



Automated Lung Segmentation on Chest Computed Tomography Images with Extensive Lung Parenchymal Abnormalities Using a Deep Neural Network

Seung-Jin Yoo, MD¹, Soon Ho Yoon, MD, PhD², Jong Hyuk Lee, MD², Ki Hwan Kim, MD³,
Hyoung In Choi, MD⁴, Sang Joon Park, PhD^{2, 5}, Jin Mo Goo, MD, PhD²

¹Department of Radiology, Hanyang University Medical Center, Hanyang University College of Medicine, Seoul, Korea; ²Department of Radiology, Seoul National University Hospital, Seoul National College of Medicine, Seoul, Korea; ³Department of Radiology, Myongji Hospital, Goyang, Korea; ⁴Korean Armed Forces Capital Hospital, Seongnam, Korea; ⁵MEDICALIP Co. Ltd., Seoul, Korea

Objective: We aimed to develop a deep neural network for segmenting lung parenchyma with extensive pathological conditions on non-contrast chest computed tomography (CT) images.

Materials and Methods: Thin-section non-contrast chest CT images from 203 patients (115 males, 88 females; age range, 31–89 years) between January 2017 and May 2017 were included in the study, of which 150 cases had extensive lung parenchymal disease involving more than 40% of the parenchymal area. Parenchymal diseases included interstitial lung disease (ILD), emphysema, nontuberculous mycobacterial lung disease, tuberculous destroyed lung, pneumonia, lung cancer, and other diseases. Five experienced radiologists manually drew the margin of the lungs, slice by slice, on CT images. The dataset used to develop the network consisted of 157 cases for training, 20 cases for development, and 26 cases for internal validation. Two-dimensional (2D) U-Net and three-dimensional (3D) U-Net models were used for the task. The network was trained to segment the lung parenchyma as a whole and segment the right and left lung separately. The University Hospitals of Geneva ILD dataset, which contained high-resolution CT images of ILD, was used for external validation.

Results: The Dice similarity coefficients for internal validation were $99.6 \pm 0.3\%$ (2D U-Net whole lung model), $99.5 \pm 0.3\%$ (2D U-Net separate lung model), $99.4 \pm 0.5\%$ (3D U-Net whole lung model), and $99.4 \pm 0.5\%$ (3D U-Net separate lung model). The Dice similarity coefficients for the external validation dataset were $98.4 \pm 1.0\%$ (2D U-Net whole lung model) and $98.4 \pm 1.0\%$ (2D U-Net separate lung model). In 31 cases, where the extent of ILD was larger than 75% of the lung parenchymal area, the Dice similarity coefficients were $97.9 \pm 1.3\%$ (2D U-Net whole lung model) and $98.0 \pm 1.2\%$ (2D U-Net separate lung model).

Conclusion: The deep neural network achieved excellent performance in automatically delineating the boundaries of lung parenchyma with extensive pathological conditions on non-contrast chest CT images.

Keywords: *Deep learning; Artificial intelligence; Lung; Computed tomography; Interstitial lung diseases*

INTRODUCTION

Computer-aided diagnosis (CAD) for chest computed tomography (CT) images is widely used to detect and analyze various lung parenchymal diseases including lung nodules (1, 2), interstitial lung disease (ILD) (3),

and emphysema (4, 5). Precise segmentation of lung parenchymal areas from CT images is a prerequisite to automatically quantify such lung parenchymal diseases (6).

Conventional automated lung segmentation, using methods such as thresholding and region growing, is effective for normal chest CT images due to the distinct

Received: March 20, 2020 **Revised:** May 31, 2020 **Accepted:** June 28, 2020

Corresponding author: Soon Ho Yoon, MD, PhD, Department of Radiology, Seoul National University Hospital, Seoul National College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.

• E-mail: yshoka@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

difference of CT attenuation between the lung parenchyma and chest wall (7). However, automated lung segmentation is challenging on chest CT images for extensive lung parenchymal diseases, particularly for those with subpleural lung pathologies (7).

Non-contrast chest CT, especially low-dose chest CT (LDCT), is one of the most commonly used CT protocols, as it provides sufficient image quality for radiologists and clinicians to evaluate lung parenchymal abnormalities with relatively low radiation exposure. LDCT is currently used in lung cancer screening and routine clinical practice, particularly for patients in need of repetitive follow-up CT studies due to chronic lung diseases (8). However, automated lung segmentation for chronic lung diseases is potentially challenging using LDCT due to its inherent high level of image noise.

The purpose of this study was to develop a deep neural network for segmenting lung parenchyma with extensive pathological conditions on non-contrast chest CT images, primarily including LDCT.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board and the requirement for patient consent was waived (IRB No. 1902-103-101).

CT Datasets

We retrospectively collected 193 LDCT scans from patients who had visited respiratory physicians and had undergone thin-section chest CT scan at a single institution between January 2017 and May 2017, including 53 LDCT scans without diffuse lung parenchymal disease and 140 LDCT scans with diffuse lung parenchymal disease. Because LDCT was preferentially performed for lung cancer screening and serial follow-up for chronic, indolent lung diseases in outpatient settings, we included 10 additional cases of standard-dose chest CT scans in the emergency department to enrich the dataset for acute severe lung diseases, such as acute respiratory distress syndrome, acute exacerbation of ILD, extensive lung malignancy, and atelectasis. Extensive lung parenchymal disease was defined as lung disease involving more than 40% of the lung parenchymal area on chest CT images (9, 10), including ILD (49 cases), emphysema (36 cases), nontuberculous mycobacterial lung disease (23 cases), tuberculous destroyed lung (15 cases), pneumonia (9 cases), lung cancer (4 cases), and other

diseases (14 cases) (Fig. 1).

CT Acquisition

All 203 CT scans were acquired with one of the following multi-detector CT scanners: Somatom Definition and Somatom Force (Siemens Healthineers); Brilliance 64, IQon spectral CT, Brilliance iCT Elite, and Ingenuity (Philips Healthcare); Aquilion One (Canon Medical Systems); and Discovery CT 750 HD (GE Healthcare). All CT examinations were performed with tube voltage of 70–150 kVp and tube current of 25–185 mAs with a volume CT dose index of 0.52–2.92 mGy (low-dose) and 2.83–14.2 mGy (standard dose). Axial images were reconstructed with sharp reconstruction kernel at 1.0-mm slice thickness. Details are provided in Supplementary Table 1.

Manual Lung Segmentation

One of five experienced board-certified radiologists (15, 10, 7, 6, and 5 years of clinical experience with chest CT interpretation, respectively) participated in preparing the reference mask of chest CT images. After uploading the CT images to a commercially available software program (MEDIP, version 1.3.2.0, Medical IP), lung parenchymal areas were initially segmented using a threshold below -400 to -500 Hounsfield unit. Then, the radiologists reviewed the result of the initial segmentation, and adjusted the boundary of the mask correctly by creating a free-drawing region-of-interest in every axial CT image slice. If needed, the radiologists additionally reviewed the coronal and sagittal images to capture the exact boundary of the lung mask, particularly for apical and basal lung areas. The radiologists modified the CT window settings, as needed. The lateral boundary of the lung mask was the outmost end of the subpleural lung, and the medial boundary of the lung mask was the innermost end of the lung parenchyma abutting the mediastinum. The radiologists included the lobar to subsegmental bronchi, arteries, and veins in the lung mask, while excluding the bilateral main bronchi, main pulmonary arteries, and veins. Any lung parenchymal pathologies, including subpleural lesions, such as honeycombing or fibrotic lesions in cases of idiopathic pulmonary fibrosis and pneumonic consolidations, were included in the lung mask. Pleural pathologies, including pleural calcifications, thickening, or pleural effusions, were excluded. A final review of the lung mask, with any modifications, was performed by two of the five radiologists in consensus.

