Rapid quantification of COVID-19 pneumonia burden from computed tomography with convolutional long short-term memory networks

Aditya Killekar,[a] Kajetan Grodecki,[b] Andrew Lin,[a] Sebastien Cadet,[a] Priscilla McElhinney,[a] Aryabod Razipour,[a] Cato Chan,[a] Barry D. Pressman,[a] Peter Julien,[a] Peter Chen,[a] Judit Simon,[c] Pal Maurovich-Horvat,[c] Nicola Gaibazzi,[d] Udit Thakur,[e] Elisabetta Mancini,[f] Cecilia Agalbato,[f] Jiro Munechika,[g] Hidenari Matsumoto,[g] Roberto Menè,[h,i] Gianfranco Parati,[h,i] Franco Cernigliaro,[h,i] Nitesh Nerlekar,[e] Camilla Torlasco,[h,i] Gianluca Pontone,[f] Damini Dey,[a] and Piotr Slomka[a,*]

[a]Cedars-Sinai Medical Center, Department of Medicine (Division of Artificial Intelligence in Medicine), Biomedical Sciences and Imaging, Los Angeles, California, United States
[b]Medical University of Warsaw, Warsaw, Poland
[c]Semmelweis University, Budapest, Hungary
[d]Azienda Ospedaliero-Universitaria di Parma, Parma, Italy
[e]Monash Health, Melbourne, Victoria, Australia
[f]University of Milan, Centro Cardiologico Monzino IRCCS, Milan, Italy
[g]Showa University School of Medicine, Tokyo, Japan
[h]IRCCS Istituto Auxologico Italiano, Department of Cardiovascular, Neural and Metabolic Sciences, Milan, Italy
[i]University of Milano-Bicocca, Department of Medicine and Surgery, Milan, Italy

Abstract

Purpose: Quantitative lung measures derived from computed tomography (CT) have been demonstrated to improve prognostication in coronavirus disease 2019 (COVID-19) patients but are not part of clinical routine because the required manual segmentation of lung lesions is prohibitively time consuming. We aim to automatically segment ground-glass opacities and high opacities (comprising consolidation and pleural effusion).

Approach: We propose a new fully automated deep-learning framework for fast multi-class segmentation of lung lesions in COVID-19 pneumonia from both contrast and non-contrast CT images using convolutional long short-term memory (ConvLSTM) networks. Utilizing the expert annotations, model training was performed using five-fold cross-validation to segment COVID-19 lesions. The performance of the method was evaluated on CT datasets from 197 patients with a positive reverse transcription polymerase chain reaction test result for SARS-CoV-2, 68 unseen test cases, and 695 independent controls.

Results: Strong agreement between expert manual and automatic segmentation was obtained for lung lesions with a Dice score of 0.89 ± 0.07; excellent correlations of 0.93 and 0.98 for ground-glass opacity (GGO) and high opacity volumes, respectively, were obtained. In the external testing set of 68 patients, we observed a Dice score of 0.89 ± 0.06 as well as excellent correlations of 0.99 and 0.98 for GGO and high opacity volumes, respectively. Computations for a CT scan comprising 120 slices were performed under 3 s on a computer equipped with an NVIDIA TITAN RTX GPU. Diagnostically, the automated quantification of the lung burden % discriminate COVID-19 patients from controls with an area under the receiver operating curve of 0.96 (0.95–0.98).

*Address all correspondence to Piotr Slomka, Piotr.Slomka@cshs.org

Conclusions: Our method allows for the rapid fully automated quantitative measurement of the pneumonia burden from CT, which can be used to rapidly assess the severity of COVID-19 pneumonia on chest CT.

 [DOI: 10.1117/1.JMI.9.5.054001]

Keywords: coronavirus disease 2019; computed tomography imaging; deep learning; image processing; lesion segmentation; supervised learning.

Paper 22060GR received Mar. 4, 2022; accepted for publication Aug. 16, 2022; published online Sep. 6, 2022.

1 Introduction

Coronavirus disease 2019 (COVID-19) is a global pandemic and public health crisis of catastrophic proportions, with over 437 million confirmed cases worldwide as of March 2, 2022.[1] Although the vaccines are now available, they are not 100% effective; new strains are emerging and immunization coverage varies significantly between the world regions due to socioeconomic differences. It is likely that vaccine boosters will be necessary, and continuous monitoring for the disease will be needed. Although the diagnosis of COVID-19 relies on a reverse transcription polymerase chain reaction (RT-PCR) test in respiratory tract specimens, computed tomography (CT) remains the central modality in disease staging.[2–5] Specific CT lung features include peripheral and bilateral ground-glass opacities (GGOs), with round and other specific morphology as well as peripheral consolidations, and increasing extension of such opacities has been associated with the risk of critical illness.[6–8] Although conventional visual scoring of the COVID-19 pneumonia extent correlates with clinical disease severity, it requires proficiency in cardiothoracic imaging and ignores lesion features, such as volumes, density, or inhomogeneity.[9,10] On the other hand, CT-derived quantitative lung measures are not part of the clinical routine, despite being demonstrated to improve prognostication in COVID-19 patients, due to prohibitively time-consuming manual segmentation of the lung lesions required for computation.[11–13] The chest CT is currently indicated in COVID-19 patients with moderate or severe respiratory symptoms and high pretest probability of infection, or any other clinical scenario requiring rapid triage. Importantly, over 15 million chest CT (including cardiac CT) are performed a year in the Unites States for indications not related to COVID-19.[14] Additionally, every thoracic Positron Emission Tomography (PET)/CT and Single Photon Emission Computed Tomography (SPECT)/CT scan (including myocardial perfusion imaging) will include Computed Tomography Attenuation Correction (CTAC) covering the lung area. Parenchymal opacification associated with COVID-19 can be potentially seen on these exams. Critically, in the coming months and years, it is likely that COVID-19 changes may often be an incidental finding on chest CT performed for other diseases in asymptomatic COVID-19 patients. These incidental findings may also be on CTAC maps often acquired in conjunction with myocardial perfusion SPECT and PET MPI. Indeed, some first reports of such incidental findings have been reported on PET/CT in the *Journal of Nuclear Medicine* (April 2020) by Albano et al.[15] in Italy, followed by others.[16–19] It is worth noting that these CTAC scans are not routinely reviewed for other abnormalities and are often viewed with window and level settings not set for review of lung abnormalities. Thus, a rapid automated alert system for COVID-19 related abnormalities would be of great benefit in such situations.

2 Related Work

Deep learning, a class of artificial intelligence (AI), has shown to be very effective for automated object detection and image classification from a wide range of data. Myriad AI systems have been introduced to aid radiologists in the detection of lung involvement in COVID-19, with several presenting the potential to improve the performance of junior radiologists to the senior level.[12,20] Bai et al.[21] developed a classification network to differentiate between COVID-19 pneumonia and other pneumonia and achieved good performance in diagnosing

the disease, achieving an area under the receiver operating characteristic (AUROC) of 0.95. They provided a heatmap in an effort to explain the model predictions, but it will be of great importance in disease staging and prognosis if the model can pin-point the legions accurately. This shortcoming was addressed by Zhang et al.,[20] who developed a system that can diagnose the disease, segment the lungs and lesions into several classes, and be used to evaluate drug treatment effects. They developed a two-stage segmentation network for segmenting lesions in lungs from CT slices, experimenting with various segmentation frameworks, and adopted DeepLabv3 as the backbone for its better segmentation performance. The model was evaluated using mean Dice coefficient and pixel accuracy by five-fold cross-validation test, achieving a 0.587 mean Dice score.

On the other hand, Fan et al.[22] developed a novel COVID-19 lung infection segmentation network that combines high-level features using a parallel partial decoder to generate a global map as initial guidance for further steps. To establish a relationship between lesion boundaries, they used their novel implicit recurrent reverse attention modules. The final training loss comprised weighted binary cross-entropy applied at different stages of the network and weighted intersection over union loss. The authors went beyond to address the shortage of expert annotations by modifying their training strategy to accommodate semi-supervised learning into their model. Although this model does not perform multi-class segmentation by itself, it can separate the lesions into two classes using UNet and their model output as guidance for segmentation, achieving a mean Dice score of 0.541.

