


CHEST



# Value of CT quantification in progressive fibrosing interstitial lung disease: a deep learning approach

Seok Young Koh<sup>1</sup>, Jong Hyuk Lee<sup>1</sup>, Hyungin Park<sup>1</sup> and Jin Mo Goo<sup>1,2,3,4\*</sup> 

## Abstract

**Objectives** To evaluate the relationship of changes in the deep learning–based CT quantification of interstitial lung disease (ILD) with changes in forced vital capacity (FVC) and visual assessments of ILD progression, and to investigate their prognostic implications.

**Methods** This study included ILD patients with CT scans at intervals of over 2 years between January 2015 and June 2021. Deep learning–based texture analysis software was used to segment ILD findings on CT images (fibrosis: reticular opacity + honeycombing cysts; total ILD extent: ground-glass opacity + fibrosis). Patients were grouped according to the absolute decline of predicted FVC (< 5%, 5–10%, and ≥ 10%) and ILD progression assessed by thoracic radiologists, and their quantification results were compared among these groups. The associations between quantification results and survival were evaluated using multivariable Cox regression analysis.

**Results** In total, 468 patients (239 men; 64 ± 9.5 years) were included. Fibrosis and total ILD extents more increased in patients with larger FVC decline ( $p < .001$  in both). Patients with ILD progression had higher fibrosis and total ILD extent increases than those without ILD progression ( $p < .001$  in both). Increases in fibrosis and total ILD extent were significant prognostic factors when adjusted for absolute FVC declines of ≥ 5% (hazard ratio [HR] 1.844,  $p = .01$  for fibrosis; HR 2.484,  $p < .001$  for total ILD extent) and ≥ 10% (HR 2.918,  $p < .001$  for fibrosis; HR 3.125,  $p < .001$  for total ILD extent).

**Conclusion** Changes in ILD CT quantification correlated with changes in FVC and visual assessment of ILD progression, and they were independent prognostic factors in ILD patients.

**Clinical relevance statement** Quantifying the CT features of interstitial lung disease using deep learning techniques could play a key role in defining and predicting the prognosis of progressive fibrosing interstitial lung disease.

## Key Points

- Radiologic findings on high-resolution CT are important in diagnosing progressive fibrosing interstitial lung disease.
- Deep learning–based quantification results for fibrosis and total interstitial lung disease extents correlated with the decline in forced vital capacity and visual assessments of interstitial lung disease progression, and emerged as independent prognostic factors.
- Deep learning–based interstitial lung disease CT quantification can play a key role in diagnosing and prognosticating progressive fibrosing interstitial lung disease.

Seok Young Koh and Jong Hyuk Lee contributed equally to this research as first authors.

\*Correspondence:

Jin Mo Goo

jmgoo@snu.ac.kr

Full list of author information is available at the end of the article

**Keywords** Lung diseases (Interstitial), Quantitative evaluation, Deep learning, Prognosis, Multidetector computed tomography

## Introduction

Progressive fibrosing interstitial lung disease (PF-ILD) comprises a heterogeneous group of disease entities characterized by a relentless decline in lung function. It encompasses idiopathic pulmonary fibrosis (IPF) [1–4] and non-IPF ILDs such as chronic hypersensitivity pneumonitis, connective tissue disease–associated ILDs (CTD-ILDs), and idiopathic nonspecific interstitial pneumonia (NSIP) with a progressive fibrotic phenotype despite immunosuppressive therapy and the removal of disease-promoting triggers [5, 6]. Based on the INBUILD study [7], which investigated the impact of nintedanib, an antifibrotic drug, on non-IPF ILDs, a diagnostic consensus of PF-ILD has been reached that considers a combination of deteriorating lung function, CT findings, and patient symptoms over 24 months despite conventional treatment. The specific diagnostic criteria are as follows: a relative decline in forced vital capacity (FVC)  $\geq 10\%$ , a relative decline in FVC  $\geq 5\%$  accompanied by increased fibrosis on high-resolution CT (HRCT) or worsening of symptoms, or worsening symptoms with increased fibrosis on HRCT [5]. As these diagnostic criteria show, fibrosis on HRCT is an important parameter for diagnosing PF-ILD, along with a decline in FVC [8–15]. Indeed, extensive traction bronchiectasis or honeycombing on HRCT is a risk factor for PF-ILD [5, 16, 17]. However, the interpretation of ILD imaging findings on HRCT has largely relied on experts' subjective judgment, which introduces the possibility of inter- and intra-observer variability and can lead to diagnostic limitations, requiring objective quantification [2, 7, 18–21].

Recent artificial intelligence models have demonstrated diagnostic capabilities comparable to those of experts in various medical fields [22–25]. In thoracic radiology, the analysis of chest imaging for disease pattern recognition, segmentation, and diagnosis has been the focus of extensive research [26–30]. Likewise, in the field of ILD, several studies have utilized quantitative software tools to assess ILD CT features, including fibrosis, and analyze image patterns [31–33]. For instance, Park et al investigated the correlation between the quantification of fibrosis extent on CT and FVC decline in IPF patients. The study used in-house software trained with 1200 typical patterns to automatically quantify six patterns (normal, ground-glass opacity [GGO], reticular opacity, honeycombing, consolidation, and emphysema) [31]. In this study, we aimed to evaluate changes in deep learning–based CT quantification of ILD according to changes in FVC and

visual assessments of ILD progression, and to investigate their prognostic implications.

## Materials and methods

This retrospective study was approved by the Institutional Review Board of Seoul National University Hospital, and the requirement for written informed consent was waived (IRB No. H-2112-040-1279). The study sample of this study has not been reported.

