



# Generative Adversarial Network-Based Image Conversion Among Different Computed Tomography Protocols and Vendors: Effects on Accuracy and Variability in Quantifying Regional Disease Patterns of Interstitial Lung Disease

Hye Jeon Hwang<sup>1</sup>, Hyunjong Kim<sup>2</sup>, Joon Beom Seo<sup>1</sup>, Jong Chul Ye<sup>3</sup>, Gyutaek Oh<sup>4</sup>, Sang Min Lee<sup>1</sup>, Ryoungwoo Jang<sup>1</sup>, Jihye Yun<sup>1</sup>, Namkug Kim<sup>1</sup>, Hee Jun Park<sup>5</sup>, Ho Yun Lee<sup>6,7</sup>, Soon Ho Yoon<sup>8</sup>, Kyung Eun Shin<sup>9</sup>, Jae Wook Lee<sup>9</sup>, Woocheol Kwon<sup>10,11</sup>, Joo Sung Sun<sup>12</sup>, Seulgi You<sup>12</sup>, Myung Hee Chung<sup>13</sup>, Bo Mi Gil<sup>13</sup>, Jae-Kwang Lim<sup>14</sup>, Youkyung Lee<sup>15</sup>, Su Jin Hong<sup>15</sup>, Yo Won Choi<sup>16</sup>

<sup>1</sup>Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

<sup>2</sup>Robotics Program, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

<sup>3</sup>Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

<sup>4</sup>Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

<sup>5</sup>Coreline Soft, Co., Ltd, Seoul, Republic of Korea

<sup>6</sup>Department of Radiology and Center for Imaging Science, Samsung Medical Center, School of Medicine, Sungkyunkwan University, Seoul, Republic of Korea

<sup>7</sup>Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea

<sup>8</sup>Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

<sup>9</sup>Department of Radiology, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea

<sup>10</sup>Department of Radiology, Ewha Womans University Seoul Hospital, Seoul, Republic of Korea

<sup>11</sup>Department of Radiology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

<sup>12</sup>Department of Radiology, Ajou University School of Medicine, Suwon, Republic of Korea

<sup>13</sup>Department of Radiology, Bucheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

<sup>14</sup>Department of Radiology, Kyungpook National University School of Medicine, Daegu, Republic of Korea

<sup>15</sup>Department of Radiology, Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri, Republic of Korea

<sup>16</sup>Department of Radiology, Hanyang University Seoul Hospital, Hanyang University College of Medicine, Seoul, Republic of Korea

**Objective:** To assess whether computed tomography (CT) conversion across different scan parameters and manufacturers using a routable generative adversarial network (RouteGAN) can improve the accuracy and variability in quantifying interstitial lung disease (ILD) using a deep learning-based automated software.

**Materials and Methods:** This study included patients with ILD who underwent thin-section CT. Unmatched CT images obtained using scanners from four manufacturers (vendors A-D), standard- or low-radiation doses, and sharp or medium kernels were classified into groups 1–7 according to acquisition conditions. CT images in groups 2–7 were converted into the target CT style (Group 1: vendor A, standard dose, and sharp kernel) using a RouteGAN. ILD was quantified on original and converted CT images using a deep learning-based software (Aview, Coreline Soft). The accuracy of quantification was analyzed using the dice similarity coefficient (DSC) and pixel-wise overlap accuracy metrics against manual quantification by a radiologist. Five radiologists evaluated quantification accuracy using a 10-point visual scoring system.

**Results:** Three hundred and fifty CT slices from 150 patients (mean age: 67.6 ± 10.7 years; 56 females) were included. The overlap accuracies for quantifying total abnormalities in groups 2–7 improved after CT conversion (original vs. converted: 0.63 vs. 0.68 for DSC, 0.66 vs. 0.70 for pixel-wise recall, and 0.68 vs. 0.73 for pixel-wise precision;  $P < 0.002$  for all). The DSCs of

**Received:** January 31, 2023 **Revised:** June 12, 2023 **Accepted:** June 18, 2023

**Corresponding author:** Joon Beom Seo, MD, PhD, Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 86 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Republic of Korea

• E-mail: seojb@amc.seoul.kr

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.

fibrosis score, honeycombing, and reticulation significantly increased after CT conversion (0.32 vs. 0.64, 0.19 vs. 0.47, and 0.23 vs. 0.54,  $P < 0.002$  for all), whereas those of ground-glass opacity, consolidation, and emphysema did not change significantly or decreased slightly. The radiologists' scores were significantly higher ( $P < 0.001$ ) and less variable on converted CT.

**Conclusion:** CT conversion using a RouteGAN can improve the accuracy and variability of CT images obtained using different scan parameters and manufacturers in deep learning-based quantification of ILD.

**Keywords:** Interstitial lung disease; Computed tomography; Quantification; Artificial intelligence

## INTRODUCTION

High-resolution computed tomography (CT) is essential for evaluating interstitial lung disease (ILD) [1-4].

However, objective and reproducible assessments of ILD by radiologists are limited, with wide variations in image interpretation [5,6]. Consequently, several automated quantitative imaging methods have been developed using either histogram- or texture-based analysis [7-12].

Quantitative assessment of ILD on CT is comparable to visual assessment of ILD on CT and correlates with measures of disease severity and pulmonary function [7,13-15]. The use of ILD quantification for prognostication and mortality prediction has been demonstrated in previous studies [16-18]. However, major issues have limited its widespread use. Various technical parameters related to CT acquisition, such as reconstruction kernel, reconstruction method, and radiation dose, can affect quantitative results [19-22]. Inherent CT characteristics, including texture, which vary depending on the manufacturer, may also affect quantification. Inconsistencies in these factors cause variations in quantification results, hindering the use of quantitative CT analysis in retrospective, longitudinal, or multicenter clinical studies.

Recently, attempts have been made to transform or standardize CT images using deep learning technology [23,24]. Previous studies have demonstrated the potential of convolutional neural network (CNN)-based CT conversion to reduce the effects of different reconstruction kernels on the emphysema index and radiomics [25,26]. However, the previous method required paired CT images obtained from the same raw data for algorithm training, and possibly performed CT conversion only between features of the paired dataset. Therefore, CT conversion across scanners with different manufacturers and acquisition parameters could not be performed. A previous study developed a CT conversion algorithm using a routable generative adversarial network (RouteGAN) that uses unsupervised image-to-image translation with unpaired CT data to convert CT images

across different scan parameters and manufacturers [24]. However, the effects of CT-style conversion across variable scan parameters and vendors on ILD quantification in chest CT images have not been evaluated.

This study aimed to investigate whether the proposed CT conversion algorithm using a RouteGAN can improve the accuracy and variability of ILD quantification using a deep learning-based automated software on chest CT images across various parameters and scanners from various manufacturers.