Similarly, Chaganti et al.[11] also developed a system for binary segmentation of CT abnormalities related to COVID-19. They trained two different models: one for segmenting lung lobes and another for lesions. The lung segmentation model was trained using a deep image-to-image network, and the lesion segmentation model was trained using a UNet-like architecture. The lesion segmentation model performs binary segmentation, that is, all of the lesions (GGOs and consolidations) were treated as one class during training and later separated into two classes by thresholding the voxels at -200 Hounsfield units (HU) during inference. Finally, they introduced two measures for evaluating the severity of the disease: percentage of opacity and percentage of high opacitiy. The overall performance was evaluated using Pearson correlation between the severity measures.

Gao et al.[23] developed a dual-branch combination network for joint binary segmentation and classification of COVID-19 using CT images. They proposed a lesion attention module to improve the sensitivity of the model in detecting small lesions. The lesion attention module is also used to interpret model predictions for the assessment of classification results. They achieved a Dice score of 0.835 on an internal test set in segmenting the lesions and an AUROC of 0.9771 in classifying COVID-19 patients.

The work presented in this paper builds on previous research to explore the quantitative prognostication and disease staging by segmenting the COVID-19 lesions into multiple classes. Earlier work focused on segmentation using one slice in the CT at a time, whereas we focus on benefiting from additional information about the anatomy and the lesions in several adjacent slices. However, most three-dimensional (3D) medical segmentation networks consume a lot of memory in storing the intermediate features for skip connections[24,25] making them difficult to implement in low-end clinical systems. To this end, we adopt the state-of-the-art segmentation network by Tao et al.[26] and replace the attention from multi-scale input to attention from adjacent slices using long short-term memory (LSTM) recurrent network,[27] which are well-known for their long data sequence/series processing capabilities. We do so to imitate a radiologist reviewing adjacent slices of a CT scan and aggregate lesion information while making manual annotations. We employ a specific variant of the LSTM network known as the convolutional long short-term memory (ConvLSTM) network,[28] which is capable of handling images directly. ConvLSTM operates directly on images, facilitating rapid segmentation and accurate 3D quantification of the disease involvement of lung lesions in COVID-19 pneumonia from both contrast and non-contrast CT images. ConvLSTM networks have the capability of preserving relevant features while simultaneous dismissing irrelevant ones in the form of the feedback loop, which translates into a memory-sparing strategy for the holistic analysis of the images.

3 Dataset

3.1 *Patient Population*

The cohort used in this study comprised 264 patients, who underwent chest CT and had a positive RT-PCR test result for SARS-CoV-2. A total of 197 patients were included in the training cohort (N_{cov}), and 68 were used for external validation (N_{ext}). Datasets for 187 out of 197 patients from the training cohort were collected from the prospective, international, multicenter registry involving centers from North America [Cedars-Sinai Medical Center, Los Angeles ($n = 75$)], Europe [Centro Cardiologico Monzino ($n = 64$), and Istituto Auxologico Italiano ($n = 17$); both Milan, Italy], Australia [Monash Medical Centre, Victoria, Australia ($n = 6$)], and Asia [Showa Medical University, Tokyo, Japan ($n = 25$)], where either non-contrast ($n = 157$) or contrast-enhanced ($n = 30$) chest CT was performed to aid in the triage of patients with a high clinical suspicion for COVID-19, in the setting of a pending RTPCR test or comorbidities associated with severe illness from COVID-19. The population is given in Table 1. Datasets for the remaining 10 COVID-19 patients were derived from an open-access repository of non-contrast CT images; therefore, no clinical data were provided for this cohort. Out of 31,560 transverse slices available, 15,588 had lesions. The external testing cohort comprised 68 non-contrast CT scans of COVID-19 patients: about 50 from an open-access repository[29] and 18 additional ones from Italy (Centro Cardiologico Monzino). There were 12,102 transverse slices available in this cohort, and 6,503 had lesions (Table 2). All data were deidentified prior to being enrolled in this study. The CT images from each patient and the clinical database were fully anonymized and transferred to one coordinating center for core lab

Table 1 Patient baseline characteristics and imaging data in a training cohort.

Baseline characteristics	**N = 187**
Age, years	61 ± 16
Men	123 (65.7)
Body mass index	26.8 ± 5.3
Current smoker	22 (11.7)
Former smoker	10 (5.3)
History of lung disease	19 (10.1)
Image characteristics	**$N_{cov} = 197$**
CT scanner	—
Aquilion ONE	73 (37.0)
GE revolution	13 (6.6)
GE discovery CT750 HD	37 (18.8)
LightSpeed VCT	36 (18.3)
Brilliance iCT	28 (14.2)
Unknown	10 (5.1)
CT type	—
Non-contrast	167 (84.8)
CT pulmonary angiography	30 (15.2)

Note: The data presented in the table are as *n* (%) or mean ± SD.

Table 2 Image findings.

Cohorts	No. of patients	No. of lesions	No. of lesion slices	
			Ground glass opacity	High opacity
COVID-19 positive (N_{cov})	197	31560	15375	6933
External testing (N_{ext})	68	12102	5181	1834
Controls ($N_{control}$)	695	113422	0	0

analysis. The study was conducted with the approval of local institutional review boards (Cedars-Sinai Medical Center IRB# study 617) and written informed consent was waived for fully anonymized data analysis.

3.2 *Ground Truth Generation*

Images were analyzed at the Cedars-Sinai Medical Center core laboratory by two physicians (K.G. and A.L.) with 3 and 8 years of experience in chest CT, respectively, and who were blinded to clinical data. A standard lung window (width of 1500 HU and level of −400 HU) was used. Lung abnormalities were segmented using semi-automated research software (FusionQuant Lung v1.0, Cedars-Sinai Medical Center, Los Angeles, California). These included GGO, consolidation, or pleural effusion according to the Fleischner Society lexicon. Consolidation and pleural effusion were collectively segmented as high-opacity to facilitate the training of the model due to a limited number of slices involving these lesions. Chronic lung abnormalities, such as emphysema or fibrosis, were excluded from segmentation, based on correlation with previous imaging and/or a consensus reading. GGO was defined as hazy opacities that did not obscure the underlying bronchial structures or pulmonary vessels; consolidation as opacification obscuring the underlying bronchial structures or pulmonary vessels; and pleural effusion as a fluid collection in the pleural cavity. The total pneumonia volume was calculated by summing the volumes of the GGO and consolidation components. The total pneumonia burden was calculated as (total pneumonia volume/total lung volume) × 100%. Difficult cases of quantitative analysis were resolved by consensus.

3.3 *Controls Dataset*

Additionally, to assess the diagnostic performance of the methods trained and tested with controls (without any lung abnormalities), we utilized a set of $N_{\text{control}} = 695$ cases from the national lung screen trial (NLST)[30] with normal lung scans. The population characteristics are described in Table 3.

Table 3 NLST controls baseline characteristics.

Baseline characteristics	$N_{control} = 695$
Age, years	59 ± 4
Men	395 (56.8)
Body mass index	29.0 ± 5.3
Current smoker	246 (35.4)
Former smoker	366 (52.7)
History of lung disease	88 (12.7)

NOTE: The data presented in the table are as *n* (%) or mean ± SD.

4 Proposed Method

The objective is to learn the function $\mathbf{\Phi}(\cdot)$ to classify each CT voxel into one of following three classes: GGOs, high opacities, and background. This act of differentiating regions based on their semantic properties is called semantic segmentation.

$$\mathbf{\Phi}\colon\ \mathbf{I} \rightarrow \mathbf{\Phi}(\mathbf{I}), \tag{1}$$

where $\mathbf{I}$ is a set of aligned consecutive CT slices such that $\mathbf{I} \in \mathbb{R}^{H\times W\times F}$. H, W, and F denote the height, width, and cardinality of the input sequence $\mathbf{I}$, respectively, with F being referred to as buffer size elsewhere in the paper. In Sec. 4.1, we introduce the data preprocessing technique used in our method. In Sec. 4.2, we explain in detail the functioning of each block of our network architecture. Finally, in Secs. 4.3 and 4.4, we introduce the loss functions[31] and optimization techniques used in our method.