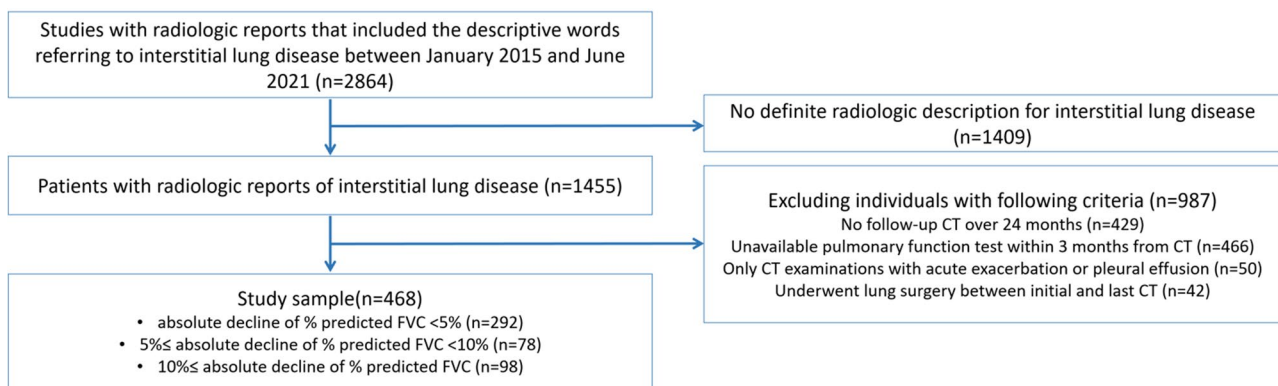
### Study sample

This retrospective study was performed at a tertiary referral hospital. All CT exams were searched for keywords describing ILDs in radiological reports (e.g., “reticular opacity,” “reticulation,” “traction bronchiectasis,” “honeycombing,” “interstitial fibrosis,” “interstitial lung disease,” “ILD,” “idiopathic pulmonary fibrosis,” “IPF,” “usual interstitial pneumonia,” “UIP,” “nonspecific interstitial pneumonia,” “NSIP,” “hypersensitivity pneumonitis,” “HP,” “interstitial lung abnormality,” or “ILA”). This search yielded 2864 patients. The search was conducted throughout December 2021, and all selected CT exams were performed from January 2015 to June 2021.

One radiology resident (S.Y.K. with 4 years of experience in ILD imaging) reviewed the radiologic reports and excluded patients for whom the radiologic reports did not indicate ILD ( $n = 1409$ ). Since PF-ILD is defined as a decline of PFT results or an increase in fibrosis on CT over 24 months, we excluded patients whose CT examinations' intervals were less than 24 months ( $n = 429$ ) or who did not have available PFT results within 3 months from the CT scans ( $n = 466$ ). Next, one thoracic radiologist (J.H.L. with 11 years of experience in ILD imaging) and a radiology resident (S.Y.K.) reviewed the remaining patients' CT images on a picture archiving and communication system and excluded patients whose CT images showed acute exacerbation of ILD or a significant amount of pleural effusion that precluded reviewing ILD findings of the lung parenchyma ( $n = 50$ ), as well as patients who underwent lung surgery between the initial and follow-up CT images ( $n = 42$ ). The flowchart for the selection process is described in Fig. 1 and the CT image acquisition information is suggested in Supplementary Text.

### Quantitative analysis of the CT images

Commercially available deep learning–based texture analysis software (AVIEW Lung Texture version 1.1.43.7, Core-Line Soft) was used. This tool performs fully automatic



**Fig. 1** Flowchart for the selection process of the study population

segmentation of the total lung parenchyma and each finding (emphysema, GGO, consolidation, reticular opacity, and honeycombing cysts) on input CT images and provides quantification results as percentages (%) of the total lung volume. We extracted reticular opacity, GGO, and honeycombing cysts as ILD CT findings. We defined (a) fibrosis extent as the sum of reticular opacity and honeycombing cysts and (b) the total ILD extent as the sum of GGO, reticular opacity, and honeycombing cysts. One thoracic radiologist (J.H.L.) applied this tool to all images of the study sample and confirmed the completeness of the segmentation of the software. No additional manual modifications were made to the segmentation results. Changes in each ILD finding, fibrosis extent, and the total ILD extent between initial and follow-up CT images were then calculated. The software's segmentation performances were reported to be in substantial concordance with those of thoracic radiologists [34, 35].

### Clinical information

The following clinical information was recorded by searching electronic medical records: age, sex, clinical diagnosis for ILD, FVC, and predicted FVC results obtained within 3 months from the initial and follow-up CT examinations, according to the American Thoracic Society guidelines [36]. FVC was expressed as volume (L) and percentage (%) of the predicted normal value. In this study, the absolute decline of % predicted FVC between initial and follow-up PFTs was used according to the recent guideline [37].

To evaluate the prognostic value of CT quantification results in PF-ILD, survival status and date of death were acquired from a database of the Ministry of the Interior and Safety, Korea. We assessed patients' overall survival (OS), defined as the interval from the date of initial CT examination to the date of death from any cause. Survival time was censored on September 28, 2022. For patients

who died, the time of censoring was defined as the date of death.

### Visual assessment of ILD progression

ILD progression on CT images was visually assessed by thoracic radiologists as a reference standard independent of PFT results. Two thoracic radiologists (J.H.L. and H.I.P., with 6 years of experience in ILD imaging) independently reviewed the initial and follow-up CT images of the study sample side-by-side and determined the progression of ILD in a binary manner (progression or stability) based on the following criteria [37]: (a) increased extent of severity of traction bronchiectasis; (b) new ground-glass opacity with traction bronchiectasis; (c) new or increased extent or coarseness of reticular opacity; (d) new or increased honeycombing; (e) increased lung volume loss. If their evaluations agreed with each other, the consensus was used. If their evaluations differed, a senior thoracic radiologist (J.M.G. with 32 years of experience in ILD imaging) adjudicated the progression of ILD.

### Statistical analyses

Baseline characteristics and ILD CT quantification results are given as mean with standard deviation or median with ranges according to normality testing.

The absolute decline of the predicted FVC (%) between initial and follow-up PFTs was categorized as < 5%, ≥ 5% and <10%, or ≥ 10%. Changes in the ILD quantification between initial and follow-up CT images were compared across these three FVC groups using the Kruskal-Wallis test. The ILD quantification changes were also compared according to the thoracic radiologists' visual assessment of ILD progression, using the Wilcoxon rank sum test. The Spearman correlation coefficient was used to evaluate the relationship between ILD quantification changes and the absolute decline of % predicted FVC (both treated as continuous variables).

We calculated the optimal threshold of ILD quantification changes for predicting absolute declines of the predicted FVC of 5% and 10%, as well as all-cause mortality, using Youden's index and the minimum *p*-value approach for all-cause mortality.