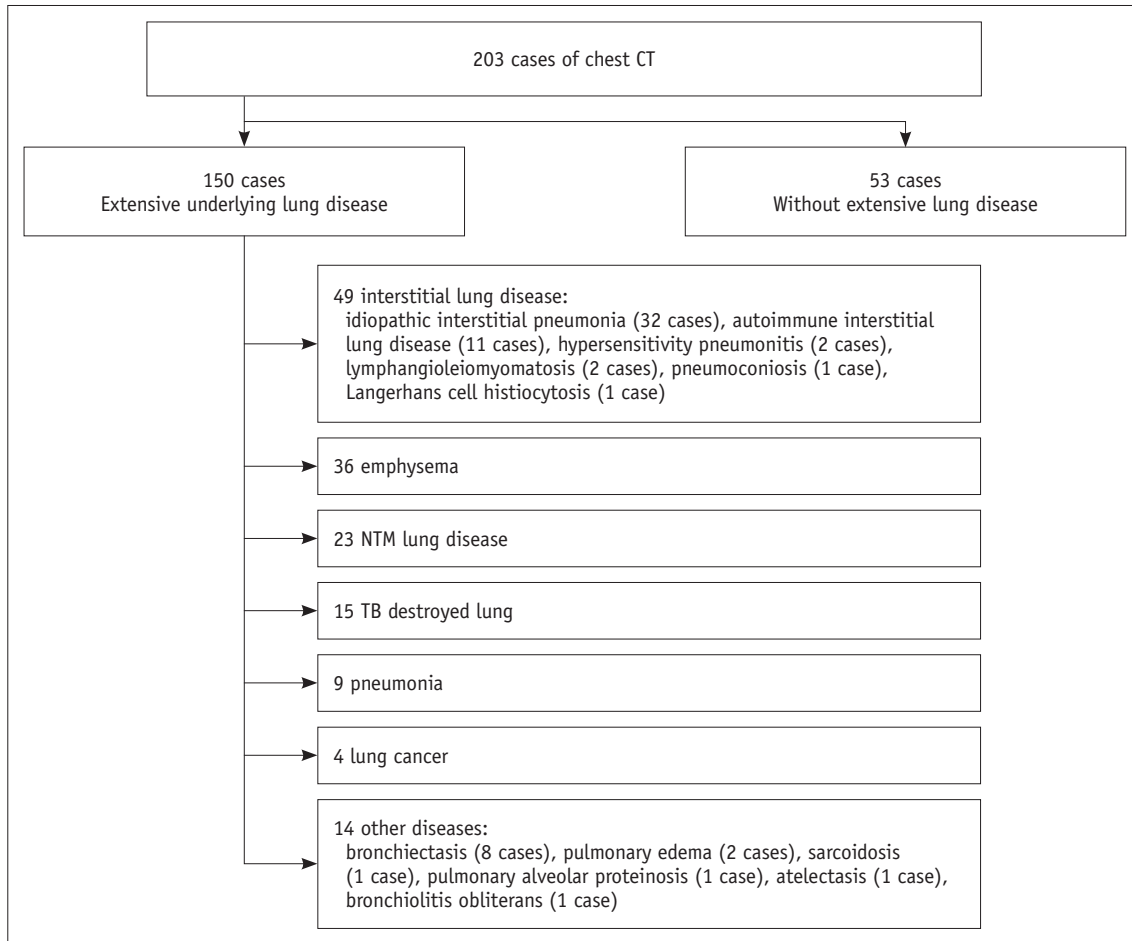


Fig. 1. Distribution of cases for the development of deep neural networks. CT = computed tomography, NTM = nontuberculous mycobacterium, TB = tuberculosis

Deep Learning-Based Training and Validation

Altogether, 203 CT scans were randomly assigned to one of the three following data sets: training set, 157 cases; tuning set, 20 cases; and internal validation set, 26 cases. Data were normalized with the lung window setting, using two-dimensional (2D) and three-dimensional (3D) U-Net models. In total, 42306 slices of axial data with existing lung areas were selected, while 8609 negative samples consisted of randomly selected slices without lung areas.

Our 2D U-Net received an input size of 512 x 512 x 1 and consisted of initial convolutions, four encoders, four decoders, and a final convolution. Except for the final convolution, which was a 1 x 1 convolution, every convolutional layer consisted of a 3 x 3 convolution followed by batch normalization (11) and the rectified linear unit (ReLU) activation function (12). For decoders, up-sampling with bilinear interpolation was used, followed by concatenation to conserve information before down-sampling (Fig. 2).

Our 3D U-Net model received an input size of 512 x 512 x 8 and used three encoders and three decoders. Except for the final convolution, which was a 1 x 1 x 1 convolution, every convolution layer consisted of a 3 x 3 x 3 convolution followed by ReLU and group normalization (13). The first encoder used 1 x 2 x 2 max pooling to preserve data in the z-axis, whereas the second and third encoders used 2 x 2 x 2 max pooling. For decoders, up-sampling with trilinear interpolation was used.

The He et al. (14) initialization method was used for weight initialization. Both models used the softmax function in the final layer and were trained using the stochastic gradient descent algorithm and the cross entropy loss function. After completion of training, the tuning dataset was used to choose the best weight, which was saved after each epoch.

We applied two types of training for the deep neural networks. First, the reference masks of a whole lung were used as input for training (whole lung model). Second, we

