we modified the classical C-SVM and constructed a cost-sensitive

TABLE I. The lung regions where the radiological features are extracted.

Categories	Lung regions
Lung lobes	Left upper lobe
	Left lower lobe
	Right upper lobe
	Right middle lobe
	Right lower lobe.
Bronchopulmonary segments	Left upper lobe/apical posterior segment
	Left upper lobe/anterior segment
	Left upper lobe/superior lingular segment
	Left upper lobe/inferior lingular segment
	Left lower lobe/superior segment
	Left lower lobe/anteromedial basal segment
	Left lower lobe/lateral basal segment
	Left lower lobe/posterior basal segment
	Right upper lobe/apical segment
	Right upper lobe/posterior segment
	Right upper lobe/anterior segment
	Right middle lobe/lateral segment
	Right middle lobe/medial segment
	Right lower lobe/superior segment
	Right lower lobe/medial basal segment
	Right lower lobe/anterior basal segment
	Right lower lobe/lateral basal segment
	Right lower lobe/posterior basal segment

learning criterion to learn the hyperplane of the classifier as follows:

$$\min_{\boldsymbol{\omega},b} \frac{1}{2}\| \mathbf{w} \|^2 + C \sum_{i=1}^{N_1} \frac{1}{N_1}\xi_i + C \sum_{i=N_1+1}^{N_1+N_2} \frac{1}{N_2}\xi_i$$
$$s.t. \, y_i(\mathbf{w}^T \varphi(\mathbf{x}_i) + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0, \quad i = 1, \dots, N$$

where N_1 and N_2 are the sample numbers of the severe and nonsevere groups, respectively, and $N = N_1 + N_2$ is the total number of samples. The Gaussian kernel function was used in our experiments.

2.H. Statistical analysis and evaluation metrics

Statistical analysis was performed by R version 3.6.1 (R Project for Statistical Computing, Vienna, Austria). Because a majority of the continuous data did not follow a normal distribution, they were expressed as the median and interquartile range (IQR, 25th and 75th percentiles).

In severity prediction, accuracy, sensitivity, specificity, and AUC were used as metrics to evaluate the classification performance. The AUC was used as the criterion for selecting the optimal hyper-parameters for SVM.

3. RESULT

3.A. Visualization of delineating infection regions

To demonstrate the effectiveness, Fig. 4 shows typical cases of COVID-19 infection in three different stages: early stage, progressive stage and peak stage.[32] Coronal images without and with overlaid segmentation are presented in

(a) A fifty-eight years old male (early stage)

(b) A fifty-six years old female (progressive stage)

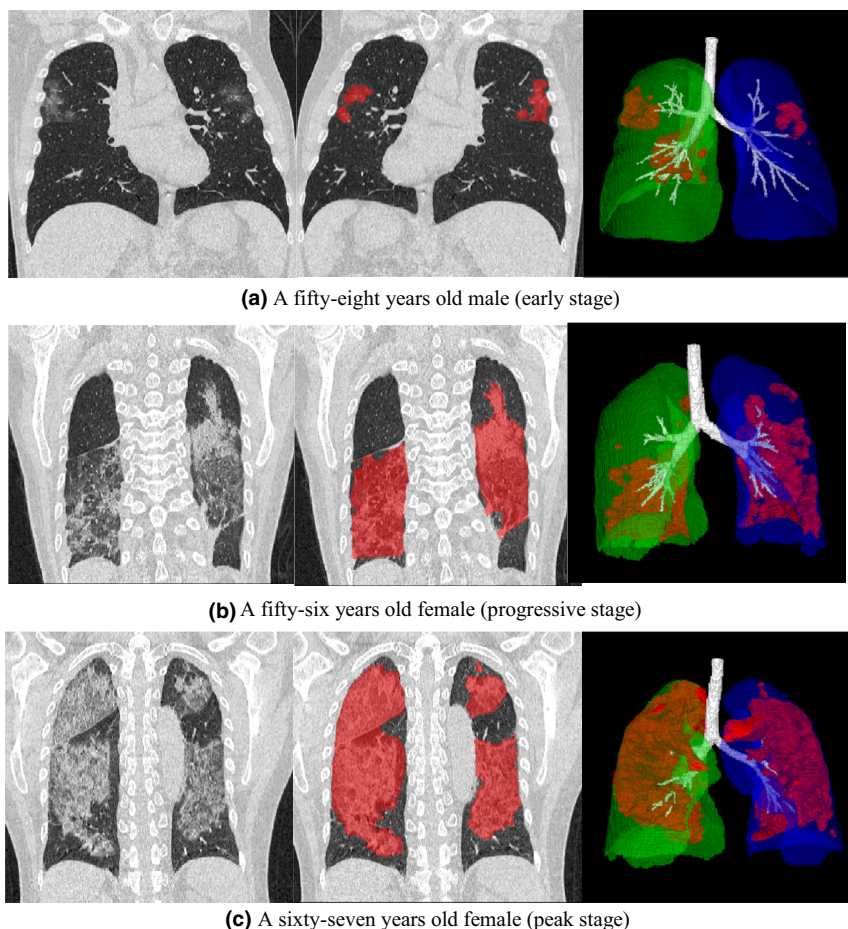**(c)** A sixty-seven years old female (peak stage)

FIG. 4. Typical infection segmentation results of computed tomography (CT) scans of three COVID-19 patients. Rows 1–3: early, progressive and peak stages. Columns 1–3: CT image, CT scans overlaid with infection segmentation, and 3D rendering of segmented infections. (a) CT of a fifty-eight years old male in the early stage; (b) CT of a fifty-six years old feamale in the progressive stage; (c) CT of a fifty-seven years old feamale in the peak stage. [Color figure can be viewed at wileyonlinelibrary.com]

parallel for comparison. In addition, 3D rendering of each case is also provided to give a more vivid understanding of COVID-19 infection within the lung. All three cases show that the contours delineated by the deep learning system match well with the visible lesion boundaries in CT scans.

