



A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases

Taiki Furukawa^{1,2,3}  | Shintaro Oyama^{2,3} | Hideo Yokota^{2,4} | Yasuhiro Kondoh⁵  |
Kensuke Kataoka⁵ | Takeshi Johkoh⁶ | Junya Fukuoka⁷ | Naozumi Hashimoto¹ |
Koiji Sakamoto¹ | Yoshimune Shiratori³ | Yoshinori Hasegawa⁸

¹Department of Respiratory Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan

²Image Processing Research Team, RIKEN Center for Advanced Photonics, Wako, Japan

³Medical IT Center, Nagoya University Hospital, Nagoya, Japan

⁴Advanced Data Science Project, Information R&D and Strategy Headquarters, RIKEN, Wako, Japan

⁵Department of Respiratory Medicine and Allergy, Tosei General Hospital, Seto, Japan

⁶Department of Radiology, Kansai Rosai Hospital, Amagasaki, Japan

⁷Department of Pathology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

⁸Nagoya Medical Center, National Hospitalization Organization, Nagoya, Japan

Correspondence

Taiki Furukawa
Email: tfuru@med.nagoya-u.ac.jp

Funding information

JSPS KAKENHI, Grant/Award Number: JP19110253; The Hori Science and Arts Foundation; The Japanese Respiratory Foundation

Associate Editor: Michael Keane; **Senior Editor:** Yuben Moodley

Abstract

Background and objective: Idiopathic pulmonary fibrosis (IPF) has poor prognosis, and the multidisciplinary diagnostic agreement is low. Moreover, surgical lung biopsies pose comorbidity risks. Therefore, using data from non-invasive tests usually employed to assess interstitial lung diseases (ILDs), we aimed to develop an automated algorithm combining deep learning and machine learning that would be capable of detecting and differentiating IPF from other ILDs.

Methods: We retrospectively analysed consecutive patients presenting with ILD between April 2007 and July 2017. Deep learning was used for semantic image segmentation of HRCT based on the corresponding labelled images. A diagnostic algorithm was then trained using the semantic results and non-invasive findings. Diagnostic accuracy was assessed using five-fold cross-validation.

Results: In total, 646,800 HRCT images and the corresponding labelled images were acquired from 1068 patients with ILD, of whom 42.7% had IPF. The average segmentation accuracy was 96.1%. The machine learning algorithm had an average diagnostic accuracy of 83.6%, with high sensitivity, specificity and kappa coefficient values (80.7%, 85.8% and 0.665, respectively). Using Cox hazard analysis, IPF diagnosed using this algorithm was a significant prognostic factor (hazard ratio, 2.593; 95% CI, 2.069–3.250; $p < 0.001$). Diagnostic accuracy was good even in patients with usual interstitial pneumonia patterns on HRCT and those with surgical lung biopsies.

Conclusion: Using data from non-invasive examinations, the combined deep learning and machine learning algorithm accurately, easily and quickly diagnosed IPF in a population with various ILDs.

KEYWORDS

computed tomography, deep learning, diagnosis, idiopathic pulmonary fibrosis, interstitial lung disease, machine learning

INTRODUCTION

Interstitial lung disease (ILD) is a heterogeneous and challenging group of pulmonary disorders with varied prognoses and management options.^{1–3} Among ILDs, idiopathic pulmonary fibrosis (IPF) is characterized as a chronic, progressively worsening fibrotic lung disease of unknown aetiology with limited prognosis.⁴ With the recently proven efficacy of antifibrotic therapies, IPF diagnostic accuracy has become crucial.³

However, the diagnostic agreement on ILD is poor, even among respiratory physicians.^{5,6} In a large international study, the interobserver agreement on IPF diagnosis as evaluated by Cohen's kappa coefficient (κ) was 0.53 among experts and only 0.41 among respiratory physicians.⁷ The gold standard for diagnosing ILD is a dynamic integrated approach using multidisciplinary discussion (MDD), with close communication among clinicians, radiologists and pathologists.² MDD improves interobserver agreement⁸ and diagnostic confidence.⁶ The diagnostic agreement on IPF

among international MDD teams is 0.60.⁶ However, diagnosing all patients with ILD by MDD may be unfeasible due to time–space limitations and a lack of ILD experts. Moreover, MDD participants should be experienced ILD experts.⁷ Surgical lung biopsy (SLB) also accurately diagnoses ILD; however, it is not performed in cases of patient refusal, significant medical history or comorbidity risks such as acute exacerbation. Therefore, a screening tool to diagnose IPF without MDD or SLB is needed.⁹

Because HRCT images provide clues for IPF diagnosis,¹⁰ several automated ILD recognition systems using HRCT images have been developed.^{11–13} Machine learning and deep learning have gradually been used in various fields. In ILD imaging, a deep learning algorithm for categorizing HRCT images according to the HRCT classification¹⁰ had 73% accuracy, equivalent to that of chest radiologists.¹⁴ However, this algorithm had some problems, such as the comprehensibility of the result, and the output was only a classification of the HRCT patterns, not an ILD diagnosis.