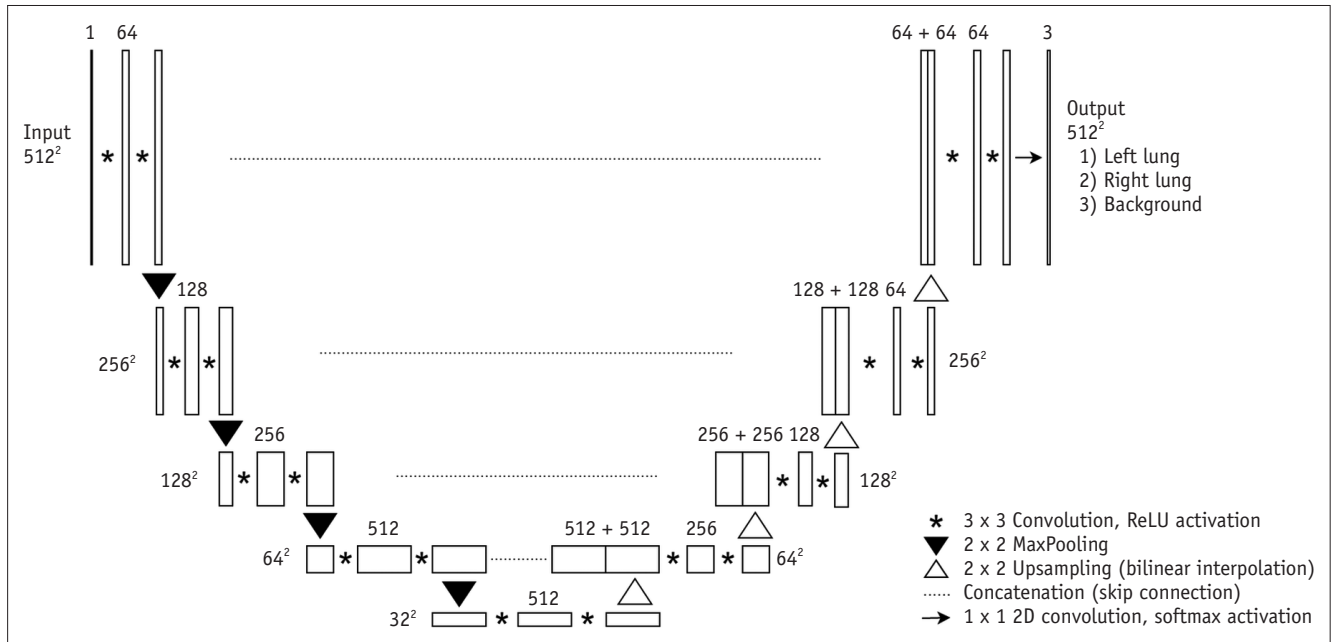


Fig. 2. 2D U-Net architecture used for lung segmentation. Our 2D U-Net consists of four down-sampling, and four up-sampling steps. Every step except for the final convolution, consists of two consecutive 3 x 3 convolution followed by batch normalization and ReLu activation function and 1 x 1 convolution with softmax activation was performed at final convolution. 2D = two-dimensional, ReLu = rectified linear unit

separated the reference masks of a whole lung into right and left lung masks (separate lung model). Then, the right lung masks were horizontally flipped onto the left side. The left lung masks and the flipped right lung masks were used to train the deep neural network to extract the left lung and vice versa.

External Validation

For external validation, we used the University Hospitals of Geneva (HUG)-ILD dataset (15), which consisted of 109 annotated CT scans with advanced ILD. Chest CT scans in the HUG-ILD dataset generally have a slice thickness of 1–2 mm with a 10–15 mm slice interval, corresponding to a high-resolution CT protocol. In some scans, the lung masks provided as the ground truth were inaccurate. We excluded cases in which the CT images had profound respiratory motion artifacts ($n = 3$) or incomplete ground truth lung segmentation ($n = 4$). In three cases, CT scans were divided into two separate series. Therefore, six of the 102 CT scans were merged into three CT scans. In total, 99 cases were included for external validation. Since the ground truth masks in the HUG-ILD dataset included the trachea and the main and lobar bronchi in the lung mask, a technician generated an airway mask using the region growing method, and the airway mask was combined with our U-Net-driven lung mask to assess the accuracy of lung segmentation.

Statistical Analysis

We compared the lung mask obtained through manual segmentation with that generated with deep-learning based automated segmentation using the Dice similarity coefficient (DSC), sensitivity, positive predictive value (PPV), and the Hausdorff distance (16). DSC considers correctly segmented areas, incorrectly segmented areas, and missing target areas to measure the performance of a classifier. PPV is an accuracy measurement that reflects the proportion of correctly predicted areas compared to all predicted areas. The Hausdorff distance is the largest value of all distances from a point in one set to the closest point in the other set. Measuring the Hausdorff distance for all pixels of the lung mask required substantial time and computing power resources. Accordingly, we calculated the Hausdorff distance based on a randomly selected sample of 1% of the pixels in the lung mask. The Hausdorff distance calculated using these randomly selected pixels between the same masks was not zero due to discrepancies in the location of the randomly selected pixels, inevitably resulting in a certain degree of error in the distance. We calculated the Hausdorff distance between the same two masks, and then subtracted the distance from the Hausdorff distance between the manual and deep-learning-driven masks.

DSC, sensitivity, PPV, and Hausdorff distance were analyzed with the internal and external validation datasets

divided by the extent of damage in the pathological lung on CT images (disease severity) into two categories (underlying lung disease involving $\leq 40\%$ or $> 40\%$ of the lung parenchymal area), three categories (underlying lung disease involving $\leq 25\%$, $> 25\%$ but $\leq 75\%$, and $> 75\%$ of the lung parenchymal area), and disease category.

Differences in the mask volume calculated with the manual and U-Net segmentation masks were compared using two-way analysis of variance with Bonferroni correction for multiple comparisons. Intraclass correlation coefficients (ICCs) were calculated between the manual segmentation masks and the 2D and 3D U-Net segmentation masks. Bland-

Altman plots were used to evaluate the differences in mask volumes between the manual segmentation masks and the 2D and 3D U-Net segmentation masks.

SPSS version 25 (IBM Corp.) was used for all statistical analyses.

RESULTS

The basic characteristics of the internal and external validation datasets are shown in Table 1.

Regarding the 2D U-Net model, DSC, sensitivity, PPV, and Hausdorff distance of the internal validation set were $99.6 \pm$

Table 1. Demographics and Clinic-Radiologic Characteristics of the Datasets

Characteristics	Training Dataset	Tuning Dataset	Internal Validation Datasets	External Validation Datasets
Number of subjects*	157	20	26	99
Age (years) [†]	69.64 \pm 9.60	68.60 \pm 11.36	68.73 \pm 8.71	58.35 \pm 20.15
Sex (M:F)	90:67	8:12	17:9	61:38
CT parameters				
Tube voltage (kVp)	70–150	100–120	100–120	N/A
Effective mAs (mAs)	15–185	15–150	15–150	N/A
Slice thickness (mm)	1–1.25	1–1.25	1	1–2
Reconstruction interval (mm)	1–1.25	1–1.25	1	10–15
CTDIvol (mGy)	0.52–14.2	0.52–2.92	0.52–2.17	N/A
Underlying extensive lung disease ($> 40\%$)*				
ILD	38	6	5	51
Emphysema	25	3	8	0
NTM lung disease	21	1	1	0
TB destroyed lung	10	3	2	0
Pneumonia	8	1	0	3
Lung cancer	3	0	1	0
Others	9	2	3	11
Without extensive lung disease ($\leq 40\%$)*				
ILD	1	0	1	17
Emphysema	3	0	0	0
NTM lung disease	3	0	0	0
TB destroyed lung	4	1	0	0
Pneumonia	0	0	0	1
Lung cancer	0	0	0	0
Others	32	3	5	16
Pleural effusion				
None*	147	18	26	85
Unilateral*	5	1	0	5
Bilateral	5	1	0	9
Anterior junctional line thickness (mm)*				
≤ 2	38	6	9	36
> 2	119	14	17	63

*Data are number of subjects, [†]Data are mean \pm standard deviation. ILD = interstitial lung disease, NTM = nontuberculous mycobacterium, TB = tuberculosis

0.3%, $99.5 \pm 0.3\%$, $99.6 \pm 0.3\%$, and 17.70 ± 6.62 pixels for the whole lung model and $99.5 \pm 0.3\%$, $99.5 \pm 0.3\%$, $99.5 \pm 0.4\%$, and 18.29 ± 6.51 pixels for the separate lung model, respectively (Table 2). Regarding the 3D U-Net model, DSC, sensitivity, PPV, and Hausdorff distance of the internal validation dataset were $99.4 \pm 0.5\%$, $99.1 \pm 0.9\%$, $99.7 \pm 0.2\%$, and 18.75 ± 7.48 pixels for the whole lung model, and $99.4 \pm 0.5\%$, $99.1 \pm 0.8\%$, $99.6 \pm 0.3\%$, and 18.16 ± 7.48 pixels for the separate lung model, respectively (Table 2).