3.B. Quantitative evaluation on segmentation and measurement accuracy

Table II shows the statistics of these evaluations on segmenting the infection regions in the lung. The average Dice similarity coefficient is 91.6% ± 10.0% (median 92.2%, IQR 89.0%–94.6%, range 9.6%–98.1%). The mean POI estimation errors are 0.3% for the whole lung, 0.5% for lung lobes, and 0.8% for bronchopulmonary segments. About 86.7% of lung-lobe POIs and 81.6% of bronchopulmonary-segment POIs are accurately estimated with differences equal or <1%.

Typical failure cases are shown in Fig. 5. They are CT scans with very minor symptoms. These cases show small lesions on CT scans, for example, a small area of GGO shown in the bottom of lung. In such cases, our algorithm may miss the lesion, thus causing a small Dice ratio.

Table III lists the quantitative comparison results of inter-rater variability analysis between two radiologists. The average Dice similarity coefficient between the two radiologists is 96.1% ± 3.5% (median 97.2%, IQR 95.4%–98.3%, range 86.5%–99.0%). The mean POI estimation difference is 0.2% for whole lung, 0.3% for lung lobes, and 0.4% for bronchopulmonary segments. About 91.4% of lung-lobe POIs and 85.9% of bronchopulmonary-segment POIs are consistently estimated with equal or less than 1% difference.

By comparing Tables II and III, it can be seen that the segmentation and measurement errors of the deep learning system are close to the inter-rater variability. This demonstrates the precision level of using deep learning to quantify the COVID-19 infection in CT scans.

Table IV compares the results of VB-Net and U-Net on segmenting the COVID-19 infection regions in the 300 infection scans. Figure 6 further gives some comparison results. The results in both Table IV and Fig. 6 indicate that our VB-Net outperforms the U-Net on segmenting the COVID-19 infection regions in the CT scans.

Based on the segmentation results, we can explore the quantitative lesion distribution specifically related to

TABLE II. Quantitative evaluation of the deep learning segmentation system on the validation dataset. The Dice coefficients, and POI estimation error in the whole lung, lung lobes and bronchopulmonary segments, were calculated to assess the automatic segmentation accuracy. * indicates no significant difference between automatic and manual ground-truth segmentations of the validation dataset according to paired *t*-test.

| Accuracy metrics	Mean	Standard deviation	Median	25% IQR	75% IQR	Number of infected samples
Dice Similarity Coefficient	91.6%	10.0%	92.2%	89.0%	94.6%	300
POI Error (The whole lung)*	0.3%	0.4%	0.1%	0.0%	0.4%	300
POI Error (Left upper lobe)	0.4%	1.0%	0.1%	0.0%	0.4%	233
POI Error (Left lower lobe)*	0.7%	1.6%	0.3%	0.1%	1.0%	267
POI Error (Right upper lobe)	0.3%	0.7%	0.1%	0.0%	0.5%	213
POI Error (Right middle lobe)	0.3%	0.7%	0.1%	0.0%	0.5%	204
POI Error (Right lower lobe)	0.6%	1.1%	0.3%	0.1%	0.9%	275
POI Error (Left upper lobe/apical posterior)	0.5%	1.0%	0.1%	0.0%	0.5%	189
POI Error (Left upper lobe/anterior)	0.5%	1.2%	0.2%	0.0%	0.5%	158
POI Error (Left upper lobe/superior lingular)	0.7%	1.7%	0.2%	0.0%	0.9%	192
POI Error (Left upper lobe/inferior lingular)	0.7%	1.8%	0.2%	0.0%	0.8%	175
POI Error (Left lower lobe/superior)*	0.9%	2.1%	0.4%	0.1%	1.2%	224
POI Error (Left lower lobe/anteromedial basal)	0.6%	1.4%	0.2%	0.0%	0.8%	209
POI Error (Left lower lobe/lateral basal)*	1.1%	2.5%	0.5%	0.1%	1.7%	228
POI Error (Left lower lobe/posterior basal)*	1.1%	2.4%	0.5%	0.1%	1.6%	233
POI Error (Right upper lobe/apical)	0.4%	1.1%	0.1%	0.0%	0.5%	142
POI Error (Right upper lobe/posterior)	0.7%	1.7%	0.2%	0.0%	0.8%	186
POI Error (Right upper lobe/anterior)	0.4%	1.1%	0.1%	0.0%	0.9%	151
POI Error (Right middle lobe/lateral)	0.6%	1.5%	0.1%	0.0%	0.6%	183
POI Error (Right middle lobe/medial)*	0.3%	0.8%	0.1%	0.0%	0.4%	167
POI Error (Right lower lobe/superior)	0.9%	1.9%	0.4%	0.1%	1.4%	233
POI Error (Right lower lobe/medial basal)*	0.6%	1.4%	0.3%	0.1%	0.9%	162
POI Error (Right lower lobe/anterior basal)	0.6%	1.4%	0.1%	0.0%	0.9%	210
POI Error (Right lower lobe/lateral basal)	0.9%	1.8%	0.4%	0.1%	1.2%	236
POI Error (Right lower lobe/posterior basal)	1.0%	2.0%	0.5%	0.1%	1.6%	249
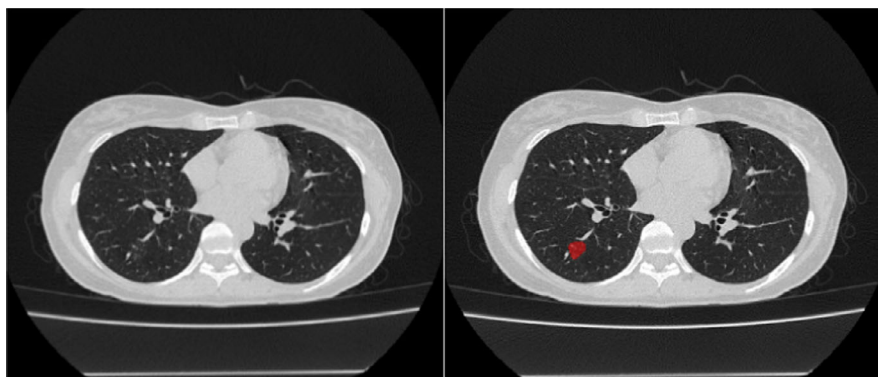
FIG. 5. A typical failure case with small lesions. (a) computed tomography scan with very minor symptom. (b) ground-truth segmentation result. [Color figure can be viewed at wileyonlinelibrary.com]