## MATERIALS AND METHODS

### Datasets

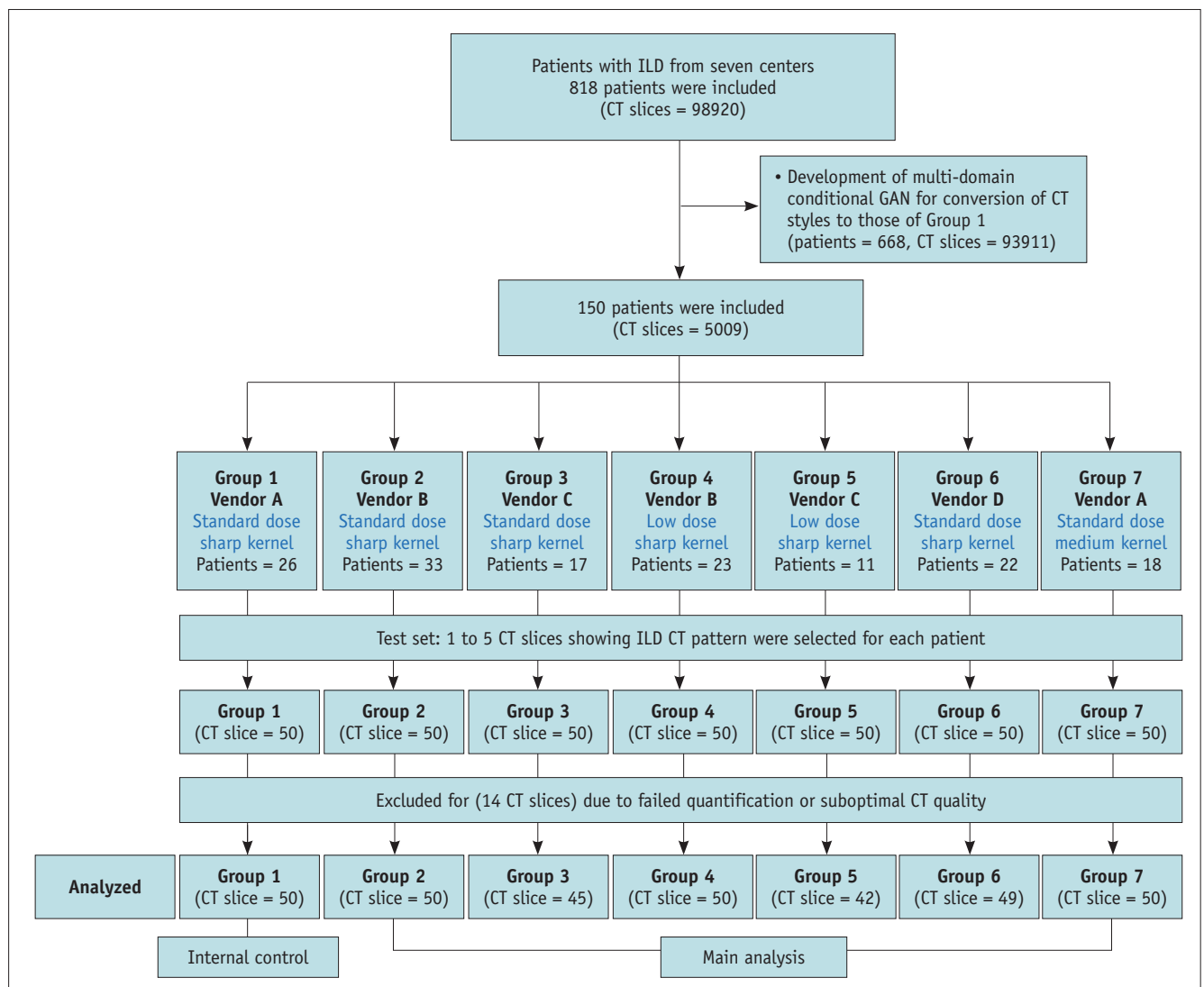
This retrospective multicenter study was approved by the institutional review board of each participating institution and written informed consent was waived. Patients with ILD who underwent chest CT between January 2007 and March 2020 were identified from seven tertiary referral centers. The inclusion criteria were as follows: 1) diagnosis of ILD with or without known causes verified through multidisciplinary discussion among experienced clinical experts, radiologists, and pathologists, following the diagnostic guidelines at each institution [27,28]; 2) available thin-section non-enhanced chest CT obtained with standard- or low-radiation dose and reconstructed with sharp or other kernels; and 3) no combined complications, such as pneumonia, lung cancer, or acute exacerbation at the time of CT acquisition. CT images obtained using scanners from four manufacturers, standard or low radiation doses, and sharp or medium kernels were collected and classified into seven groups according to the acquisition conditions (Group 1: vendor A, standard dose, and sharp kernel; Group 2: vendor B, standard dose, and sharp kernel; Group 3: vendor C, standard dose, and sharp kernel; Group 4: vendor B, low dose, and sharp kernel; Group 5: vendor C, low dose, and sharp kernel; Group 6: vendor D, standard dose, and sharp kernel; and Group 7: vendor A, standard dose, and medium kernel). The CT protocols, scanners, and manufacturers for

each group are summarized in Supplementary Table 1. A total of 98920 CT slices were initially collected from 818 patients, of which 93911 CT slices from 668 patients were randomly selected and used to develop the CT conversion algorithm in a previous study [24]. The remaining 5009 CT slices from 150 individual patients (mean age:  $67.6 \pm 10.7$  years; 56 females) were included in this study to avoid overlap with the study population of the previous study [24]. A thoracic radiologist (H.J.H.; 12 years of experience in chest radiology) who was blinded to all measurements and results in the study selected 350 CT slices (50 per group) from 150 patients for the test set (Fig. 1) based

on the following visual review criteria: 1) slices showed at least 10% extent of the ILD CT pattern; 2) slices showed no parenchymal abnormalities other than the ILD CT pattern; and 3) when selecting multiple slices from a patient, the slices were at least 10 slices apart. Full details of the data collection are described in the Supplementary Material.

**CT Conversion and ILD Segmentation/Quantification on CT**

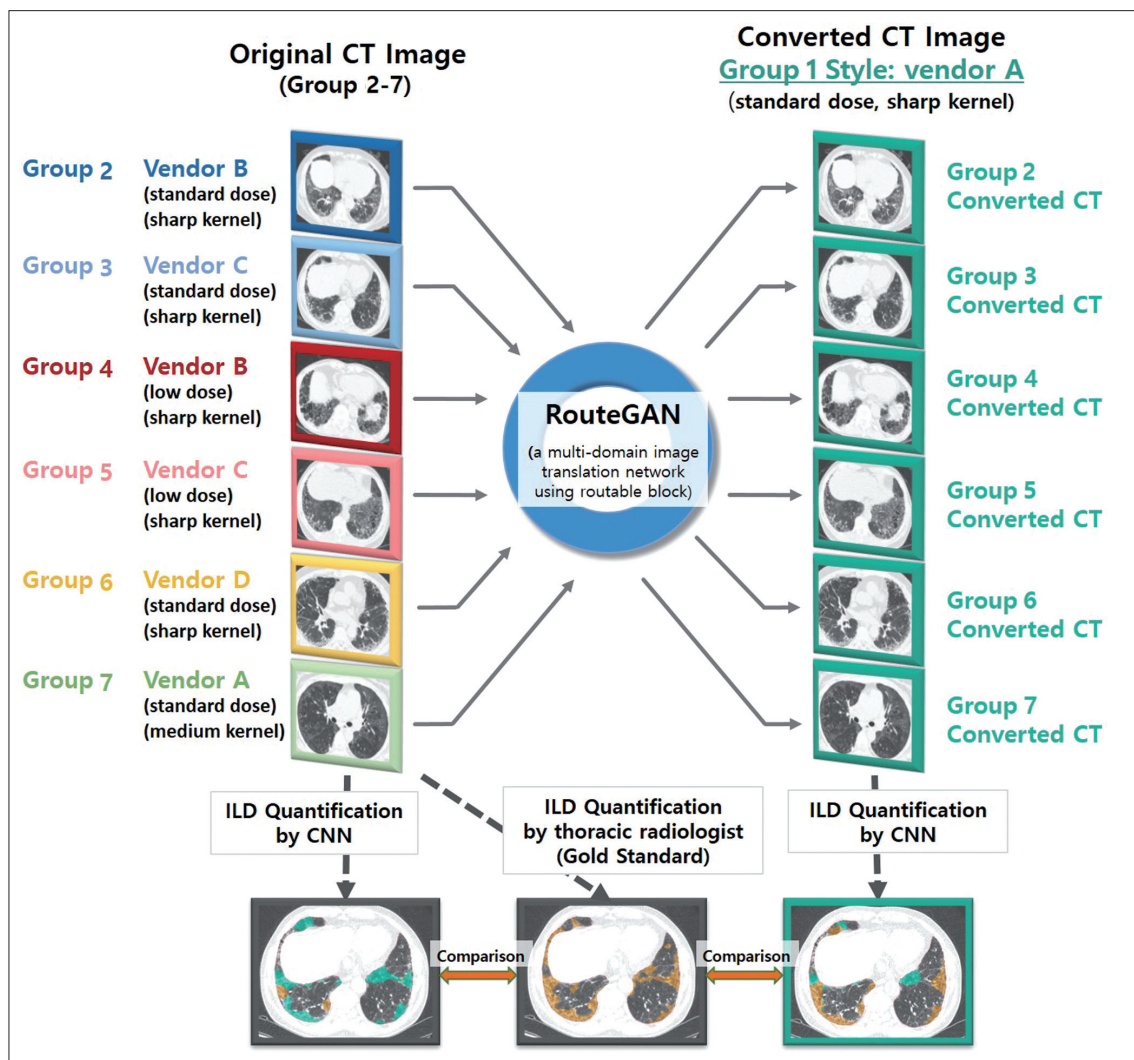
We used a RouteGAN algorithm that uses unsupervised image-to-image translation with unmatched CT images [24]. This network can translate unpaired CT images from one acquisition protocol to another within the training set



