Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs

Hirotaka Nishikiori [1], Koji Kuronuma[1], Kenichi Hirota[2], Naoya Yama[3], Tomohiro Suzuki[4], Maki Onodera[3], Koichi Onodera[3], Kimiyuki Ikeda[1], Yuki Mori[1], Yuichiro Asai[1], Yuzo Takagi[5], Seiwa Honda[4], Hirofumi Ohnishi[6], Masamitsu Hatakenaka[3], Hiroki Takahashi[1] and Hirofumi Chiba[1]

[1]Department of Respiratory Medicine and Allergology, Sapporo Medical University School of Medicine, Sapporo, Japan. [2]Department of Medical Information Planning, Sapporo Medical University Hospital, Sapporo, Japan. [3]Department of Diagnostic Radiology, Sapporo Medical University School of Medicine, Sapporo, Japan. [4]M3 Inc., Tokyo, Japan. [5]SEEDSUPPLY Inc., Kanagawa, Japan. [6]Department of Public Health, Sapporo Medical University School of Medicine, Sapporo, Japan.

Corresponding author: Koji Kuronuma (kuronumak@sapmed.ac.jp)

Shareable abstract (@ERSpublications)
A deep-learning algorithm was developed to detect fibrotic interstitial lung disease using chest radiographs. The algorithm's detection capability was noninferior to that of doctors, including pulmonologists and radiologists. https://bit.ly/3SAClW2

Abstract

Background Antifibrotic therapies are available to treat chronic fibrosing interstitial lung diseases (CF-ILDs), including idiopathic pulmonary fibrosis. Early use of these treatments is recommended to slow deterioration of respiratory function and to prevent acute exacerbation. However, identifying patients in the early stages of CF-ILD using chest radiographs is challenging. In this study, we developed and tested a deep-learning algorithm to detect CF-ILD using chest radiograph images.

Method From the image archive of Sapporo Medical University Hospital, 653 chest radiographs from 263 patients with CF-ILDs and 506 from 506 patients without CF-ILD were identified; 921 were used for deep learning and 238 were used for algorithm testing. The algorithm was designed to output a numerical score ranging from 0 to 1, representing the probability of CF-ILD. Using the testing dataset, the algorithm's capability to identify CF-ILD was compared with that of doctors. A second dataset, in which CF-ILD was confirmed using computed tomography images, was used to further evaluate the algorithm's performance.

Results The area under the receiver operating characteristic curve, which indicates the algorithm's detection capability, was 0.979. Using a score cut-off of 0.267, the sensitivity and specificity of detection were 0.896 and 1.000, respectively. These data showed that the algorithm's performance was noninferior to that of doctors, including pulmonologists and radiologists; performance was verified using the second dataset.

Conclusions We developed a deep-learning algorithm to detect CF-ILDs using chest radiograph images. The algorithm's detection capability was noninferior to that of doctors.

Introduction

Interstitial lung diseases (ILDs) are a heterogeneous group of distinctive lung disorders that are classified according to their cause [1]. Some ILDs present with fibrosis and may be intractable and resistant to anti-inflammatory and/or immunosuppressive agents. Among the disorders classified as chronic fibrosing ILD (CF-ILD), idiopathic pulmonary fibrosis (IPF) is most common [2]; IPF is treated using antifibrotic drugs [3, 4]. A recent clinical trial reported that the antifibrotic drug nintedanib was effective for treating progressive fibrosing ILDs (other than IPF) that are resistant to ILD type-specific conventional treatment [5]. Early use of antifibrotic drugs is critical, to slow the deterioration of respiratory function and to prevent acute exacerbation [4, 5].

There is a considerable diagnostic delay in patients with IPF. HOYER *et al.* [6] reported a median diagnostic delay of 2.1 years that was mainly attributable to time from onset of symptoms in patients until first

healthcare contact, time from contact with the first general practitioner until further referral and time from the first visit to a community hospital until ILD centre referral. These findings indicate that it is crucial for physicians who are not ILD specialists to suspect CF-ILDs on chest radiographs in the early stage of disease and to recommend patients to referral centres when appropriate. However, early-stage identification is challenging.

Computer-aided detection/diagnosis (CAD) supports the detection and/or diagnosis of abnormalities and/or diseases by identifying the presence and location of lesions on medical images. Over the past 15 years, CAD algorithms able to detect lung nodules [7], pulmonary tuberculosis [8] and other pulmonary diseases [9, 10] using chest radiographs have been developed using the deep-learning method. In this study, we developed and tested a deep-learning algorithm to help doctors detect CF-ILDs using chest radiographs of patients with computed tomography (CT) image-confirmed signs of fibrosis.

Material and methods

Study design

This retrospective study was conducted using patient data collected at the Sapporo Medical University Hospital (Sapporo, Japan) to develop and test a deep-learning algorithm using patient chest radiographs to support the detection of CF-ILDs. This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the institutional review board of the Sapporo Medical University Hospital (approval number 322-56, 23 June 2020).