4.1 *Data Preprocessing*

CT scans from different scanners or with different reconstruction parameters may have different appearance (as seen in column 1 of Fig. 1) and contain voxel values (HU) ranging between -1024 to $+3071$ for a 12-bit scan. Therefore, there is a need for homogenizing the data before we train or infer from it. The input stack of CT images $\mathbf{I}$ are first cropped to the body region of the middle-most image and resized to 256×256. Because we have a very small dataset to train on, we randomly augment the data with rotation of $[-10\ \text{deg}, +10\ \text{deg}]$, translation of up to 10-pixels in the x- and y-directions, and scaling of [0.9, 1.05] times. Finally, we normalize the data by clipping the Hounsfield units between -1024 to $+600$ (expert reader's lung window), followed by a voxel intensity scaling technique called standardization or Z-score normalization.

$$\mathbf{I} = \begin{cases} -1024, & \mathbf{I} < -1024 \\ \mathbf{I}, & -1024 \leq \mathbf{I} \leq +600 \\ +600, & \mathbf{I} > +600 \end{cases}, \tag{2}$$

$$\mathbf{I}_{\text{std}} = \frac{\mathbf{I} - \mu}{\sigma}, \tag{3}$$

where μ is the mean of all of the HU values of voxels in the lung region of the training set and σ is its standard deviation. For simplicity, we refer to $\mathbf{I}_{\text{std}}$ as $\mathbf{I}$ in the rest of the paper.

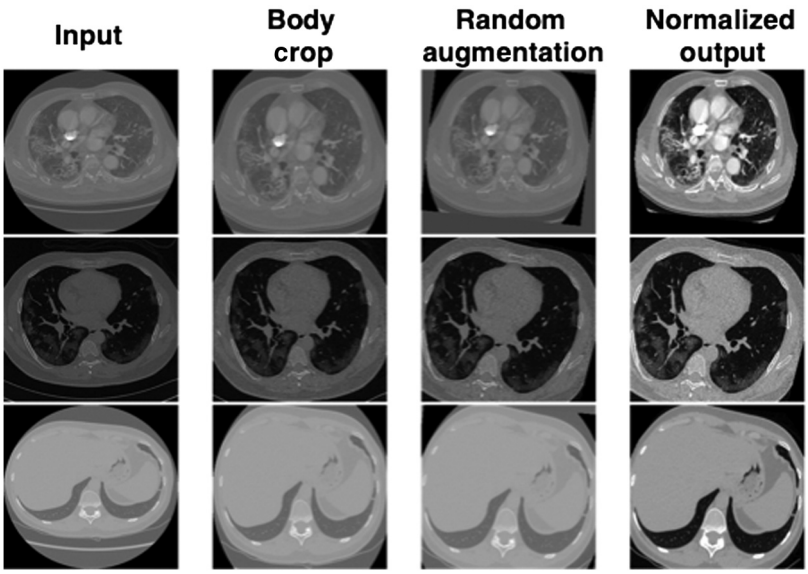

Fig. 1 Order of data preprocessing from input $\mathbf{I}$ (left) to processed output $\mathbf{I}_{\text{std}}$ (right).

To crop the scan to the body region, we threshold the scan at −500 HU and create a binary mask followed by a series of morphological operations: closing, erosion, dilation, etc., and obtain a bounding box around the largest object in the threshold scan. Transferring this bounding box, the original scan returns the body cropped input scan (shown in column 2 of Fig. 1).

4.2 *Network Architecture*

The network architecture, shown in Fig. 2, is inspired by the hierarchical multi-scale attention for semantic segmentation[26] with major changes in the attention branch. Instead of the attention branch looking at the input at various different scales as in Ref. 26, we formulate the attention branch to focus on adjacent slices to aggregate information about the lesions/anatomy from the neighboring slices using a ConvLSTM in the attention branch of the network to improve lesion recognition.

4.2.1 *Main branch* $\mathbf{\Phi}_{\text{main}}$

All of the larger and easy-to-classify lesions are segmented by this branch of the network. It consists of two trainable blocks: the dense block $\Phi^{\text{dense}}_{\text{main}}$, also referred to as Trunk elsewhere in the paper, and the segmentation block $\Phi^{\text{seg}}_{\text{main}}$. Throughout this paper, the subscript of Φ represents the branch name, and the superscript represents the block in that branch.

$$\mathbf{S}_{\text{main}} = \Phi_{\text{main}}(\mathbf{I_k}), \quad = \Phi^{\text{seg}}_{\text{main}}(\text{scale\_up}(\Phi^{\text{dense}}_{\text{main}}(\mathbf{I_k}))), \tag{4}$$

where $\mathbf{S}_{\text{main}} \in \mathbb{R}^{H\times W\times C}$ are the output features from the segmentation block 1, C is the number output classes, scale_up re-scales the features back to the input size using bilinear interpolation, and $\mathbf{I_k}$ is the k'th slice in the input CT stack I, typically the middle most slice.

Dense block $\mathbf{\Phi}^{\text{dense}}_{\text{main}}$. This is the feature extraction block that extracts 256 feature maps of size 64×64 from input $\mathbf{I}$. It is made up of the first dense block of DenseNet121.[32] The reason for choosing a DenseNet for feature extraction is its ability to strengthen feature propagation and mitigate the vanishing-gradient problem, as well as its reduced number of trainable parameters.

Segmentation block 1 $\mathbf{\Phi}^{\text{seg}}_{\text{main}}$. This block is downstream to the dense block. It uses the 256 up-scaled feature maps from $\Phi^{\text{dense}}_{\text{main}}$ as input and classifies each voxel into one of three classes. This block is composed of three convolutional sub-blocks: the first two are made up of 3×3

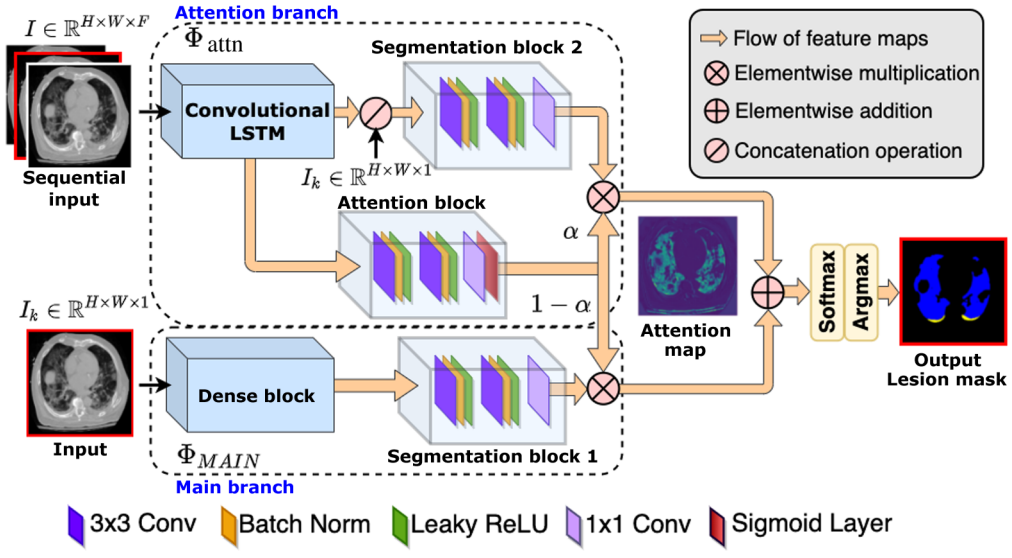

Fig. 2 Framework of the proposed method.

convolutional layers followed by a batch normalization layer and a leaky ReLU layer and the final sub-block is just a 1×1 convolutional layer (see segmentation block in Fig. 2).