COVID-19. According to recent literature,[33,34] COVID-19 infection happens more frequently in lower lobes of the lung. However, so far no researches have reported quantitatively the severity of COVID-19 infection in each lung lobe and bronchopulmonary segment. Based on the segmentation results, the POIs of lung lobes and bronchopulmonary segments can be automatically calculated. Thus, statistics of infection distribution can be summarized in a large-scale dataset, for example, 300 CT scans in our study. Figure 6 shows the boxplots of these POIs calculated from 300 CT scans of COVID-19 patients in Shanghai. Figure 7(a) shows that the mean POIs of left and right lower lobes are higher than those of other lobes, which coincides with the findings reported in Refs. [33,34].

3.C. Human-involved-model-iterations (HIMI) strategy

Two metrics were used to evaluate the HIMI strategy. First, the time of manual contouring was recorded to

TABLE III. Inter-rater variability analysis between two radiologists on randomly sampled 10 CT cases. The Dice coefficients, and POI difference in whole lung, lung lobes and bronchopulmonary segments, were estimated to serve as the reference for assessing the automatic segmentation accuracy. * indicates no significant difference between contouring results of two radiologists on the validation dataset according to paired *t*-test.

Inter-rater variability metrics	Mean	Standard deviation	Median	25% IQR	75% IQR	Number of infected samples
Dice Similarity Coefficient	96.1%	3.5%	97.2%	95.4%	98.3%	10
POI Error (Whole lung)*	0.2%	0.1%	0.2%	0.1%	0.2%	10
POI Error (Left upper lobe)*	0.4%	0.7%	0.1%	0.0%	0.3%	7
POI Error (Left lower lobe)*	0.2%	0.2%	0.3%	0.0%	0.4%	7
POI Error (Right upper lobe)*	0.3%	0.5%	0.1%	0.1%	0.3%	6
POI Error (Right middle lobe)*	0.3%	0.5%	0.1%	0.0%	0.1%	6
POI Error (Right lower lobe)*	0.2%	0.2%	0.2%	0.0%	0.3%	9
POI Error (Left upper lobe/apical posterior)*	0.9%	1.1%	0.2%	0.0%	1.2%	5
POI Error (Left upper lobe/anterior)*	0.9%	0.8%	0.4%	0.3%	1.2%	3
POI Error (Left upper lobe/superior lingular)*	0.6%	0.9%	0.0%	0.0%	0.6%	7
POI Error (Left upper lobe/inferior lingular)*	0.2%	0.2%	0.1%	0.0%	0.3%	4
POI Error (Left lower lobe/superior)*	0.1%	0.1%	0.2%	0.1%	0.2%	4
POI Error (Left lower lobe/anteromedial basal)*	0.2%	0.1%	0.3%	0.2%	0.3%	5
POI Error (Left lower lobe/lateral basal)*	0.3%	0.4%	0.2%	0.0%	0.4%	6
POI Error (Left lower lobe/posterior basal)*	0.6%	0.5%	0.4%	0.2%	0.7%	6
POI Error (Right upper lobe/apical)*	0.5%	0.7%	0.2%	0.0%	0.6%	5
POI Error (Right upper lobe/posterior)*	0.5%	0.5%	0.2%	0.1%	0.8%	5
POI Error (Right upper lobe/anterior)*	0.5%	0.9%	0.1%	0.0%	0.2%	5
POI Error (Right middle lobe/lateral)*	0.2%	0.3%	0.1%	0.0%	0.2%	6
POI Error (Right middle lobe/medial)*	0.3%	0.6%	0.1%	0.0%	0.1%	5
POI Error (Right lower lobe/superior)*	0.4%	0.4%	0.2%	0.1%	0.7%	7
POI Error (Right lower lobe/medial basal)*	0.5%	0.3%	0.6%	0.3%	0.8%	4
POI Error (Right lower lobe/anterior basal)*	0.2%	0.3%	0.1%	0.0%	0.1%	8
POI Error (Right lower lobe/lateral basal)*	0.2%	0.2%	0.1%	0.1%	0.2%	7
POI Error (Right lower lobe/posterior basal)*	0.3%	0.5%	0.1%	0.0%	0.2%	7

TABLE IV. Comparison on Dice values of VB-Net and U-Net in segmenting infections of the lung.

Network	Mean	Standard deviation	Median	25% IQR	75% IQR	Number of testing samples
U-Net	87.3%	10.1%	89.5%	85.6%	93.2%	300
VB-Net	91.6%	10.0%	92.2%	89.0%	94.6%	300

compare labeling time of a CT scan with the deep learning model. Second, the segmentation accuracy of deep learning models at different iterations was assessed to see whether the accuracy improves with more annotated training data. Table V shows the labeling time and segmentation accuracy at different iterations. Without any assistance of deep learning, it takes 211.3 ± 52.6 min to contour COVID-19 infection regions on one CT scan. The contouring time drops dramatically to 31.1 ± 8.1 min with the assistance of the first deep learning model trained with 36 annotated CT scans. It further drops to 12.0 ± 2.9 min with 114 annotated data, and to 4.7 ± 1.1 with 249 annotated data. Meanwhile, the segmentation accuracy of deep learning models was evaluated using

Dice similarity coefficient on the entire validation set including 300 scans. It improves from 85.1 ± 11.4%, to 91.0 ± 9.6%, and to 91.6% ± 10.0 with more training data added.