Datasets

The first dataset included 263 patients with CF-ILD and 506 patients without CF-ILD who visited Sapporo Medical University Hospital between 1 January 2003 and 30 November 2018. The specific CF-ILD diseases represented among these patients included IPF, nonspecific interstitial pneumonia, unclassifiable idiopathic interstitial pneumonia, idiopathic pleuroparenchymal fibroelastosis, fibrotic hypersensitivity pneumonitis, rheumatoid arthritis-associated ILD, collagen vascular diseases and antineutrophil cytoplasmic antibody-related ILD. CF-ILDs were diagnosed by multidisciplinary discussion based on diagnostic guidelines for each disease [11–13]. Between one and three chest radiographs from each patient with CF-ILD were used (653 images in total: 518 in the learning dataset (combined training and validation datasets) and 135 in the testing dataset). For all CF-ILD patients, signs of fibrosis were confirmed using a CT image (for some patients, chest radiograph and CT images were taken on different days). Control images included 96 chest radiographs from patients diagnosed with other pulmonary diseases (learning dataset n=79; testing dataset n=17) and 410 from patients without lung disease (learning dataset n=324; testing dataset n=86) (supplementary figure S1).

The proportion of patients with CF-ILDs who had chest radiographs in the first dataset (over half of chest radiographs in this study) was not comparable with the prevalence in the clinical setting. Therefore, a second dataset was assembled. For this confirmatory dataset, we identified consecutive patients who visited Sapporo Medical University Hospital between 1 January and 31 December 2019, and who had available chest radiograph and chest CT images (taken on the same day). Patients aged <19 years; those who had chest radiographs not taken from the posterior-anterior view or in the standing position; and those who overlapped with the first dataset were excluded. A total of 1280 chest radiographs of 1280 patients were included in the confirmatory dataset.

Development of a deep-learning method for CF-ILD detection

The deep convolutional neural network (DCNN) was utilised to develop the algorithm. DenseNet121 [14] was used for the feature extraction layer and the WILDCAT [15] network was used for the discriminant layer. A PyTorch [16] framework with an NVIDIA GeForce 1080 graphics processing unit (NVIDIA, Santa Clara, CA, USA) was used for DCNN implementation. The learning chest radiograph dataset was separated into training and validation datasets using the k-fold cross validation method (k=4). Patient identification number was used to separate chest radiographs to ensure that multiple chest radiographs from the same patient were not distributed across the datasets. Before training, image contrasts were normalised to have an intensity range of −1 to 1 and resized to 876×876 pixels. During training, image augmentation processes, such as scaling and rotation, were also performed. Training was performed using focal loss functions [17] (parameters $\alpha=0.25$ and $\gamma=2.0$) with stochastic gradient descent optimisation (learning rate parameter $\alpha=0.01$). Training was done in two stages: first, only the weight of the discriminant layer was trained, and then the overall weight was trained.

Testing the detecting capability of the algorithm

Using the testing dataset, the algorithm was used to output a numerical value from 0 to 1 (probability of CF-ILD score) for each chest radiograph. A receiver operating characteristic (ROC) curve (which is a plot of the sensitivity and the false positive rate (1 − specificity) for every score cut-off) was created and the area under the curve (AUC) was calculated. The optimal threshold was determined using the Youden index; sensitivity and specificity were calculated.

Comparing the detecting capabilities of the algorithm versus doctors

Using the testing dataset, the detecting capability of the algorithm was compared with that of doctors (radiologists with >5 years' experience, n=5; pulmonologists with >10 years' experience, n=8). A high-resolution liquid crystal display monitor was used for image interpretation. The distance from the doctor's eyes to the monitor was set to ∼50 cm and the reading time was set to ∼5 s per image.

Performance testing the algorithm using the confirmatory dataset

Using the second dataset, the algorithm was used to interpret chest radiographs and output a score. For confirmation, one of four pairs of radiologists and pulmonologists (both >10 years' experience) interpreted each chest CT image and classified it as follows: no abnormal findings, CF-ILDs suspected/diagnosed or other abnormal findings. If the interpreters recorded differing classifications for the same image, the two doctors would discuss until they reached consensus. A ROC curve was created using CT data (first, CF-ILD-positive versus CF-ILD-negative; second, CF-ILD-positive versus no abnormal findings) and the AUC was calculated.

Extent of fibrosis on chest CT in the confirmatory dataset

Radiologists and pulmonologists classified CF-ILD suspected/diagnosed CT images into five grades according to the extent of the fibrotic area (grade 0: 0%; grade 1: >0–10%; grade 2: >10–25%; grade 3: >25–50%; grade 4: >50%; examples are shown in supplementary figure S2) at three different levels (upper level: above the aortic arch; middle level: 2 cm below the tracheal bifurcation; lower level: above the diaphragm) of each side (right and left). We calculated the average of fibrosis grade points for each level and for the three levels combined (formulae shown in supplementary table S1). Next, we created ROC curves (CF-ILD-positive versus no abnormal findings) according to the different extents of fibrosis (average of grade points: $\leqslant 1$, >1–2, >2–3, >3) and calculated the AUC of each ROC curve.

Statistical analysis

The sensitivity and specificity of detection in the first dataset were compared between the algorithm and doctors using the Wilcoxon signed-rank test. SPSS Statistics 27 (IBM) was used for statistical analysis.

