

Novel 3D-based deep learning for classification of acute exacerbation of idiopathic pulmonary fibrosis using high-resolution CT

Xinmei Huang,^{1,2} Wufei Si,³ Xu Ye,¹ Yichao Zhao,⁴ Huimin Gu,⁵ Mingrui Zhang,⁴ Shufei Wu,⁴ Yanchen Shi,⁴ Xianhua Gui,^{1,2} Yonglong Xiao,^{1,2} Mengshu Cao^{1,2,4,5}

To cite: Huang X, Si W, Ye X, et al. Novel 3D-based deep learning for classification of acute exacerbation of idiopathic pulmonary fibrosis using high-resolution CT. *BMJ Open Respir Res* 2024;**11**:e002226. doi:10.1136/bmjresp-2023-002226

XH and WS contributed equally.

Received 1 December 2023
Accepted 28 February 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Mengshu Cao;
mengshuca@nju.edu.cn,

Dr Yonglong Xiao;
yonglong11a@163.com and

Dr Xianhua Gui;
xianxian.xian@163.com

ABSTRACT

Purpose Acute exacerbation of idiopathic pulmonary fibrosis (AE-IPF) is the primary cause of death in patients with IPF, characterised by diffuse, bilateral ground-glass opacification on high-resolution CT (HRCT). This study proposes a three-dimensional (3D)-based deep learning algorithm for classifying AE-IPF using HRCT images.

Materials and methods A novel 3D-based deep learning algorithm, SlowFast, was developed by applying a database of 306 HRCT scans obtained from two centres. The scans were divided into four separate subsets (training set, n=105; internal validation set, n=26; temporal test set 1, n=79; and geographical test set 2, n=96). The final training data set consisted of 1050 samples with 33600 images for algorithm training. Algorithm performance was evaluated using accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic (ROC) curve and weighted κ coefficient.

Results The accuracy of the algorithm in classifying AE-IPF on the test sets 1 and 2 was 93.9% and 86.5%, respectively. Interobserver agreements between the algorithm and the majority opinion of the radiologists were good ($\kappa_w=0.90$ for test set 1 and $\kappa_w=0.73$ for test set 2, respectively). The ROC accuracy of the algorithm for classifying AE-IPF on the test sets 1 and 2 was 0.96 and 0.92, respectively. The algorithm performance was superior to visual analysis in accurately diagnosing radiological findings. Furthermore, the algorithm's categorisation was a significant predictor of IPF progression.

Conclusions The deep learning algorithm provides high auxiliary diagnostic efficiency in patients with AE-IPF and may serve as a useful clinical aid for diagnosis.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive pulmonary disease of unknown aetiology.¹ A subset of patients with IPF experience a significant minority develop episodes of acute clinical respiratory worsening, termed acute exacerbations of IPF (AE-IPF).² AE-IPF is difficult to predict or prevent and precedes approximately half of the IPF-related deaths, with a mean survival of

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Acute exacerbation of idiopathic pulmonary fibrosis (AE-IPF) is a significant cause of death in patients with IPF, characterised by clinically significant respiratory deterioration and diffuse bilateral ground-glass opacification on high-resolution CT (HRCT) scans. However, radiological evaluation of AE-IPF remains challenging and is subject to substantial interobserver variability. Currently, several deep learning models have been applied to diagnostic support of fibrotic interstitial lung disease, however, most research has focused on deep learning models based on two-dimensional data, with limited research exploring deep learning for AE-IPF diagnosis.

WHAT THIS STUDY ADDS

⇒ This study investigated that the innovative three-dimensional video-sequence methodology called SlowFast, is proposed to classify acute exacerbation in patients with IPF on HRCT scans. The accuracy of the algorithm in predicting radiological diagnosis was superior to that of thoracic radiologists (area under the curve=0.96), with excellent interobserver agreement ($\kappa_w=0.90$).

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The study provides a valuable contribution to the field by demonstrating the potential of deep learning algorithms to provide low-cost, consistent patient stratification and assistance for radiological decisions of AE-IPF.