Considering comprehensiveness, a deep learning semantic segmentation method assigns disease-specific labels, such as honeycombing, to every pixel of computed tomography (CT) images and visually shows the disease area on the images. Thus, we used semantic HRCT findings and non-invasive clinical examination data to improve the diagnostic accuracy of deep learning algorithms because MDD teams achieve a diagnosis by integrating these data.

Herein, we aimed to develop an algorithm for diagnosing IPF in a sample of various ILDs using non-invasive examination data, clinical information and CT imaging findings available in daily clinical practice. For improved comprehensibility of the algorithm, lesion recognition on CT by a semantic segmentation method was also performed.

METHODS

Data sets and data collection

For algorithm development, we retrospectively analysed medical records of 1068 consecutive patients with chronic ILD initially evaluated between April 2007 and March 2017 at an ILD referral centre in Japan. ILD was diagnosed through MDD according to the 2018 IPF guidelines, 2013 idiopathic interstitial pneumonia statement and other corresponding disease guidelines.^{2,4,15} This MDD team included specialists from other ILD centres. All patients' diagnoses were confirmed in December 2018. The usual interstitial pneumonia (UIP) categorization of the HRCT images was based on the 2018 HRCT criteria and was made by two clinicians with 33 and 21 years' experience, respectively, and/or a thoracic radiologist with 30 years' experience, who were blinded to the patients' clinical course and examination data.⁴

The Institutional Review Board of Nagoya University Graduate School of Medicine (IRB No. 2017-0263-2)

SUMMARY AT A GLANCE

Our comprehensible combined deep learning and machine algorithm can be used to easily, rapidly and non-invasively diagnose and differentiate idiopathic pulmonary fibrosis from various interstitial lung diseases using non-invasive examinations and HRCT, with high accuracy, sensitivity, specificity and kappa coefficient values.

approved this study. The informed consent requirement was waived.

Eligible patients were initially evaluated by non-invasive examinations, such as patient characteristics, pulmonary function tests, bronchoalveolar lavage and serologic tests, as is usual for ILD assessments. Chest HRCT images with 0.5–0.625-mm slice thickness at end-inspiration in the supine position were obtained at the initial evaluation. The duration from initial evaluation to last attendance or death was recorded.

Development and validation data sets

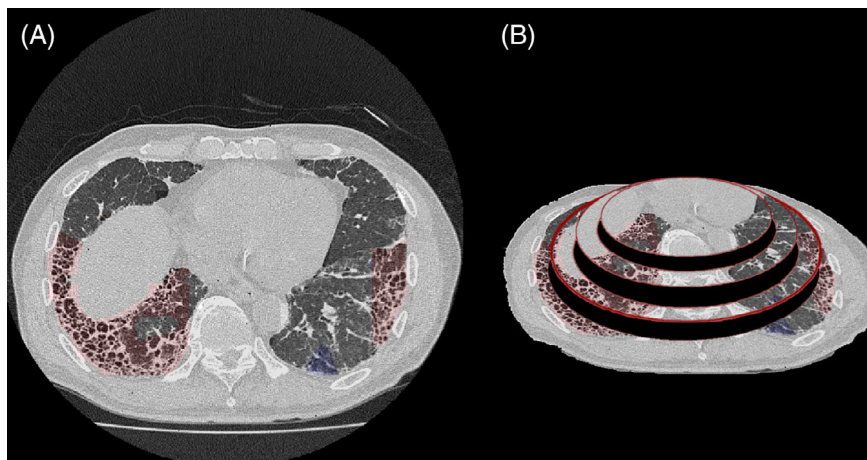
We adopted a five-fold cross-validation scheme to ensure the validity of the results. The data set was randomly split into five non-overlapping sets and each time four fifths of the data were used for training and the remainder was tested for model performance; this process was repeated five times. As a principal performance metric, we used the balanced accuracy averaged over the five-fold for the test set. Missing values in clinical examination data were appropriately imputed (Appendix S1 in the Supporting Information). On axial images, the lung HRCT data were divided into four zones per patient (upper, middle, lower and basal; Appendix S1 in the Supporting Information) based on a previous report.¹⁶ CT images were augmented with random filter and rotation effects. Subsequently, each augmented image was divided into 25 images.