Results

Characteristics of patients included in the first dataset

Patients with CF-ILD in the first dataset had a mean±SD age of 69.8±8.8 years (learning dataset 70.5±8.0 years; testing dataset 67.0±11.1 years); 33.8% of patients (learning dataset 38.6%; testing dataset

TABLE 1 Number of patients with chronic fibrosing interstitial lung disease (CF-ILD) and number of chest radiograph images included in the first dataset according to disease type and inclusion in learning or testing dataset

	Learning dataset	Testing dataset
CF-ILD	210 (518)	53 (135)
IPF	140 (342)	37 (92)
Fibrotic NSIP	24 (62)	8 (21)
Unclassifiable IIP	15 (39)	5 (14)
PPFE	6 (15)	
Fibrotic HP	2 (5)	
CVD-ILD	12 (27)	2 (5)
RA-ILD	8 (20)	
ANCA-related ILD	3 (8)	1 (3)

Data are presented as n patients (n chest radiograph images). IPF: idiopathic pulmonary fibrosis; NSIP: nonspecific interstitial pneumonia; IIP: idiopathic interstitial pneumonia; PPFE: idiopathic pleuroparenchymal fibroelastosis; HP: hypersensitivity pneumonitis; CVD-ILD: collagen vascular disease-associated interstitial lung disease; RA-ILD: rheumatoid arthritis-associated interstitial lung disease; ANCA: antineutrophil cytoplasmic antibodies.
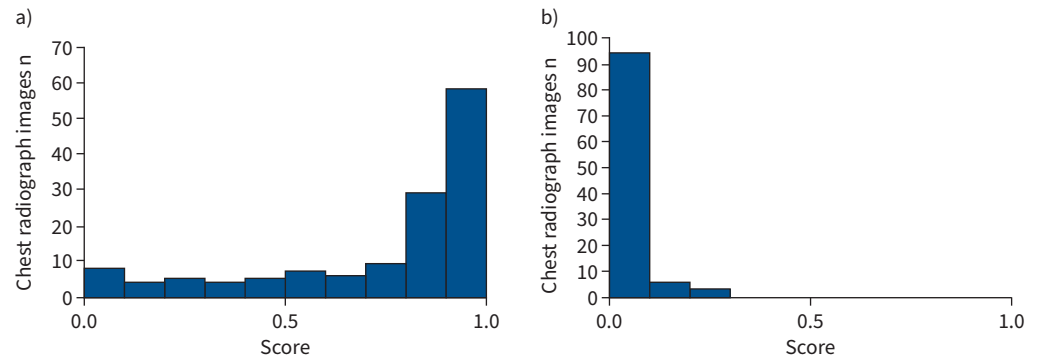
FIGURE 1 Distribution of the number of chest radiograph images according to the algorithm-assigned score among a) chronic fibrosing interstitial lung disease (CF-ILD)-positive chest radiographs and b) CF-ILD-negative chest radiographs included in the testing dataset.

17.0%) were female. In the control dataset, the mean±SD age was 58.7±14.7 years (learning dataset 58.5±14.5 years; testing dataset 59.6±15.3 years); 71.5% of patients (learning dataset 74.2%; testing dataset 61.2%) were female. CF-ILD diagnosis is shown in table 1. Lung diseases diagnosed in patients who were included in the control dataset are shown in supplementary table S2.

Detection capability of the algorithm

The distribution of chest radiograph images included in the testing dataset according to the algorithm-assigned score for those who were confirmed to be CF-ILD-positive or CF-ILD-negative is shown in figure 1a and b, respectively. Figure 2 shows the ROC curve, which represents the detecting capability of the algorithm. The AUC was 0.979 (95% CI 0.961–0.998). Using a score cut-off of 0.267, the sensitivity and specificity of the algorithm were 0.896 and 1.000, respectively.

Performance of the algorithm versus doctors

Among the 13 doctors who interpreted chest radiographs, the median (range) sensitivity and specificity were 0.837 (0.467–0.904) and 0.990 (0.864–1.000), respectively. The respective median (range) sensitivity
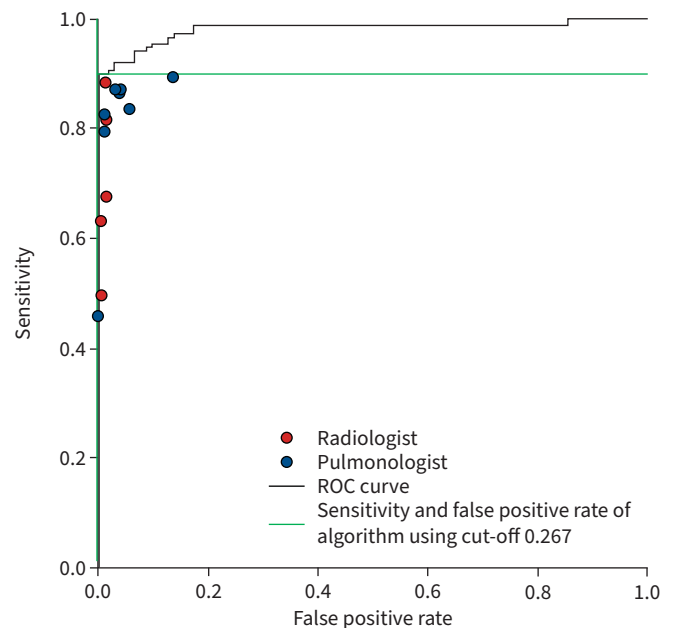

FIGURE 2 Comparison of the sensitivity and false positive rate of the algorithm and doctors using images from the testing dataset. ROC: receiver operating characteristic.

and specificity were 0.681 (0.496–0.889) and 0.990 (0.990–1.000) among radiologists and 0.859 (0.467–0.904) and 0.966 (0.864–1.000) among pulmonologists. These results are plotted in figure 2. All data points of doctors were on or below the ROC curve representing the detection capability of the algorithm. These findings demonstrate that the performance of the algorithm was noninferior to that of doctors. Furthermore, analysis using the Wilcoxon signed-rank test revealed that the algorithm was superior for both sensitivity and specificity (p<0.01 for both). Figure 3 shows representative paired chest radiograph and CT images from the testing dataset.