From Table V, it can be seen that the segmentation accuracy is improved after each iteration, which greatly reduces human intervention and thus reduces significantly the time of annotation and labeling.

3.D. Clinical use of quantification

The best accuracy of severity prediction was 73.4% ± 1.3% when using MOI across 5 lung lobes as features. Besides, the accuracy of severity prediction was 72.0% ± 2.7% when using POI across 18 bronchopulmonary segments in the study. These observations indicate the potential clinical application of quantification technique on severity prediction.

Besides, our results show an accuracy of 63.3% when only PSI was used in severity prediction. This is lower than using the quantified radiological features such as POI and MOI. This shows that quantified radiological features are more informative to predict severity of COVID-19 than PSI.
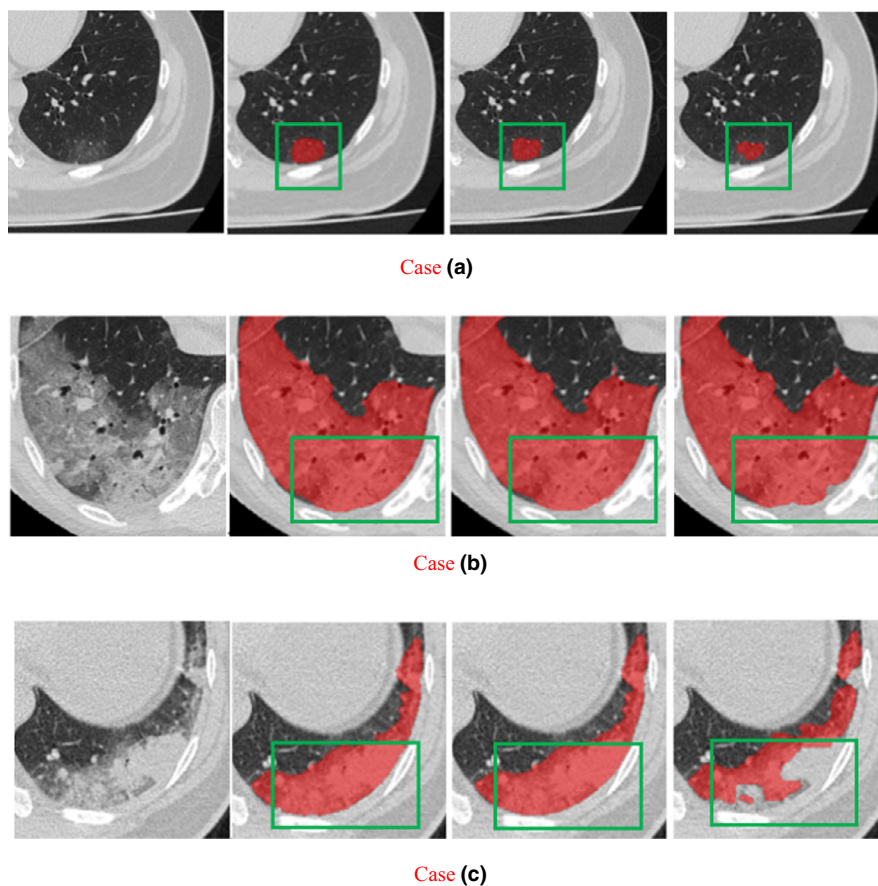
Fig. 6. Comparison of segmentation results by VB-Net and U-Net on three cases. First column shows original images, and the second column shows the ground-truth segmentations. The segmentation results by VB-Net and U-Net are shown in the third and fourth columns, respectively. Green boxes in each case indicate regions with large segmentation differences by VB-Net and U-Net. (a) Comparison of segmentation results on case 1. (b) Comparison of segmentation results on case 2. (c) Comparison of segmentation results on case 3. [Color figure can be viewed at wileyonlinelibrary.com]

4. DISCUSSION

Computed tomography imaging has become an efficient tool for both screening COVID-19 patients and assessing the severity of COVID-19. However, radiologists lack a computerized tool to accurately quantify the severity of COVID-19, for example, the percentage of infection in the whole lung. In the literature, deep learning has become a popular method in medical image analysis and has been used in analyzing diffuse lung diseases on CT.[35,36] In this work, we explored deep learning to segment COVID-19 infection regions within lungs on CT images. The accurate segmentation provides quantitative information that is necessary to track disease progression and analyze longitude changes of COVID-19 during the entire treatment period.[37] A research of severity prediction on the first day of enrollment is reported based on quantification results. We believe that this deep learning system for COVID-19 quantification will open up many new research directions of interest in this community.

In our research, we have obtained more than 72% accuracy in severity prediction based on the quantified radiological features. In clinical practice, more information such as clinical features is available, which provides more information about patients. Besides, patients often take CT scan every

3–5 days, which provides longitudinal changing information about progression of the disease. All these kinds of information can be used in the future to improve the accuracy of severity prediction.

One potential research application of this system is to quantify longitudinal changes in the follow-up CT scans of COVID-19 patients. Hospitalized patients with confirmed COVID-19 typically take a CT examination every 3–5 days. As currently there is no effective medicine to target COVID-19, most patients recover with different degrees of supportive medicine intervention. Given many such patients, it is interesting to see how disease progresses under different clinical management. Figure 8 gives a case with three follow-up CT scans. With infection region segmented, the changes of infection volume as well as consolidation and ground glass opacities can be easily visualized using surface rendering technique.