**Fig. 1.** Flowchart of patient inclusion for the study. A total of 98920 CT slices were initially collected from 818 patients, of which 93911 CT slices from 668 patients were randomly selected and used to develop the CT conversion algorithm in a previous study. The remaining 5009 CT slices from 150 individual patients were included in this study. A thoracic radiologist, who was blinded to any measurements or results in the study, selected 350 CT slices (50 per group) from 150 patients for the test set, based on the visual review criteria. ILD = interstitial lung disease, CT = computed tomography, GAN = generative adversarial network

(including across different scanner types and acquisition parameter differences). Group 1 images were set as the target CT style for CT conversion because the CNN-based ILD quantification software used in this study was developed mainly using CT images with the same acquisition protocol and scanner manufacturer as those in group 1 [29,30]. The original CT slices of Group 2 to 7 were converted into the target CT style using the RouteGAN (Fig. 2). Regional CT patterns of ILD, including honeycombing, reticulation, ground-glass opacity (GGO), consolidation, emphysema,

and normal patterns, were segmented and quantified on both the original and converted CT slices of groups 2 to 7 using a deep learning-based quantification software (Aview, Coreline Soft) [29,30]. The original CT slices of group 1 were also analyzed using the automated software to serve as an internal control. The total abnormalities were calculated as the sum of the extent of honeycombing, reticulation, GGO, consolidation, and emphysema, and the fibrosis score was defined as the sum of the extent of honeycombing and reticulation. In addition, a thoracic radiologist (H.J.H.; 12



**Fig. 2.** Schematic flow diagram of the conversion of original computed tomography (CT) images to the target CT style (Group 1 CT style) using a routable generative adversarial network (RouteGAN). All 300 CT slices, 50 CT slices for each of the six groups, were converted to Group 1 CT style using the RouteGAN to assess the effect of CT style conversion on the quantification of CT patterns of interstitial lung disease (ILD). Quantification of regional CT patterns of ILD was performed on both original and converted CT images using deep learning-based ILD quantification. For the reference standard quantification of ILD, a thoracic radiologist who was blinded to the quantification results of the software manually drew the six CT patterns of ILD on the original CT slices. The quantifications on the original or converted CT images were compared with the radiologist's manual quantifications, which were used as the reference standard. CNN = convolutional neural network

years of experience in chest radiology) who was blinded to the automated segmentation drew each CT pattern on the original CT slices of groups 1 to 7 using a manual drawing tool of the Aview software (Coreline Soft) to perform ILD segmentation to be used as the reference standard.

### Visual Assessment of the Accuracy of ILD Segmentation on CT

Five thoracic radiologists (H.Y.L., J.S.S., W.C.K., S.H.Y., and J.K.L. with 14, 17, 17, 8, and 9 years of experience in chest radiology, respectively) independently assessed the results of the automated segmentation of ILD on the original and converted CTs and the reference standard results manually drawn by a thoracic radiologist on the images in groups 2–7. The radiologists were blinded to whether the results were obtained by the automated software or a human (i.e., the reference standard) and whether the images were original or converted CT images. The radiologists scored the ILD segmentation accuracy on each image using a 10-point scale, considering the total extent of the ILD and each ILD pattern (see Supplementary Material for details).

### Statistical Analysis

The accuracy of segmentation/quantification was analyzed using measures of spatial overlap accuracy and visual accuracy scores from the five radiologists. For the measures of spatial overlap accuracy, the dice similarity coefficient (DSC) and pixel-wise recall and precision were obtained by comparing ILD quantification on the original or converted CT images with the radiologist's reference standard. These metrics were calculated for each individual CT pattern, total abnormalities, and fibrosis score as follows:

$$DSC_A = \frac{2|S \cap R|}{|S| + |R|},$$

$$\text{pixel-wise recall}_A = \frac{|\text{pixel number of } S \cap R|}{|\text{pixel number of } R|}, \text{ and}$$

$$\text{pixel-wise precision}_A = \frac{|\text{pixel number of } S \cap R|}{|\text{pixel number of } S|},$$

where *S* is the area quantified as pattern *A* by the software in the original or converted CT, and *R* is the area quantified as *A* in the reference standard [31]. The differences in these metrics between the original and converted CT images were evaluated using paired *t*-tests. The average visual scores of the five radiologists were compared among CT

types by a repeated-measures analysis performed using PROC MIXED in SAS (version 9.4; SAS Institute). The lack of independence between repeated observations by the same person was accounted for by including repeated statements in statistical routines. The effects of CT conversion with the RouteGAN on the variability of segmentation/quantification were not analyzed using any formal statistical tests. Statistical significance set as *P* < 0.05. Bonferroni corrections were used to account for multiple tests.

## RESULTS

### Patient Characteristics

Among the 350 CT slices from 150 patients of seven groups (mean age: 67.6 ± 10.7 years; 56 females), which were initially selected for the test dataset, 14 were excluded because of suboptimal CT quality or failed quantification, and 336 CT slices were finally analyzed. Patient characteristics are summarized in Table 1.

### Spatial Overlap Accuracy of the Automated ILD Quantification on Original and Converted CT Images

Table 2 shows the overall overlap accuracies of the quantified areas on the original and converted CT images in Group 2 to 7. All overlap accuracy metrics for total abnormalities were significantly higher after CT conversion (original vs. converted: 0.63 vs. 0.68 for DSC, 0.66 vs. 0.70 for pixel-wise recall, and 0.68 vs. 0.73 for pixel-wise precision; all *P* < 0.002). The overall DSCs of fibrosis score, honeycombing, and reticulation were significantly higher after CT conversion (original vs. converted: 0.32 vs. 0.64 for fibrosis score, 0.19 vs. 0.47 for honeycombing, and 0.23 vs. 0.54 for reticulation; all *P* < 0.002). For other CT patterns, the overall DSCs did not change significantly for GGO and emphysema and was slightly lower for consolidation after CT conversion (original vs. converted: 0.08 vs. 0.08 for GGO, *P* = 0.631; 0.14 vs. 0.12 for emphysema, *P* = 0.037; and 0.14 vs. 0.07 for consolidation, *P* < 0.002). The overlap accuracies of the original CT images in Group 2 to 7 were lower than that in Group 1 for all CT patterns, except the pixel-wise precision of reticulation (Supplementary Table 2).

In the pixel-wise analysis, more pixels from more CT slices were correctly classified according to the fibrosis score, honeycombing, and reticulation after CT conversion (Fig. 3 and Supplementary Fig. 1, Supplementary Table 3). However, fewer pixels were classified as GGO, consolidation, or emphysema in the converted CT quantification (Fig. 3 and



**Table 1.** Patient characteristics in the test set

Characteristics	Total	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
		(Vendor A, standard dose, sharp kernel)	(Vendor B, standard dose, sharp kernel)	(Vendor C, standard dose, sharp kernel)	(Vendor B, low dose, sharp kernel)	(Vendor C, low dose, sharp kernel)	(Vendor D, standard dose, sharp kernel)	(Vendor A, standard dose, medium kernel)
Patients	150	26	33	17	23	11	22	18
CT slices (analyzed)	350 (336)	50 (50)	50 (50)	50 (45)	50 (50)	50 (42)	50 (49)	50 (50)
Age, yr	67.6 ± 10.7*							
Sex, M:F	94:56							
Disease								
UIP	76 (50.7)	13 (50.0)	17 (51.5)	12 (70.6)	10 (43.5)	7 (63.6)	6 (27.3)	11(61.1)
NSIP	45 (30.0)	5 (19.2)	10 (30.3)	3 (17.6)	10 (43.5)	3 (27.3)	7 (31.8)	7 (38.9)
COP	14 (9.3)	6 (23.1)	2 (6.1)	1 (5.9)	0 (0)	1 (9.1)	4 (18.2)	0 (0)
CHP	3 (2.0)	0 (0)	2 (6.1)	0 (0)	0 (0)	0 (0)	1 (4.5)	0 (0)
Smoking-related ILD	12 (8.0)	2 (7.7)	2 (6.1)	1 (5.9)	3 (13.0)	0 (0)	4 (18.2)	0 (0)

Data are numbers of patients (%) or CT slices, unless specified otherwise.

\*Data are expressed as mean ± standard deviation.