4.2.2 *Attention branch* $\mathbf{\Phi}_{\text{attn}}$

All of the errors made by the main branch in ambiguous and difficult to segment parts of the lesions are corrected by the attention branch using information from adjacent slices (shown in Fig. 3). The attention branch comprises a sequential processor $\Phi_{\text{attn}}^{\text{clstm}}$, a segmentation block $\Phi_{\text{attn}}^{\text{seg}}$, and a self-attention block $\Phi_{\text{attn}}^{\text{attn}}$.

$$\alpha = \Phi_{\text{attn}}^{\text{attn}}(\Phi_{\text{attn}}^{\text{clstm}}(\mathbf{I})), \tag{5}$$

where α is the self-attention.

$$\mathbf{S}_{\text{attn}} = \Phi_{\text{attn}}(\mathbf{I}), \quad = \Phi_{\text{attn}}^{\text{seg}}(\mathbf{I}_{\mathbf{k}} \oslash \Phi_{\text{attn}}^{\text{clstm}}(\mathbf{I})), \tag{6}$$

where $\mathbf{S}_{\text{attn}} \in \mathbb{R}^{H \times W \times C}$ are the output features from the segmentation block 2 and C is the number output classes.

Convolutional LSTM $\mathbf{\Phi}_{\text{attn}}^{\text{clstm}}$. We used ConvLSTM[33] for processing sequential data. The ConvLSTM block allows for imitating a radiologist reviewing adjacent slices of a CT scan and aggregate lesion information from adjacent slices to detect lung abnormalities and ensure appropriate annotations.

Segmentation block 2 $\mathbf{\Phi}_{\text{attn}}^{\text{seg}}$. This block is structurally identical to segmentation block 1, except for the input layer. It takes in the main segmentation slice concatenated with ConvLSTM output as the input.

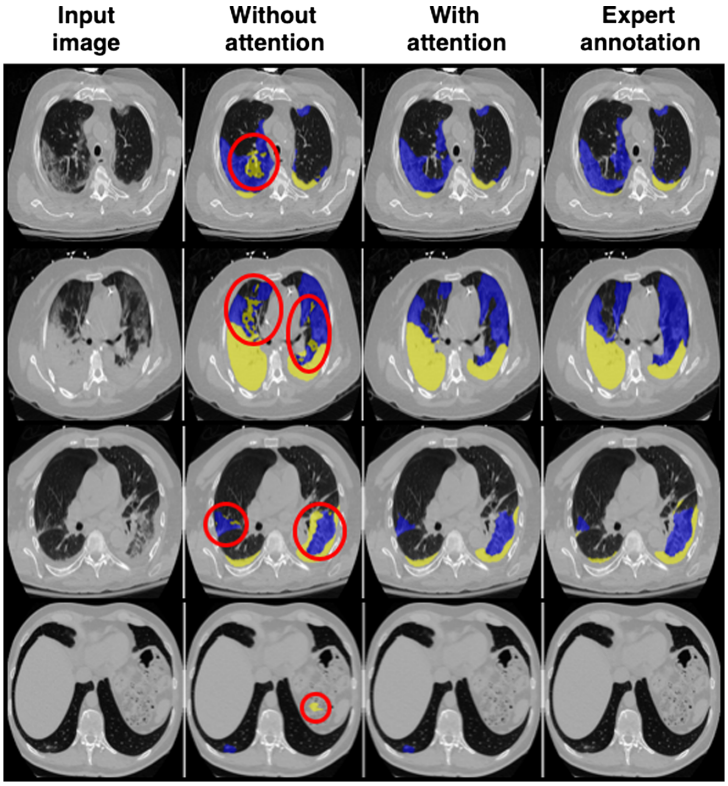

Fig. 3 Intermediate output showing error correction by the attention branch for four different cases in each row. Blue indicates GGOs, and yellow indicates high opacities. Errors are encircled in red.

Attention block $\Phi_{\text{attn}}^{\text{attn}}$. As in Ref. 26, we also adopt an attention mechanism to combine multi-branch outputs ($\mathbf{S}_{\text{main}}$ and $\mathbf{S}_{\text{attn}}$) together at a pixel level. The attention block is identical to the segmentation block in structure with the only difference being that the final 1×1 convolutional layer is followed by a sigmoid layer. This block takes in the output of the ConvLSTM block, as shown in Eq. (5), as input and learns to pixel-wise weight (α) the outputs from the two branches to produce the final prediction [Eq. (7)].

The final prediction is given by the following equation in which the argmax is taken over the channel dimension:

$$\mathbf{S}_{\text{out}} = \arg \max_{C}(\sigma((1-\alpha)\mathbf{S}_{\text{main}} + \alpha\mathbf{S}_{\text{attn}})), \tag{7}$$

where $\sigma : \mathbb{R}^C \rightarrow (0,1)^C$ is the Softmax over the channel dimension.

4.3 *Loss Function*

In our training, we utilize a combination of focal loss[31] and Visual Geometry Group (VGG) loss.[34] The focal loss compensates for the imbalance between background, GGO, and high opacity classes. The importance for each of the classes in focal loss was set to [0.1, 1.0, 1.0], respectively, and the focusing parameter γ was set to 3. This focusing parameter in the focal loss allows the model to penalize the hard to classify samples more than the easy ones. We tap into the low-level features in the VGG network to compute the VGG loss, which represent edge information, for better segmentation output. These losses are weighted equally ($\lambda = 1.0$) during training

$$\mathcal{L} = \mathcal{L}_{\text{focal}} + \lambda\mathcal{L}_{\text{vgg}}. \tag{8}$$

4.4 *Optimization*

The model parameters were optimized using an Adam (adaptive moment estimation) optimizer[35] with initial learning rate of 10^{-3}, weight decay of 10^{-6}, and training batch size of 32. All of the model parameters were initialized using Xavier initialization,[36] except for the dense block, which was initialized using the weights pre-trained on ImageNet.[37] To avoid over-fitting while fully train the model, we use a popular learning rate scheduler called ReduceOnPlateau (Fig. 4).[38] In this technique, a metric (validation loss, accuracy, etc.) is continuously monitored throughout the training. If no improvement is seen in the tracked metric for "patience" number of epochs/iterations, the current learning rate is then reduced by the given "factor." The training continues as usual until the learning rate is reduced beyond a certain minimum (10^{-7}). As soon as the

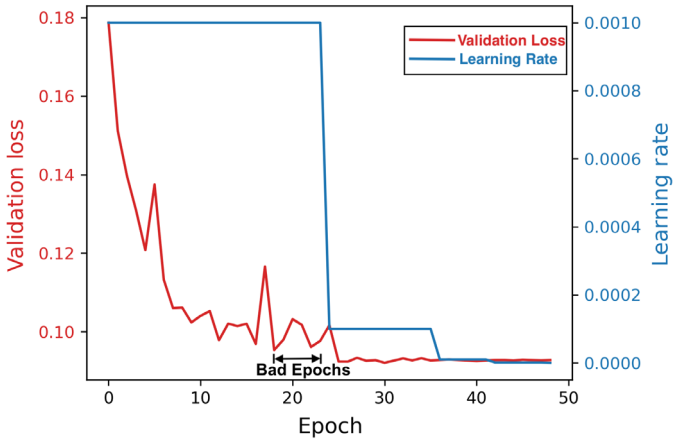

Fig. 4 Reduce-on-plateau learning rate scheduler. Bad epochs refers to the number of epochs for which the validation loss has not decreased.

learning rate hits this minimum, the training is stopped, saving the model at the last best validation metric step. In our experiment, the parameters factor and patience were set to 0.1 and 5, respectively.

4.5 *Implementation*

We trained the model using the Pytorch (v1.7.1) deep-learning framework and incorporated research CT lung analysis software (Deep Lung) written in C++. The training was performed on an NVIDIA TITAN RTX 24GB GPU with a tenth generation Intel Core i9 CPU. Deep Lung can be used with or without the GPU acceleration.

5 Experimental Evaluation

5.1 *Five-Fold Cross-Validation*

The primary endpoint of this study was the performance of the deep-learning method compared with the evaluation by the expert reader. The model is extensively evaluated using the Dice similarity coefficient for structural similarities. The reported Dice score is the mean of per-patient Dice scores computed over all slices in the scan. We also show the quantitative performance on volumes using the Bland–Altman plot and coefficient of determination R^2 (Pearson correlation). To perform a robust non-biased evaluation of the framework, five-fold cross-validation was used, using five independently trained identical models and five exclusive hold-out sets, each of 20%. The whole cohort of $N_{\text{cov}} = 197$ cases was split into five subsets called folds. For each fold of the five-fold cross-validation, the following data splits were used: (1) training split (125 or 126 cases) was used to train the ConvLSTM; (2) alidation split (32 cases) was defined to tune the network, select optimal hyperparameters, and verify that there was no over-fitting; and (3) test split (39 or 40 cases) was used for the evaluation of the method. The final results were obtained by concatenating the results from five test subsets. Thus, the overall test population was 197, referred to as internal test set further in the paper. We also test our model on an unseen external dataset consisting of $N_{\text{ext}} = 68$ patients.