Moreover, infection distribution can be analyzed further down to the bronchopulmonary segment level, as shown in Fig. 7(b). To the best of our knowledge, this is the first work that reveals the COVID-19 distribution in bronchopulmonary segments in terms of a large-scale patient CT data. Our results show that the following segments are often infected by COVID-19 (as indicated with decreasing mean POI): right
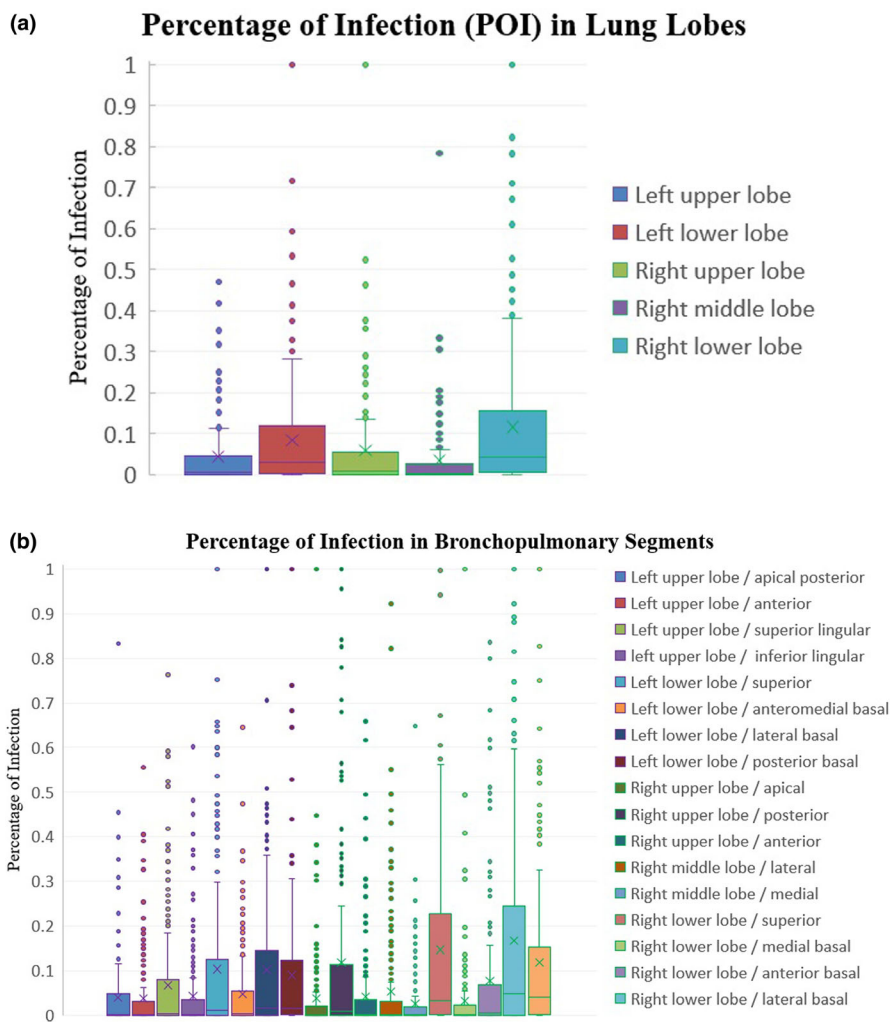
(a)

Percentage of Infection (POI) in Lung Lobes

- Left upper lobe
- Left lower lobe
- Right upper lobe
- Right middle lobe
- Right lower lobe

(b)

Percentage of Infection in Bronchopulmonary Segments

- Left upper lobe / apical posterior
- Left upper lobe / anterior
- Left upper lobe / superior lingular
- left upper lobe / inferior lingular
- Left lower lobe / superior
- Left lower lobe / anteromedial basal
- Left lower lobe / lateral basal
- Left lower lobe / posterior basal
- Right upper lobe / apical
- Right upper lobe / posterior
- Right upper lobe / anterior
- Right middle lobe / lateral
- Right middle lobe / medial
- Right lower lobe / superior
- Right lower lobe / medial basal
- Right lower lobe / anterior basal
- Right lower lobe / lateral basal

Fɪɢ. 7. The box-and-whisker plots of POIs in five different lung lobes (a) and 18 different bronchopulmonary segments (b) on 300 validation computed tomography (CT) scans of COVID-19 patients. The bottom and top of each box represent the 25th and the 75th percentile, respectively. The line in the box indicates the 50th percentile or the median value. [Color figure can be viewed at wileyonlinelibrary.com]

Tᴀʙʟᴇ V. Validation of the human-involved-model-iterations (HIMI) strategy. Manual time indicates the manual labeling/correction time without DL or with different DL models. Accuracy indicates the segmentation accuracy of DL models. '# of Images' indicates the number of training images used in training each DL model.

Time (min)	Without DL (min)	First iteration (min)	Second iteration (min)	Third iteration (min)
Manual time	211.3 ± 52.6	31.1 ± 8.1	12 ± 2.9	4.7 ± 1.1
Accuracy (DSC)	N/A	85.1 ± 11.4%	91.0 ± 9.6%	91.6%±10.0%
# of Images	N/A	36	114	249

lower lobe — lateral basal, right lower lobe — superior, right lower lobe — posterior basal, left lower lobe — lateral basal, left lower lobe — superior, left lower lobe — posterior basal, and right upper lobe — posterior. Using HIMI strategy in

training the segmentation network is a novel feature of our system. Existing AI-based systems for automatic quantitative assessment always require a large amount of annotation CT data, whereas collecting the annotated data is very expensive or even difficult. Moreover, these AI systems are always trained as a black box to users, who however always want to know what has happened behind the model. Our experimental results indicate that the HIMI strategy makes the manual annotation process faster with the assistance of deep learning models. Also, the HIMI strategy makes the system more comprehensible. That is, with manual intervention in HIMI, the radiologists are aware of how good the system performs in the training process. Besides, the HIMI strategy helps radiologists accustomed to the AI system because they are involved in the training process. It integrates professional knowledge from radiologists in an interactive way.