Multivariable Cox proportional hazard regression analyses were performed to evaluate the prognostic implications of deep learning-based CT quantification results in PF-ILD. Since 5% and 10% absolute decline in FVC are accepted as the standard for defining PF-ILD, Cox analyses were performed using those criteria and adding a visual assessment for ILD progression and changes in each ILD finding, fibrosis extent, or the total ILD extent, adjusted for sex, age, and initial predicted FVC (%). ILD quantification changes were treated as categorical variables using the optimal threshold described above. For these various Cox regression models, we calculated the C-index using Uno's concordance statistic and statistically compared them with the baseline model consisting of age, sex, initial predicted FVC (%), and absolute decline of FVC  $\geq$  5% or 10%.

As a sensitivity analysis, we separately performed the above survival analyses (multivariable Cox proportional hazard regression analyses and C-index calculation) in subgroups with total ILD extents on initial CT of  $<$  10% and  $\geq$  10%.

All statistical analyses were performed using SPSS version 21.0 (IBM Corp.), SAS version 9.4 (SAS Institute Inc.), and R version 3.6.1. (R Project for Statistical Computing), and a *p*-value of  $<$  .05 was considered to indicate statistical significance.

## Results

### Baseline characteristics of study sample

In total, 468 individuals (239 men and 229 women;  $64 \pm 9.5$  years) were included in this study. The clinical diagnoses were IPF ( $n = 220$ , 47%), CTD-ILD ( $n = 148$ , 31.6%), idiopathic NSIP ( $n = 22$ , 4.7%), chronic hypersensitivity pneumonitis ( $n = 5$ , 1.1%), cryptogenic organizing pneumonia ( $n = 3$ , 0.6%), desquamative interstitial pneumonia ( $n = 1$ , 0.2%), and unclassifiable ILD ( $n = 69$ , 14.7%) (Table 1). The initial and follow-up % predicted FVCs were  $81.3 \pm 18.1\%$  and  $78.3 \pm 20.5\%$ , respectively. Furthermore, 292 (62.4%), 78 (16.7%), and 98 (20.9%) patients were categorized as having declines of  $<$  5%,  $\geq$  5% and  $<$  10%, and  $\geq$  10% in % predicted FVC, respectively. The radiologists determined that 266 patients had ILD progression (56.8%) (Supplementary Table 1). In a total follow-up of 75 months (interquartile range [IQR], 59–84 months), 102 individuals (21.8%) died.

### CT quantification results of ILD

The median interval between the initial and follow-up CT scans was 44 months (IQR 33–56 months) (Table 1). On initial CT, the quantification results were 3.9% for

reticular opacity (range 0.1–24.9%), 1.0% for GGO (range 0.04–31.3%), 2.0% for honeycombing cysts (range 0.1–30.0%), 6.6% for fibrosis extent (range 0.2–42.7%), and 8.7% for total ILD extent (range 0.3–51.8%). On follow-up CT, the values were 4.9% for reticular opacity (range 0.04–41.1%), 0.9% for GGO (range 0.04–63.0%), 2.9% for honeycombing cysts (range 0.1–52.2%), 6.7% for fibrosis extent (range 0.2–76.3%), and 12.1% for total ILD extent (range 0.3–84.5%). Representative cases of CT quantification are presented in Figs. 2 and 3. The Spearman correlation coefficients between the CT quantification results and FVC decline are presented in Supplementary Table 2.

### Change in CT quantification according to FVC decline and visual assessment of ILD progression

According to groups categorized by the absolute decline of % predicted FVC, each ILD finding (reticular opacity 0.3%, 1.4%, and 3.3% in the  $<$  5%,  $\geq$  5% and  $<$  10%, and  $\geq$  10% groups, respectively; GGO –0.2%, 0.1%, and 0.4%, respectively; honeycombing cysts 0.3%, 0.9%, and 3.3%, respectively), fibrosis extent (0.9%, 2.9%, and 8.1%, respectively), and total ILD extents (0.9%, 4.2%, and 9.6%, respectively) were significantly different; the increases in CT findings were significantly higher in the group with a large decrease in FVC (all *p*-values  $<$  .001) (Table 2). The subgroup analysis (total ILD extent in initial CT,  $<$  10% and  $\geq$  10%) is presented in Supplementary Tables 3 and 4, respectively.

According to the groups of ILD progression assessed by visual assessments, each ILD finding (reticular opacity –0.1% versus 1.9% in the absence and presence of ILD progression, respectively, *p*  $<$  .001; GGO –0.1% versus –0.01%, *p* = .04; honeycombing cysts 0.1% versus 1.7%, *p*  $<$  .001), fibrosis extent (0.2% versus 4.9%, *p*  $<$  .001), and total ILD extent (0.2% versus 5.3%, *p*  $<$  .001) were significantly different and significant increases in CT findings were observed in the group with visually assessed ILD progression (Table 3).

### Cox regression analysis

Cox regression analysis results for all-cause mortality are described in Table 4. After adjustment for age, sex, and initial % predicted FVC, increases in quantification values for reticular opacity (hazard ratio [HR] 1.577, 95% confidence interval [CI] 1.04, 2.393, *p* = .03), honeycombing cysts (HR 1.739, 95% CI 1.142, 2.648, *p* = .01), fibrosis extent (HR 1.844, 95% CI 1.165, 2.919, *p* = .01), or total ILD extent (HR 2.484, 95% CI 1.565, 3.943, *p*  $<$  .001) were significant prognostic factors for a  $\geq$  5% absolute decline of FVC. However, the addition of a visual assessment of ILD progression (HR 1.159, 95% CI 0.693, 1.939, *p* = .57) or increase in GGO (HR 1.364, 95% CI 0.899, 2.07, *p* = .15) was not significant.