3–4 months.^{3,4} The most common radiological feature in patients with AE-IPF is the presence of new ground-glass opacities (GGO) superimposed on subpleural reticular and honeycomb-like densities.² There have been several studies identified that high-resolution CT (HRCT) scan plays a central role in the adequate diagnosis and early intervention of AE-IPF.^{5–7} However, radiological evaluation of AE-IPF remains challenging and is susceptible to significant variability between observers,

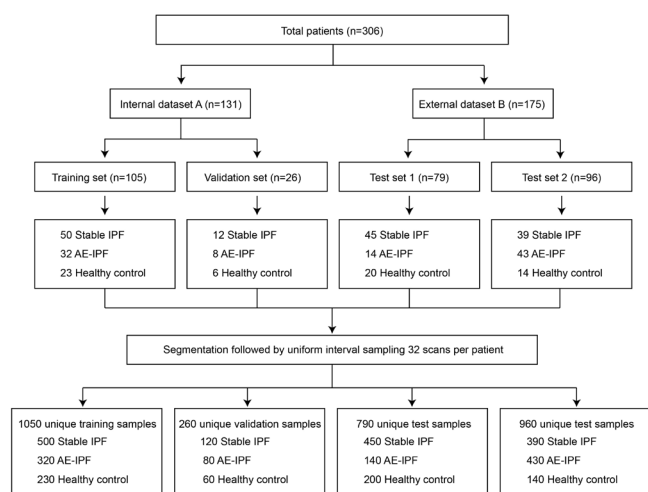


Figure 1 Data set flowchart of the study design. The flowchart diagram illustrates the division of the total cohort of HRCT scans into training, validation and test cohorts, followed by image segmentation and resampling. AE-IPF, acute exacerbation of idiopathic pulmonary fibrosis; IPF, idiopathic pulmonary fibrosis.

even among experienced radiologists.^{7 8} Therefore, developing better methods for HRCT scans detection and disease classification generated by deep learning algorithms, has the potential to improve the radiological diagnosis of AE-IPF.

Deep learning, a subset of artificial intelligence (AI) technology that efficiently identifies patterns in high dimensional data, has recently entered an accelerated phase in medical image interpretation.^{9–11} Currently, several two-dimensional (2D)-data-based deep learning models (Convolution neural networks, Recurrent neural networks, Deep belief networks, etc) have been successfully applied to diagnostic support of fibrotic interstitial lung disease (ILD), early detection of clinically significant fibrotic lung disease and prediction of progressive fibrotic lung disease.^{9 12 13} However, despite the apparent benefits of these models for fibrotic lung disease, there are still some critical limitations, including algorithm performance, data heterogeneity and constraints, relative opacity of neural networks (black box phenomenon) and lack of histopathological reference standard.¹⁴ More importantly, the 2D-based models randomly select four segmented axial HRCT image slices from a range of 250–450 axial image slices per patient for algorithm training, which inevitably leads to the loss of some HRCT image information.^{9 15} To address this research gap, this study proposes a three-dimensional (3D) video-sequence-based methodology called SlowFast, which incorporates a novel algorithm for detecting HRCT scans. Generally, the SlowFast algorithm uses a slow, high-definition CNN (Fast pathway) to analyse the static content of a video, while in parallel, a fast, low-resolution CNN (Slow pathway) is used to analyse the dynamic content.^{16 17} HRCT scans consist of large amounts of ordered high-resolution images,¹⁸ making them highly suitable for deep learning

models based on 3D data. However, most research so far has focused on deep learning models based on 2D data, with limited research exploring deep learning for AE-IPF diagnosis. This study, for the first time, investigated that the 3D video-sequence-based algorithm SlowFast is proposed to classify acute exacerbation in patients with IPF on HRCT scans.

MATERIALS AND METHODS

Patient and public involvement statement

The patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Data split

For model pretraining, an internal data set A comprising 131 patients with HRCT scans taken between December 2015 and December 2018 was obtained from Nanjing Drum Tower Hospital, consisting of 62 cases of stable IPF, 40 cases of AE-IPF and 29 cases of healthy controls. For model validation, an external data set B comprising 175 patients with HRCT taken between January 2019 and December 2022 was obtained from Nanjing Drum Tower Hospital and Nanjing Traditional Chinese Medicine Hospital, consisting of 84 cases of stable IPF, 57 cases of AE-IPF and 34 cases of healthy controls.

The inclusion criteria were: (1) the availability of HRCT with slices thickness of less than 1.5 mm, and each HRCT showing evidence of diagnoses.^{3 19 20} For AE-IPF, the diagnostic evidence was characterised by the presence of new bilateral GGO and/or consolidation, superimposed on a background pattern consistent with the usual interstitial pneumonia pattern. (2) Other clinical data meet the diagnostic criteria.^{3 19 20} For stable-IPF, diagnostic criteria include stable clinical symptoms, HRCT imaging and pulmonary function tests for at least 1-month prior to inclusion. The exclusion criterion was the use of contrast enhancement.

The ground truth labels were proved by four thoracic radiologists/respirologists (with 5–20 years of experience in diagnosing ILD). The total data set size was 3060 samples with 97920 images (204–351 axial image slices per patient). The internal data set A was split into a training set (n=105) and a validation set (n=26). The external data set B was split into two sets: test set 1 (n=79) and test set 2 (n=96), which were used for temporal validation of the model (figure 1).