Data preprocessing

For input to the algorithm, we used only blood tests, pulmonary function tests, patient characteristics and CT images, which are routine non-invasive clinical tests and evaluations. No histopathological images, histological findings, CT reading results or CT pattern readings by radiologists were used.

In data preprocessing, HRCT images were labelled as follows by a clinician with 10 years' experience, who was independent of the MDD team: IPF, non-IPF, lung and extrapulmonary area. Briefly, areas of honeycombing and/or reticular pattern with peripheral traction bronchiectasis were

FIGURE 1 Examples of semantic segmentation results and elliptically distributed labelled images. (A) A labelled image generated by the trained algorithm. Red and blue areas are suggestive of idiopathic pulmonary fibrosis (IPF) and non-IPF, respectively. (B) A labelled image generated from the semantic segmentation is elliptically distributed and summarized based on the disease-specific area.



classified as suggestive of IPF. Areas suggestive of non-IPF were predominant consolidation, extensive pure ground-glass opacity, extensive mosaic attenuation and/or diffuse nodules or cysts.

Combination of machine learning algorithms

We built a comprehensible machine learning model for IPF diagnosis in two steps. As information on both disease lesion semantics and location on HRCT are important for ILD diagnosis, we first used the deep fully convolutional neural network FCN-Alexnet¹⁷ for semantic segmentation of HRCT images, with labelled images as training data (Appendix S1 in the Supporting Information).¹⁷ The network parameters were trained by minimizing the categorical cross-entropy using the Stochastic Gradient Descent optimizer with back-propagation with a learning rate of 0.001 and batch size of 1024 (Appendix S1 in the Supporting Information).

Second, the semantic image results (Figure 1A) were elliptically divided into five inner to outer layers, and the respective IPF/non-IPF area ratios were calculated (Figure 1B). Based on disease-specific ratios of distribution areas and other clinical data (Appendix S1 in the Supporting Information), support vector machines, a type of binary machine learning classifiers, were trained to diagnose IPF using data standardization and a linear kernel (Appendix S1 in the Supporting Information). Appendix S1 in the Supporting Information shows the variables used to build machine learning models. Even if SLBs were performed, the pathological findings were not used to develop the model. Finally, the combination algorithm was assessed for diagnostic accuracy using five-fold cross-validation with each of the test sets.

Statistical analysis

Data preprocessing, machine learning algorithm development and other statistical analyses were performed using

MATLAB 9.3 (MathWorks, Natick, MA) and R statistical software (version 3.6.1, R Foundation Inc., Vienna, Austria). The HRCT semantic segmentation model was implemented using the open-source software Keras based on the machine learning library TensorFlow, running on NVIDIA Tesla V100 Volta graphics processor units, for the input image cropped as 128×128 pixels.

Continuous data are presented as the mean \pm SD. Categorical variables are reported as frequency (%). Between-group differences were assessed using the two-sided *t*-test or chi-square test as appropriate. The Cox proportional hazard analysis results are presented as estimated hazard ratios (HRs) with 95% CIs, with adjustment for baseline age, sex and percent forced vital capacity. Harrell's *C*-statistic was used to evaluate the ability of each Cox proportional hazards model to predict mortality. Cumulative survival probabilities were plotted with the Kaplan–Meier method. Validity of the algorithm diagnosis for mortality prediction was evaluated by the Cox hazards and Kaplan–Meier methods in all patients, patients with UIP patterns on CT and patients without SLB. *p* values of less than 0.05 were considered statistically significant.

RESULTS

Of the 1068 eligible patients with ILD, 456 (42.7%) were diagnosed with IPF by MDD. Overall, 238 (22.3%), 234 (21.9%) and 28 (2.6%) patients had unclassifiable idiopathic interstitial pneumonia, connective tissue disease-related ILD and hypersensitivity pneumonia, respectively (Table S1 in the Supporting Information). Table 1 shows the patients' demographic and clinical data. The median survival time was 86.5 months (95% CI, 77.8–102.2 months).

The final HRCT data set comprised 646,800 unique images for algorithm training. Using these and the corresponding labelled images, semantic segmentation training was conducted. Averaging over all folds, the patient numbers were 854 for training and 213 for testing. After semantic segmentation, the semantic results were obtained