**Table 1** Baseline characteristics of the study sample and CT quantification values

| Variables   | N = 468                          |                 |
|---|----------------------------------|-----------------|
| Age (years)   | 64 ± 9.5                         |                 |
| Sex (male)  | 239 (51.1%)                      |                 |
| Clinical diagnosis  |                                  |                 |
| IPF   | 220 (47%)                        |                 |
| CTD-ILD   | 148 (31.6%)                      |                 |
| Idiopathic NSIP   | 22 (4.7%)                        |                 |
| Chronic hypersensitivity pneumonitis                              | 5 (1.1%)                         |                 |
| Cryptogenic organizing pneumonia                                  | 3 (0.6%)                         |                 |
| Desquamative interstitial pneumonia                               | 1 (0.2%)                         |                 |
| Unclassifiable ILD  | 69 (14.7%)                       |                 |
| Pulmonary function test   | Initial                          | Follow-up       |
| FVC (L)   | 2.7 ± 0.8                        | 2.5 ± 0.9       |
| % predicted FVC (%)   | 81.3 ± 18.1                      | 78.3 ± 20.5     |
| Absolute decline in predicted FVC < 5% during follow-up           | 292 (62.4%)                      |                 |
| Absolute decline in predicted FVC ≥ 5% but < 10% during follow-up | 78 (16.7%)                       |                 |
| Absolute decline in FVC predicted value ≥ 10% during follow-up    | 98 (20.9%)                       |                 |
| CT quantification results   | Initial                          | Follow-up       |
| Reticular opacity (%)   | 3.9 (0.1–24.9)                   | 4.9 (0.04–41.1) |
| GGO (%)   | 1.0 (0.04–31.3)                  | 0.9 (0.04–63.0) |
| Honeycombing cysts (%)  | 2.0 (0.1–30.0)                   | 2.9 (0.1–52.2)  |
| Fibrosis extent (%) <sup>a</sup>                                  | 6.6 (0.2–42.7)                   | 6.7 (0.2–76.3)  |
| Total ILD extent (%) <sup>b</sup>                                 | 8.7 (0.3–51.8)                   | 12.1 (0.3–84.5) |
| CT interval (months)  | 44 (interquartile range: 33, 56) |                 |
| ILD progression assessed by visual assessment                     | 266 (56.8%)                      |                 |
| Overall follow-up (months)  | 75 (interquartile range: 59, 84) |                 |
| Death   | 102 (21.8%)                      |                 |

IPF idiopathic pulmonary fibrosis, CTD-ILD connective tissue disease-associated interstitial lung disease, NSIP nonspecific interstitial pneumonia, ILD interstitial lung disease, FVC forced vital capacity, GGO ground-glass opacity

<sup>a</sup> Sum of reticular opacity and honeycombing cysts

<sup>b</sup> Sum of reticular opacity, GGO, and honeycombing cysts

For a ≥ 10% absolute decline of FVC, a visual assessment of ILD progression (HR 1.64, 95% CI 1.007, 2.67,  $p = .047$ ), increases in each ILD extent (reticular opacity, HR 1.85, 95% CI 1.21, 2.828,  $p = .01$ ; GGO, HR 2.04, 95% CI 1.358, 3.064,  $p < .001$ ; honeycombing cysts, HR 1.932, 95% CI 1.286, 2.903,  $p = .002$ ), fibrosis extent (HR 2.918, 95% CI 1.906, 4.468,  $p < .001$ ), and total ILD extent (HR 3.125, 95% CI 2.014, 4.849,  $p < .001$ ) were all significant prognostic factors.

The optimal thresholds for increases in ILD quantification are described in Supplementary Table 5. The results of Cox regression analyses in the subgroup analysis (total ILD extent on initial CT, < 10% and ≥ 10%) are described in Supplementary Table 6 and Supplementary Table 7, respectively.

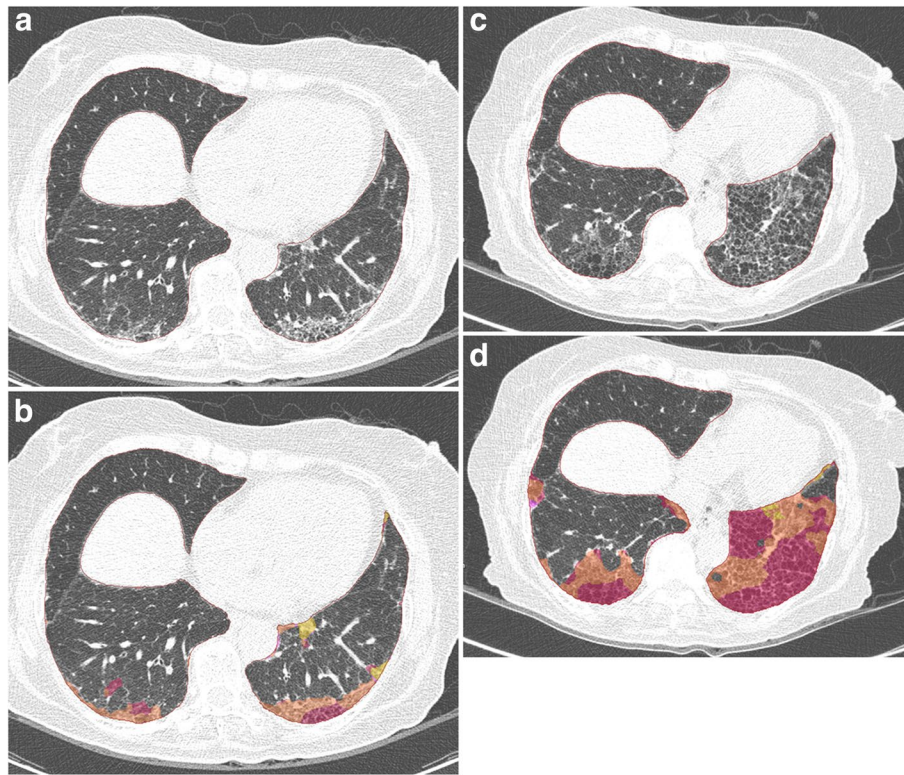
#### Added value of ILD CT quantification results for the outcome.

A model with age, sex, initial % predicted FVC, and absolute decline of FVC ≥ 5% had a C-index of 0.772 (Table 5). When total ILD extent was added to this

model, the model's C-index significantly increased to 0.795 ( $p = .04$ ). A model with age, sex, initial % predicted FVC, and absolute decline of FVC ≥ 10% had a C-index of 0.755. Adding fibrosis extent or total ILD extent significantly increased the C-index to 0.785 ( $p = .02$ ) and 0.787 ( $p = .01$ ), respectively. The corresponding results of the subgroup analysis (total ILD extent on initial CT, < 10% and ≥ 10%) are described in Supplementary Tables 8 and 9, respectively.