CT = computed tomography, M = male, F = female, UIP = usual interstitial pneumonia, NSIP = nonspecific interstitial pneumonia, COP = cryptogenic organizing pneumonia, CHP = chronic hypersensitivity pneumonitis, ILD = interstitial lung disease

**Table 2.** Spatial overlap accuracies of the automated ILD quantifications on the original and converted CT images in Group 2 to 7

CT pattern	Spatial overlap accuracy metric								
	DSC			Pixel-wise recall			Pixel-wise precision <sup>†</sup>		
	Original CT (1)	Converted CT (2)	Difference [(2)-(1)]	Original CT (1)	Converted CT (2)	Difference [(2)-(1)]	Original CT (1)	Converted CT (2)	Difference <sup>†</sup> [(2)-(1)]
Total abnormalities	0.63 ± 0.20	0.68 ± 0.18	0.06*	0.66 ± 0.23	0.70 ± 0.20	0.04*	0.68 ± 0.22	0.73 ± 0.19	0.05*
Fibrosis score	0.32 ± 0.28	0.64 ± 0.21	0.32*	0.26 ± 0.26	0.62 ± 0.23	0.37*	0.80 ± 0.27	0.74 ± 0.22	-0.06*
Honeycombing	0.19 ± 0.26	0.47 ± 0.31	0.28*	0.15 ± 0.24	0.46 ± 0.32	0.30*	0.68 ± 0.41	0.63 ± 0.37	-0.04
Reticulation	0.23 ± 0.22	0.54 ± 0.20	0.31*	0.19 ± 0.21	0.54 ± 0.23	0.37*	0.58 ± 0.34	0.64 ± 0.24	0.06*
Ground-glass opacity	0.08 ± 0.18	0.08 ± 0.17	-0.00	0.42 ± 0.41	0.19 ± 0.30	-0.23*	0.03 ± 0.10	0.04 ± 0.14	0.01
Consolidation	0.14 ± 0.24	0.07 ± 0.20	-0.07*	0.16 ± 0.28	0.07 ± 0.18	-0.09*	0.12 ± 0.26	0.12 ± 0.29	0.00
Emphysema	0.14 ± 0.25	0.12 ± 0.23	-0.02	0.13 ± 0.22	0.09 ± 0.19	-0.03	0.19 ± 0.37	0.23 ± 0.39	0.01

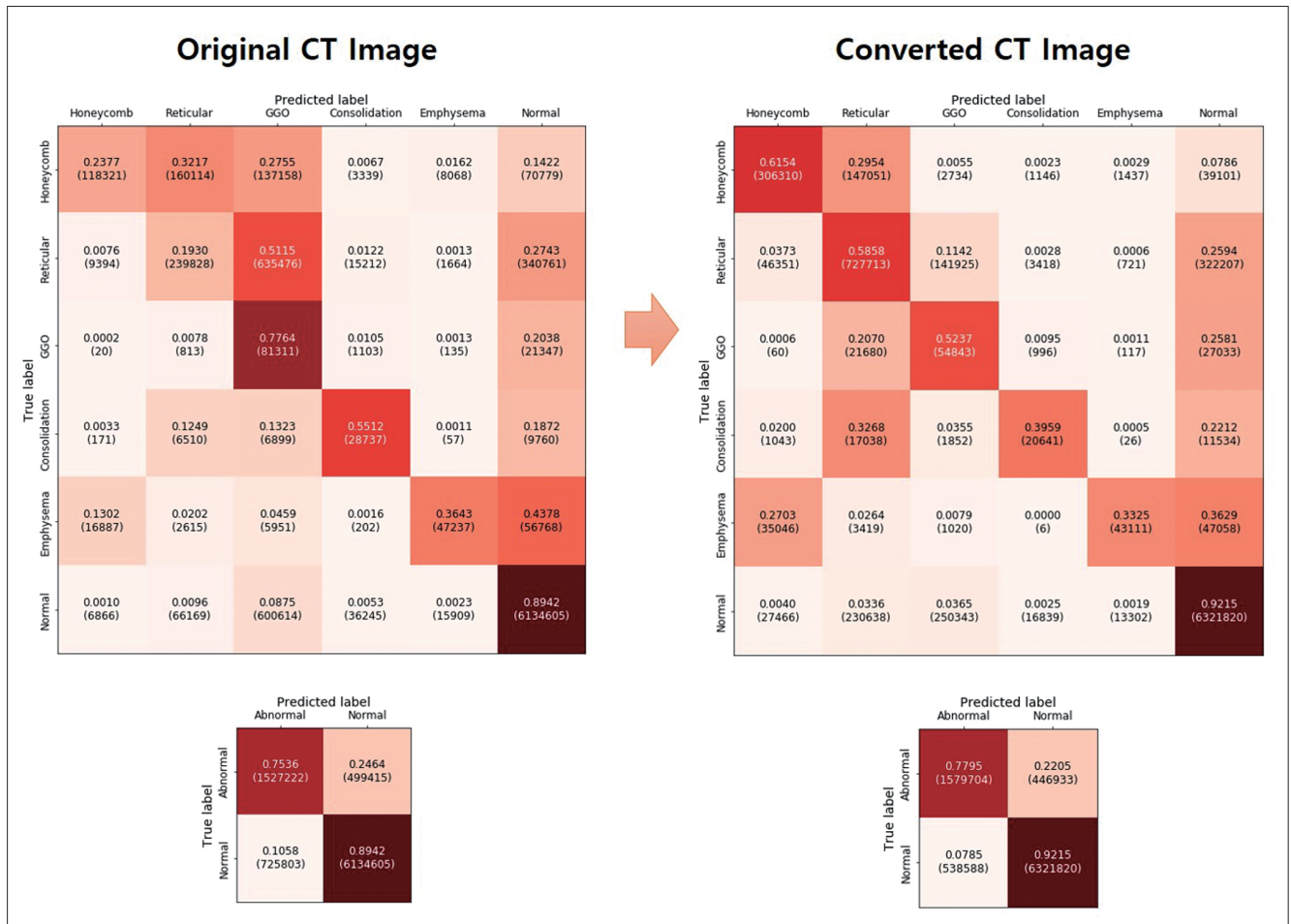
Data are presented as the mean ± standard deviation of the spatial overlap accuracy metrics of the CT slices.

\*Indicates the difference in the overlap accuracy metrics between (1) and (2) is statistically significant ( $P < 0.002$ ).  $P$ -values were calculated using paired  $t$ -tests. Significance level of 0.002 takes into account the Bonferroni correction for multiple tests, <sup>†</sup>The mean pixel-wise precisions of the original and converted CT images were calculated for those CT slices for which pixel-wise precision was available for both the original and converted CT images, <sup>†</sup>The difference in pixel-wise precision between the original and converted CT images and its statistical significance were calculated for those CT slices for which pixel-wise precision was available for both the original and converted CT images.

ILD = interstitial lung disease, CT = computed tomography, DSC = dice similarity coefficient

Supplementary Fig. 1). The pixel-wise recall of fibrosis score, reticulation, and honeycombing were significantly higher after CT conversion (original vs. converted: 0.26 vs. 0.62 for fibrosis score, 0.19 vs. 0.54 for reticulation, and 0.15 vs. 0.46 for honeycombing, all  $P < 0.002$ ). After CT conversion, the pixel-wise precision of reticulation was slightly higher (original vs. converted: 0.58 vs. 0.64,  $P < 0.002$ ), whereas the fibrosis score and honeycombing were slightly lower (original vs. converted: 0.80 vs. 0.74 for fibrosis score,  $P < 0.002$ ; and 0.68 vs. 0.63 for honeycombing,  $P = 0.076$ ),

despite the fact that more pixels were classified into these patterns on the converted than on the original CT. The pixel-wise recall of GGO and consolidation were significantly lower, whereas that of emphysema was not significantly different after CT conversion (original vs. converted: 0.42 vs. 0.19 for GGO,  $P < 0.002$ ; 0.16 vs. 0.07 for consolidation,  $P < 0.002$ ; and 0.13 vs. 0.09 for emphysema,  $P = 0.008$ ) (Table 2). The pixel-wise precisions of these patterns were not significantly different after CT conversion. The increase in the pixel-wise recall of the fibrosis score, honeycombing,



**Fig. 3.** Confusion matrixes of pixel-wise analysis of interstitial lung disease (ILD) quantification on the original and converted computed tomography (CT) images in group 2 to 7 in comparison with the radiologist's quantifications. The predicted labels of the original and converted CT quantifications of ILD are shown along the x-axis, and true labels (i.e., the reference standard quantification by a thoracic radiologist) are shown along the y-axis. Confusion matrixes show the ratio of  $\frac{|\text{pixel number of } S \cap R|}{|\text{pixel number of } R|}$ , where S is the area quantified by a software as one of the six patterns in the original or converted CT images, and R is the area quantified as one of six patterns in the reference standard. The numbers in parentheses are pixel numbers of  $S \cap R$ . Many pixels that were incorrectly classified as ground-glass opacity (GGO) or reticulation on the original CT images were correctly classified as honeycombing or reticulation on the converted CT images. For GGO patterns, some pixels incorrectly classified as reticulation after CT conversion were correctly classified as GGO on the original CT images.

and reticulation was greater than the decrease in the pixel-wise recall of GGO, consolidation, and emphysema.