Testing the algorithm's performance using the second dataset

Of the 1280 chest radiographs that were included in the second dataset, 367 were from patients who had diagnosed or suspected CF-ILD (including patients in whom CF-ILD and other lung abnormalities overlapped), 596 were from patients who had abnormal findings other than CF-ILD (excluding patients in whom CF-ILD and other lung abnormalities overlapped) and 317 were from patients who had no abnormal findings on their chest CT images. The mean±SD age of patients was 66.5±13.6 years (patients with diagnosed or suspected CF-ILD 69.8±10.4 years; patients with abnormal findings other than CF-ILD 67.1±13.8 years; patients without abnormal findings 61.6±14.9 years); 47.7% of patients (patients with diagnosed or suspected CF-ILD 41.7%; patients with abnormal findings other than CF-ILD 45.6%; patients without abnormal findings 58.4%) were female. The distributions of chest radiograph images according to the algorithm-assigned score for those diagnosed with or suspected of having CF-ILD, those with abnormal CT findings other than CF-ILD and those with no abnormal CT findings are shown in figure 4a,b, and c, respectively. The ROC curve based on the findings reported using the CT images for CF-ILD-positive *versus* CF-ILD-negative patients is shown in figure 5a and that for CF-ILD-positive
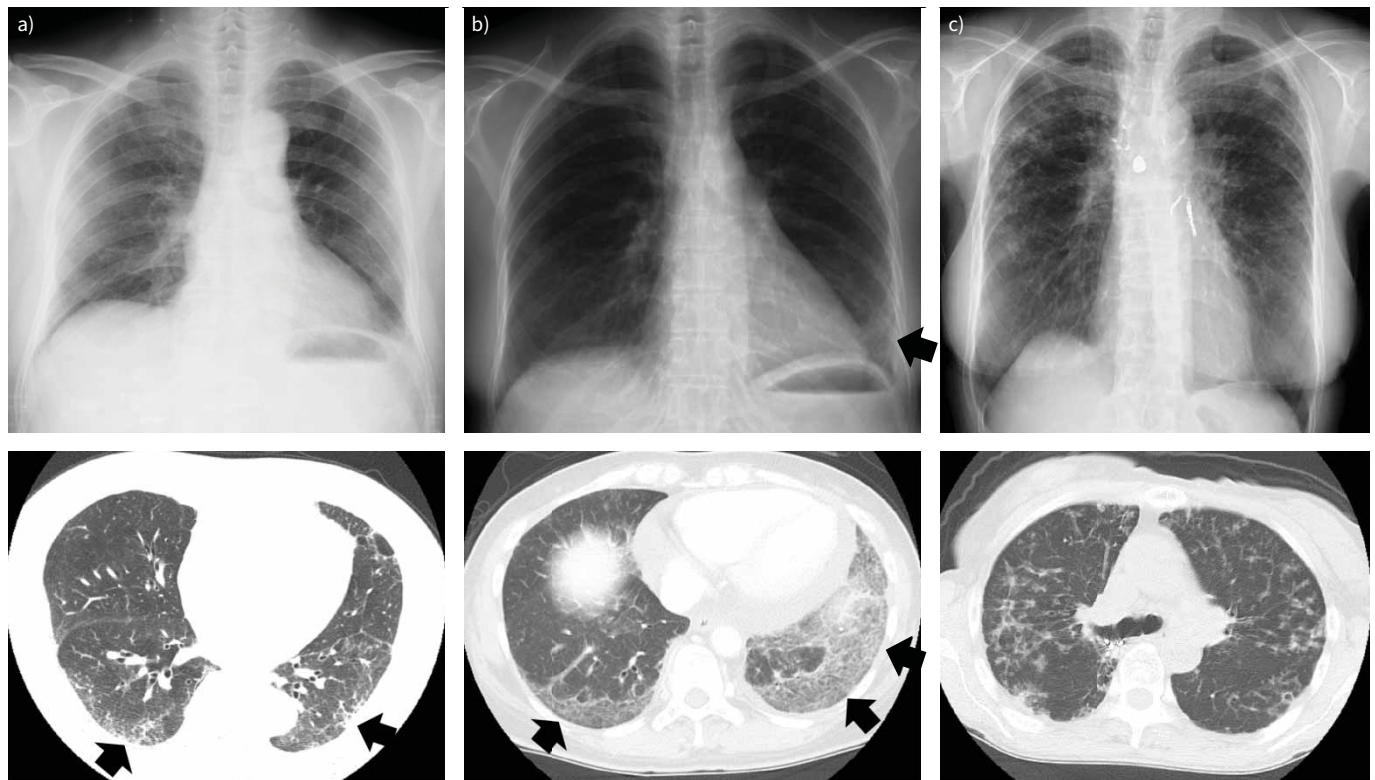