TABLE 1 Baseline characteristics of the data sets

	Diagnosis by MDD		<i>p</i> value
	IPF	Non-IPF	
Participants, <i>N</i>	456	612	
Age, years	67.9 ± 8.2	64.3 ± 10.6	<0.001
Male sex, <i>N</i> (%)	363 (80%)	310 (51%)	<0.001
BMI, kg/m ²	23.2 ± 3.5	22.4 ± 3.9	<0.001
Pack-years ^a	33.2 ± 50.3	20.1 ± 29.0	<0.001
Surgical lung biopsy	138 (30%)	263 (43%)	<0.001
FVC, % pred	78.0 ± 20.5	83.8 ± 22.4	<0.001
FEV ₁ /FVC	85.6 ± 7.9	82.7 ± 9.2	<0.001
DL _{CO} , % pred ^b	58.8 ± 21.3	67.4 ± 22.8	<0.001
KL-6, U/ml	1394 ± 991	1386 ± 1457	0.91
BALF findings ^c			
Total cell count, 10 ⁵ /ml	1.68 ± 1.16	1.79 ± 1.69	0.23
Macrophages, %	86.3 ± 15.7	75.5 ± 25.1	<0.001
Lymphocytes, %	6.3 ± 7.9	14.7 ± 19.1	<0.001
Neutrophils, %	4.5 ± 11.9	6.3 ± 14.5	0.04
Eosinophils, %	1.6 ± 4.1	2.4 ± 5.9	0.01
CD4/CD8	2.57 ± 2.33	2.23 ± 3.28	0.06

Note: Data are presented as mean ± SD or number (%). Autoantibody data are not shown.

Abbreviations: % pred, percent predicted; BALF, bronchoalveolar lavage fluid; DL_{CO}, diffusion capacity for carbon monoxide; FEV₁, forced expiratory volume in the first second; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; KL-6, serum Krebs von den Lungen-6; MDD, multidisciplinary discussion; pack-year, a measure of the amount a person has smoked.

^aNine hundred and twenty-six patients.

^bOne thousand and thirty-four patients.

^cNine hundred and thirty patients.

as labelled slice images (Figure 1A). The average test accuracy for semantic segmentation was 96.1% (Table 2). The average accuracy of IPF diagnosis increased from 65.3% without distribution to 79.7%, with the elliptically

distributed IPF/non-IPF ratio calculated from the semantic results (Figure 1B and Table 2). Moreover, using these results and the clinical data (Appendix S1 in the Supporting Information), the algorithm exhibited high accuracy, sensitivity, specificity, positive predictive value, negative predictive value and kappa coefficient (83.6%, 80.7%, 85.8%, 80.9%, 85.6% and 0.665, respectively). Table S1 in the Supporting Information shows the algorithmic prediction frequency agreement with MDD diagnosis.

Cox hazard analysis showed that IPF diagnosed by the machine learning algorithm (AI-IPF) was a significant prognostic factor (HR, 2.634; 95% CI, 2.104–3.298; *p* < 0.001; *C*-statistic, 0.765), as was IPF diagnosed by MDD (MDD-IPF; Table 3). Moreover, AI-IPF had a prognostic discriminatory ability equivalent to that of MDD-IPF (Figure 2). Patients with mismatched MDD and AI diagnoses showed similar prognosis (Figure S1 in the Supporting Information).

Among patients with MDD-IPF, 88 (19.3%) were categorized as AI-non-IPF by the algorithm (Table S2 in the Supporting Information), and the AI-IPF group had a significantly worse survival probability compared to the AI-non-IPF group (HR, 2.213; 95% CI, 1.536–3.189; *p* < 0.001; Figure 3). AI-IPF had a higher ratio of the presence of honeycombing than AI-non-IPF (69.0% vs. 35.2%, *p* < 0.001); however, the presence of honeycombing was not significantly associated with a poor prognosis (HR, 1.210; 95% CI, 0.930–1.574; *p* = 0.155) (Appendix S2 in the Supporting Information).

Even in cases with SLB that were difficult to diagnose without SLB, the algorithm accuracy was 81.0%. Moreover, in patients with UIP patterns or UIP + probable UIP patterns on HRCT, our algorithm showed a high diagnostic accuracy (83.6% and 80.2%, respectively) and prognostic discrimination ability (*C*-index, 0.725 and 0.720, respectively; Table S3 in the Supporting Information). Moreover, in patients with UIP pattern on HRCT or histopathology, our algorithm labelled 87.5% of MDD-IPF as AI-IPF.

TABLE 2 Accuracy of test data for semantic segmentation and diagnostic results using SVMs based on HRCT and other clinical data

Cross-validation	Semantic segmentation accuracy for ILD findings ^a		Diagnostic accuracy ^b		
	Using HRCT (%)		Using semantic result (%)	With elliptical distribution (%)	Addition of clinical data (%)
Set 1	96.122		66.512	81.860	83.721
Set 2	95.993		64.455	81.517	84.834
Set 3	96.095		62.500	77.315	79.167
Set 4	95.956		67.606	78.404	84.507
Set 5	96.329		65.566	79.717	85.849
Mean	96.099		65.328	79.763	83.616

Note: The accuracy of semantic segmentation is the pixel-based accuracy following ILD labelling of images. 'Using semantic result' indicates diagnostic accuracy across SVM results using semantic results from HRCT and MDD diagnosis of IPF. 'With elliptical distribution' indicates diagnostic accuracy across SVM results using semantic results from HRCT and MDD diagnosis of IPF, with the addition of elliptical distribution. 'Addition of clinical data' indicates diagnostic accuracy across SVM results using semantic results from HRCT and MDD diagnosis of IPF, with the addition of clinical data as well as elliptical distribution.