### Spatial Overlap Accuracy of the Automated ILD Quantification Areas in Each Group

The DSCs of total abnormalities in the original CT were not significantly different among Group 2 to 7, whereas the DSC of the fibrosis score in Group 7 was significantly lower than that in all other groups (all  $P < 0.05$ ). In all groups, the overlap accuracy metrics for total abnormalities were higher after CT conversion; however, the mean differences between the original and converted CT images were not statistically significant in groups 2, 5, and 7 for pixel-

wise recall ( $P = 0.009, 0.096, \text{ and } 0.467$ , respectively) and in groups 5 and 6 for pixel-wise precision ( $P = 0.130 \text{ and } 0.112$ , respectively) (Table 3). For the fibrosis score, the DSC and pixel-wise recall were significantly higher in all groups after CT conversion ( $P < 0.008$ ). The pixel-wise precision of the fibrosis score was slightly lower in all groups, but not statistically significant in groups 2, 5, 6, and 7 after CT conversion; however, the increase in pixel-wise recall was higher than the decrease in pixel-wise precision in all groups.

### Visual Accuracy Scores for ILD Segmentation

The radiologists visually scored the accuracy of ILD

**Table 3.** Spatial overlap accuracy in the automated quantification of total abnormalities and fibrosis score on the original and converted CT images in each group

CT patterns	Group	Spatial overlap accuracy metric									
		DSC			Pixel-wise recall			Pixel-wise precision <sup>†</sup>			Difference <sup>‡</sup> [(2)-(1)]
		Original CT (1)	Converted CT (2)	Difference [(2)-(1)]	Original CT (1)	Converted CT (2)	Difference [(2)-(1)]	Original CT (1)	Converted CT (2)	Difference [(2)-(1)]	
Total abnormalities	Group 2	0.62 ± 0.23	0.68 ± 0.18	0.06*	0.61 ± 0.25	0.64 ± 0.22	0.04	0.71 ± 0.16	0.79 ± 0.14	0.05*	0.05*
	Group 3	0.63 ± 0.21	0.70 ± 0.19	0.05*	0.64 ± 0.25	0.69 ± 0.24	0.05*	0.75 ± 0.18	0.79 ± 0.15	0.04*	0.04*
	Group 4	0.56 ± 0.21	0.65 ± 0.20	0.06*	0.66 ± 0.22	0.72 ± 0.20	0.06*	0.63 ± 0.26	0.67 ± 0.25	0.04*	0.04*
	Group 5	0.64 ± 0.17	0.68 ± 0.16	0.04*	0.71 ± 0.19	0.73 ± 0.16	0.02	0.65 ± 0.22	0.67 ± 0.21	0.02	0.02
	Group 6	0.68 ± 0.16	0.73 ± 0.13	0.05*	0.67 ± 0.22	0.72 ± 0.18	0.05*	0.76 ± 0.15	0.79 ± 0.13	0.04	0.04
	Group 7	0.57 ± 0.19	0.65 ± 0.17	0.08*	0.67 ± 0.20	0.68 ± 0.20	0.01	0.59 ± 0.24	0.69 ± 0.21	0.10*	0.10*
Fibrosis score	Group 2	0.30 ± 0.24	0.65 ± 0.20	0.36*	0.21 ± 0.19	0.62 ± 0.22	0.41*	0.83 ± 0.28	0.78 ± 0.21	-0.03	-0.03
	Group 3	0.45 ± 0.28	0.65 ± 0.24	0.50*	0.36 ± 0.26	0.63 ± 0.25	0.27*	0.84 ± 0.19	0.76 ± 0.23	-0.08*	-0.08*
	Group 4	0.46 ± 0.22	0.65 ± 0.19	0.18*	0.39 ± 0.24	0.69 ± 0.21	0.30*	0.76 ± 0.26	0.67 ± 0.26	-0.08*	-0.08*
	Group 5	0.42 ± 0.28	0.62 ± 0.23	0.20*	0.34 ± 0.17	0.61 ± 0.24	0.27*	0.76 ± 0.28	0.70 ± 0.21	-0.05	-0.05
	Group 6	0.29 ± 0.31	0.63 ± 0.21	0.34*	0.23 ± 0.27	0.60 ± 0.26	0.37*	0.81 ± 0.29	0.75 ± 0.20	-0.05	-0.05
	Group 7	0.04 ± 0.09	0.64 ± 0.19	0.59*	0.02 ± 0.60	0.58 ± 0.21	0.55*	0.83 ± 0.33	0.78 ± 0.16	-0.02	-0.02

Data are presented as the mean ± standard deviation of the spatial overlap accuracy metrics of the CT slices.

\*Indicates that the difference in the overlap accuracy metrics between (1) and (2) is statistically significant ( $P < 0.008$ ).  $P$ -values were calculated using paired  $t$ -tests. Significance level of 0.008 takes into account the Bonferroni correction for multiple tests. †The mean pixel-wise precision of the original and converted CT images was calculated for those CT slices for which pixel-wise precision was available for both the original and converted CT images. ‡The difference in pixel-wise precision between the original and converted CT images and its statistical significance were calculated for those CT slices for which pixel-wise precision was available for both the original and converted CT images. CT = computed tomography, DSC = dice similarity coefficient



CT Image Conversion for Quantification of ILD

segmentation in each of the three types of CT images (original CT, converted CT, and reference standard) in 286 CT slices from groups 2 to 7, and a total of 858 CT slices were evaluated. The overall visual accuracy scores differed

significantly across the CT types ( $P < 0.001$ ) (Table 4); the highest score was for the reference standard, followed by the converted CT and the original CT (Figs. 4, 5). The overall mean accuracy scores were 8.54 for the reference standard,

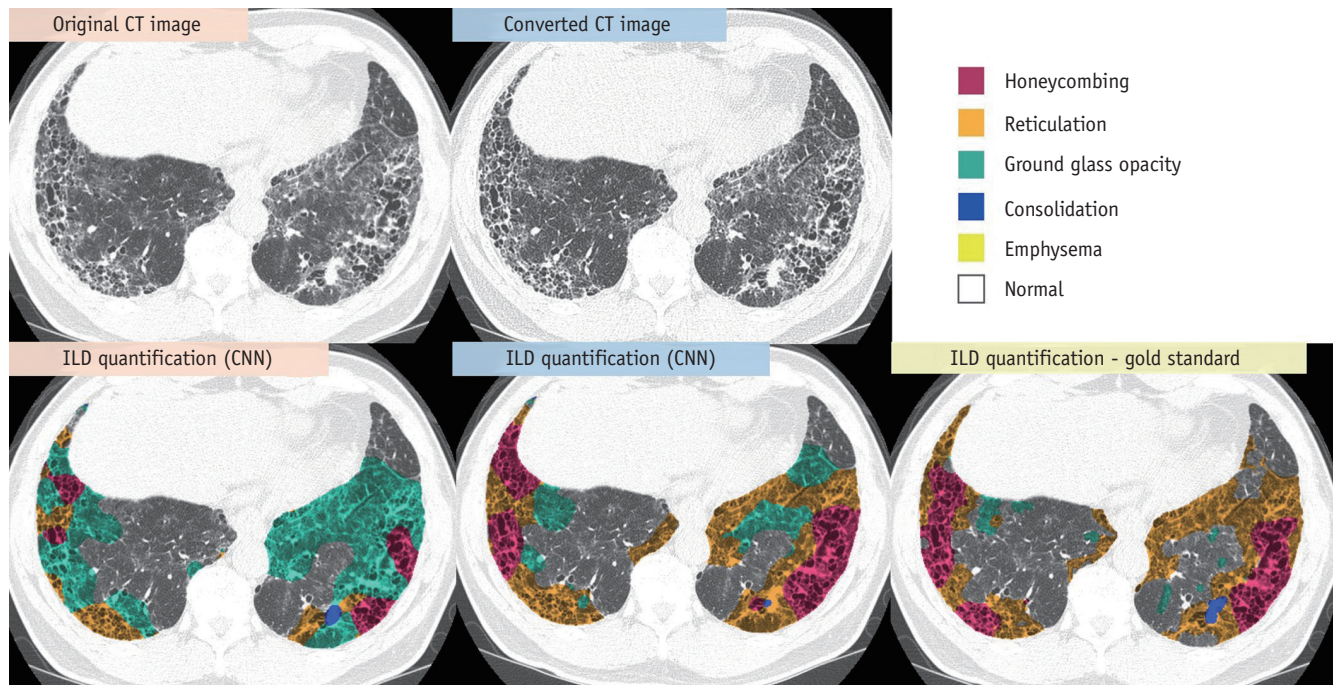
**Table 4.** Visual accuracy scores in the ILD CT segmentation