FIGURE 3 Representative paired chest radiograph and computed tomography (CT) images from patients in the testing dataset. a) A male in his seventies had a chest radiograph that appeared normal, but his chest CT showed subtle fibrosis in the dorsal subpleural area. He was diagnosed with idiopathic pulmonary fibrosis (IPF) having a probable usual interstitial pneumonia pattern according to the 2018 American Thoracic Society, European Respiratory Society, Japanese Respiratory Society and Latin American Thoracic Society clinical practice guideline for IPF [11]. His chest radiograph received a score of 0.853 from the algorithm. b) A female in her forties with collagen vascular disease-associated interstitial lung disease had chest radiograph and chest CT images that showed fibrosis in the lower subpleural area. Her chest radiograph received a score of 0.634 from the algorithm. c) The chest radiograph of this female in her seventies who was diagnosed with nontuberculous mycobacteriosis received a score of 0.260 from the algorithm.
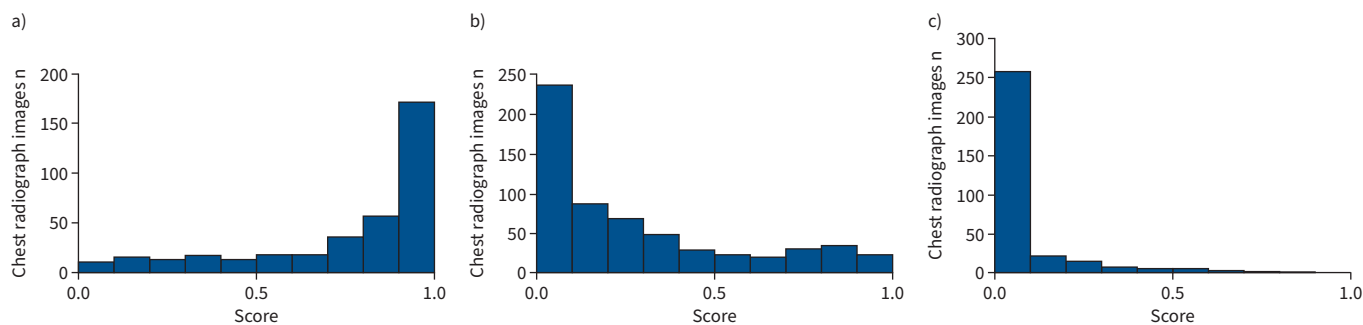
FIGURE 4 Distribution of the number of chest radiograph images according to algorithm score among patients in the second dataset a) with suspected/diagnosed chronic fibrosing interstitial lung disease (CF-ILD), b) who had abnormal computed tomography findings other than CF-ILD or c) who had no abnormal findings.

patients *versus* those with no abnormal findings is shown in figure 5b. The AUC was 0.910 (95% CI 0.893–0.926) for the CF-ILD-positive *versus* CF-ILD-negative ROC and 0.970 (95% CI 0.960–0.981) for the CF-ILD-positive patients *versus* those with no abnormal findings ROC. Figure 6 shows representative paired chest radiograph and CT images that were scored as false positive by the algorithm.

Algorithm performance by the extent of fibrosis

Supplementary table S3 shows the number of CF-ILD-positive CT images by the fibrosis grade for each level and for the three levels combined. Figure 7 shows the ROC curves by the extent of fibrosis on the lower level and the three levels combined among CF-ILD-positive patients *versus* those with no abnormal findings. As fibrosis grade decreased, AUC decreased; even with an average fibrosis grade point of $\leqslant 1$ (fibrosis-occupied area 0–10%), the AUC was 0.947 and 0.949 for the lower level and the three levels combined, respectively.

Discussion

Recently, a CAD system has been developed for chest images using the deep-learning method. In the field of ILD, research has been focused on the accurate diagnosis of ILDs using chest CT images; to our
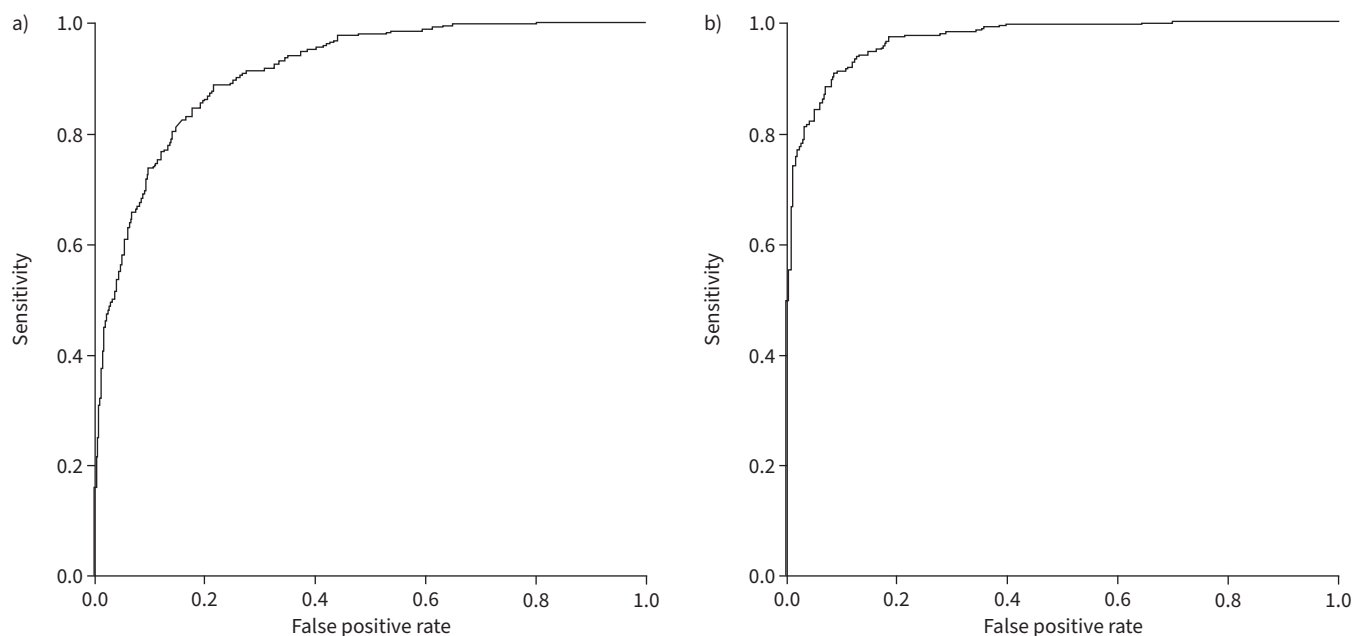