Both SVM and LASSO are effective classifiers in COVID-19 applications. In Ref. [37], Shi et al. constructed a least absolute shrinkage and selection operator (LASSO)
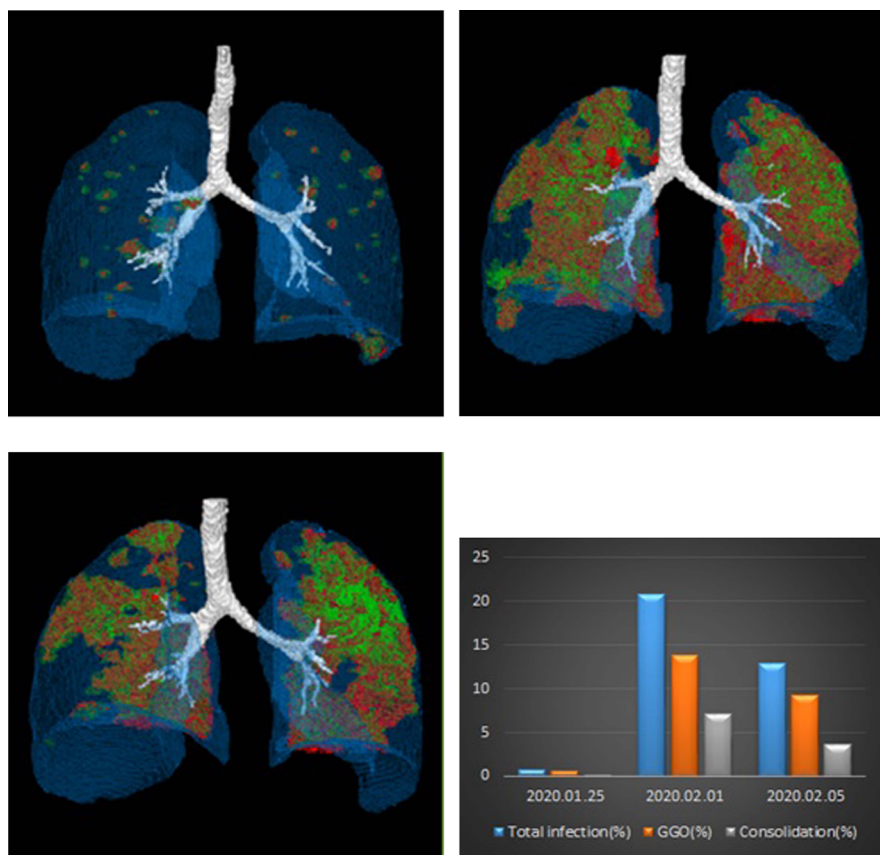
FIG. 8. The follow-up study results of a forty-six female patient. Green and red colors indicate ground glass and consolidation opacities, respectively. The POI values show the progression and gradual recovery of the patient from Jan 25th, Feb 1st, to Feb 5th 2020. [Color figure can be viewed at wileyonlinelibrary.com]

logistic regression model for severity prediction with 24 clinico-radiological features. The 24 clinico-radiological features were reduced to five potential predictors, which include age, lactate dehydrogenase (LDH), C-reactive protein (CRP), CD4 + T cell counts and MOI in the whole lung. Different from their method that used both clinical and radiological features, we used all the quantified radiological features of five lung lobes and 18 bronchopulmonary segments and further constructed classifier based on SVM. Compared with the LASSO logistic regression in Ref. [37], our method is a nonlinear classification method with stronger discriminant capability. Besides, we consider the imbalanced issue in COVID-19 severity prediction. Both factors make our method superior to the LASSO logistic regression model in Ref. [37].

Conclusively, chest CT has played a key role not only in the diagnosis and treatment of COVID-19 but also in evaluating both disease progression and therapeutic efficacy.[13,38–40] However, the role of CT in identifying COVID-19 is still controversial. Some researchers gave a critical review and questioned the role of CT in identifying COVID-19.[41] In the early outbreak of COVID-19 in China, American College of Radiology also showed a reserving attitude towards the use of CT in identifying and screening COVID-19. This is mainly because the CT manifestations of COVID-19 are generally not specific, and overlap with other infections, including influenza, H1N1, SARS and MERS.[42] Different from these studies, our study focuses on COVID-19 severity prediction, in which all the COVID-19 patients are confirmed and have typical CT manifestations of COVID-19. Besides, the key technical issue of severity prediction is to find discriminant features to predict the progress of the disease, which focuses more on the evolution of the CT features over a period of time. Our preliminary results in Section 3D also show the effectiveness of CT on the severity prediction of COVID-19.

It is worth noting the limitations of our work in several aspects:

First, the validation CT datasets were collected in one center, which may not be representative of all COVID-19 patients in other geographic areas. The generalization of the deep learning system needs to be further validated on multi-center datasets.