#### Discussion

This study revealed that larger FVC declines (≥ 10% and ≥ 5%, which are currently used for the diagnosis and prognosis of PF-ILD) and a visually assessed ILD progression were associated with significant increases in quantification results for each ILD finding, fibrosis extent, and total ILD extent (all  $p$ -values < .05). Cox regression analysis for all-cause mortality showed that increases in quantified ILD CT features were all significant prognostic



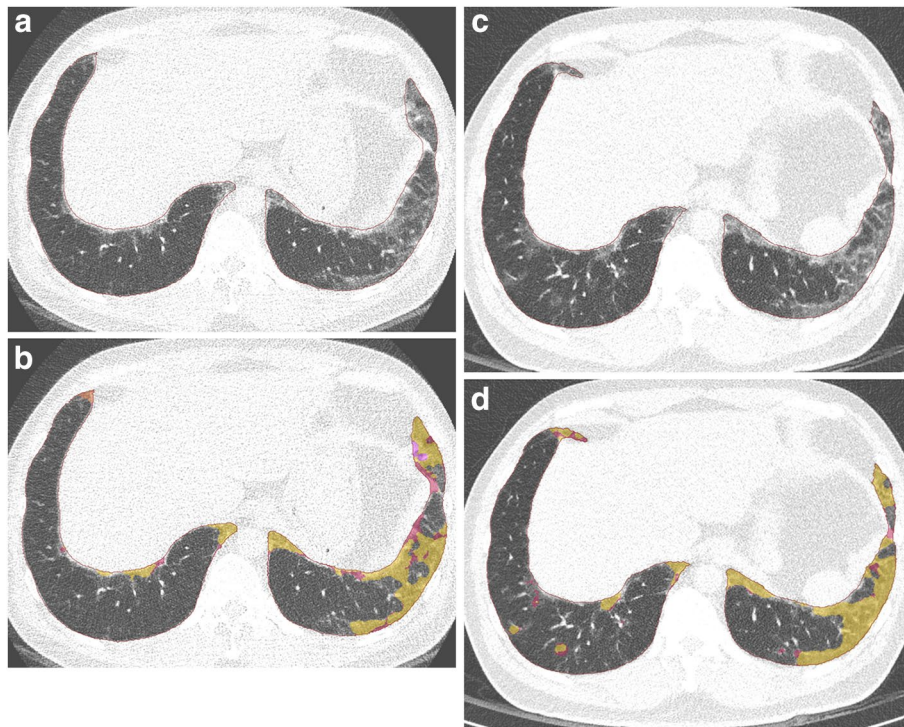
**Fig. 2** Representative case of progressive fibrosing interstitial lung disease (PF-ILD) from a 69-year-old woman with connective tissue disease. **a** Unenhanced axial CT images in the initial scan and **(b)** segmentation overlaying the initial CT images show the segmentation results of ground-glass opacity (GGO) (yellow), reticular opacity (orange), and honeycombing cysts (red). The fibrosis and total ILD extents were 8.6% (reticular opacity 6.3%, honeycombing cysts 2.3%) and 9.7% (GGO 1.1%, reticular opacity 6.3%, honeycombing cysts 2.3%), respectively. **c** Unenhanced axial CT images in a 59-month follow-up scan and **(d)** segmentation overlaying the follow-up CT images. The fibrosis and total ILD extents were 26.6% (reticular opacity 12.9%, honeycombing cysts 13.7%) and 27.0% (GGO 0.4%, reticular opacity 12.9%, honeycombing cysts 13.7%), respectively. The patient had an absolute decline of FVC of 25% during this period, and thoracic radiologists visually determined that these two CT series indicated ILD progression. Eighty-five months after the initial CT scan, the patient died

factors ( $p$ -values  $< .05$ ) even with adjustment for FVC decline. However, a visual assessment of progression was not a significant prognostic factor for a  $\geq 5\%$  decline of FVC. Additionally, the inclusion of fibrosis extent and total ILD extent increased the C-index compared to the models that only included age, sex, initial % predicted FVC, and FVC decline (for  $\geq 5\%$ , total ILD extent, not fibrosis extent, was significant). This finding highlights the added value of quantifying ILD CT features in predicting the outcomes of ILD.

Several studies have aimed to objectively quantify fibrosis findings on CT. Some studies evaluated the extent of fibrosis using a semi-quantitative scoring method on CT in patients with IPF and demonstrated a significant correlation with survival. Lynch et al [38] evaluated 315 patients with IPF from 58 institutions, and at least two radiologists assessed the extent and severity of reticular opacity, GGO, honeycombing, and attenuation on CT, and then scored them according to the percentage

of the affected lung. A higher fibrosis score, combining reticular opacity and honeycombing, was associated with a higher mortality rate. Furthermore, Sumikawa et al [39] evaluated CT scans of 98 patients with definite usual interstitial pneumonia (UIP) based on ATS/ERS criteria and scored the presence, extent, and distribution of reticular opacity, consolidation, honeycombing, and traction bronchiectasis in each lung zone. They calculated the percentage of the affected lung for each zone, scored them, and defined a fibrosis score, and found that traction bronchiectasis and fibrosis scores were associated with the prognosis.

However, it is noteworthy that these semi-quantitative methods still have limitations because they rely on expert's visual assessment [39]. Hence, there was a need for automated quantification of ILD CT features. Indeed, recent studies have used automated quantification systems to evaluate the degree of fibrosis. Park et al [31] investigated the correlation between the



**Fig. 3** Representative case of stable interstitial lung disease (ILD) from a 64-year-old woman diagnosed with nonspecific interstitial pneumonia. **a** Unenhanced axial CT images in the initial scan and **(b)** segmentation overlaying the initial CT images show the segmentation results of ground-glass opacity (GGO) (yellow), reticular opacity (orange), and honeycombing cysts (red). The fibrosis and total ILD extents were 2.1% (reticular opacity 1.6%, honeycombing cysts 0.5%) and 3.9% (GGO 1.8%, reticular opacity 1.6%, honeycombing cysts 0.5%), respectively. **c** Unenhanced axial CT images in the 60-month follow-up scan and **(d)** segmentation overlaying the follow-up CT images. The fibrosis and total ILD extents were 2.8% (reticular opacity 2.2%, honeycombing cysts 0.6%) and 4.6% (GGO 1.8%, reticular opacity 2.2%, honeycombing cysts 0.6%), respectively. The patient had no absolute decline in FVC (less than 0%, due to measurement variability) during this period, and thoracic radiologists visually determined that these two series showed ILD stability. The patient survived as of September 2022 (82 months later)