5.2 *Diagnostic Per-Patient Performance*

To assess the diagnostic performance of the convLSTM on a per-patient basis, we trained our model utilizing an additional $N^{\text{train}}_{\text{control}} = 197$ NLST controls (read as number of controls in training) during the five-fold cross-validation, making the total training cases $N = N_{\text{cov}} + N^{\text{train}}_{\text{control}} = 394$. An additional set of $N^{\text{test}}_{\text{control}} = 498$ normal NLST cases (read as number of controls in testing), added during testing, were evaluated with the best fold model from the five-fold cross-validation. Thus, the total normal NLST cases included in experiment sums to $N_{\text{control}} = N^{\text{train}}_{\text{control}} + N^{\text{test}}_{\text{control}} = 695$. Each normal case was evaluated with the model, which did not include these cases for training. We report the specificity at 95% sensitivity for the convLSTM models trained with and without additional controls. The diagnostic sensitivity and specificity was compared using McNemar's test[39] on paired measurements.

6 Results

6.1 *Ablation Study*

Table 4 shows how the results are affected by altering different building blocks of our model.[40] We select the model with the best validation Dice score (mean of GGO and high opacity) for the final evaluation. The model configurations with the highest and lowest performances are highlighted in green and orange, respectively. We experimentally found that the best results were obtained at buffer size $F = 3$.

Table 4 Ablation study on fold-1 for model selection ($N_{cov} = 197$). Highest and lowest performances are highlighted in bold and italic, respectively.

Trunk	Main branch input dim	Buffer size (F)	Feature merge	Validation (Dice score)			
				Background	Ground glass opacity	High opacity	Mean
Dense	3	5	Add	0.9964	0.6727	0.5690	0.6208
Dense	3	3	Add	0.9967	0.6854	0.6172	**0.6513**
Dense	1	3	Add	0.9966	0.6938	0.5958	0.6448
ResNet50	3	3	Add	0.9962	0.6801	0.5221	*0.6011*
Dense	3	3	Max	0.9967	0.6849	0.5788	0.6319

6.2 *Model Comparison*

In Table 5, we show the performance of our model as compared with UNet2D and UNet3D across five-folds ($N_{cov} = 197$). For fair comparison, UNet2D and UNet3D were trained with an identical training setup to our model, i.e., the same loss function, optimizer, learning rate strategy, and training fold splits. The performance is measured with two main metrics: mean Dice score and compute resource utilization. The mean Dice score reported in Table 5 gives the binarized mean Dice score per class. The computation time and memory are calculated for 128 CT slices and 16 CT slices, respectively, on an Nvidia Titan RTX GPU and Intel i9 CPU using Pytorch Profiler.[41] In Fig. 5, we show the significance of our results using the Wilcoxon signed-rank test. We see that our model outperforms UNet2D ($p = 0.001$) and Unet3d ($p < 0.0001$) in segmenting high-opacities, has a comparable performance to UNet2D ($p = 0.22$) in segmenting GGOs, and significantly outperforms UNet3D ($p < 0.0001$) in segmenting GGOs. But the major advantages of our model over the other two are in terms of computational resources as follows:

1. It is nearly 1.3× and 6.8× faster than UNet2D and UNet3D, respectively, on the GPU.
2. It is 2.1× and 1.2× faster than UNet2D and UNet3D, respectively, on the CPU.

Hence, it is can be easily deployed on less powerful machines in clinical setups.

Model complexity in terms of number of trainable parameters and the required tera floating point operations (TFLOPs) is shown in Table 6.

6.3 *Lesion Quantification in the Internal Testing (N_{cov} = 197) and External Testing (N_{ext} = 98) Cohorts*

In Table 7, we present the interquartile range (IQR) and coefficient of determination (R^2) on volumes between expert and automatic segmentation along with the overall per-patient mean

Table 5 Model comparisons on $N_{cov} = 197$ (UNet2D, UNet3D, and our). Best performance is highlighted in bold.

Model	Mean Dice score (five-fold test set)		Model inference[a]		
	Ground glass opacity	High opacity	CPU time (s)	GPU time (ms)	Memory (Gb)
UNet2D	0.9152 ± 0.0526	0.9427 ± 0.0662	91.02	1379.00	**3.60**
UNet3D	0.8949 ± 0.0555	0.9395 ± 0.0679	59.81	6986.67	13.08
Our	**0.9171 ± 0.0502**	**0.9473 ± 0.0611**	**45.01**	**1040.29**	6.65

Note: The data preprocessing is the same for all models and takes about 2.52 s for 128 CT slices.
[a]Time for 128 slices and memory for 16 slices.

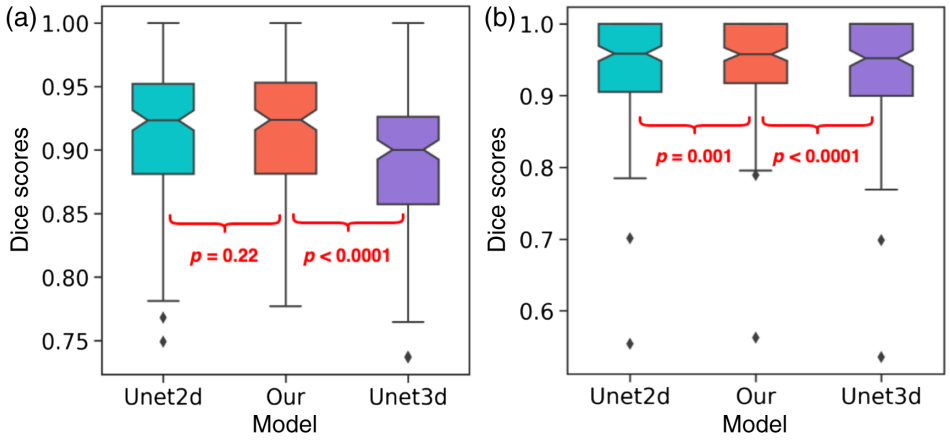

Fig. 5 Box plot for $N_{cov} = 197$ cases displaying the significance of Dice scores between models using the Wilcoxon signed-rank test for (a) GGO and (b) high opacity.

Table 6 Model complexity (UNet2D, UNet3D, and our).

Model	No. of trainable parameters	TFLOPs
UNet2D	17,267,523	0.641
UNet3D	16,318,821	5.143
Our	788,683	0.492

Note: For TFLOPs, the lower the better.

Table 7 Our model performance on $N_{cov} = 197$ and $N_{ext} = 68$.

		Ground glass opacity		High opacity	
		Expert	Automatic	Expert	Automatic
Internal testing dataset ($N_{cov} = 197$)	Median (ml)	288.80	325.71	10.53	8.68
	IQR (ml)	84.74–723.33	89.54–739.71	0–150.42	0.21–94.99
	R^2	0.8664 ($p < 0.001$)		0.9537 ($p < 0.001$)	
	Dice score	0.8918 ± 0.0668			
External testing dataset ($N_{ext} = 68$)	Median (ml)	76.51	74.37	0	0.25
	IQR (ml)	26.42–150.34	27.48–150.03	0–0	0–4.23
	R^2	0.9716 ($p < 0.001$)		0.9529 ($p < 0.001$)	
	Dice score	0.8938 ± 0.0552			

Note: IQR, interquratile range; ml, milliliter; R^2, coefficient of determination.

Dice score for both internal as well as external test datasets. In the internal test set ($N_{cov} = 197$), no significant difference between volumes of expert and automatic segmentations was observed for GGOs ($p = 0.3612$). Similarly, no significant difference between volumes of expert and automatic segmentations was observed for GGOs ($p = 0.1563$) or high opacities in the external test set ($N_{ext} = 98$).