Regarding the external validation using the HUG-ILD dataset, the 2D U-Net model showed DSC, sensitivity, PPV, and Hausdorff distance values of $98.4 \pm 1.0\%$, $98.7 \pm 1.3\%$, $98.1 \pm 1.5\%$, and 7.66 ± 3.93 pixels for the whole lung model and $98.4 \pm 1.0\%$, $98.7 \pm 1.1\%$, $98.0 \pm 1.6\%$, and 7.59 ± 3.69 pixels for the separate lung model, respectively (Table 3). The 3D U-Net models showed DSC, sensitivity, PPV, and Hausdorff distance values of $95.3 \pm 3.1\%$, $98.0 \pm 1.9\%$, $92.8 \pm 4.6\%$, and 15.58 ± 5.60 pixels for the whole lung model and $96.1 \pm 2.2\%$, $98.1 \pm 1.9\%$, $94.3 \pm 3.5\%$, and 11.67 ± 4.84 pixels for the separate lung model, respectively (Table 3).

Subgroup analyses of the internal and external validation datasets are summarized in Supplementary Tables 2 and 3.

The mean DSC of the 2D U-Net whole and separate lung models was high in cases with underlying lung disease involving $\leq 25\%$ of the lung parenchymal area in the internal (99.7% and 99.7% , respectively) and external

datasets (98.9% and 98.9% , respectively), and in cases with underlying lung disease occupying more than 75% of the lung parenchymal area in the internal (99.3% and 99.4% , respectively) and external validation sets (97.9% and 98.0% , respectively) (Supplementary Tables 2, 3). The mean DSC of the 3D U-Net whole and separate lung models in cases with underlying lung disease occupying $> 75\%$ of the lung parenchymal area was lower in the external validation dataset (93.7% and 94.8% , respectively) than in the internal validation dataset (99.2% and 99.2% , respectively) (Supplementary Tables 2, 3).

The mean DSC of the internal validation dataset divided into seven disease categories was over 98.8% in all models (Supplementary Table 2). In the external validation dataset, the performance of the 2D U-Net model was excellent in all categories, with mean DSC over 96.8% ; however, the performance of the 3D U-Net model was good but unsatisfactory, with mean DSC over 92.8% (Supplementary Table 3).

Two-way analysis of variance of lung volumes among the 2D and 3D whole and separate lung models showed no significant difference in either the internal validation ($p = 0.997$) or the external validation dataset ($p = 0.784$). ICCs of lung volumes between the manually segmented masks and each set of deep-learning-driven masks are shown in Supplementary Table 4.

The percentage difference and limits of agreement of volumes between the manually segmented (ground truth)

Table 2. Dice Score, Sensitivity, PPV, and Hausdorff Distance of 2D and 3D U-Net Whole Lung and Separate Lung Training Model in Internal Validation Set

	2D U-Net		3D U-Net	
	Whole Lung Model	Separate Lung Model	Whole Lung Model	Separate Lung Model
DSC (%)	99.6 ± 0.3	99.5 ± 0.3	99.4 ± 0.5	99.4 ± 0.5
Sensitivity (%)	99.5 ± 0.3	99.5 ± 0.3	99.1 ± 0.9	99.1 ± 0.8
PPV (%)	99.6 ± 0.3	99.5 ± 0.4	99.7 ± 0.2	99.6 ± 0.3
Hausdorff distance (pixels)	17.70 ± 6.62	18.29 ± 6.51	18.75 ± 7.48	18.16 ± 7.48

Data are mean \pm standard deviation. DSC = Dice similarity coefficient, PPV = positive predictive value, 2D = two-dimensional, 3D = three-dimensional

Table 3. Dice Score, Sensitivity, PPV and Hausdorff Distance of 2D and 3D U-Net Whole Lung and Separate Lung Training Model in HUG-ILD External Validation Set

	2D U-Net		3D U-Net	
	Whole Lung Model	Separate Lung Model	Whole Lung Model	Separate Lung Model
DSC (%)	98.4 ± 1.0	98.4 ± 1.0	95.3 ± 3.1	96.1 ± 2.2
Sensitivity (%)	98.7 ± 1.3	98.7 ± 1.1	98.0 ± 1.9	98.1 ± 1.9
PPV (%)	98.1 ± 1.5	98.0 ± 1.6	92.8 ± 4.6	94.3 ± 3.5
Hausdorff distance (pixels)	7.66 ± 3.93	7.59 ± 3.69	15.58 ± 5.60	11.67 ± 4.84

Data are mean \pm standard deviation. HUG = University Hospitals of Geneva

masks and the 2D whole lung, 2D separate lung, 3D whole lung, and 3D separate lung models were 0.1% (-0.4, 0.6), 0.0% (-0.6, 0.6), 0.6% (-1.1, 2.3), and 0.5% (-0.9, 1.9), respectively, in the internal validation set, and -0.6% (-4.2, 3.0), -0.7% (-4.4, 2.9), -5.7% (-9.3, -2.1), and -4.0% (-7.7, -0.4), respectively, in the external validation set. The 2D U-Net model showed better agreement in both the internal and external datasets. Bland-Altman plots showing differences between the volumes of the manually segmented lung masks and each set of automatically segmented masks are presented in Figures 3 and 4.

Regarding the separation of anterior junctional line

thickness less than 2 mm, the 2D separate and whole lung models completely separated seven out of the nine cases (77.8%) in the internal validation dataset with a thin anterior junctional line in the full scan range on CT. In the remaining two cases, the anterior junctional line was incompletely demarcated in several axial scans. In the 3D separate lung model, three out of the nine cases were completely demarcated. In the 3D whole lung model, anterior junctional line segmentation was partially incomplete in all nine cases. In the external dataset, 36 cases had an anterior junctional line thickness of less than 2 mm. When the 2D separate lung model was applied, the