Abbreviations: ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; MDD, multidisciplinary discussion; SVM, support vector machine.

^aThe accuracy of every pixel of the semantic results for HRCT images against the corresponding disease-specific labelled images.

^bThe accuracy of the algorithmic diagnosis of IPF against the MDD diagnosis.

TABLE 3 Cox hazard analysis for mortality from IPF versus non-IPF: IPF diagnosis by humans versus that by machine

	HR (95% CI) ^a	p value	C-statistic
MDD-IPF	2.711 (2.162–3.400)	<0.001	0.762
AI-IPF	2.593 (2.069–3.250)	<0.001	0.765

Abbreviations: AI-IPF, IPF diagnosed by the machine learning algorithm; HR, hazard ratio; IPF, idiopathic pulmonary fibrosis; MDD-IPF, IPF diagnosed by multidisciplinary discussion.

^aAdjusted for age, sex and percent forced vital capacity at baseline.

DISCUSSION

Using non-invasive examination data and CT images available in clinical practice, we developed a comprehensive machine learning algorithm that can diagnose and differentiate IPF from all ILDs with an accuracy equivalent to that of MDD diagnosis. Currently, there are several machine learning models using CT images or molecules as input,¹⁸ which outputs CT patterns mentioned in the IPF guidelines,¹⁴ indicates pathological UIP¹⁹ or shows CT

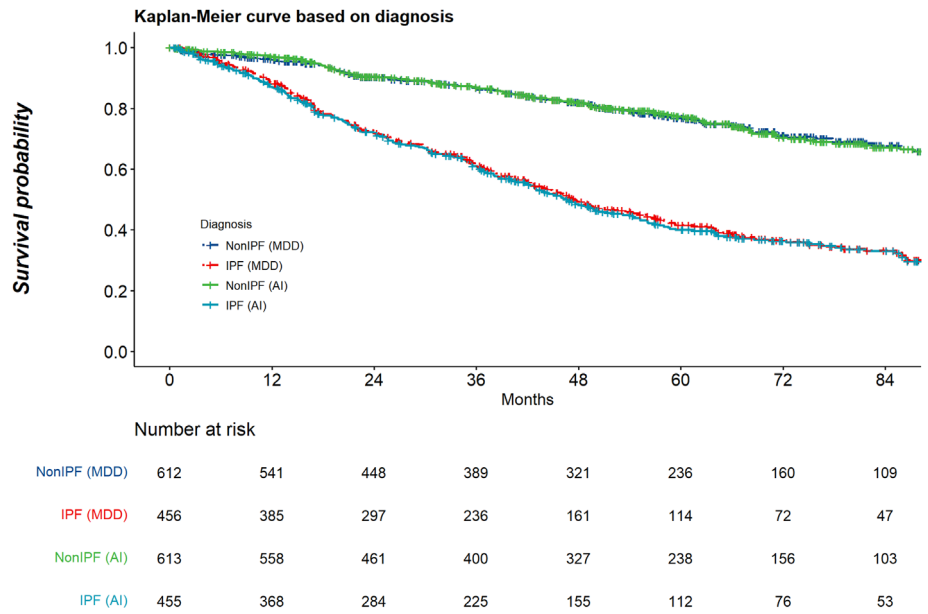


FIGURE 2 Kaplan–Meier curve showing survival differences between IPF and non-IPF diagnosed by MDD and the AI. AI, the newly developed algorithm; IPF, idiopathic pulmonary fibrosis; MDD, multidisciplinary discussion