The Bland–Altman analysis on the internal test set demonstrated a low bias of 0.56 (Fig. 6) and 18.61 ml (Fig. 7) for GGOs and high opacities, respectively. Similarly, the Bland-Altman analysis on the external test set demonstrated a low bias of 7.16 (Fig. 8) and 2.92 ml (Fig. 9) for

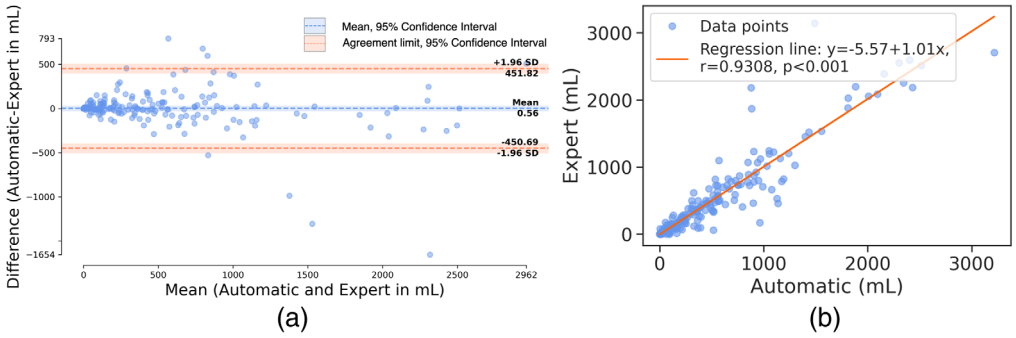

Fig. 6 Expert and automatic quantification of GGO in testing cohort ($N_{cov} = 197$). (a) Bland–Altman plot and (b) best fitting regression line.

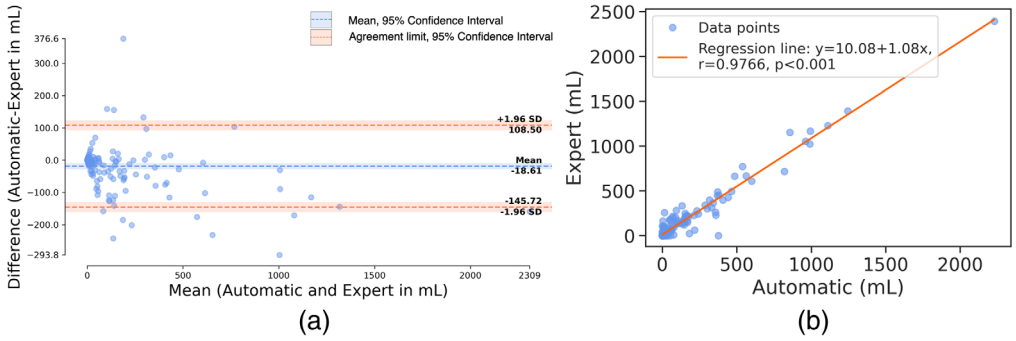

Fig. 7 Expert and automatic quantification of high opacity in testing cohort ($N_{cov} = 197$). (a) Bland–Altman plot and (b) best fitting regression line.

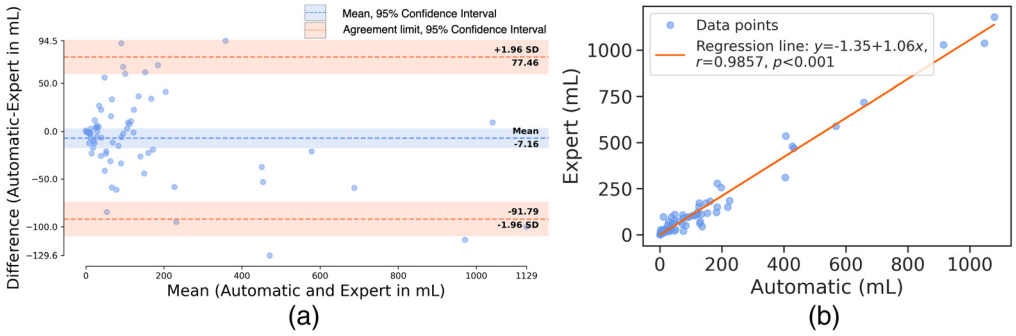

Fig. 8 Expert and automatic quantification of GGO in external unseen testing cohort ($N_{ext} = 68$). (a) Bland–Altman plot and (b) best fitting regression line.

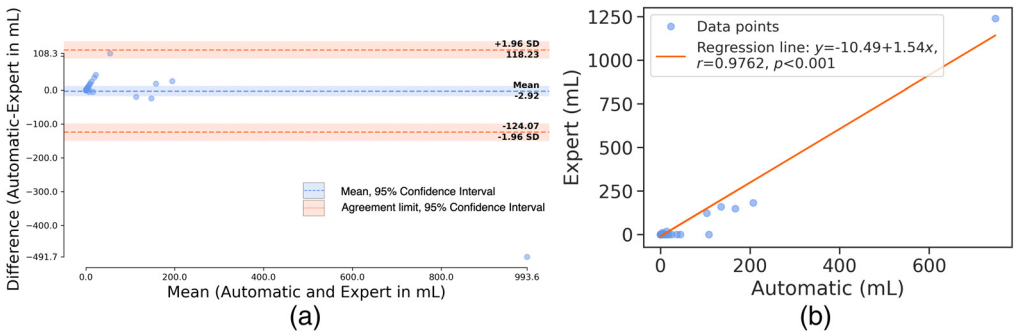

Fig. 9 Expert and automatic quantification of high opacity in external unseen testing cohort ($N_{ext} = 68$). (a) Bland–Altman plot and (b) best fitting regression line.

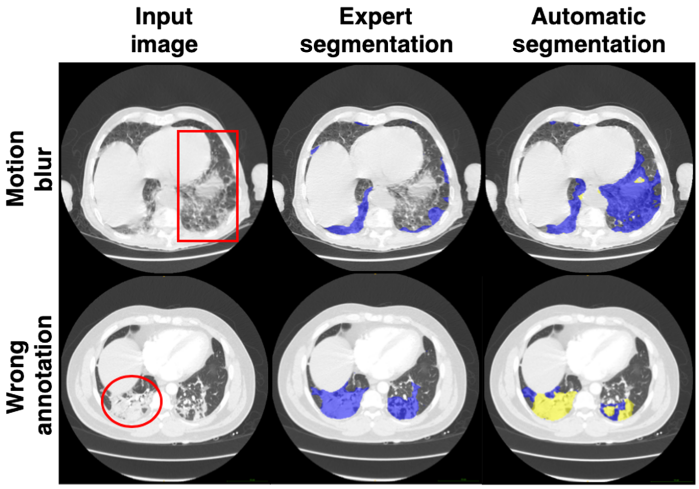

Fig. 10 Samples of extreme outliers from Bland–Altman plots. Highlighted red rectangle and ellipse are the areas of mis-classifications. Blue indicates GGOs, and yellow indicates high opacities.

GGOs and high opacities, respectively. After further analyzing the anomalies (cases outside the limit of agreement) in the Bland–Altman plots (Figs. 6 and 7), we observed that some input scans were corrupted due to various reasons including motion artefacts, errors in expert annotations, etc., as shown in Fig. 10. Thus, a significant ($p < 0.001$) difference in high opacity volumes between expert and automatic segmentations was observed in the internal test set.

The internal testing cohort consisted of 30 contrast enhanced and 167 non-contrast CT scans. We observed no significant difference ($p = 0.2137$) between the mean Dice scores calculated for segmentations from non-contrast and contrast-enhanced CT scans, which were (0.8939 ± 0.0663) and (0.8801 ± 0.0682), respectively.

6.4 *Diagnostic Comparison*

We trained the same convLSTM model with and without additional $N_{\text{control}}^{\text{train}} = 197$ controls and tested them with five-fold cross-validation. We also tested an additional $N_{\text{control}}^{\text{test}} = 498$ unseen controls with the best performing model from five-fold cross-validation. The AUROC with and without NLST in training was 0.965 and 0.959, respectively, but they did not reach significance. However, McNemar's test results (Table 8) show that the model trained with an additional $N_{\text{control}}^{\text{train}} = 197$ NLST cases significantly increased the specificity at 95% sensitivity of the model. Thus, adding NLST controls to the training decreased the false positive rate in diagnosis. The overall per-patient mean Dice score also improved drastically, as shown in Table 8.