	Original CT (automated)	Converted CT (automated)	Reference standard (original, manual)	<i>P</i>	Post-hoc test
Overall (Group 2 to 7)	6.35 ± 2.49	7.64 ± 1.94	8.54 ± 1.60	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 2	6.04 ± 2.57	7.52 ± 2.13	8.79 ± 1.42	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 3	6.81 ± 2.32	7.54 ± 2.16	8.59 ± 1.34	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 4	6.78 ± 2.07	7.56 ± 1.84	8.39 ± 1.53	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 5	7.11 ± 1.96	7.72 ± 1.77	8.25 ± 1.79	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 6	6.79 ± 2.31	7.93 ± 1.58	8.39 ± 2.00	< 0.001*	RS > Conv > ORIG <sup>†</sup>
Group 7	4.74 ± 2.80	7.56 ± 2.09	8.77 ± 1.39	< 0.001*	RS > Conv > ORIG <sup>†</sup>

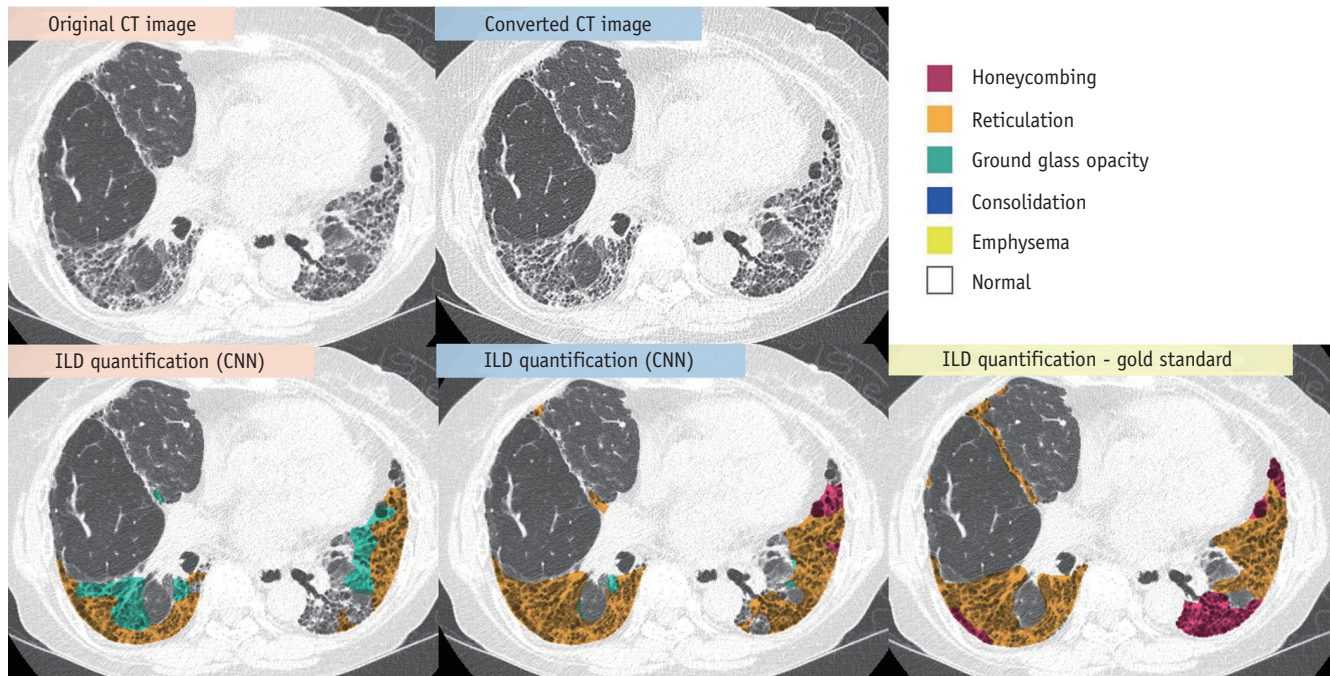
Data are presented as mean ± standard deviation of the average of the visual scores of five readers of each CT slice. Visual accuracy scores are defined as the degree of agreement with the readers' subjective segmentation of ILD CT patterns: 1 = agreement from 0 to 9%, 2 = 10% to 19%, 3 = 20% to 29%, 4 = 30% to 39%, 5 = 40% to 49%, 6 = 50% to 59%, 7 = 60% to 69%, 8 = 70% to 79%, 9 = 80% to 89%, and 10 = 90% to 100%.

\**P*-value was calculated using a repeated-measures analysis, <sup>†</sup>Indicates that the post-hoc tests among RS, Conv, and ORIG are statistically significant.

ILD = interstitial lung disease, CT = computed tomography, RS = reference standard, Conv = Converted, ORIG = original



**Fig. 4.** Example of the conversion of group 3 computed tomography (CT) slices of a man with interstitial lung disease (ILD) into the target CT style using a routable generative adversarial network, and the quantification of ILD on the original CT, converted CT, and radiologist's reference standard. The dice similarity coefficient (DSC) values of the quantified total abnormalities, fibrosis score, honeycombing, reticulation, and ground-glass opacity (GGO) on the original CT images were 0.86, 0.44, 0.50, 0.18, and 0.04, respectively, with the radiologist's quantification used as the reference standard. After CT image conversion, the DSC values were 0.88, 0.88, 0.78, 0.71, and 0.18 for total abnormalities, fibrosis score, honeycombing, reticulation, and GGO, respectively, on the converted CT images, which were higher than on the original CT images. The five radiologists' mean visual accuracy scores of the segmentations on the original, converted, and reference standard images were 5.40 ± 1.35, 6.80 ± 1.17, and 8.80 ± 1.17, respectively. CNN = convolutional neural network



**Fig. 5.** Example of the conversion of a group 4 computed tomography (CT) slice of a male patient with interstitial lung disease (ILD) into the target CT style using a routable generative adversarial network, and the quantification of ILD on each CT image. On the original CT images, the dice similarity coefficient (DSC) values of total abnormalities, fibrosis score, honeycombing, and reticulation quantifications were 0.78, 0.59, 0.00, and 0.60, respectively. After CT image conversion, the DSC values on the converted CT images were 0.86, 0.85, 0.14, and 0.76 for total abnormalities, fibrosis score, honeycombing, and reticulation, respectively. The five radiologists' mean visual accuracy scores of the segmentations on the original, converted, and reference standard images were  $7.20 \pm 1.60$ ,  $8.60 \pm 0.80$ , and  $9.00 \pm 0.89$ , respectively. CNN = convolutional neural network

which is close to a score of 9, indicating 80%–89% agreement with the radiologists' segmentation; 7.64 for the converted CT, indicating 70%–79% agreement; and 6.35 for the original CT, indicating 60%–69% agreement. Visual accuracy scores were significantly highest for the reference standard, followed by the converted CT and original CT in all groups (all  $P < 0.001$ ). The average visual scores of Group 2 to 7 showed a wide range from 4.74 to 7.11 on the original CT, but became similar among the groups, ranging from 7.52 to 7.93, with decreased variability after CT conversion (Table 4).

## DISCUSSION

The quantification of ILD is dependent on the characteristics of CT images, which are affected by technical parameters and manufacturers. Our study quantitatively and visually demonstrated that the use of a RouteGAN to convert CT images with various acquisition conditions to CT images with a particular acquisition condition could improve the accuracy and variability of deep learning-based quantification of regional CT patterns of ILD, such as fibrosis score, honeycombing, and reticulation, as well as

the total abnormalities of ILD.

Previous studies using texture-based quantification have demonstrated the value of quantitative ILD imaging [8,13-18]. However, these studies were conducted in a single or a few institutions. CT quantification of ILD is based on complex texture analysis of the relationships between adjacent pixels and CT densities, and is therefore inherently sensitive to variations caused by technical parameters. This limitation is an important hurdle in the application of quantitative ILD imaging to various CT datasets. In our study, the accuracy of a deep learning-based ILD quantification system developed using mainly CT images with particular technical parameters was limited when applied to various parameters and manufacturers.