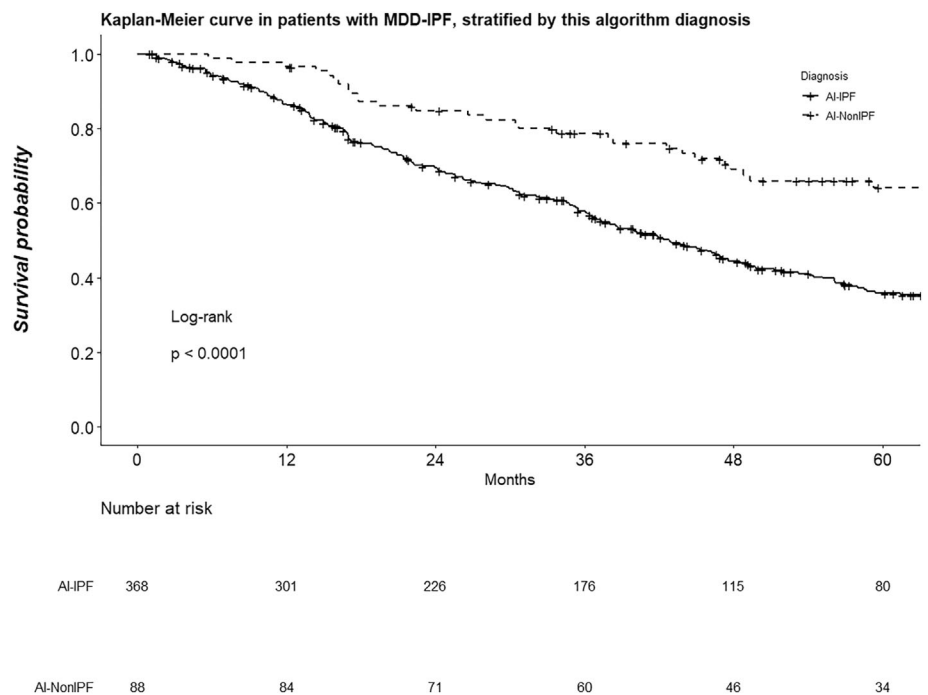


FIGURE 3 Kaplan–Meier curve of patients with IPF as determined by MDD, stratified by machine learning algorithm diagnosis. AI-IPF, IPF diagnosed by the newly developed algorithm; IPF, idiopathic pulmonary fibrosis; MDD-IPF, IPF diagnosed by multidisciplinary discussion

findings such as ground-glass opacities,^{11,13,20–23} based on each image^{13,14,20} or each pixel of an image^{21,22}; however, there is no model to diagnose IPF and integrate CT images and clinical information. We are the first to develop a multi-modal AI to diagnose IPF and generate labelled images as a diagnostic basis by integrating analysis of clinical information, examination and CT images similar to clinicians, without input from ILD experts, MDD or SLB. The prognostic discrimination ability of the algorithm was equivalent to that of MDD (Table 3). Moreover, the diagnostic accuracy was good even in patients with UIP patterns on HRCT and in patients with SLB, even though the input was limited to data derived from non-invasive methods. Even among patients with MDD-IPF, those with AI-IPF had a worse mortality rate. Therefore, by combining our algorithm with MDD, IPF and/or groups with poor prognosis may be identified quickly and easily.

As HRCT pattern is important in diagnosing IPF,⁴ and an early IPF diagnosis is crucial for early treatment to preserve lung function,²⁴ researchers have tested automated HRCT-based systems that can identify ILD findings (e.g., ground-glass opacity).¹² Recently, machine recognition of ILD findings using deep learning exhibited 82% accuracy²¹; our algorithm has a higher accuracy of 96%, attributable to the larger training data used, with more than six times the number of patients enrolled in previous studies. This is consistent with the fact that deep learning generally requires voluminous training data of good quality. Another reason might be the choice of segmentation labels used with our algorithm, which aimed to indicate whether an HRCT finding was suggestive of IPF, thereby facilitating diagnosis.

Deep learning can be limited by the incomprehensibility of its judgement process. However, our algorithm yielded high-accuracy labelled images, providing a comprehensible model with which to diagnose IPF that identifies the patterns used to make the diagnosis, where the algorithm recognized these patterns on the HRCT images. Thus, the algorithm can both generate semantically labelled images and diagnose IPF, potentially enabling it to play an important role in clinical settings. The diagnostic accuracy was good even in cases with UIP patterns on HRCT, suggesting that our algorithm not only recognized CT patterns but also the integrated clinical information. Because perfect ILD lesions segmentation on HRCT is difficult, good segmental deep learning results using manually labelled images may have limited impact; however, this study's most important finding was the high accuracy of IPF diagnosis using HRCT segmental results. Our algorithm showed a higher diagnostic agreement in IPF diagnosis ($\kappa = 0.67$) than the international MDD teams ($\kappa = 0.53$)⁶ and respiratory physicians ($\kappa = 0.41$).⁷ This may be due to the segmentation and elliptical division of the segmented images into outer and inner layers, which mimic the diagnostic process used by physicians. This is useful for particularly evaluating honeycombing, the key observation for IPF diagnosis, because this feature appears in the subpleural area.¹⁵

To date, using MDD to diagnose all patients with ILD is challenging because of the above-mentioned MDD-associated problems. However, our algorithm can be used as a screening tool by general physicians and non-expert respiratory physicians. On combining this algorithm with MDD, ILD diagnosis could be improved so that opportunities for suitable treatment (e.g., with antifibrotic drugs) would not be missed. Moreover, our algorithm showed good diagnostic accuracy even in patients with SLB; thus, a potential benefit of our algorithm is that its use may allow some patients to avoid SLB, thereby reducing the risk of acute exacerbation and death.