Image preprocessing and resampling

Data set for semantic segmentation: the semantic segmentation data set consists of 512 HRCT slicers selected from the data set A with uniform sampling. The original images, initially sized at 512×512 pixels, were cropped to 320×320 and labelled by the four radiologists/respirologists using the graphical image annotation tool LabelMe.²¹ The data set was split into training, validation and test set, with

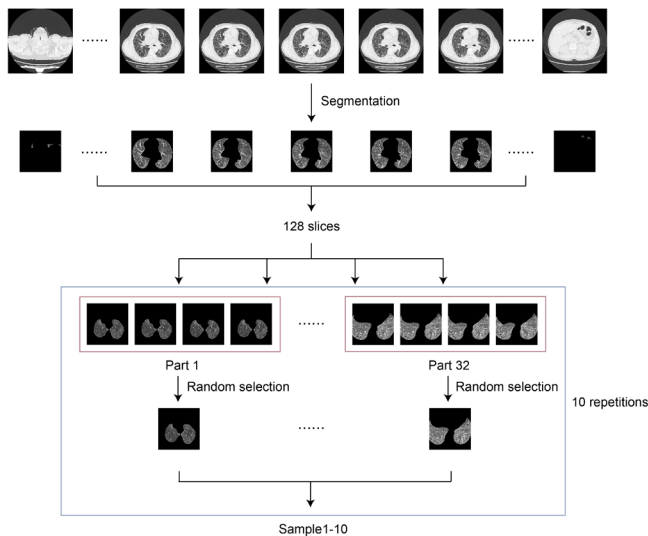


Figure 2 Flowchart of the training framework. The samples from high-resolution CT scans after segmentation were concatenated and fed to the video classification network SlowFast.

proportions of 0.70:0.15:0.15, containing 358, 77 and 77 images, respectively. The segmentation model DeepLabV3+ was trained using this data set to remove redundant information from the original HRCT scans.

Data set for video classification: after being segmented by the segmentation model, the HRCT scans of 306 patients were ready to generate the video classification data set. For each patient, 128 consecutive scans from the middle section were selected and divided into 32 equal parts. Then, 1 scan was randomly chosen from each part to form a learnable sample consisting of 32 slices. To train the video classification model SlowFast, a sample with 32 images is proper. Samples were randomly drawn 10 times from each patient series yielding a total of 3060 samples. Finally, the 3060 samples were split into the

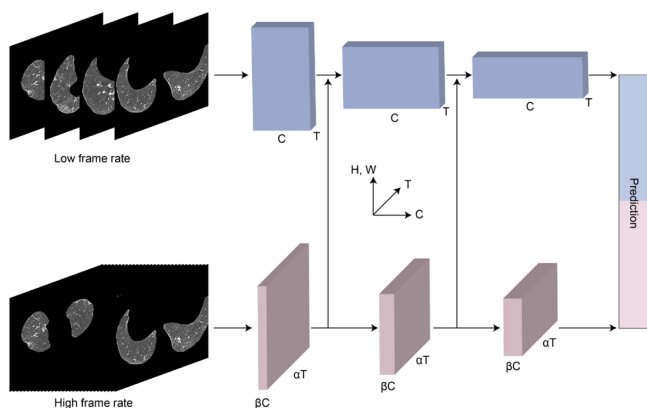


Figure 3 The architecture of deep learning algorithm. The slow (top) path uses lower frame rates to sample frames from the high-resolution CT scans. The fast (bottom) path uses higher frame rates for sampling, while including a fraction of the channels used by the slow path.

train, validation, test set 1 and test set 2 with 1050, 260, 790 and 960 samples, respectively (figure 2).

Data augmentation: to increase the size and diversity of the data set, data augmentation techniques were employed during the preprocessing stage, including Flip, Rotate and Dropout. Horizontal flipping was applied to each image with a 50% probability to generate additional images by creating mirror images of the originals. Additionally, each image was randomly rotated by a degree between -20° and $+20^\circ$ to simulate different viewing angles and orientations. Finally, the Dropout technique was used to randomly select and set 5% of pixels in each image to zero.

Algorithm development

Two neural networks were used in this work. The semantic segmentation network DeepLabV3+ was used to separate the lung area from the original HRCT scan. The resulting segmented image was then used for the subsequent classification task. The video classification network SlowFast was used to make a diagnosis using the extracted image sequence from the segmented results.¹⁷ The final output was a prediction of the diagnosis category, which included stable IPF, AE-IPF or healthy control (figure 3). The algorithms in the study were developed using PyTorch framework (V.1.9.0 with CUDA V.10.2), using 4 NVIDIA V100 GPUs. For more details, the model was trained for 60 epochs with a batch size of 16. Adam was chosen as the optimiser with learning rate $1e-3$ and weight decay $1e-4$. The classical cross-entropy loss was used as the loss function.