**Table 2** Comparison of interstitial lung disease (ILD) CT quantification results according to the absolute decline of % predicted forced vital capacity (FVC)

| Increase in CT finding            | Absolute decline of % predicted FVC |                         |                   | p-value* |
|-----------------------------------|-------------------------------------|-------------------------|-------------------|----------|
|                                   | < 5% (n = 292)                      | ≥ 5% and < 10% (n = 78) | ≥ 10% (n = 98)    |          |
| Reticular opacity (%)             | 0.3 (−12.5, 15.5)                   | 1.4 (−15.3, 15.6)       | 3.3 (−10.7, 39.5) | < .001   |
| GGO (%)                           | −0.2 (−17.5, 29.3)                  | 0.1 (−10.2, 52.2)       | 0.4 (−18.2, 33.4) | < .001   |
| Honeycombing cysts (%)            | 0.3 (−11.9, 40.1)                   | 0.9 (−10.9, 25.1)       | 3.3 (−4.7, 35.1)  | < .001   |
| Fibrosis extent (%) <sup>a</sup>  | 0.9 (−16.6, 42.9)                   | 2.9 (−15.0, 30.9)       | 8.1 (−6.9, 45.1)  | < .001   |
| Total ILD extent (%) <sup>b</sup> | 0.9 (−21.0, 63.0)                   | 4.2 (−16.0, 41.9)       | 9.6 (−11.4, 78.5) | < .001   |

Minimum and maximum values are shown in parentheses

Normality tests were performed using the Kolmogorov-Smirnov test

ILD interstitial lung disease, FVC forced vital capacity, GGO ground-glass opacity

\*The Kruskal-Wallis test was used for statistical comparisons

<sup>a</sup> Sum of reticular opacity and honeycombing cysts

<sup>b</sup> Sum of reticular opacity, GGO, and honeycombing cysts

**Table 3** Comparison of interstitial lung disease (ILD) CT quantification results according to the visual assessment for ILD progression by thoracic radiologists

| Increase in CT finding            | Absence of ILD progression (n = 202) | Presence of ILD progression (n = 266) | p-value* |
|-----------------------------------|--------------------------------------|---------------------------------------|----------|
| Reticular opacity (%)             | -0.1 (-15.3, 8.4)                    | 1.9 (-10.7, 39.5)                     | < .001   |
| GGO (%)                           | -0.1 (-17.5, 52.2)                   | -0.01 (-18.2, 33.4)                   | .04      |
| Honeycombing cysts (%)            | 0.1 (-11.9, 16.1)                    | 1.7 (-8.4, 40.1)                      | < .001   |
| Fibrosis extent (%) <sup>a</sup>  | 0.2 (-16.6, 12.7)                    | 4.9 (-6.9, 45.1)                      | < .001   |
| Total ILD extent (%) <sup>b</sup> | 0.2 (-21.0, 41.9)                    | 5.3 (-16.0, 78.5)                     | < .001   |

Minimum and maximum values are shown in the parentheses

Normality tests were performed using the Kolmogorov-Smirnov test

ILD interstitial lung disease, GGO ground-glass opacity

\*The Wilcoxon rank sum test was used for statistical comparisons

<sup>a</sup> Sum of reticular opacity and honeycombing cysts

<sup>b</sup> Sum of reticular opacity, GGO, and honeycombing cysts

**Table 4** Cox regression analysis for all-cause mortality with adjustment of age, sex, and initial % predicted forced vital capacity (FVC)

| Variables   | Hazard ratio         | p-value |
|---|----------------------|---------|
| Absolute decline of FVC ≥ 5%                              | 3.445 (2.171, 5.467) | < .001  |
| Visual assessment for ILD progression (reference: no)     | 1.159 (0.693, 1.939) | .57     |
| Absolute decline of FVC ≥ 5%                              | 3.120 (1.997, 4.876) | < .001  |
| Increase in reticular opacity (reference: no)             | 1.577 (1.040, 2.393) | .03     |
| Absolute decline of FVC ≥ 5%                              | 3.454 (2.253, 5.293) | < .001  |
| Increase in GGO (reference: no)                           | 1.364 (0.899, 2.070) | .15     |
| Absolute decline of FVC ≥ 5%                              | 3.307 (2.157, 5.072) | < .001  |
| Increase in honeycombing cysts (reference: no)            | 1.739 (1.142, 2.648) | .01     |
| Absolute decline of FVC ≥ 5%                              | 3.030 (1.951, 4.708) | < .001  |
| Increase in fibrosis extent (reference: no) <sup>a</sup>  | 1.844 (1.165, 2.919) | .01     |
| Absolute decline of FVC ≥ 5%                              | 2.624 (1.676, 4.109) | < .001  |
| Increase in total ILD extent (reference: no) <sup>b</sup> | 2.484 (1.565, 3.943) | < .001  |
| Absolute decline of FVC ≥ 10%                             | 2.212 (1.464, 3.343) | < .001  |
| Visual assessment for ILD progression (reference: no)     | 1.640 (1.007, 2.670) | .047    |
| Absolute decline of FVC ≥ 10%                             | 1.970 (1.278, 3.035) | .002    |
| Increase in reticular opacity (reference: no)             | 1.850 (1.210, 2.828) | 0.01    |
| Absolute decline of FVC ≥ 10%                             | 2.179 (1.449, 3.276) | < .001  |
| Increase in GGO (reference: no)                           | 2.040 (1.358, 3.064) | < .001  |
| Absolute decline of FVC ≥ 10%                             | 2.224 (1.483, 3.336) | < .001  |
| Increase in honeycombing cysts (reference: no)            | 1.932 (1.286, 2.903) | .002    |
| Absolute decline of FVC ≥ 10%                             | 1.722 (1.125, 2.636) | .01     |
| Increase in fibrosis extent (reference: no) <sup>a</sup>  | 2.918 (1.906, 4.468) | < .001  |
| Absolute decline of FVC ≥ 10%                             | 1.652 (1.08, 2.526)  | .02     |
| Increase in total ILD extent (reference: no) <sup>b</sup> | 3.125 (2.014, 4.849) | < .001  |

Lower and upper 95% confidence intervals are shown in the parentheses

FVC forced vital capacity, ILD interstitial lung disease, GGO ground-glass opacity

CT quantification results are treated as categorical values using optimal cut-offs values by Youden's index for absolute 5% and 10% decline of % predicted FVC