Interestingly, even in the MDD-IPF group, patients classified as AI-IPF had a worse mortality rate. Our algorithm may thus contribute new information for ILD diagnosis and treatment. However, current technology cannot create a fully explainable AI, as not all the calculations and processes of this AI system can be completely understood; therefore, further study and innovation are needed.

As our study was based on a learning algorithm from a single MDD team's diagnoses and the labelled images segmented by a single clinician, our findings require cautious interpretation. Another generalizability issue is the disease proportion bias. Particularly, the proportion of hypersensitivity pneumonitis was less than that in other published cohorts (2.6% vs. 2%–47%).²⁵ While our algorithm accurately classified 82% of patients with hypersensitivity pneumonitis, it may not work well in cohorts with higher frequencies of hypersensitivity pneumonitis. Moreover, while the algorithm was trained and tested using five-fold cross-validation, validation was with a single-centre data set of patients from a single race. Thus, the results may not be generalizable to other populations. Moreover, as our study collected HRCT images with 0.5–0.625-mm slice thickness, our algorithm may not work well with images of other slice thicknesses. However, in a feasibility study (Figure S2 in the Supporting Information), our system had almost the same semantic segmentation accuracy with images of 2-mm slice thickness. Other slice thicknesses are needed for validation.

This easy-to-use and rapid machine learning algorithm based on data from non-invasive techniques can diagnose IPF with high accuracy, sensitivity, specificity, positive predictive value, negative predictive value and kappa coefficient. However, a more robust algorithm is clinically needed, further warranting multicentre studies.

AUTHOR CONTRIBUTION

Taiki Furukawa: Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); software (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).
Shintaro Oyama: Conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition (equal); investigation (equal); methodology (equal); project administration (equal); resources (equal); writing –

original draft (equal); writing – review and editing (equal). **Hideo Yokota:** Conceptualization (equal); data curation (supporting); formal analysis (equal); funding acquisition (supporting); investigation (supporting); methodology (equal); project administration (equal); visualization (supporting); writing – original draft (equal); writing – review and editing (equal). **Yasuhiro Kondoh:** Conceptualization (equal); data curation (equal); formal analysis (supporting); investigation (supporting); methodology (equal); project administration (supporting); validation (supporting); visualization (supporting); writing – original draft (equal); writing – review and editing (equal). **Kensuke Kataoka:** Conceptualization (supporting); data curation (supporting); formal analysis (supporting); investigation (supporting); methodology (supporting); project administration (supporting); validation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Takeshi Johkoh:** Conceptualization (supporting); formal analysis (supporting); investigation (supporting); methodology (supporting); project administration (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Junya Fukuoka:** Conceptualization (supporting); data curation (supporting); formal analysis (supporting); investigation (supporting); methodology (supporting); project administration (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Naozumi Hashimoto:** Conceptualization (supporting); formal analysis (supporting); project administration (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Koji Sakamoto:** Conceptualization (supporting); formal analysis (supporting); methodology (supporting); project administration (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Yoshimune Shiratori:** Conceptualization (supporting); data curation (supporting); funding acquisition (supporting); methodology (supporting); project administration (supporting); resources (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Yoshinori Hasegawa:** Conceptualization (supporting); funding acquisition (supporting); investigation (supporting); project administration (supporting); writing – original draft (supporting); writing – review and editing (supporting).

ACKNOWLEDGEMENTS

We thank Ryo Teramachi, MD (Graduate School of Medicine, Nagoya University, Nagoya, Japan) and Reoto Takei, MD (Tosei General Hospital, Nagoya, Japan) for data collection support, and Yasuhiko Yamano, PhD (Tosei General Hospital, Nagoya, Japan) for support in clinical data preprocessing.

Research funding: This work was supported by the JSPS KAKENHI (JP19110253), the Hori Science and Arts Foundation and a grant from the Japanese Respiratory Foundation. None of the sponsors had any role in study design, data collection, data analysis or manuscript preparation.