Radiologist classification

Each HRCT scan in the test sets was visually scored by 3 fellowship-trained radiologists (with 3–24 years of post-fellowship experience) on a 4-point ordinal scale corresponding to the 2018 American Thoracic Society guidelines for IPF and 2016 International Working Group Report for AE-IPF^{3 19 20} (0=AE-IPF, 1=stable IPF, 2=healthy control). The score was used to compare the diagnostic opinion of the algorithm. To align with AI diagnostic methods, radiologists were provided with complete access to all series and images for each HRCT scan in the test set, while being blinded to other medical imaging and patient history.

Statistical analysis

Statistical analysis was performed in Python (V.3.7). The performance of the algorithm was evaluated by comparing the areas under the receiver operating characteristic curves (AUCs) using the paired DeLong test. The accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were also used to assess the performance. Sensitivity, also known as the true positive rate (TPR), was calculated as the percentage of positive

Table 1 Demographic and clinical information of the included patients, split into training, validation and test sets

	Training set	Validation set	Test set 1	Test set 2	Total
Age at diagnosis (years)	67.6±8.2	69.0±4.7	70.7±10.4	69.9±9.2	69.2±8.9
Male (%)	80.5	80.0	79.7	70.7	77.0
Female (%)	19.5	20.0	20.3	29.3	23.0
Never smoker (%)	46.3	50.0	55.1	63.5	54.2
Pulmonary infections (%)	31.7	40.0	30.5	39.0	34.6
Number of AE-IPF (%)	61.0	60.0	23.7	52.4	60.1
PaO ₂ (mm Hg)	75.6±17.7	75.2±19.2	87.1±30.9	76.0±24.0	78.0±23.4
FiO ₂ (mm Hg)	0.30±0.17	0.29±0.16	0.34±0.21	0.38±0.22	0.33±0.20
PaO ₂ /FiO ₂	289.8±107.2	295.2±113.2	307.7±148.5	255.8±141.5	282.4±129.3

AE-IPF, acute exacerbation of idiopathic pulmonary fibrosis.

patients that were correctly identified. Specificity, also known as the true negative rate (TNR), was calculated as the percentage of negative patients that were correctly identified. Accuracy was the percentage of subjects with TPR and TNR. To evaluate interobserver agreement between the algorithm and radiologists, the Cohen's weighted kappa coefficient (κ_w) was used for an estimation of the probability of each diagnosis.²² Weighted κ coefficients were categorised as follows: poor ($0 < \kappa_w \leq 0.20$), fair ($0.20 < \kappa_w \leq 0.40$), moderate ($0.40 < \kappa_w \leq 0.60$), good ($0.60 < \kappa_w \leq 0.80$) and excellent ($0.80 < \kappa_w \leq 1.00$).²³ The correlations of algorithm's categorisation with physiological variables (PaO₂/FiO₂) were evaluated using logistic regression. For all comparisons, a two-sided p value threshold of 0.05 was considered statistically significant. The Python package scikit-learn V.1.0.2 was used for statistical calculation.

RESULTS

Patients characteristics

A total of 306 participants were included after applying inclusion and exclusion criteria. 146 patients were diagnosed with stable IPF, 97 with AE-IPF and 63 were healthy controls. The demographic and clinical information of patients with IPF can be found in [table 1](#). Briefly, the average age of patients undergoing HRCT was 69.2 years, with men accounting for 77.0% of the entire cohort. Among them, 54.2% were never-smokers, 34.6% had concomitant pulmonary infections and 60.1% experienced acute exacerbations. The mean PaO₂/FiO₂ ratio among patients with available blood gas analysis results was 282.4±129.3. In test set 1, 79.7% of patients were men and the mean age was 70.7 years, with 55.1% never-smokers, 30.5% having concomitant pulmonary infections and 23.7% experiencing acute exacerbations. The mean PaO₂/FiO₂ ratio for patients in test set 1 was 307.7±148.5. In test set 2, 70.7% of patients were men and the mean age was 69.9 years, with 63.5% never-smokers, 39.0% having concomitant pulmonary infections and

52.4% experiencing acute exacerbations. The mean PaO₂/FiO₂ ratio for patients in test set 2 was 255.8±141.5.