<sup>a</sup> Sum of reticular opacity and honeycombing cysts

<sup>b</sup> Sum of reticular opacity, GGO, and honeycombing cysts



**Table 5** Added value of CT quantification results to the change in absolute decline of % predicted forced vital capacity (FVC) for all-cause mortality

| Baseline model  | Add of CT quantification results          | C-index | Standard error | p-value |
|---|---|---------|----------------|---------|
| Age + sex + initial % predicted FVC + absolute decline of FVC $\geq$ 5% (C-index, 0.772)  | Visual assessment for ILD progression     | 0.772   | 0.003          | .99     |
|   | Increase in reticular opacity             | 0.784   | 0.008          | .14     |
|   | Increase in GGO                           | 0.776   | 0.005          | .43     |
|   | Increase in honeycombing cysts            | 0.779   | 0.006          | .32     |
|   | Increase in fibrosis extent <sup>a</sup>  | 0.786   | 0.008          | .09     |
|   | Increase in total ILD extent <sup>b</sup> | 0.795   | 0.011          | .04     |
| Age + sex + initial % predicted FVC + absolute decline of FVC $\geq$ 10% (C-index, 0.755) | Visual assessment for ILD progression     | 0.756   | 0.008          | .84     |
|   | Increase in reticular opacity             | 0.770   | 0.010          | .13     |
|   | Increase in GGO                           | 0.765   | 0.011          | .35     |
|   | Increase in honeycombing cysts            | 0.764   | 0.008          | .28     |
|   | Increase in fibrosis extent <sup>a</sup>  | 0.785   | 0.013          | .02     |
|   | Increase in total ILD extent <sup>b</sup> | 0.787   | 0.013          | .01     |

Uno's concordance statistic was used for the model's C-index

FVC forced vital capacity, ILD interstitial lung disease, GGO ground-glass opacity

<sup>a</sup> Sum of reticular opacity and honeycombing cysts

<sup>b</sup> Sum of reticular opacity, GGO, and honeycombing cysts

quantification of fibrosis range in CT scans of patients with IPF and FVC decline and found that only reticular opacity was a significant predictor of FVC decline of over 10% after 1 year. In another study [32], supervised classification models were used to quantify GGO, reticular opacity, and honeycombing in 57 patients with IPF. That study found that an increase in reticular opacity within 7 months was significantly associated with changes in FVC and DLco, and the total increases of reticular opacity, honeycombing, and GGO were associated with FVC changes. In a study on non-IPF ILDs, Goldin et al [33] reported that the quantification values of fibrosis, GGO, and honeycombing improved in 142 patients with scleroderma-associated ILD who were treated with cyclophosphamide or mycophenolate compared to placebo. Thus, several studies have attempted to quantify the extent of IPF or non-IPF ILDs using automated quantitative software. However, no studies to date have investigated changes in the ILD extent on CT scans based on deep learning technology, especially in relation to the FVC criteria of 5% and 10% for PF-ILD. Furthermore, no studies have examined the relationship between these factors and mortality rates. Therefore, this study is valuable insofar as it addressed this gap in the literature using a deep learning technique.

The limitations of this study include its retrospective nature, which introduces the possibility that quantitative ILD CT findings may be influenced by different CT vendor machines, and variations in CT technical parameters such as slice thickness, reconstruction protocols, and

the presence of contrast enhancement. These factors can result in heterogeneity and variations in the analysis of CT images. Additionally, the intervals between initial and follow-up CT scans were not consistently standardized. Future studies using consistent intervals between initial and follow-up CT might obtain more reliable and comparable results. Furthermore, this study utilized only one quantification software tool, and there is currently no evidence for the consistency of results when using other quantification tools. Further research should address this issue. Finally, we did not conduct direct comparative analyses between the deep learning-based CT quantification results and the visual assessment by radiologists. Since the visual assessment by radiologists is regarded as the reference standard for determining ILD progression in real-world clinical settings, we assessed the changes in ILD CT features according to this reference standard (Table 3). Nevertheless, it is important to acknowledge that this study has limitations in that we did not verify the added value of ILD quantification results to radiologists' assessments in determining ILD progression.

In conclusion, we demonstrated that patients with larger FVC declines, using the PF-ILD criteria of FVC decline  $\geq$  10% and  $\geq$  5%, exhibited a significant increase in quantified ILD CT findings using a deep learning technique. Additionally, quantified ILD features were identified as independent predictors of the prognosis of ILD. Consequently, the quantification of ILD CT features using deep learning might be valuable in the future, serving as an important method of defining and predicting the prognosis of PF-ILD.

**Abbreviations**

|         |  |
|---------|--|
| CI      | Confidence interval  |
| CTD-ILD | Connective tissue disease-associated interstitial lung disease |
| DLco    | Diffusing capacity of the lung for carbon monoxide             |
| FVC     | Forced vital capacity  |
| GGO     | Ground-glass opacity   |
| HR      | Hazard ratio   |
| HRCT    | High-resolution CT   |
| ILD     | Interstitial lung disease                                      |
| IQR     | Interquartile range  |
| IPF     | Idiopathic pulmonary fibrosis                                  |
| NSIP    | Nonspecific interstitial pneumonia                             |
| PF-ILD  | Progressive fibrosing interstitial lung disease                |
| PFT     | Pulmonary function test  |
| UIP     | Usual interstitial pneumonia                                   |

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1007/s00330-023-10483-9>.

**ESM 1** (PDF 310 kb)

**Acknowledgements**

The authors would like to acknowledge Andrew Dombrowski, Ph.D. (Compecs, Inc.) for his assistance in English editing.

**Funding**

The authors state that this work has not received any funding.

**Declarations****Guarantor**

The scientific guarantor of this publication is Jin Mo Goo.

**Conflict of interest**

The authors of this manuscript declare relationships with the following companies:

Activities not related to the present article—Jin Mo Goo reports research grants from LG Electronics and CoreLine Soft; Jong Hyuk Lee reports research grants from CoreLine Soft and consulting fee from Radisen.

Other authors (Seok Young Koh, Hyungin Park) have no conflicts of interest.

**Statistics and biometry**

One of the authors has significant statistical expertise (Jong Hyuk Lee).

**Informed consent**

Written informed consent was waived by the Institutional Review Board.

**Ethical approval**

The Institutional Review Board of Seoul National University Hospital approval was obtained (IRB No. H-2112-040-1279).

**Study subjects or cohorts overlap**

The study participants have not been reported before.