Second, the system is developed to quantify COVID-19 infections only, and it may not be applicable for quantifying other types of pneumonia, for example, bacterial pneumonia. Figure 9 shows some examples on the application of our model to other lung diseases. From the figure, one may observe that the model trained with COVID-19 CT images is able to detect similar symptoms (i.e., ground glass opacities)
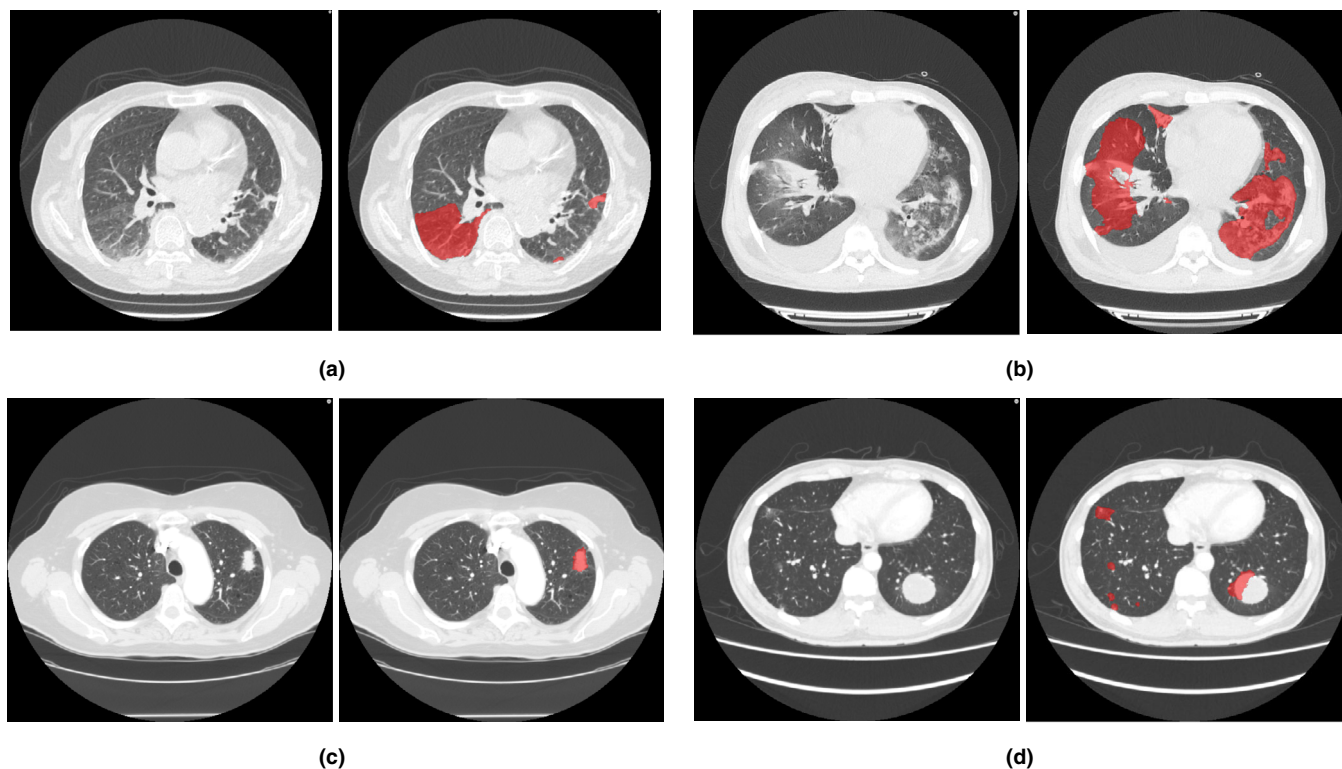
FIG. 9. Examples on the application of our model on other lung diseases. (a) computed tomography (CT) of one patient infected with bacterial pneumonia (Left: CT image, Right: CT image overlaid with auto-segmentation by our model); (b) CT of one patient infected with mycotic pneumonia (Left: CT image, Right: CT image overlaid with auto-segmentation by our model); (c) CT of one patient with lung cancer (small tumor in LIDC dataset) (Left: CT image, Right: CT image overlaid with auto-segmentation by our model); (d) CT of one patient with lung cancer (large tumor in LIDC dataset) (Left: CT image, Right: CT image overlaid with auto-segmentation by our model). [Color figure can be viewed at wileyonlinelibrary.com]

in CT images from other lung diseases. However, if a tumor or an infection contains a large portion of homogenous consolidation, the model would fail to detect the complete contour as shown in Fig. 9(b) (right lung) and Fig. 9(d) (large tumor in LIDC dataset[43]). Since most of COVID-19 infections in CT consist of ground glass opacities and sometimes a small portion of inhomogeneous consolidation, there is seldom a large portion of homogenous consolidation associated with COVID-19. Therefore, the model fails to recognize this pattern, which often appears in bacterial pneumonia or lung cancer.

Finally, in our future work, we will extend the system to quantify severity of other pneumonia using advanced machine learning methods such as transfer learning and deep ensemble learning.[44]

One may argue that the typical cases in Fig. 5 indicate that the segmentation network may miss small infection in the lung, implying that the proposed segmentation network would not be helpful for studying disease progression starting from early stage. In our opinion, the failure cases of the segmentation model are due to the following two reasons. First, the GGO is very light and small, and the contrast is insufficient for the algorithm to draw accurate contour of infection. Second, the cases with small infection are minority in our dataset. As our algorithm is learning-based and data-driven, it did not do quite well on such cases. In the future, such data will be specifically collected to address this problem.

With this automatic DL-based segmentation, many studies on quantifying imaging metrics and correlating them with syndromes, epidemiology, and treatment responses could further reveal insights about imaging markers and findings towards improved diagnosis and treatment for COVID-19.

ACKNOWLEDGMENT

CONFLICT OF INTEREST

The authors have no conflict to disclose.

*Fei Shan, Yaozong Gao and Jun Wang contributed equally to this work.
a)Author to whom correspondence should be addressed. Electronic mail: Dinggang.Shen@gmail.com(DS); shiyuxin@shphc.org.cn(YS).

REFERENCES

1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727–733.

2. Tan W, Zhao X, Ma X, et al. A novel coronavirus genome identified in a cluster of pneumonia cases - Wuhan, China 2019–2020. *China CDC Weekly*. 2020;2:61–62.