Classification performance

Deep learning algorithm SlowFast and radiologist performance were evaluated using AUC, accuracy, sensitivity (TPR), specificity (TNR), PPV and NPV. The SlowFast model achieved AUC scores of 0.96 and 0.92 for classifying AE-IPF in test sets 1 and 2, respectively, while the mean radiologist achieved an AUC of 0.91 and 0.81, respectively ([figure 4A,B](#)). Furthermore, in test set 1, SlowFast achieved an accuracy of 93.9%, with a sensitivity of 90.0%, specificity of 95.7%, PPV of 90.0% and NPV of 95.7%. In test set 2, SlowFast achieved an accuracy of 86.5%, with a sensitivity of 80.9%, specificity of 91.8%, PPV of 90.5% and NPV of 83.3% ([table 2](#)). In comparison, in test set 1, radiologists achieved an accuracy of 77.8±9.1%, with a sensitivity, specificity, PPV and NPV of 59.8±12.2%, 87.8±15.2%, 81.1±15.5% and 78.0±7.0%, respectively. In test set 2, radiologists achieved an accuracy of 76.4±8.0%, with a sensitivity, specificity, PPV and NPV of 69.6±34.0%, 82.6±18.3%, 83.7±15.6% and 80.3±16.4%, respectively. For classifying stable-IPF, the SlowFast model achieved an AUC of 0.97 and 0.91 in test sets 1 and 2, respectively ([figure 4C,D](#)). Moreover, the model achieved accuracy, sensitivity, specificity, PPV NPV of 93.9%, 93.9%, 93.9%, 93.9%, 93.9% in test set 1 and 86.5%, 81.6%, 89.7%, 83.8%, 88.1% in test set 2, respectively ([table 2](#)). And radiologists achieved an AUC of 0.87 and 0.79 ([figure 4C,D](#)), and accuracy, sensitivity, specificity, PPV, NPV of 77.8±9.1%, 72.7±24.5%, 84.2±12.9%, 85.4±8.6% and 76.1±14.9% in test set 1, and 76.4±8.0%, 74.6±23.7%, 77.6±25.8%, 75.4±18.2% and 85.4±11.0% in test set 2, respectively ([table 2](#)).

[Figure 4E](#) shows an example of HRCT that was accurately identified as AE-IPF by the SlowFast model but misclassified as an alternative diagnosis by two radiologists. There was one case that was accurately classified

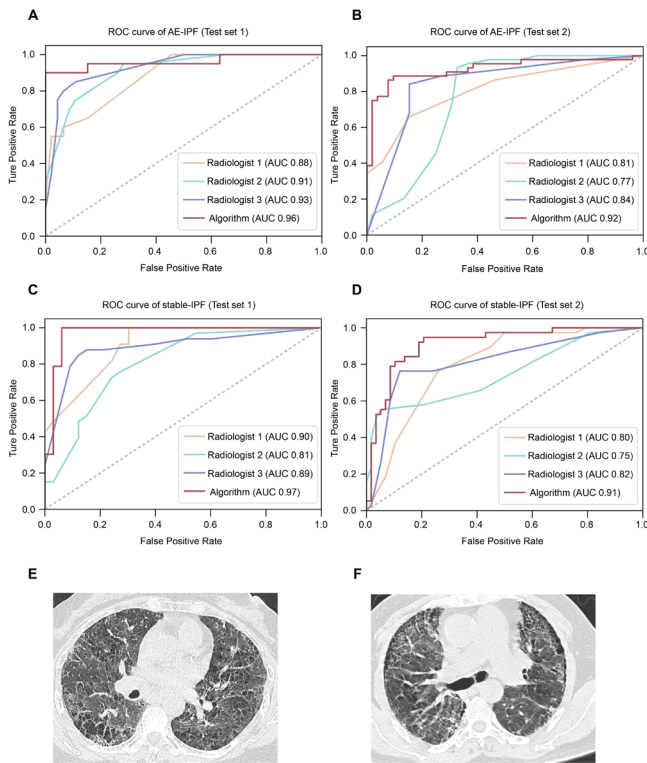


Figure 4 Classification performance of algorithm and radiologists for each HRCT scan on the test set. (A and B) ROC curves of algorithm and individual radiologists for AE-IPF in test set 1(A) and test set 2(B). (C and D) ROC curves of algorithm and individual radiologists for stable-IPF in test set 1(C) and test set 2(D). (E) Selected slices from an HRCT scan were correctly classified as AE-IPF by the algorithm but incorrectly classified by two radiologists. (F) Selected slices from an HRCT scan were correctly classified as AE-IPF by the radiologists but incorrectly classified by the algorithm. AE-IPF, acute exacerbations of idiopathic pulmonary fibrosis; AUC, area under the curve; HRCT, high-resolution CT; ROC, receiver operating characteristic.

as AE-IPF by radiologists but misclassified by SlowFast model (figure 4F).