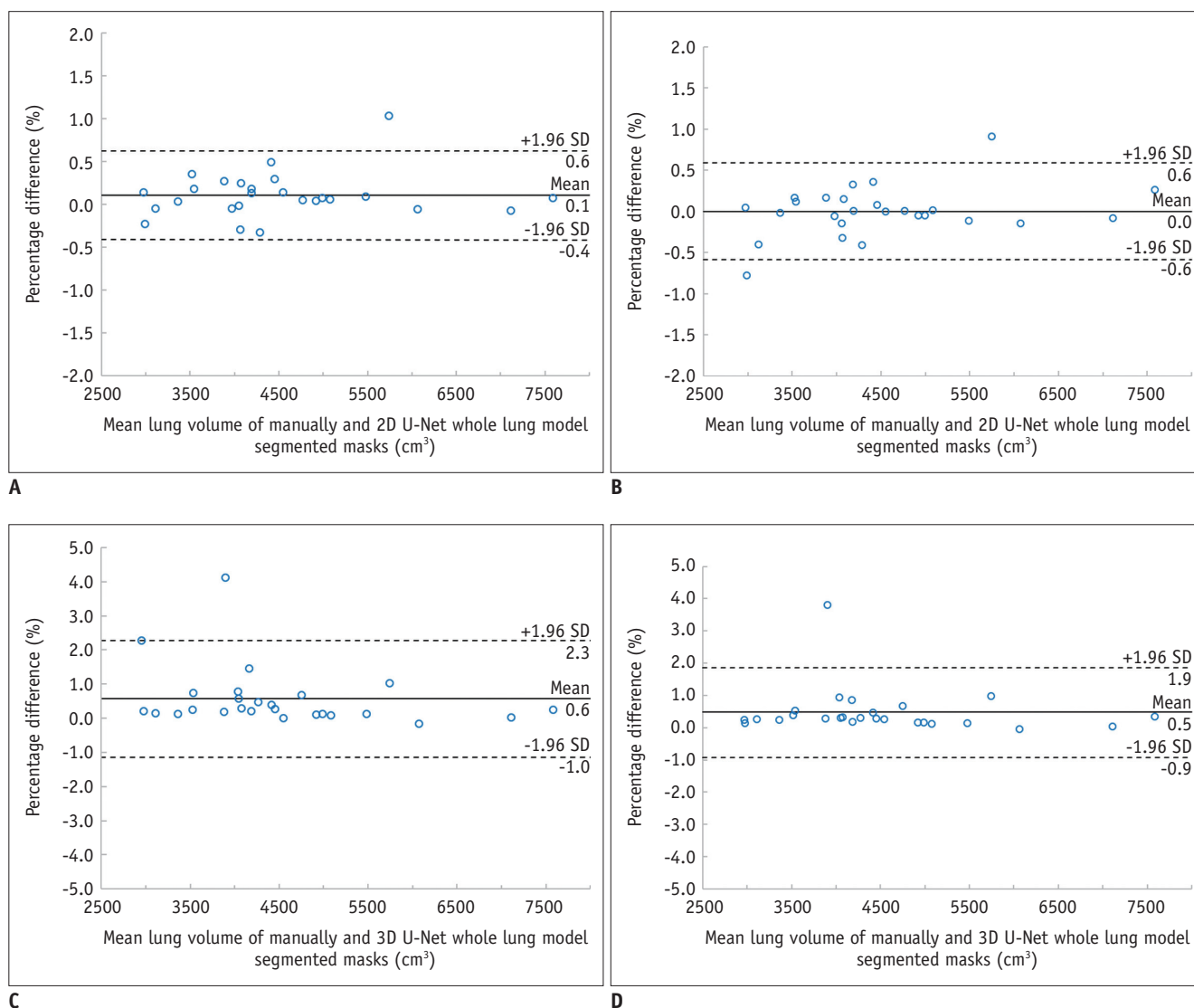


Fig. 3. Bland-Altman plots of volumes of 2D U-Net whole lung model (A), 2D U-Net separate lung model (B), 3D U-Net whole lung model (C), and 3D U-Net separate lung model (D) applied in the internal validation dataset. The solid line represents mean of volume percentage differences and dashed lines represent the limits of agreements (1.96 times SD). The percentage difference and limits of agreement of volumes between the manually segmented (ground truth) masks and the 2D whole lung, 2D separate lung, 3D whole lung, and 3D separate lung models were 0.1% (-0.4, 0.6), 0.0% (-0.6, 0.6), 0.6% (-1.1, 2.3), and 0.5% (-0.9, 1.9), respectively, suggesting high performance of the U-Net. SD = standard deviation, 3D = three-dimensional

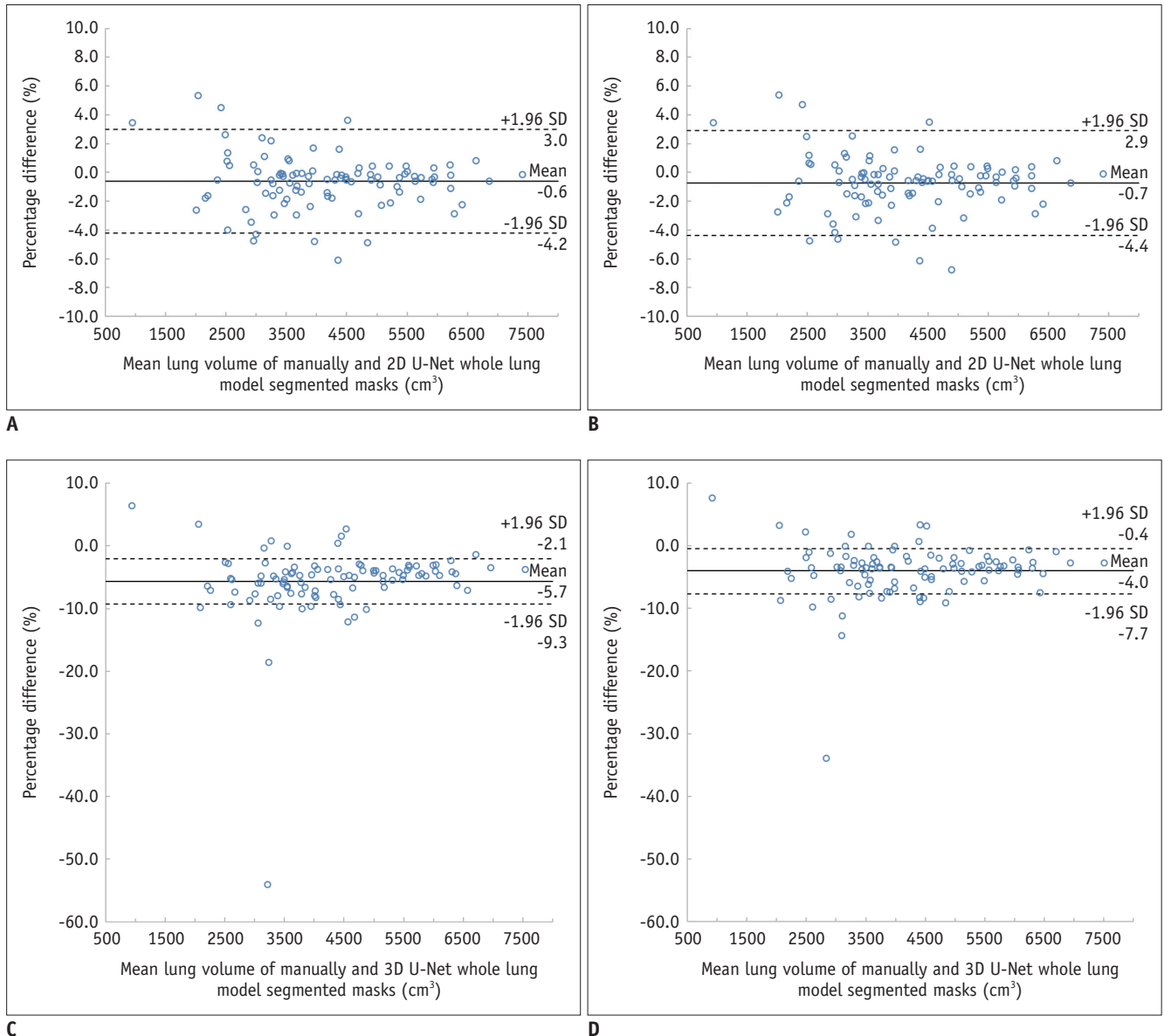


Fig. 4. Bland-Altman plots of volumes of 2D U-Net whole lung model (A), 2D U-Net separate lung model (B), 3D U-Net whole lung model (C), and 3D U-Net separate lung model (D) applied in the external validation dataset. The solid line represents mean of volume percentage differences and dashed lines represent the limits of agreements (1.96 times SD). The percentage difference and limits of agreement of volumes between the provided ground truth and the 2D whole lung, 2D separate lung, 3D whole lung, and 3D separate lung models were -0.6% (-4.2, 3.0), -0.7% (-4.4, 2.9), -5.7% (-9.3, -2.1), and -4.0% (-7.7, -0.4), respectively.

anterior junctional line was completely demarcated in 28 cases (77.8%).