FIGURE 5 Receiver operating characteristic curves representing the detecting capability of the algorithm using images from the second dataset for a) chronic fibrosing interstitial lung disease (CF-ILD)-positive *versus* CF-ILD-negative patients and b) patients who were CF-ILD-positive *versus* those with no abnormal findings.
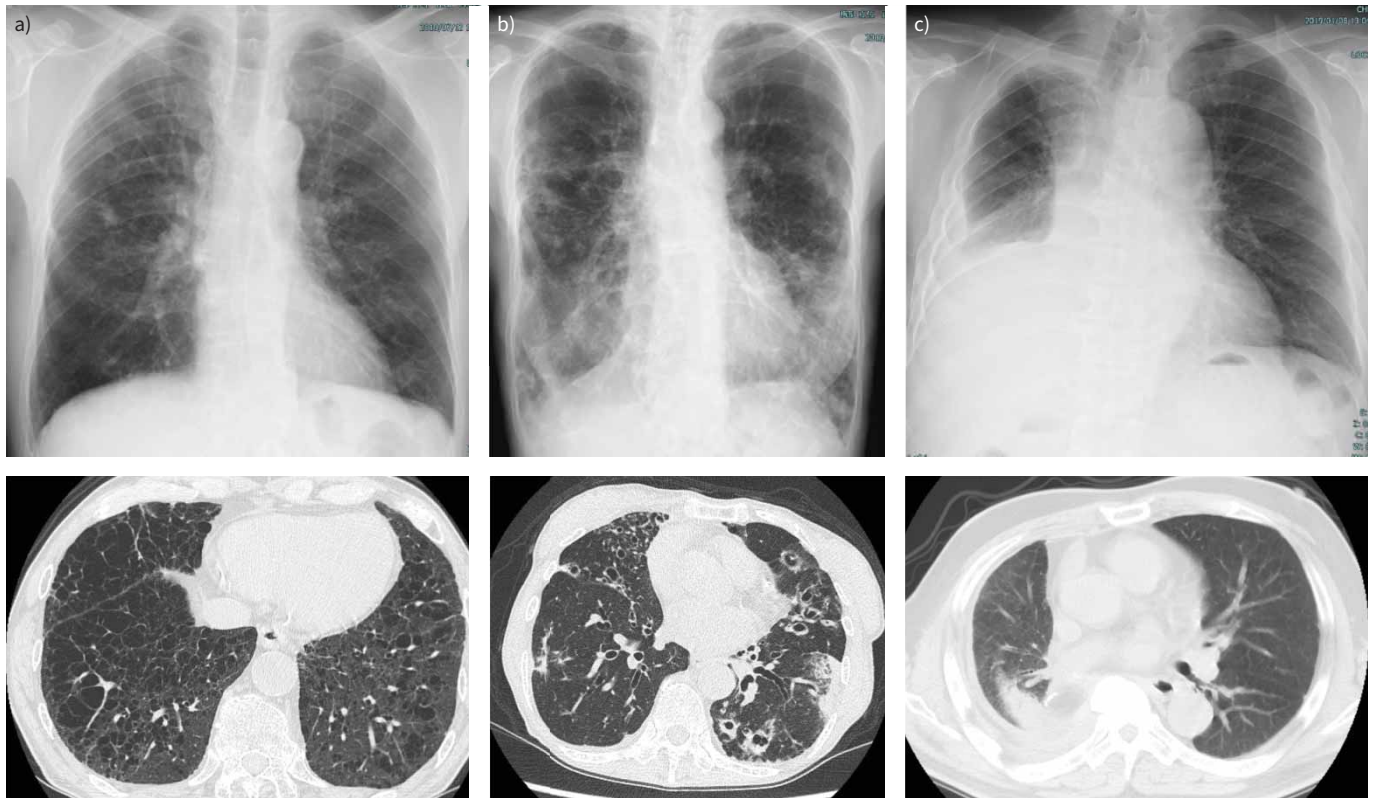
FIGURE 6 Representative paired chest radiograph and chronic fibrosing interstitial lung disease (CF-ILD)-negative computed tomography (CT) images that received a high score by the algorithm (second dataset). a) A male in his seventies with pulmonary emphysema. His chest CT showed many pulmonary cysts in the lower dorsal area, but without fibrosis. His chest radiograph received a score of 0.983 from the algorithm. b) A male in his sixties who was diagnosed with nontuberculous mycobacteriosis. His chest CT showed multiple nodules in the distal lung area and bronchiectasis in the lower lung area. His chest radiograph received a score of 0.810 from the algorithm. c) A male in his seventies who had undergone a right upper lobectomy due to lung cancer; his cancer had relapsed. His chest radiograph received a score of 0.929 from the algorithm.

knowledge, there are no notable reports on diagnosis using chest radiographs. WALSH et al. [18] developed an algorithm to classify chest CT images according to ILD type. They reported that their algorithm and a radiologist (majority opinion) similarly classified ILDs; in terms of accuracy, the algorithm outperformed over half of thoracic radiologists [18]. Although an ILD diagnosis requires discussion within a multidisciplinary team [12, 19], there are few facilities in which clinicians, radiologists and pathologists who are specialised in ILDs are available. Therefore, the development of CAD using deep-learning algorithms that can replace multidisciplinary discussion would be greatly beneficial for ILD diagnosis.