7 Discussion

We developed and evaluated a novel deep-learning ConvLSTM network approach for fully automatic quantification of the COVID-19 pneumonia burden from both non-contrast and contrast-

Table 8 Diagnostic performance on $N_{\text{total}} = N_{\text{cov}} + N_{\text{control}} = 892$ NLST patients.

NLST in training	AUROC	Dice score[a]	Sensitivity/specificity	McNemar's test		
				χ^2 statistics	$\chi^2_{1,0.05}$	*p*-value
No	0.959	0.9813 ± 0.0398	95.0%/70.8%	30.22	3.841	<0.0001
Yes	0.965	0.9803 ± 0.0433	**95.0%/77.3%**			

Note: AUROC, area under the receiver operating characteristic.
Best performance is highlighted in bold.
[a]Includes GGO, high opacity, and background.

enhanced chest CT. To the best of our knowledge, ConvLSTM has not been applied before for segmentation of medical imaging data. We demonstrated that automatic pneumonia burden quantification by the proposed method shows strong agreement with expert manual measurements and rapid performance that is suitable for clinical deployment. Although vaccines have been developed to protect from COVID-19, the incidental findings of COVID-19 abnormalities due to imperfect vaccination rates and new strains will be a mainstay of medical practice. This method will provide a 'real-time' detection of parenchymal opacifications associated with COVID-19 to the physician and aid image-based triage to optimize the distribution of resources during the pandemic. Figure 11 shows the lesion annotations (expert and automatic) in 3D for one of the patients in test set.

The evolution of deep-learning applications for COVID-19 is reflecting the changing role of CT imaging during the pandemic. Initially, when RT-PCR testing was unavailable or delayed, chest CT was used as a surrogate tool to identify suspected COVID-19 cases.[42] AI-assisted image analysis could improve the diagnostic accuracy of junior doctors in differentiating COVID-19 from other chest diseases including community-acquired pneumonia and facilitate prompt isolation of patients with suspected SARS-CoV-2 infection.[20,43]

Currently, when RT-PCR testing is widely available with timely results, rapid quantification of the pneumonia burden from chest CT as proposed here can aid prognostication and disease staging in patients with COVID-19. As demonstrated in prior investigations, increasing attenuation of GGO and a higher proportion of consolidation in the total pneumonia burden had prognostic value, thus underscoring the importance of utilizing all CT information for training the patients.[13,44] Manual segmentation of the lung lesions is, however, challenging and prohibitively time-consuming task due to complex appearances and ambiguous boundaries of the opacities.[45] To automate the segmentation of respective lung lesions in COVID-19, several different segmentation networks have been introduced.[11,20,22,46] Most of these tend to consume a lot of

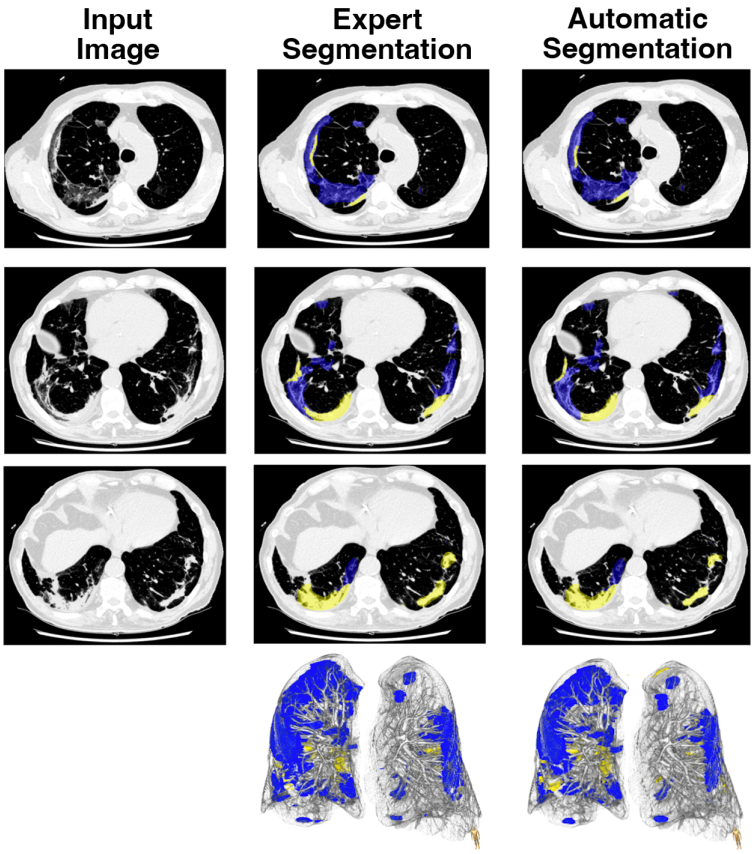

Fig. 11 Qualitative comparison between expert and automatic segmentations of the lung lesions using our system. Blue represents GGO, and yellow represents high-opacity. The last row is the 3D representation. The Dice score coefficient for this patient was 0.792.

memory in storing the intermediate features for skip connections, and it may be favorable to use several input slices to improve the performance of semantic segmentation tasks.[24,25] We propose the application of ConvLSTM, presenting the potential to outperform other neural networks in capturing the spatio-temporal correlations, due to its capability of preserving relevant features with simultaneous dismission of irrelevant ones in the form of the feedback loop for the memory-sparing strategy and holistic analysis of the images.[28] It has been found that ConvLSTM localized at the input end allowed for effectively capturing the global information and optimizing the model performance.

Automated segmentation of lung lesions with ConvLSTM networks offers a solution to generating big data with limited human resources and minimal hardware requirements. Because results of segmentation are presented to the human reader for visual inspection, eventual corrections enable the implementation of a human-in-the-loop strategy to reduce the annotation effort and provide high-volume training datasets to improve the performance of deep-learning models.[45] Furthermore, objective and repeatable quantification of the pneumonia burden might aid the evaluation of the disease progression and assist the tomographic monitoring of different treatment responses.

Our study had several limitations. First, different patient profiles and treatment protocols between countries may have resulted in heterogeneity in COVID-19 pneumonia severity. Second, most of the CT scans were acquired during the hospital admission; therefore, availability of the slices with high-opacity (consolidations and plural effusion), representing a peak stage of the disease, was limited. Finally, training and external validation datasets comprised a relatively low number of patients manually segmented by two expert readers; however, to mitigate this, we have utilized repeated testing that has allowed us to evaluate expected average performance of the model.

In our experiments, we have a diverse multi-center cohort not typically available for training. But for future research, in experiments with limited availability of expertly annotated data, it is desirable to incorporate advanced data augmentation techniques as proposed in Refs. 47 and 48 and regularization techniques[49] for better model generalization and for mitigating the issue of over-fitting.

8 Conclusions

We proposed and evaluated a deep-learning method based on convolutional LSTM and Hierarchical multi-scale attention network for fully automated quantification of the pneumonia burden in COVID-19 patients from both non-contrast and contrast-enhanced CT datasets. The proposed method provided rapid segmentation of lung lesions with strong agreement with manual segmentation and may represent a robust tool to generate big data with an accuracy similar to that of an expert reader. The model generalized very well on unseen external datasets. In our proposed method, the attention network using ConvLSTM largely helps with error correction in segmentation and can be used in other segmentation tasks in which one can leverage information from adjacent slices of the scan.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments

This research was supported by Cedars-Sinai COVID-19 funding. This research was also supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH; R01HL133616). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Kajetan Grodecki was supported by the Foundation for Polish Science (FNP). IRCCS Istituto Auxologico Italiano research was supported by the Italian Ministry of Health. We thank the National Lung Screening Trial (NLST) consortium for supporting our research by providing us with valuable data. A preliminary version[50] of this work with a subset of patients was presented at SPIE Medical Imaging 2022.