Interobserver agreement

We used Cohen's weighted kappa coefficient (κ_w) to assess interobserver agreement between the algorithm and radiologists for each diagnostic category.²² We categorised weighted kappa coefficient as follows: poor ($0 < \kappa_w \leq 0.20$), fair ($0.20 < \kappa_w \leq 0.40$), moderate ($0.40 < \kappa_w \leq 0.60$), good ($0.60 < \kappa_w \leq 0.80$) and excellent ($0.80 < \kappa_w \leq 1.00$). In test set 1, median interobserver agreement between the algorithm and the majority opinion of the radiologists was excellent (weighted κ , $\kappa_w = 0.90$), and between each of the thoracic radiologists and the majority opinion of the radiologists was good (weighted κ , $\kappa_w = 0.65 \pm 0.13$) table 3. Similarly, in test set 2, the algorithm had a good interobserver agreement with the majority opinion of the radiologists (weighted κ , $\kappa_w = 0.73$). Additionally, the median interobserver agreement between each of the thoracic

Table 2 Performance of algorithm and radiologists on the test set 1 and test set 2

Data sets	Pattern	Accuracy (%)		Sensitivity (%)		Specificity (%)		PPV (%)		NPV (%)	
		Algorithm	Radiologists	Algorithm	Radiologists	Algorithm	Radiologists	Algorithm	Radiologists	Algorithm	Radiologists
Test set 1	Stable-IPF	93.9	77.8 (68.2–86.4)	93.9	72.7 (44.4–87.9)	93.9	84.2 (71.0–96.7)	93.9	85.4 (76.9–94.1)	93.9	76.1 (59.2–87.5)
	AE-IPF	93.9	77.8 (68.2–86.4)	90.0	59.8 (45.8–68.2)	95.7	87.8 (70.3–97.6)	90.0	81.1 (63.3–88.2)	95.7	78.0 (72.2–85.7)
Test set 2	Stable-IPF	86.5	76.4 (67.7–83.3)	81.6	74.6 (50.0–97.4)	89.7	77.6 (48.3–96.6)	83.8	75.4 (55.2–90.5)	88.1	85.4 (74.7–96.6)
	AE-IPF	86.5	76.4 (67.7–83.3)	80.9	69.6 (31.1–95.5)	91.8	82.6 (63.5–100.0)	90.5	83.7 (68.9–100.0)	83.3	80.3 (62.2–94.3)

AE-IPF, acute exacerbation of idiopathic pulmonary fibrosis; IPF, idiopathic pulmonary fibrosis; NPV, negative predictive value; PPV, positive predictive value.

Table 3 Weighted κ values between the algorithm, thoracic radiologists and radiologists' majority opinion

	Algorithm	Radiologist 1	Radiologist 2	Radiologist 3	P value
Majority of test set 1	0.90	0.65	0.52	0.78	<0.01
Majority of test set 2	0.73	0.50	0.69	0.88	<0.05

radiologists and the majority opinion of the radiologists was good (weighted κ , $\kappa_w=0.69\pm 0.19$) [table 3](#).

Correlation between the model's categorisation and prognosis

To investigate the prognostic value of the deep learning model in AE-IPF, logistic regression was used to assess the correlations between the model's categorisation and PaO₂/FiO₂ ratio, which is a prognostic factors of AE-IPF.²⁴ The result indicated that both the model's categorisation and the radiologists' majority opinion were the significant predictors of the disease severity of IPF (p<0.0001, OR=1.007, 95% CI=1.003 to 1.010 for the model's categorisation, and p<0.0001, OR=1.007, 95% CI=1.004 to 1.011 for radiologists' majority opinion, respectively) ([figure 5A,B](#)).

DISCUSSION

In this study, we investigated the potential of the deep learning algorithm SlowFast to classify AE-IPF using HRCT scans. In addition, we compare the performance of this deep learning model to the diagnostic performance of radiologists. Our study suggested that the model provided almost instantaneous reporting with accuracy and reproducibility comparable to human experts (AUC 0.96 vs 0.91 in test set 1, AUC 0.92 vs 0.81 in test set 2). As a fatal complication of IPF, the accurate classification of AE-IPF plays a crucial role in improving prognostication, directing patient treatment and facilitating research.^{25–27} AE-IPF shares similar pathophysiological characteristics with acute respiratory distress syndrome, which can be triggered by COVID-19 and is considered one of the major causes of increased mortality.^{28 29} However, due to the complicated clinical course, making accurate diagnoses for patients with AE-IPF is a significant challenge for clinicians.³⁰ Notably, HRCT-based deep learning models and diagnostic biomarkers for ILDs have garnered

widespread attention in the precision medicine diagnosis of IPF.^{31–35} Under these circumstances, the importance of using deep learning models to assist in the accurate diagnosis of AE-IPF by radiologists becomes evident, which could provide cheap and consistent patient stratification for clinical trials, thereby reducing failures during screening and costs. Moreover, there is a pressing clinical need to identify contributing or alternative causes of decline in patients with GGO and/or consolidation on a background of IPF.³⁶ Therefore, predicting the future functional decline or the occurrence of AE-IPF would be a valuable and unmet objective, which could be addressed by the application of advanced deep learning techniques to the analysis of HRCT scans.