DISCUSSION

Our study analyzed 203 cases of non-contrast chest CT images, of which 193 were LDCT scans, performed using CT machines from various vendors. One hundred and fifty cases had extensive underlying lung disease involving more than 40% of the lung parenchymal area. Manual lung

segmentation for building the ground truth, which was a time-consuming process, although it enabled precise establishment of training datasets, was performed by board-certified radiologists. We used 2D and 3D deep learning algorithms that were trained in two different ways (whole lung training and separate lung training). As a result, DSC of the internal validation dataset was 99.4–99.6% and of the external dataset was 95.3–98.4%. Our model achieved high performance in both internal and external validation datasets.

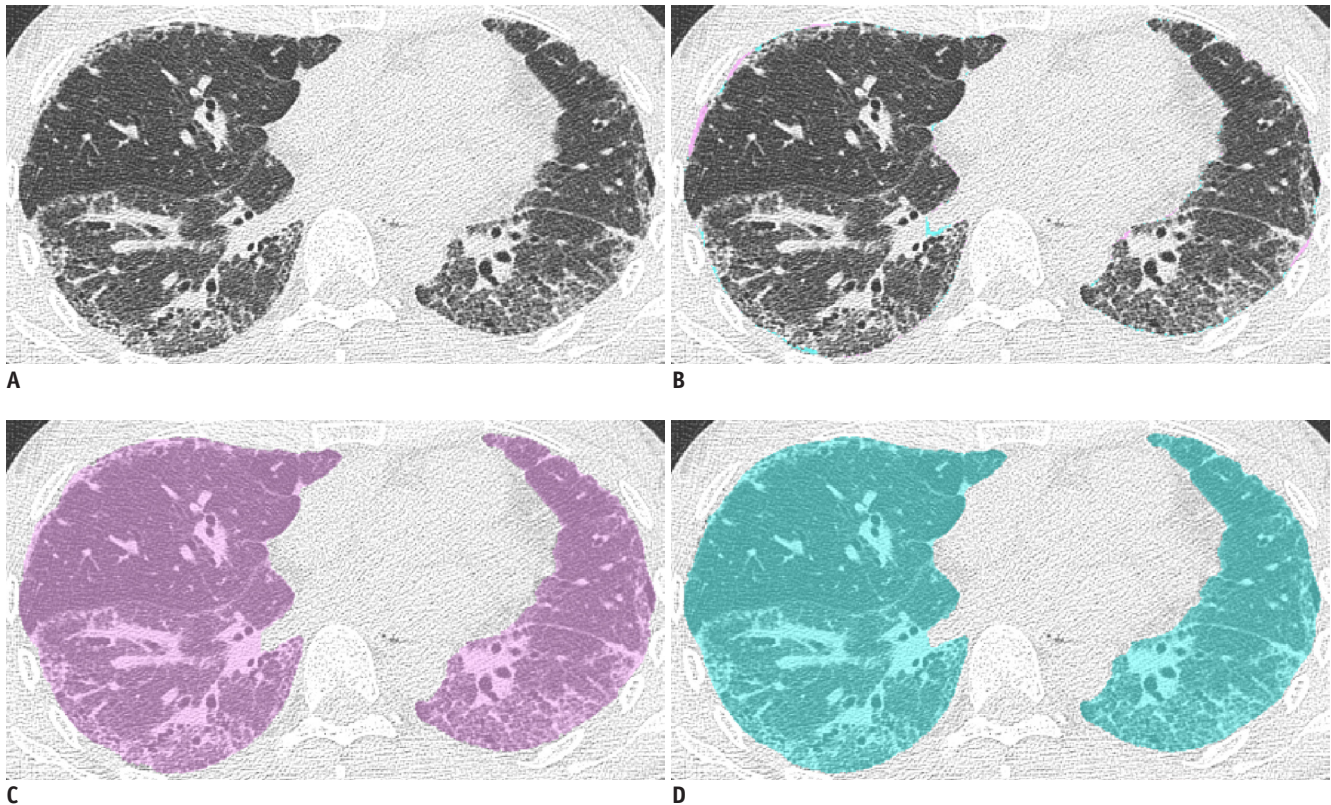


Fig. 5. Representative images of a 41-year old female with systemic sclerosis-associated interstitial lung disease in the internal validation dataset.

Chest CT image showing peripheral reticular and ground-glass opacities manifesting as a nonspecific interstitial pneumonia pattern (A). Manual lung mask (C) and segmented lung mask by 2D U-Net separate lung model (D) match almost perfectly on subtracted mask of manual and 2D U-Net (B). The Dice similarity coefficient between the masks was 99.7%.

Demand for automatic detection and analysis of pulmonary disease in chest CT images has increased as medical technology has improved. Automatic segmentation of the lung field in CT images has been applied for analysis of various diffuse pulmonary diseases including emphysema (4, 5), ILD (3), and infectious diseases, such as Coronavirus Disease 2019 (17). This CAD process is based on two steps: 1) extraction of lung field and 2) identification of lung disease from CT images (6). Therefore, precise segmentation of lung-field with automated lung segmentation algorithms is a prerequisite for radiologists to acquire further quantitative values from CT images, such as total lung volume and extent of the pathologic lung. Consequently, classification of the severity of the underlying lung disease or determination of the normal lung parenchymal volume (18) may be possible, which can be useful for clinicians.

Accurate segmentation of lung regions in the presence of severe pathologies is challenging. Pulagam et al. (19) applied a thresholding-based algorithm with a modified convexity algorithm on 60 high-resolution CT scans with

underlying honeycombing, reticular pattern, ground glass opacities, pleural plaques, and emphysema, resulting in a mean DSC of 98.6%. Harrison et al. (20) applied a fully convolutional network (FCN)-based deep-learning algorithm to chest CT scans with infections, ILD, and chronic obstructive pulmonary disease, obtaining a mean DSC of $98.5 \pm 1.1\%$. Alves et al. (21) also applied an FCN-based deep-learning algorithm to the HUG-ILD dataset, and obtained a DSC of $98.7 \pm 0.9\%$ (21). Our model achieved generally higher DSCs in internal validation (99.4–99.6%) and even in scans with extensive underlying lung disease involving more than 40% of the lung field with a DSC of 99.3–99.5% (Fig. 5).

For the whole lung model, DSC, sensitivity, and PPV were higher than in those reported in previous studies with a similar framework. Nevertheless, we discovered that separation of the anterior junctional line was unsatisfactory in the whole lung model. The anterior junctional line is a landmark separating the right from the left lung in the anteromedial aspect, formed by apposition of the visceral