To apply quantification to various CT datasets, standardization of the input images to the features of the optimized CT parameters for quantitative imaging is a viable option. Recently, deep learning-based CT conversion using a supervised learning method was applied to the emphysema index and radiomics in different reconstruction kernels [25,26]. However, the previous method requires paired CT datasets and possibly performs only between the features



of paired datasets. A RouteGAN, which uses unsupervised image-to-image translation with unpaired CT images, enables image conversion simultaneously across different parameters, including reconstruction kernel, radiation dose, and scanner manufacturer. Although this method requires a relatively large number of CT images for training, it is easier to prepare multiple training datasets because the images can be unpaired according to parameters and manufacturers and unlabeled with disease patterns. Furthermore, this method enables the addition of datasets after initial algorithm training, thereby possibly broadening the scope of conversion to various CT parameters [24]. Recently, Lee et al. [32] applied CT image conversion using a generative adversarial network (GAN) in radiomics and showed that the synthesis of standard CT images using a GAN enhances the reproducibility of radiomics features across various CT protocols and scanners in a phantom setting. We believe that this method may promote the use of quantitative ILD imaging in clinical practice and in longitudinal multicenter studies with CT images under various acquisition conditions and its use in the assessment of incidental ILD or interstitial lung abnormalities detected using non-dedicated CT protocols.

Another approach for applying ILD quantification to various CT datasets involves training the quantification algorithm with a heterogeneous array of technical parameters [29,33], which does not require a CT conversion process. However, collecting a large dataset with varied CT parameters and manufacturers for patients with particular diseases may be challenging. In addition, the manual labeling of each CT pattern of ILD is required for algorithm training, which would be difficult and time-consuming to perform, even for expert radiologists.

This study implemented several metrics, including the DSC and pixel-wise analysis [31], to evaluate the quantification accuracy for multiple CT patterns. We converted the CT images of Group 2 to 7 into the CT style of Group 1 using a RouteGAN, resulting in increased extents of the segmented fibrosis score, honeycomb, and reticulation and decreased extents of the segmented GGO and consolidation. These trends may be the main cause of the improvement or non-improvement of the spatial overlap metrics in this study. All overlap metrics showed improved quantification of total abnormalities and reticulation after CT conversion but not necessarily an improvement in other patterns. However, the decrease in the overlap metrics were either not statistically significant or were much smaller than the increase in the DSC and pixel-wise recall of fibrosis, honeycombing,

and reticulation. The negative results in improving the quantification of emphysema and consolidation could be caused by insufficient training on CT conversion for kernel and radiation dose factors, which may affect the texture of these patterns. In addition, the relatively small structural size of the emphysema or consolidation could affect DSC after CT conversion [34]. Improvement in quantification after CT conversion may be limited to certain CT patterns of ILD. However, given that the total extent of ILD and CT patterns, such as fibrosis score, honeycombing, and reticulation, are important for evaluating disease severity and outcome prediction in ILD, our method may be helpful for the accurate evaluation of ILD on CT images under various acquisition conditions.

Our study had several limitations. First, we used a single-expert-based manual segmentation of each CT pattern as the reference standard for evaluating spatial overlap accuracy. Using two CT images of each ILD patient in different acquisition settings would be ideal; however, this is challenging because of unnecessary radiation exposure. Manual segmentation is inherently subjective and biased. However, readings by five thoracic radiologists showed approximately 80%–89% agreement in subjective segmentation, suggesting that the use of a single radiologist to establish the reference standard did not bias our results. Second, our study was conducted on only seven groups of CT images classified according to acquisition conditions that are most widely used in ILD evaluation. However, other CT parameters should be evaluated to expand the applicability of our results. Third, we conducted our study on individual slices rather than on patients, and parenchymal abnormalities were not observed or were minimal in some slices, making it difficult to evaluate the CT conversion effect by DSC or pixel-wise analysis on a per-patient basis. In addition, the ILD quantification software used analyzes on a per-slice basis of the whole CT volume and not by 3D volume. To evaluate the impact of CT conversion on the clinical relevance of ILD quantification, an analysis of the whole CT volume would be more appropriate. Fourth, we analyzed a small number of CT slices, which may have limited the statistical power. Fifth, ILD quantification was performed using a single software package (Aview, Coreline Soft). Our results should be confirmed using other automated segmentation tools. Finally, the effect of CT conversion on the clinical application of ILD quantification was not evaluated. Previous studies have reported the clinical value of quantitative ILD imaging [13,17,35]. However, its usefulness across various

CT scans remains unknown and should be evaluated in future studies.

In conclusion, CT image conversion using a RouteGAN can improve the accuracy and variability of CT images obtained using different parameters and scanners with different manufacturers in deep-learning-based automatedILD quantification. This method is expected to provide a feasible quantification that is robust to variations in CT parameters and machine settings, enabling quantitative analysis across inconsistent CT datasets, such as those resulting from retrospective, longitudinal, or multicenter studies, as well as those acquired in clinical practice.

## Supplement

The Supplement is available with this article at <https://doi.org/10.3348/kjr.2023.0088>.

### Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

### Conflicts of Interest

Joon Beom Seo, Namkug Kim, and Ho Yun Lee, contributing editors of the *Korean Journal of Radiology*, were not involved in the editorial evaluation or decision to publish this article. Jong Chul Ye, Hyunjong Kim, Joon Beom Seo, Sang Min Lee, and Hye Jeon Hwang hold a patent for tomography image processing method using single neural network based on unsupervised learning for image standardization and apparatus therefor (Patent No. KR-10-2021-0040878). In this study, this patented item was used. Joon Beom Seo and Namkug Kim hold a patent on a method for an automatic classifier of lung diseases (Patent No. KR-10-0998630) and have received royalty payments from Coreline Soft, Co., Ltd. Joon Beom Seo, Namkug Kim, Sang Min Lee holds stock/stock options in Coreline Soft, Co., Ltd., Korea. Hee Jun Park was an employee of Coreline Soft, Co., Ltd., Korea. All remaining authors have declared no conflicts of interest.

### Author Contributions

Conceptualization: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye. Data curation: Hye Jeon Hwang, Hee Jun Park, Ryoungwoo Jang, Jihye Yun. Formal analysis: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jihye Yun.

Funding acquisition: Sang Min Lee. Investigation: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye, Hyunjong Kim, Gyutaek Oh, Ryoungwoo Jang, Jihye Yun. Methodology: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye, Jihye Yun. Project administration: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee. Resources: Hye Jeon Hwang, Sang Min Lee, Hee Jun Park, Ho Yun Lee, Soon Ho Yoon, Kyung Eun Shin, Jae Wook Lee, Woocheol Kwon, Joo Sung Sun, Seulgi You, Myung Hee Chung, Bo Mi Gil, Jae-Kwang Lim, Youkyung Lee, Su Jin Hong, Yo Won Choi. Software: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye, Hyunjong Kim, Gyutaek Oh, Ryoungwoo Jang, Jihye Yun, Namkug Kim, Hee Jun Park. Supervision: Joon Beom Seo, Jong Chul Ye, Ho Yun Lee, Soon Ho Yoon. Validation: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye, Ho Yun Lee, Soon Ho Yoon, Kyung Eun Shin, Woocheol Kwon, Joo Sung Sun, Myung Hee Chung, Yo Won Choi. Visualization: Joon Beom Seo, Hye Jeon Hwang, Sang Min Lee, Jong Chul Ye, Hyunjong Kim, Gyutaek Oh, Ryoungwoo Jang, Jihye Yun, Namkug Kim, Hee Jun Park. Writing—original draft: Joon Beom Seo, Hye Jeon Hwang. Writing—review & editing: Jong Chul Ye, Ho Yun Lee, Soon Ho Yoon, Kyung Eun Shin, Jae Wook Lee, Woocheol Kwon, Joo Sung Sun, Seulgi You, Myung Hee Chung, Bo Mi Gil, Jae-Kwang Lim, Youkyung Lee, Su Jin Hong, Yo Won Choi.