CONFLICTS OF INTEREST

Yasuhiro Kondoh reports personal fees from Asahi Kasei Pharma Corp., Boehringer Ingelheim, Co., Ltd., Eisai Inc., KYORIN Pharmaceutical Co., Ltd., Novartis Pharma K.K. and Shionogi & Co. outside the submitted work. Koji Sakamoto reports grants from the Pulmonary Fibrosis Foundation and Novartis Pharma outside the submitted work. Yoshinori Hasegawa reports personal fees for speaking activity from Boehringer Ingelheim Japan outside the submitted work. The other authors have nothing to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

HUMAN ETHICS APPROVAL DECLARATION

This study was performed in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Nagoya University Graduate School of Medicine (approval: 2017-0263-2). Adult participant consent was waived by the Institutional Review Board due to the retrospective nature of this study.

ORCID

Taiki Furukawa  <https://orcid.org/0000-0002-3522-1845>

Yasuhiro Kondoh  <https://orcid.org/0000-0001-7456-5459>

REFERENCES

1. American Thoracic Society, European Respiratory Society. American Thoracic Society/European Respiratory Society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med.* 2002;165:277–304.
2. 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;188:733–48.
3. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, et al. An official ATS/ERS/JRS/ALAT clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline. *Am J Respir Crit Care Med.* 2015;192:e3–19.
4. 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.
5. Walsh SL, Calandriello L, Sverzellati N, Wells AU, Hansell DM. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax.* 2016;71:45–51.
6. Walsh SL, Wells AU, Desai SR, Poletti V, Piciucchi S, Dubini A, et al. Multicentre evaluation of multidisciplinary team meeting agreement on diagnosis in diffuse parenchymal lung disease: a case-cohort study. *Lancet Respir Med.* 2016;4:557–65.
7. Walsh SLF, Maher TM, Kolb M, Poletti V, Nusser R, Richeldi L, et al. Diagnostic accuracy of a clinical diagnosis of idiopathic pulmonary fibrosis: an international case-cohort study. *Eur Respir J.* 2017;50:1700936.
8. Flaherty KR, King TE Jr, Raghu G, Lynch JP 3rd, Colby TV, Travis WD, et al. Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med.* 2004;170:904–10.

9. Kondoh Y, Cottin V, Brown KK. Recent lessons learned in the management of acute exacerbation of idiopathic pulmonary fibrosis. *Eur Respir Rev*. 2017;26:170050.
10. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med*. 2018;6:138–53.
11. Zavaletta VA, Bartholmai BJ, Robb RA. High resolution multidetector CT-aided tissue analysis and quantification of lung fibrosis. *Acad Radiol*. 2007;14:772–87.
12. Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, et al. Automated quantification of radiologic patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J*. 2014;43:204–12.
13. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging*. 2016;35:1207–16.
14. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med*. 2018;6:837–45.
15. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. 2018;198:e44–68.
16. Sumikawa H, Johkoh T, Colby TV, Ichikado K, Suga M, Taniguchi H, et al. Computed tomography findings in pathological usual interstitial pneumonia: relationship to survival. *Am J Respir Crit Care Med*. 2008;177:433–9.
17. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39:640–51.
18. Raghu G, Flaherty KR, Lederer DJ, Lynch DA, Colby TV, Myers JL, et al. Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study. *Lancet Respir Med*. 2019;7:487–96.
19. Shaish H, Ahmed FS, Lederer D, D'Souza B, Armenta P, Salvatore M, et al. Deep learning of computed tomography virtual wedge resection for prediction of histologic usual interstitial pneumonitis. *Ann Am Thorac Soc*. 2021;18:51–9.
20. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35:1285–98.
21. Anthimopoulos M, Christodoulidis S, Ebner L, Geiser T, Christe A, Mougiakakou S. Semantic segmentation of pathological lung tissue with dilated fully convolutional networks. *IEEE J Biomed Health Inform*. 2019;23:714–22.
22. Handa T, Tanizawa K, Oguma T, Uozumi R, Watanabe K, Tanabe N, et al. Novel artificial intelligence-based technology for chest computed tomography analysis of idiopathic pulmonary fibrosis. *Ann Am Thorac Soc*. 2022;19:399–406.
23. 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;43:204–12.
24. Kolb M, Richeldi L, Behr J, Maher TM, Tang W, Stowasser S, et al. Nintedanib in patients with idiopathic pulmonary fibrosis and preserved lung volume. *Thorax*. 2017;72:340–6.
25. Raghu G, Remy-Jardin M, Ryerson CJ, Myers JL, Kreuter M, Vasakova M, et al. Diagnosis of hypersensitivity pneumonitis in adults. An official ATS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med*. 2020;202:e36–69.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Furukawa T, Oyama S, Yokota H, Kondoh Y, Kataoka K, Johkoh T, et al. A comprehensible machine learning tool to differentially diagnose idiopathic pulmonary fibrosis from other chronic interstitial lung diseases. *Respirology*. 2022. <https://doi.org/10.1111/resp.14310>