Previous research has explored the potential of deep learning algorithms to classify fibrotic lung disease on chest HRCT scans. Walsh *et al* developed a deep learning algorithm for classifying usual interstitial pneumonia (UIP) on HRCT based on the neural network architecture, which achieved human-level accuracy (76.4% vs 70.7% on the test set).¹⁴ Alex *et al* employed a custom deep learning algorithm to predict histopathological diagnosis (UIP vs non-UIP) from chest CT patterns, which provided better diagnostic performance than visual evaluation (AUC 0.87 vs 0.80; p=0.03).³⁷ In addition, Kim *et al* applied content-based image retrieval (CBIR) to improve the diagnostic accuracy for patients with ILD (before vs after CBIR, 46.1% vs 60.9%).³⁸ In the study by Tzouveleakis *et al*, a machine learning software system (Imbio V.1.4.2.) was used to evaluate HRCT in patients with non-IPF ILDs receiving mycophenolate mofetil. The software demonstrated similar performance to specialist radiologists, indicating its potential as a valuable diagnostic and prognostic tool (ICC 0.73 vs 0.88).³⁹ In addition to providing diagnostic support, some studies also focus on the early detection or prediction of progressive fibrotic lung disease. In the study by Agarwala *et al*, a deep learning framework was conducted to automatically identify ILD patterns in HRCT images, achieving an 86% success rate and 74% sensitivity in sections with lung fibrosis.⁴⁰ Besides, Simon *et al* developed a deep learning algorithm SOFIA and demonstrated that it improved outcome prediction in patients with progressive fibrotic lung disease when compared with radiologist evaluation (HR, 1.73; p<0.0001; 95% CI=1.40 to 2.14).¹⁵ Notably, researchers have endeavoured to address two major barriers in the management of ILD: the diagnosis of disease subtypes and the predicting of patient prognosis. Yang *et al* employed RadImageNet pretrained models to diagnose five types of ILD and a transformer model to determine a

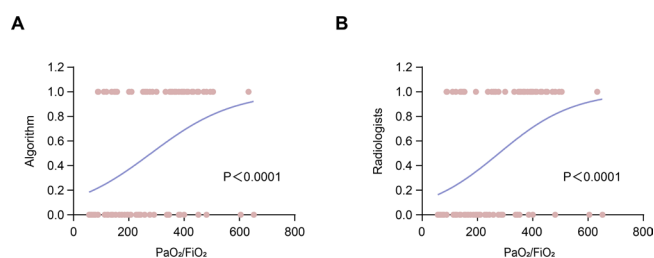


Figure 5 Logistic regression curves of the algorithm and radiologists. (A) The logistic regression curve of the model's categorisation and PaO₂/FiO₂ ratio. (B) The logistic regression curve of the radiologists' categorisation and PaO₂/FiO₂ ratio.

patient's 3-year survival rate, which proves to be a useful tool to distinguish ILD subcategories and manage the long-term progression of patients.⁴¹

However, at the moment, there has been limited research on using deep learning for the diagnosis of AE-IPF, with most deep learning models being trained on 2D data. Our study highlights several advantages of applying this deep learning model to image analysis in fibrotic lung disease. First, our model achieved more extraordinary diagnostic performance than visual evaluation. Second, we innovatively used the 3D-video-sequence-based methodology SlowFast for image analysis on HRCT scans, which provides sequential cross-sectional images of the lungs. This approach allows for more objective analysis compared with traditional 2D-image-based models, and is the first application of this 3D-video-sequence-based methodology on HRCT scans. Third, we directly compared the performance of our model with that of radiologists, and our model demonstrated the potential to outperform the established chest CT classification scheme based on visual analysis.

It should be noted that our study suffers from a few limitations. First, due to the low incidence of AE-IPF in the general patient population,⁴² the number of cases for model training and testing was small. Although we have employed external validation to confirm the transportability and generalisability of our model, we acknowledge the need for a large-scale, multicentre study of AE-IPF, which could lead to the development of more robust and effective algorithms. Second, only cases of stable IPF, AE-IPF and healthy control were covered in this study. Therefore, algorithm performance for other ILD subtypes is unknown. Further versions of the algorithm will include an extension to cover these other patterns. Despite this, including healthy control subjects in the data set provides a valuable reference point for comparison with patients with IPF. This approach may help to identify HRCT scan features that are specific to the disease and facilitate the development of more robust models. Third, the algorithm was designed to alleviate the workload, improve accuracy and enhance consistency in challenging diagnoses made by radiologists. Nevertheless, the performance of the algorithm was only benchmarked against three radiologists, which may not accurately represent the entire spectrum of human capabilities.