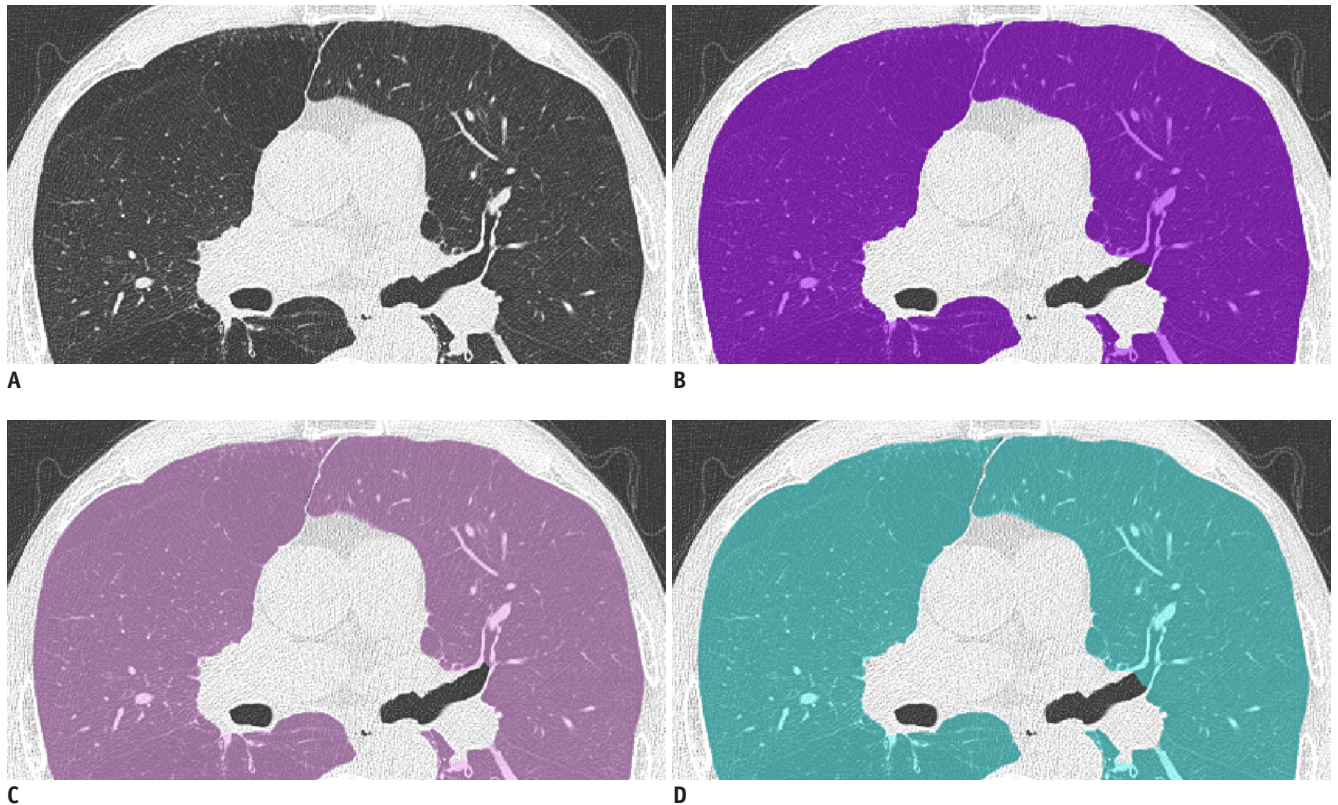


Fig. 6. Representative images of a 68-year old male patient with emphysema in the internal validation dataset. Chest CT image showing a very thin anterior junctional line due to hyperinflation (A). A segmented lung mask of 2D U-Net whole lung model (B) contains the anterior junctional line in the mask. 2D U-Net separate lung model (D) demarcates the anterior junctional line and separates the right from the left lung, as the ground truth (C).

and parietal pleura and a small amount of intervening fat (22). In patients with extensive emphysema, the anterior junctional line becomes very thin, due to hyperinflation of the lung. The thin anterior junctional line is a well-known cause of failure to automatically separate the right from the left lung (1). We developed a separate training model to overcome the weakness of the whole lung training model. The digitized results of the two training models were not significantly different; however, separation of the right from the left lung by the anterior junctional line was more satisfying in the separate training model in a case-by-case visual review (Fig. 6). Compared to the 3D U-Net model, the 2D U-Net model was superior in demarcating thin anterior junctional lines.

The external validation results were slightly inferior to the internal validation results. In scans from the HUG-ILD dataset, the trachea and main bronchi were basically included in the ground truth mask, in contrast to our models, which were trained to exclude the trachea and main bronchi. To make a comparison, we added an airway mask to our 2D U-Net lung mask. However, in most scans from the

HUG-ILD dataset, mediastinal fat tissue around the trachea was also included in the lung mask. Therefore, discrepancies were inevitable, regardless of the accuracy of lung segmentation, which led to underperformance of our model. Nevertheless, we revealed that lung segmentations obtained using our model tended to be slightly inaccurate in HUG-ILD cases with pleural effusion (Fig. 7). The number of cases with pleural effusion in the training dataset was small (only 10 of 157 cases). In some cases, our model showed more accurate lung segmentation than the ground truth of the HUG-ILD dataset, especially for the discrimination of the anterior junctional line and lung parenchyma with subpleural pathologies (Fig. 7, Supplementary Fig. 1).

Our study had several limitations. First, segmentation was insufficient in some cases with dense subpleural consolidations (Supplementary Fig. 2). However, in those cases, accurate lung segmentation was difficult even for experienced radiologists, because the attenuation of collapsed or consolidative lung and thickened pleura on LDCT without contrast enhancement is indistinct. Second, during manual lung segmentation, radiologists may have

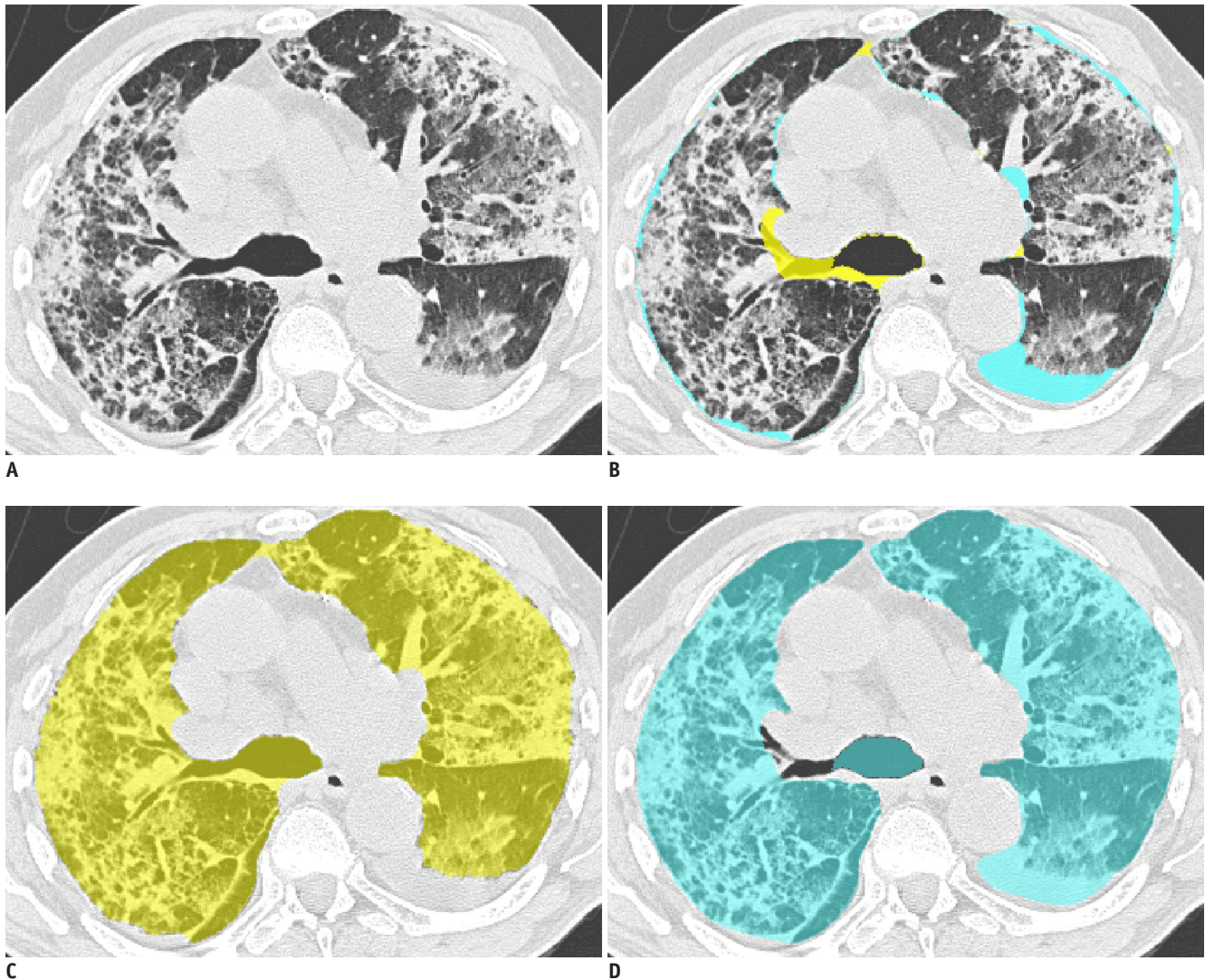


Fig. 7. Representative images of an 81-year old male suspected of having pneumonia over a pulmonary fibrosis in the external validation dataset.