References

1. "World Health Organization Coronavirus (COVID-19) Dashboard," https://covid19.who.int/table (2 March 2022).
2. B. Böger et al., "Systematic review with meta-analysis of the accuracy of diagnostic tests for covid-19," *Am. J. Infect. Control* **49**(1), 21–29 (2021).
3. M. Francone et al., "Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis," *Eur. Radiol.* **30**(12), 6808–6817 (2020).
4. F. Khatami et al., "A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis," *Sci. Rep.* **10**(1), 1–12 (2020).
5. T. C. Kwee and R. M. Kwee, "Chest CT in COVID-19: what the radiologist needs to know," *RadioGraphics* **40**(7), 1848–1865 (2020).
6. A. Bernheim et al., "Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection," *Radiology* **295**, 200463 (2020).
7. Y. Wang et al., "Temporal changes of CT findings in 90 patients with COVID-19 pneumonia: a longitudinal study," *Radiology* **296**(2), E55–E64 (2020).
8. F. Pan et al., "Time course of lung changes at chest CT during recovery from coronavirus disease 2019 (COVID-19)," *Radiology* **295**(3), 715–721 (2020).
9. K. Li et al., "CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19)," *Eur. Radiol.* **30**(8), 4407–4416 (2020).
10. R. Yang et al., "Chest CT severity score: an imaging tool for assessing severe COVID-19," *Radiol.: Cardiothorac. Imaging* **2**(2), e200047 (2020).
11. S. Chaganti et al., "Automated quantification of ct patterns associated with COVID-19 from chest CT," *Radiol.: Artif. Intell.* **2**(4), e200048 (2020).
12. C. Gieraerts et al., "Prognostic value and reproducibility of AI-assisted analysis of lung involvement in COVID-19 on low-dose submillisievert chest CT: sample size implications for clinical trials," *Radiol.: Cardiothorac. Imaging* **2**(5), e200441 (2020).
13. K. Grodecki et al., "Quantitative burden of COVID-19 pneumonia on chest CT predicts adverse outcomes: a post-hoc analysis of a prospective international registry," *Radiol.: Cardiothorac. Imaging* **2**(5), e200389 (2020).
14. A. B. De González et al., "Projected cancer risks from computed tomographic scans performed in the United States in 2007," *Arch. Internal Med.* **169**(22), 2071–2077 (2009).
15. D. Albano et al., "Incidental findings suggestive of COVID-19 in asymptomatic patients undergoing nuclear medicine procedures in a high-prevalence region," *J. Nucl. Med.* **61**(5), 632–636 (2020).
16. V. Habouzit et al., "Incidental finding of COVID-19 lung infection in 18F-FDG PET/CT: what should we do?" *Clin. Nucl. Med.* **45**, 649–651 (2020).
17. S. Neveu et al., "Incidental diagnosis of COVID-19 pneumonia on chest computed tomography," *Diagn. Intervent. Imaging* **101**(7–8), 457–461 (2020).
18. A. Pallardy et al., "Incidental findings suggestive of COVID-19 in asymptomatic cancer patients undergoing 18F-FDG PET/CT in a low prevalence region," *Eur. J. Nucl. Med. Mol. Imaging* **48**(1), 287–292 (2021).
19. R. V. Ramanan et al., "Incidental chest computed tomography findings in asymptomatic COVID-19 patients. A multicentre Indian perspective," *Indian J. Radiol. Imaging* **31**(Suppl. 1), S45 (2021).
20. K. Zhang et al., "Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography," *Cell* **181**(6), 1423–1433.e11 (2020).
21. H. X. Bai et al., "Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT," *Radiology* **296**(3), E156–E165 (2020).
22. D.-P. Fan et al., "Inf-Net: automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imaging* **39**(8), 2626–2637 (2020).
23. K. Gao et al., "Dual-branch combination network (DCN): towards accurate diagnosis and lesion segmentation of COVID-19 using CT images," *Med. Image Anal.* **67**, 101836 (2021).
24. Ö. Çiçek et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.* **9901**, 424–432 (2016).

25. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: fully convolutional neural networks for volumetric medical image segmentation," in *Fourth Int. Conf. 3D vision (3DV)*, IEEE, pp. 565–571 (2016).
26. A. Tao, K. Sapra, and B. Catanzaro, "Hierarchical multi-scale attention for semantic segmentation," arXiv:2005.10821 (2020).
27. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.* **9**(8), 1735–1780 (1997).
28. X. Shi et al., "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Proc. 28th Int. Conf. Adv. Neural Inf. Process. Syst*. (2015).
29. S. Morozov et al., "Mosmeddata: chest CT scans with COVID-19 related findings dataset," arXiv:2005.06465 (2020).
30. N. L. S. T. R. Team, "The national lung screening trial: overview and study design," *Radiology* **258**(1), 243–253 (2011).
31. T.-Y. Lin et al., "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2980–2988 (2017).
32. G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4700–4708 (2017).
33. A. Pfeuffer, K. Schulz, and K. Dietmayer, "Semantic segmentation of video sequences with convolutional LSTMs," in *IEEE Intell. Veh. Symp. (IV)*, IEEE, pp. 1441–1447 (2019).
34. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4681–4690 (2017).
35. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
36. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Thirteenth Int. Conf. Artif. Intell. and Stat., JMLR Workshop and Conf. Proc.*, pp. 249–256 (2010).
37. J. Deng et al., "ImageNet: a large-scale hierarchical image database," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, IEEE, pp. 248–255 (2009).
38. Pytorch, "Learning rate scheduler: ReduceLROnplateau," 2019, https://pytorch.org/docs/1.7.1/_modules/torch/optim/lr_scheduler.html#ReduceLROnPlateau.
39. Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika* **12**(2), 153–157 (1947).
40. R. Meyes et al., "Ablation studies in artificial neural networks," arXiv:1901.08644 (2019).
41. Pytorch, "CPU and GPU profiler," 2019, https://pytorch.org/docs/1.7.1/autograd.html#profiler.
42. T. Ai et al., "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology* **296**(2), E32–E40 (2020).
43. L. Li et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy," *Radiology* **296**(2), E65–E71 (2020).
44. K. Grodecki et al., "Epicardial adipose tissue is associated with extent of pneumonia and adverse outcomes in patients with COVID-19," *Metabolism* **115**, 154436 (2021).
45. G. Wang et al., "A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images," *IEEE Trans. Med. Imaging* **39**(8), 2653–2663 (2020).
46. A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus segnet," *BMC Med. Imaging* **21**, 19 (2021).
47. Q. Zheng et al., "A full stage data augmentation method in deep convolutional neural network for natural image classification," *Discr. Dyn. Nat. Soc.* **2020**, 4706576 (2020).
48. Q. Zheng et al., "Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification," *Neural Comput. Appl.* **33**(13), 7723–7745 (2021).
49. Q. Zheng et al., "MR-DCAE: manifold regularization-based deep convolutional autoencoder for unauthorized broadcasting identification," *Int. J. Intell. Syst.* **36**(12), 7204–7238 (2021).

50. A. Killekar et al., "COVID-19 lesion segmentation using convolutional LSTM for self-attention," *Proc. SPIE* **12032**, 120323P (2022).

Aditya Killekar is a programmer/analyst at the Cedars-Sinai Medical Center, Los Angeles, California. He received his MS degree in electrical engineering from the University of Southern California in 2018. He specializes in computer vision and machine learning. His current research interests include applications of deep learning in cardiac imaging. Apart from research, he is passionate about teaching and has served as a volunteer to educate and inspire students from various parts of Los Angeles.

Kajetan Grodecki, MD, PhD, graduated from Medical University of Warsaw and he is currently working as a cardiology resident. He is interested in non-invasive modalities to optimize interventional procedures as well as developing AI-based solutions to imporve risk stratification.

Piotr Slomka is the Director of Innovation in Imaging, Professor of Medicine and Cardiology, Division of Artificial Intelligence in Medicine, Cedars-Sinai, and Professor of Medicine In-Residence, UCLA School of Medicine. He received his PhD in medical biophysics from the University of Western Ontario. He serves as PI for an NIH R35 Outstanding Investigator Award aimed to transform the clinical utility of PET/CT in detection and management of high-risk coronary artery disease.

Biographies of the other authors are not available.