3. Phan LT, Nguyen TV, Luong QC, et al. Importation and human-to-human transmission of a novel coronavirus in Vietnam. *N Engl J Med*. 2020;382:872–874.

4. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med*. 2020;382:929–936.

5. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020;395:470–473.

6. Spina S, Marrazzo F, Migliari M, Stucchi R, Sforza A, Fumagalli R. The response of Milan's Emergency Medical System to the COVID-19 outbreak in Italy. *The Lancet*. 2020;395:e49–e50.

7. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2000;579:1–5.

8. Gorbalenya AE. Severe acute respiratory syndrome-related coronavirus–- The species and its viruses, a statement of the Coronavirus Study Group. BioRxiv; 2020.

9. WHO. Coronavirus disease 2019 (COVID-19)Situation Report – 158. https://www.who.int/docs/default-source/coronaviruse/situation-reports/ 20200626-covid-19-sitrep-158.pdf?sfvrsn=1d1aae8a_2 (accessed June 27th, 2020).

10. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. *Radiology*. 2020;296:200343.

11. 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.

12. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*. 2020;323:1061–1069.

13. Dong D, Tang Z, Wang S. The role of imaging in the detection and management of COVID-19: a review. *IEEE Rev Biomed Eng*. 2020;1. http:// doi.org/10.1109/rbme.2020.2990959

14. Gozes O, Frid-Adar M, Greenspan H, et al. Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. arXiv preprint arXiv:2003.05037; 2020.

15. 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;395:507–513.

16. Chassagnon G, Vakalopoulou M, Battistell E, et al. AI-Driven CT-based quantification, staging and short-term outcome prediction of COVID-19 pneumonia. arXiv preprint arXiv:2004.12852; 2020.

17. Liang T. The First Affiliated Hospital. In: Z. U. S. o. M, Ed. *Handbook of COVID-19 Prevention and Treatment*; 2020.

18. Shi F, Yap PT, Fan Y, et al. Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *Neuroimage*. 2010;51: 684–693.

19. Jia H, Yap PT, Shen D. Iterative multi-atlas-based multi-image segmentation with tree-based registration. *NeuroImage*. 2012;59:422–430.

20. Milletari F, Navab N, Ahmadi S-A.V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*; 2016. IEEE: 565–571.

21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016:770–778.

22. Han M, Gao Y, Zhang Y, et al. Large-scale evaluation of V-Net for organ segmentation in image guided radiation therapy. In: *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, 2019, vol. 10951: International Society for Optics and Photonics, p. 109510O.

23. Mu G, Ma Y, Han M, et al. Automatic MR kidney segmentation for autosomal dominant polycystic kidney disease. In: *Medical Imaging 2019: Computer-Aided Diagnosis*, 2019, vol. 10950: International Society for Optics and Photonics, p. 109500X.

24. Slicer Wiki. https://www.slicer.org/w/index.php?title=Main_Page&old id=62645 (accessed 2020).

25. Jacobs C, van Rikxoort EM, Twellmann T, et al. Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Med Image Anal*. 2014;18:374–384.

26. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Berlin: Springer; 2015:234–241.

27. Cao Y, Xu Z, Feng J, et al. Longitudinal assessment of COVID-19 using a deep learning–based quantitative CT pipeline: illustration of two cases. *Radiol Cardioth Imaging*. 2020;2:e200082.

28. Huang L, Han R, Ai T, et al. Serial quantitative chest CT assessment of COVID-19: deep-learning approach. *Radiol Cardioth Imaging*. 2020;2: e200075.

29. Song YS, Park CM, Park SJ, Lee SM, Jeon YK, Goo JM. Volume and mass doubling times of persistent pulmonary subsolid nodules detected in patients without known malignancy. *Radiology*. 2014;273:276–284.

30. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336:243–250.

31. Shah BA, Ahmed W, Dhobi GN, Shah NN, Khursheed SQ, Haq I. Validity of pneumonia severity index and CURB-65 severity scoring systems in community acquired pneumonia in an Indian setting. *Indian J Chest Dis All Sci*. 2010;52:9–17.

32. Pan F, Ye T, Sun P, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology*. 2020;295:200370.

33. Song F, Shi N, Shan F, et al. Emerging coronavirus 2019-nCoV pneumonia. *Radiology*. 2020;295:200274.

34. Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*. 2020;295:200463.

35. Pang T, Guo S, Zhang X, Zhao L. Automatic lung segmentation based on texture and deep features of HRCT images with interstitial lung disease. *Biomed Res Int*. 2019;2019:1–8.

36. Park B, Park H, Lee SM, Seo JB, Kim N. Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. *J Digit Imaging*. 2019;32:1019–1026.

37. Shi W, Peng X, Liu T, et al. Deep learning-based quantitative computed tomography model in predicting the severity of COVID-19: A retrospective study in 196 patients. *SSRN Elec J*. http://doi.org/10.2139/ssrn. 3546089

38. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE reviews in biomedical engineering; 2020.

39. Ouyang X, Huo J, Xia L, et al. Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans Med Imaging*. 2020;39:2595.

40. Kang H, Xia L, Yan F, et al. Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning. *IEEE Trans Med Imaging*. 2020;39:2606.

41. Raptis C, Hammer MM, Short RG, et al. Chest CT and Coronavirus disease (COVID-19): A critical review of the literature to date. AJR. American journal of roentgenology, pp. 1-4, 04/16 2020. https://doi.org/10. 2214/AJR.20.23202

42. A. C. o. R. (ACR). ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection (accessed June 20th, 2020).

43. LUNA16 - Grand Challenge. https://luna16.grand-challenge.org (accessed September 5th, 2020).

44. Suk HI, Lee SW, Shen D. Deep ensemble learning of sparse regression models for brain disease diagnosis. *Med Image Anal*. 2017;37:101–113.