Chest CT image showing multifocal patchy ground-glass opacities and consolidations with underlying bronchiectasis (A). Compared to ground truth (C), lung segmentation by 2D U-Net separate training model (D) included the pleural effusion as a lung in the left hemithorax. However, our model (D) superiorly discriminated the anterior junctional line. Mismatch is observed in the trachea and large bronchi in the subtracted mask (B). Dice similarity coefficient, sensitivity, positive predictive value and Hausdorff distance were 95.4%, 98.7%, 92.2% and 8.00 pixels, respectively.

subjectively drawn the border of the hilar structure. Finally, when comparing the 2D and 3D U-Net models, the performance of the 3D U-Net model in the external validation set was unsatisfactory. We therefore assume that the 3D U-Net model may have limited applicability in CT scans with thick image slices.

Here, we present a deep neural network for automated lung segmentation in non-contrast chest CT scans with underlying extensive lung disease. DSC, sensitivity, and PPV were higher than reported in previous relevant publications for the segmentation of CT scans of patients with various

extensive lung diseases, even in LDCT scans performed using machines from various vendors. This highly applicable method of automated lung segmentation in CT images using a deep neural network can form the basis for advanced computer-aided lung analysis in the future.

Supplementary Materials

The Data Supplement is available with this article at <https://doi.org/10.3348/kjr.2020.0318>.

Conflicts of Interest

Sang Joon Park is the CEO of Medical IP. All the other authors have no potential conflicts of interest to disclose.

Acknowledgments

The authors would like to acknowledge Andrew Dombrowski, PhD (Compecs, Inc.) for his assistance in improving the use of English in this manuscript.

ORCID iDs

Seung-Jin Yoo
<https://orcid.org/0000-0002-0779-3889>
 Soon Ho Yoon
<https://orcid.org/0000-0002-3700-0165>
 Jong Hyuk Lee
<https://orcid.org/0000-0002-9594-683X>
 Ki Hwan Kim
<https://orcid.org/0000-0002-6911-8125>
 Hyoung In Choi
<https://orcid.org/0000-0002-6784-3913>
 Sang Joon Park
<https://orcid.org/0000-0003-1013-681X>
 Jin Mo Goo
<https://orcid.org/0000-0003-1791-7942>

REFERENCES

- De Nunzio G, Tommasi E, Agrusti A, Cataldo R, De Mitri I, Favetta M, et al. Automatic lung segmentation in CT images with accurate handling of the hilar region. *J Digit Imaging* 2011;24:11-27
- Scholten ET, Jacobs C, van Ginneken B, Willeminck MJ, Kuhnigk JM, van Ooijen PM, et al. Computer-aided segmentation and volumetry of artificial ground-glass nodules at chest CT. *AJR Am J Roentgenol* 2013;201:295-300
- Wang J, Li F, Li Q. Automated segmentation of lungs with severe interstitial lung disease in CT. *Med Phys* 2009;36:4592-4599
- Tan KL, Toshiyuki T, Hidetoshi N, Akitoshi I. A neural network based computer-aided diagnosis of emphysema using CT lung images. *Proceedings of the SICE Annual Conference; 2007 Sep 17-20; Takamatsu, Japan: IEEE; p. 703-709*
- Brown MS, Kim HJ, Abtin FG, Strange C, Galperin-Aizenberg M, Pais R, et al. Emphysema lung lobe volume reduction: effects on the ipsilateral and contralateral lobes. *Eur Radiol* 2012;22:1547-1555
- Agnes SA, Anitha J, Peter JD. Automatic lung segmentation in low-dose chest CT scans using convolutional deep and wide network (CDWN). *Neural Comput & Applic* 2020;32:15845-15855
- Mansoor A, Bagci U, Foster B, Xu Z, Papadakis GZ, Folio LR, et al. Segmentation and image analysis of abnormal lungs at CT: current approaches, challenges, and future trends. *Radiographics* 2015;35:1056-1076
- Lee JW, Kim HY, Goo JM, Kim EY, Lee SJ, Kim TJ, et al. Radiological report of pilot study for the Korean lung cancer screening (K-LUCAS) project: feasibility of implementing lung imaging reporting and data system. *Korean J Radiol* 2018;19:803-808
- Romei C, Tavanti L, Sbragia P, De Liperi A, Carrozzi L, Aquilini F, et al. Idiopathic interstitial pneumonias: do HRCT criteria established by ATS/ERS/JRS/ALAT in 2011 predict disease progression and prognosis? *Radiol Med* 2015;120:930-940
- Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, et al. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology* 2011;261:274-282
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint* 2015;arXiv:1502.03167
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Proceedings neural information processing systems conference (NIPS) 2012; 2012 Dec 3-6; Nevada, USA: NIPS; p. 1097-1105*
- Wu Y, He K. Group normalization. *Proceedings of the European Conference on Computer Vision (ECCV); 2018 Sep 8-14; Munich, Germany: ECCV; p. 3-19*
- He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision (ICCV); 2015 DEC 7-13; Santiago, Chile: IEEE; p. 1026-1034*
- Depeursinge A, Vargas A, Platon A, Geissbuhler A, Poletti PA, Müller H. Building a reference multimedia database for interstitial lung diseases. *Comput Med Imaging Graph* 2012;36:227-238
- Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 2015;37:2153-2163
- Choi H, Qi X, Yoon SH, Park SJ, Lee KH, Kim JY, et al. Extension of coronavirus disease 2019 (COVID-19) on chest CT and implications for chest radiograph interpretation. *Radiology: Cardiothoracic Imaging* 2020;2:e200107
- Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology* 2020;296:E86-E96
- Pulagam AR, Kande GB, Ede VK, Inampudi RB. Automated lung segmentation from HRCT scans with diffuse parenchymal lung diseases. *J Digit Imaging* 2016;29:507-519
- Harrison AP, Xu Z, George K, Lu L, Summers RM, Mollura DJ. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In:

- Descoteaux Lena M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S. eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2017*; 2017 Sep 11-13; Quebec City, Canada: Springer; p. 621-629
21. Alves JH, Neto PMM, Oliveira LF. Extracting lungs from CT images using fully convolutional networks. *Proceedings 2018 International Joint Conference on Neural Networks (IJCNN)*; 2018 Jul 8-13; Rio de Janeiro, Brazil: IEEE; p. 1-8
22. Gibbs JM, Chandrasekhar CA, Ferguson EC, Oldham SA. Lines and stripes: where did they go?--From conventional radiography to CT. *Radiographics* 2007;27:33-48