**Methodology**

- retrospective
- diagnostic or prognostic study
- performed at one institution

**Author details**

<sup>1</sup>Department of Radiology, Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul 03080, South Korea. <sup>2</sup>Department of Radiology, Seoul National University College of Medicine, 101, Daehak-ro, Jongno-gu, Seoul 03080, South Korea. <sup>3</sup>Institute of Radiation Medicine, Seoul National University Medical Research Center, 101, Daehak-ro, Jongno-gu, Seoul 03080,

South Korea. <sup>4</sup>Cancer Research Institute, Seoul National University, 101, Daehak-ro, Jongno-gu, Seoul 03080, South Korea.

**Received: 11 July 2023 Revised: 23 October 2023 Accepted: 27 October 2023**

**Published online: 12 December 2023**

**References**

1. Raghu G, Remy-Jardin M, Myers JL et al (2018) Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 198:e44–e68
2. Lynch DA, Sverzellati N, Travis WD et al (2018) Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med* 6:138–153
3. George PM, Patterson CM, Reed AK, Thillai M (2019) Lung transplantation for idiopathic pulmonary fibrosis. *Lancet Respir Med* 7:271–282
4. Kim HJ, Perlman D, Tomic R (2015) Natural history of idiopathic pulmonary fibrosis. *Respir Med* 109:661–670
5. George PM, Spagnolo P, Kreuter M et al (2020) Progressive fibrosing interstitial lung disease: clinical uncertainties, consensus recommendations, and research priorities. *Lancet Respir Med* 8:925–934
6. Wijsenbeek M, Kreuter M, Olson A et al (2019) Progressive fibrosing interstitial lung diseases: current practice in diagnosis and management. *Curr Med Res Opin* 35:2015–2024
7. Flaherty KR, Wells AU, Cottin V et al (2019) Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 381:1718–1727
8. Goldin JG (2018) Predicting outcome in idiopathic pulmonary fibrosis using automated computed tomography analysis. *Am J Respir Crit Care Med* 198:701–702
9. Richeldi L, Ryerson CJ, Lee JS et al (2012) Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. *Thorax* 67:407–411
10. du Bois RM, Weycker D, Albera C et al (2011) Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 184:1382–1389
11. Zappala CJ, Latsi PI, Nicholson AG et al (2010) Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. *Eur Respir J* 35:830–836
12. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK (2003) Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 168:538–542
13. Jegal Y, Kim DS, Shim TS et al (2005) Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 171:639–644
14. Latsi PI, du Bois RM, Nicholson AG et al (2003) Fibrotic idiopathic interstitial pneumonia: the prognostic value of longitudinal functional trends. *Am J Respir Crit Care Med* 168:531–537
15. Flaherty KR, Mumford JA, Murray S et al (2003) Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. *Am J Respir Crit Care Med* 168:543–548
16. Salisbury ML, Gu T, Murray S et al (2019) Hypersensitivity pneumonitis: radiologic phenotypes are associated with distinct survival time and pulmonary function trajectory. *Chest* 155:699–711
17. Walsh SL, Sverzellati N, Devaraj A, Wells AU, Hansell DM (2012) Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. *Eur Radiol* 22:1672–1679
18. King TE, Bradford WZ, Castro-Bernardini S et al (2014) A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 370:2083–2092
19. Maher TM, Corte TJ, Fischer A et al (2020) Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. *Lancet Respir Med* 8:147–157
20. Richeldi L, du Bois RM, Raghu G et al (2014) Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 370:2071–2082
21. Behr J, Neuser P, Prasse A et al (2017) Exploring efficacy and safety of oral Pirfenidone for progressive, non-IPF lung fibrosis (RELIEF) - a randomized,

- double-blind, placebo-controlled, parallel group, multi-center, phase II trial. *BMC Pulm Med* 17:122
22. Ehteshami Bejnordi B, Veta M, Johannes van Diest P et al (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318:2199–2210
  23. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118
  24. Gulshan V, Peng L, Coram M et al (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316:2402–2410
  25. Ting DSW, Cheung CY-L, Lim G et al (2017) Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318:2211–2223
  26. Lee SM, Seo JB, Yun J et al (2019) Deep learning applications in chest radiography and computed tomography: current state of the art. *J Thorac Imaging* 34:75
  27. González G, Ash SY, Vegas-Sánchez-Ferrero G et al (2018) Disease staging and prognosis in smokers using deep learning in chest computed tomography. *Am J Respir Crit Care Med* 197:193–203
  28. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284:574–582
  29. Cicero M, Bilbily A, Colak E et al (2017) Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol* 52:281–287
  30. Hwang EJ, Park S, Jin K-N et al (2019) Development and validation of a deep learning-based automated detection algorithm for major thoracic diseases on chest radiographs. *JAMA Netw Open* 2:e191095
  31. Park HJ, Lee SM, Song JW et al (2016) Texture-based automated quantitative assessment of regional patterns on initial CT in patients with idiopathic pulmonary fibrosis: relationship to decline in forced vital capacity. *AJR Am J Roentgenol* 207:976–983
  32. Kim HJ, Brown MS, Chong D et al (2015) Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months. *Acad Radiol* 22:70–80
  33. Goldin JG, Kim GHJ, Tseng C-H et al (2018) Longitudinal changes in quantitative interstitial lung disease on computed tomography after immunosuppression in the Scleroderma Lung Study II. *Ann Am Thorac Soc* 15:1286–1295
  34. Park B, Park H, Lee SM, Seo JB, Kim N (2019) Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks. *J Digit Imaging* 32(6):1019–1026
  35. Kim MS, Choe J, Hwang HJ et al (2022) Interstitial lung abnormalities (ILA) on routine chest CT: comparison of radiologists' visual evaluation and automated quantification. *Eur J Radiol* 157:110564
  36. Graham BL, Steenbruggen I, Miller MR et al (2019) Standardization of Spirometry 2019 Update. An official American Thoracic Society and European Respiratory Society technical statement. *Am J Respir Crit Care Med* 200:e70–e88
  37. Raghu G, Remy-Jardin M, Richeldi L et al (2022) Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 205:e18–e47
  38. Lynch DA, Godwin JD, Safrin S et al (2005) High-resolution computed tomography in idiopathic pulmonary fibrosis: diagnosis and prognosis. *Am J Respir Crit Care Med* 172:488–493
  39. Sumikawa H, Johkoh T, Colby TV et al (2008) Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med* 177:433–439

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.