Early detection of ILDs is important. It has been reported that the antifibrotic drugs pirfenidone and nintedanib reduce the annual decline of forced vital capacity in patients with progressive fibrosing ILD (PF-ILDs), including IPF [3–5]. Furthermore, nintedanib is reported to prolong the time to the first acute exacerbation in patients with IPF and other PF-ILDs [4, 5]. However, these drugs have an insufficient effect on the improvement of respiratory function in most patients. Identifying patients with ILDs who are in the early disease stage and initiating treatment at the appropriate time can contribute to the long-term preservation of respiratory function and prevention of fatal acute exacerbations. However, the identification of early-stage ILD is difficult. Difficulty in the detection and diagnosis of ILDs leads to misdiagnosis and delays in accessing subspeciality care. LAMAS et al. [20] reported that the median time from the onset of dyspnoea to the initial evaluation at a tertiary care centre in patients with IPF was 2.2 years; delayed access was associated with a higher mortality rate. HOYER et al. [6] reported that diagnostic delays were mainly attributable to the time it took for a patient to be seen by a general practitioner, referred to a community hospital and, finally, referred to an ILD centre. If CAD can improve the detectability of ILDs on chest radiographs and provide nonspecialists with this information, diagnostic delays can be shortened, which may ultimately lead to an improvement in the prognosis of patients with ILDs.
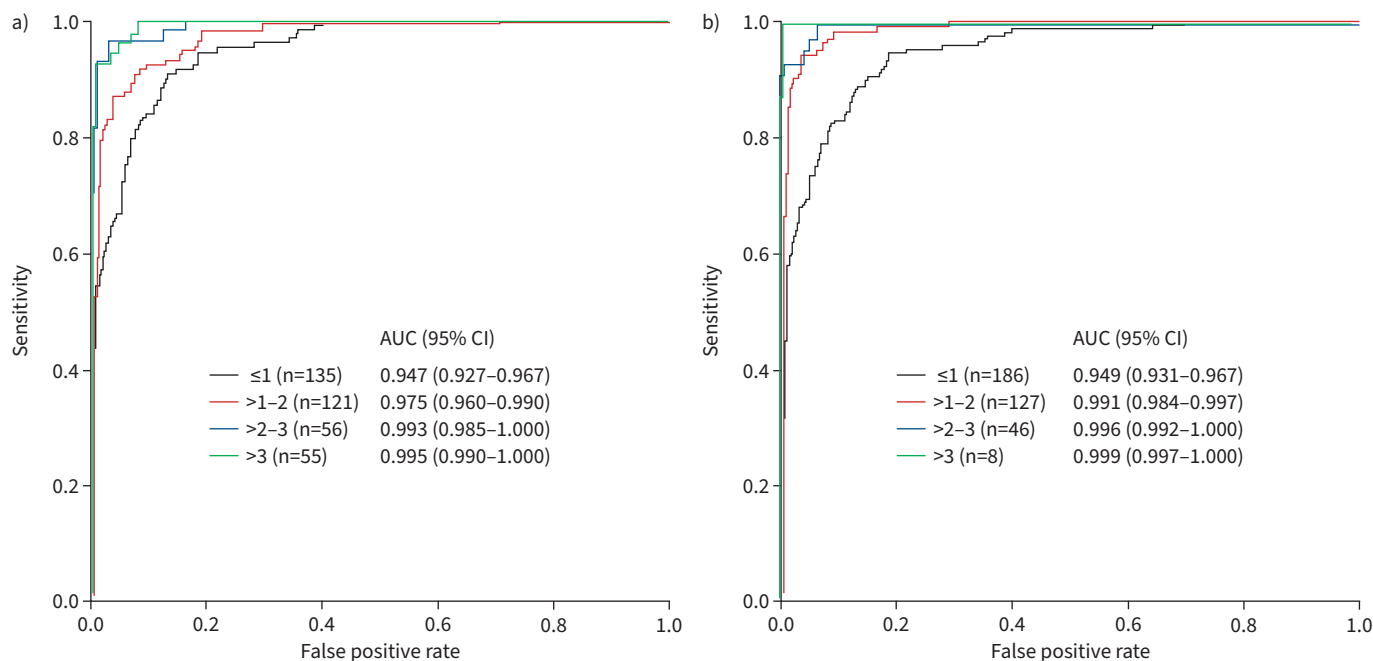
FIGURE 7 Receiver operating characteristic curves representing the detecting capability of the algorithm by the extent of fibrosis on a) the lower level and b) the three levels combined on the corresponding computed tomography (CT) images among chronic fibrosing interstitial lung disease (CF-ILD)-positive patients *versus* those without abnormal findings. Numbers of CF-ILD-positive images for each fibrosis grade point are shown. AUC: area under the curve.