### ORCID IDs

Hye Jeon Hwang

<https://orcid.org/0000-0003-3508-2870>

Hyunjong Kim

<https://orcid.org/0000-0003-1653-3410>

Joon Beom Seo

<https://orcid.org/0000-0003-0271-7884>

Jong Chul Ye

<https://orcid.org/0000-0001-9763-9609>

Gyutaek Oh

<https://orcid.org/0000-0002-7689-1538>

Sang Min Lee

<https://orcid.org/0000-0002-2173-2193>

Ryoungwoo Jang

<https://orcid.org/0000-0002-1511-7469>

Jihye Yun

<https://orcid.org/0000-0002-5233-6687>

Namkug Kim

<https://orcid.org/0000-0002-3438-2217>

Hee Jun Park

<https://orcid.org/0000-0002-7795-4801>

Ho Yun Lee

<https://orcid.org/0000-0001-9960-5648>

Soon Ho Yoon

<https://orcid.org/0000-0002-3700-0165>

Kyung Eun Shin

<https://orcid.org/0000-0003-2373-4704>

Jae Wook Lee

<https://orcid.org/0000-0002-1496-472X>

Woocheol Kwon

<https://orcid.org/0000-0003-0397-4364>

Joo Sung Sun

<https://orcid.org/0000-0002-3408-8544>

Seulgi You

<https://orcid.org/0000-0003-1136-9287>

Myung Hee Chung

<https://orcid.org/0000-0002-6141-2258>

Bo Mi Gil

<https://orcid.org/0000-0003-0576-5830>

Jae-Kwang Lim

<https://orcid.org/0000-0002-1299-9996>

Youkyung Lee

<https://orcid.org/0000-0002-0939-4574>

Su Jin Hong

<https://orcid.org/0000-0002-0634-4731>

Yo Won Choi

<https://orcid.org/0000-0003-0508-9430>

### Funding Statement

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, Republic of Korea, the Ministry of Food and Drug Safety) (Project Number: NTIS 1711138474).

### REFERENCES

- Gay SE, Kazerooni EA, Toews GB, Lynch JP 3rd, Gross BH, Cascade PN, et al. Idiopathic pulmonary fibrosis: predicting response to therapy and survival. *Am J Respir Crit Care Med* 1998;157(4 Pt 1):1063-1072
- Muller NL. Clinical value of high-resolution CT in chronic diffuse lung disease. *AJR Am J Roentgenol* 1991;157:1163-1170
- Nishimura K, Izumi T, Kitaichi M, Nagai S, Itoh H. The diagnostic accuracy of high-resolution computed tomography in diffuse infiltrative lung diseases. *Chest* 1993;104:1149-1155
- Scatarige JC, Diette GB, Haponik EF, Merriman B, Fishman EK. Utility of high-resolution CT for management of diffuse lung disease: results of a survey of U.S. pulmonary physicians. *Acad Radiol* 2003;10:167-175
- Aziz ZA, Wells AU, Hansell DM, Bain GA, Copley SJ, Desai SR, et al. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004;59:506-511
- Collins CD, Wells AU, Hansell DM, Morgan RA, MacSweeney JE, du Bois RM, et al. Observer variation in pattern type and extent of disease in fibrosing alveolitis on thin section computed tomography and chest radiography. *Clin Radiol* 1994;49:236-240
- Best AC, Lynch AM, Bozic CM, Miller D, Grunwald GK, Lynch DA. Quantitative CT indexes in idiopathic pulmonary fibrosis: relationship with physiologic impairment. *Radiology* 2003;228:407-414
- Best AC, Meng J, Lynch AM, Bozic CM, Miller D, Grunwald GK, et al. Idiopathic pulmonary fibrosis: physiologic tests, quantitative CT indexes, and CT visual scores as predictors of mortality. *Radiology* 2008;246:935-940
- Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol* 2004;59:1061-1069
- Delorme S, Keller-Reichenbecher MA, Zuna I, Schlegel W, Van Kaick G. Usual interstitial pneumonia. Quantitative assessment of high-resolution computed tomography findings by computer-assisted texture-based image analysis. *Invest Radiol* 1997;32:566-574
- Rodriguez LH, Vargas PF, Raff U, Lynch DA, Rojas GM, Moxley DM, et al. Automated discrimination and quantification of idiopathic pulmonary fibrosis from normal lung parenchyma using generalized fractal dimensions in high-resolution computed tomography images. *Acad Radiol* 1995;2:10-18
- Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol* 2007;14:772-787
- Iwasawa T, Takemura T, Okudera K, Gotoh T, Iwao Y, Kitamura H, et al. The importance of subpleural fibrosis in the prognosis of patients with idiopathic interstitial pneumonias. *Eur J Radiol* 2017;90:106-113
- Sverzellati N, Calabrò E, Chetta A, Concarì G, Larici AR, Mereu M, et al. Visual score and quantitative CT indices in pulmonary fibrosis: Relationship with physiologic impairment. *Radiol Med* 2007;112:1160-1172
- Yoon RG, Seo JB, Kim N, Lee HJ, Lee SM, Lee YK, et al. Quantitative assessment of change in regional disease patterns on serial HRCT of fibrotic interstitial pneumonia with texture-based automated quantification system. *Eur Radiol* 2013;23:692-701
- Jacob J, Bartholmai BJ, Rajagopalan S, Brun AL, Egashira R, Karwoski R, et al. Evaluation of computer-based computer tomography stratification against outcome models in connective tissue disease-related interstitial lung disease: a patient outcome study. *BMC Med* 2016;14:190
- Lee SM, Seo JB, Oh SY, Kim TH, Song JW, Lee SM, et



- al. Prediction of survival by texture-based automated quantitative assessment of regional disease patterns on CT in idiopathic pulmonary fibrosis. *Eur Radiol* 2018;28:1293-1300
18. Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 2014;43:204-212
  19. Chen-Mayer HH, Fuld MK, Hoppel B, Judy PF, Sieren JP, Guo J, et al. Standardizing CT lung density measure across scanner manufacturers. *Med Phys* 2017;44:974-985
  20. Gierada DS, Bierhals AJ, Choong CK, Bartel ST, Ritter JH, Das NA, et al. Effects of CT section thickness and reconstruction kernel on emphysema quantification relationship to the magnitude of the CT emphysema index. *Acad Radiol* 2010;17:146-156
  21. Kemerink GJ, Kruize HH, Lamers RJ, van Engelshoven JM. Density resolution in quantitative computed tomography of foam and lung. *Med Phys* 1996;23:1697-1708
  22. Madani A, De Maertelaer V, Zanen J, Gevenois PA. Pulmonary emphysema: radiation dose and section thickness at multidetector CT quantification--comparison with macroscopic and microscopic morphometry. *Radiology* 2007;243:250-257
  23. Kim J, Lee JK, Lee KM. *Accurate image super-resolution using very deep convolutional networks*. In: *2016 The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27-30; Las Vegas, USA. Danvers: The Institute of Electrical and Electronics Engineers, Inc; 2016. p. 1646-1654
  24. Kim H, Oh G, Seo JB, Hwang HJ, Lee SM, Yun J, et al. Multi-domain CT translation by a routable translation network. *Phys Med Biol* 2022;67:21
  25. Lee SM, Lee JG, Lee G, Choe J, Do KH, Kim N, et al. CT image conversion among different reconstruction kernels without a sinogram by using a convolutional neural network. *Korean J Radiol* 2019;20:295-303
  26. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology* 2019;292:365-373
  27. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788-824
  28. Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013;188:733-748
  29. Kim GB, Jung KH, Lee Y, Kim HJ, Kim N, Jun S, et al. Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease. *J Digit Imaging* 2018;31:415-424
  30. Choe J, Hwang HJ, Seo JB, Lee SM, Yun J, Kim MJ, et al. Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology* 2022;302:187-197
  31. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015;15:29
  32. Lee SB, Cho YJ, Hong Y, Jeong D, Lee J, Kim SH, et al. Deep learning-based image conversion improves the reproducibility of computed tomography radiomics features a phantom study. *Invest Radiol* 2022;57:308-317
  33. Mårtensson G, Ferreira D, Granberg T, Cavallin L, Oppedal K, Padovani A, et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Med Image Anal* 2020;66:101714
  34. Reinke A, Eisenmann M, Tizabi MD, Sudre CH, Rädtsch T, Antonelli M, et al. *Common limitations of performance metrics in biomedical image analysis*. In: *Medical Imaging with Deep Learning 2021*; 2021 Jul 7-9; Lübeck, Germany. 2021
  35. Kim GHJ, Weigt SS, Belperio JA, Brown MS, Shi Y, Lai JH, et al. Prediction of idiopathic pulmonary fibrosis progression using early quantitative changes on CT imaging for a short term of clinical 18-24-month follow-ups. *Eur Radiol* 2020;30:726-734