CAD has the potential to detect very subtle abnormalities that the human eye cannot find. As shown in figure 3a, the algorithm appears to be able to detect a shadow hidden behind the diaphragm on the chest radiograph. Over half of the chest radiographs in the second dataset showed signs of fibrosis in 0–10% of the lung area on corresponding CT images. Even for those chest radiographs, the AUC, which represents the algorithm's performance, was 0.949. Thus, these findings indicate that the algorithm is capable of detecting limited diseases.

We excluded nonfibrosing ILDs, such as cryptogenic organising pneumonia or nonfibrotic hypersensitivity pneumonitis, from the target of the algorithm because their clinical course and response to immunosuppressive therapy are completely different from those of CF-ILDs. Our algorithm had an AUC of 0.979 for the testing dataset and 0.970 for the second dataset (CF-ILD-positive *versus* no abnormal findings). In the second dataset, the AUC was smaller for CF-ILD-positive *versus* CF-ILD-negative (0.910). The causal findings of false positives were mostly other lung diseases, such as pleural effusion, pulmonary emphysema and bronchiectasis. This may have been the case because there was a limited number of chest radiographs from patients with other lung diseases (*i.e.* disease controls) in the learning dataset. Additional disease control images are needed to train the algorithm to distinguish CF-ILDs from other lung abnormalities.

Our study had some limitations. First, the number of images used in the learning dataset was relatively small. Generally, a larger learning dataset should result in better algorithm performance. Accumulated chest radiographs from patients who were accurately diagnosed with ILDs, such as those obtained from a registry trial, would result in a smarter algorithm. Second, the present study showed that our algorithm has a high capacity to detect CF-ILDs; however, further real-world studies are needed to investigate how the use of this CAD algorithm may contribute to the detection of CF-ILD by nonspecialists compared with those who do not use CAD. Third, there was chest radiograph selection bias in both the first and second datasets. In the first dataset, we identified the chest radiographs of patients with pulmonary diseases other than CF-ILDs from an enormous set of chest radiographs from patients who visited our hospital. Therefore, it was possible that we unwittingly selected typical images for each pulmonary disease as disease control images. In addition, the prevalence of CF-ILD was different from that in the real world. For these reasons, we prepared the second dataset from consecutive patients who had chest radiographs and chest CTs taken on the same day. Using this method, 29% of all patients had CF-ILDs and 47% had other

lung abnormalities. These proportions may be completely different from those seen by general practitioners in the real-world clinical setting. Finally, both the learning and testing of this algorithm were conducted at a single facility. It is uncertain whether the same performance would be observed with chest radiographs taken under different conditions and using different equipment. A clinical trial further validating our CAD algorithm in the real-world setting is warranted to investigate these issues.

In conclusion, we developed a deep-learning algorithm to detect CF-ILDs using chest radiograph images. The detection capability of the algorithm was noninferior to that of doctors.

References

1   Antoine M, Mlika M. Interstitial lung disease. *In*: StatPearls. Treasure Island, StatPearls Publishing, 2020. Available from: www.ncbi.nlm.nih.gov/books/NBK541084/

2   Sgalla G, Iovene B, Calvello M, *et al.* Idiopathic pulmonary fibrosis: pathogenesis and management. *Respir Res* 2018; 19: 32.

3   King TE Jr, Bradford WZ, Castro-Bernardini S, *et al.* A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2083–2092.

4   Richeldi L, du Bois RM, Raghu G, *et al.* Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 2014; 370: 2071–2082.

5   Flaherty KR, Wells AU, Cottin V, *et al.* Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; 381: 1718–1727.

6   Hoyer N, Prior TS, Bendstrup E, *et al.* Risk factors for diagnostic delay in idiopathic pulmonary fibrosis. *Respir Res* 2019; 20: 103.

7   Schilham AM, van Ginneken B, Loog M. A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. *Med Image Anal* 2006; 10: 247–258.

8   Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284: 574–582.

9   Parveen NR, Sathik MM. Detection of pneumonia in chest X-ray images. *J Xray Sci Technol* 2011; 19: 423–428.

10 Cicero M, Bilbily A, Colak E, *et al.* Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol* 2017; 52: 281–287.

11 Raghu G, Remy-Jardin M, Myers JL, *et al.* Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198: e44–e68.

12 Travis WD, Costabel U, Hansell DM, *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.

13   Aletaha D, Neogi T, Silman AJ, *et al.* 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010; 62: 2569–2581.

14   Huang G, Liu Z, Pleiss G, *et al.* Convolutional networks with dense connectivity. *IEEE Trans Pattern Anal Mach Intell* 2022; 44: 8704–8716.

15   Durand T, Mordan T, Thome N, *et al.* WILDCAT: weakly supervised learning of deep convNets for image classification, pointwise localization and segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2017; pp. 642–651.

16   Paszke A, Gross S, Massa F, *et al.* PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019; 32: 8024–8035.

17   Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *2017 IEEE Conference on Computer Vision* 2017; pp. 2999–3007.

18   Walsh SLF, Calandriello L, Silva M, *et al.* Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med* 2018; 6: 837–845.

19   Flaherty KR, King TE Jr, Raghu G, *et al.* Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 2004; 170: 904–910.

20   Lamas DJ, Kawut SM, Bagiella E, *et al.* Delayed access and survival in idiopathic pulmonary fibrosis: a cohort study. *Am J Respir Crit Care Med* 2011; 184: 842–847.