In conclusion, we have developed a deep learning algorithm with similar performance to a human reader for classifying AE-IPF on HRCT scans. In principle, this algorithm has the potential to provide low-cost, consistent patient stratification and assist in radiological decision-making.

Author affiliations

¹Department of Respiratory and Critical Care Medicine, Nanjing Drum Tower Hospital, The Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu, China

²Nanjing Institute of Respiratory Diseases, Nanjing, Jiangsu, China

³Purple Mountain Laboratories, Nanjing, Jiangsu, China

⁴Department of Respiratory and Critical Care Medicine, Nanjing Drum Tower Hospital Clinical College of Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

⁵Department of Respiratory and Critical Care Medicine, Nanjing Drum Tower Hospital Clinical College of Nanjing Medical University, Nanjing, Jiangsu, China

Contributors XH: conceptualisation (lead); data curation (lead); investigation (equal); methodology (equal); project administration (equal); writing—original draft (lead); writing—review and editing (lead). WS: conceptualisation (equal); formal analysis (lead); investigation (equal); methodology (equal); software (lead); writing—review and editing (supporting). XY: investigation (supporting); resources (supporting); writing—review and editing (supporting). YZ: investigation (supporting); methodology (supporting); writing—review and editing (supporting). HG: investigation (supporting); validation (lead); writing—review and editing (supporting). MZ: investigation (supporting); validation (supporting); visualisation (supporting); writing—review and editing (supporting). SW: investigation (supporting); software (supporting); writing—review and editing (supporting). YS: investigation (supporting); validation (supporting); writing—review and editing (supporting). XG: supervision (equal); writing—review and editing (supporting). YX: funding acquisition (equal); supervision (equal); writing—review and editing (supporting). MC: funding acquisition (equal); supervision (equal); visualisation (lead); writing—review and editing (supporting).

Funding This work was supported by National Natural Science Foundation of China (82070064, 81670059 and 81200049), Natural Science Found of Jiangsu province (SBK20230140), Fundings for Clinical Trials from the Nanjing University Medical School Affiliated Drum Tower Hospital (2022-LCYJ-MS-11), Special fund project for clinical research of Nanjing Drum Tower Hospital (2021-LCYJ-DBZ-06).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This retrospective study was approved by IRB of Nanjing Drum Tower Hospital of Nanjing University Medical School (approval: 23-01-18) with a waiver for written informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Martinez FJ, Collard HR, Pardo A, *et al*. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017;3:17074.
- Leuschner G, Behr J. Acute exacerbation in interstitial lung disease. *Front Med (Lausanne)* 2017;4:176.
- Collard HR, Ryerson CJ, Corte TJ, *et al*. Acute exacerbation of idiopathic pulmonary fibrosis. an international working group report. *Am J Respir Crit Care Med* 2016;194:265–75.
- Juarez MM, Chan AL, Norris AG, *et al*. Acute exacerbation of idiopathic pulmonary fibrosis—a review of current and novel pharmacotherapies. *J Thorac Dis* 2015;7:499–519.
- Kishaba T, Tamaki H, Shimaoka Y, *et al*. Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. *Lung* 2014;192:141–9.
- Fujimoto K, Taniguchi H, Johkoh T, *et al*. Acute exacerbation of idiopathic pulmonary fibrosis: high-resolution CT scores predict mortality. *Eur Radiol* 2012;22:83–92.

- 7 Hirano C, Ohshimo S, Horimasu Y, *et al.* Baseline high-resolution CT findings predict acute exacerbation of idiopathic pulmonary fibrosis: German and Japanese cohort study. *J Clin Med* 2019;8:2069.
- 8 Gruden JF. CT in idiopathic pulmonary fibrosis: diagnosis and beyond. *AJR Am J Roentgenol* 2016;206:495–507.
- 9 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;6:837–45.
- 10 De Fauw J, Ledsam JR, Romera-Paredes B, *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24:1342–50.
- 11 Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
- 12 Humphries SM, Swigris JJ, Brown KK, *et al.* Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. *Eur Respir J* 2018;52:1801384.
- 13 Putman RK, Gudmundsson G, Axelsson GT, *et al.* Imaging patterns are associated with interstitial lung abnormality progression and mortality. *Am J Respir Crit Care Med* 2019;200:175–83.
- 14 Walsh SLF, Humphries SM, Wells AU, *et al.* Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. *Lancet Respir Med* 2020;8:1144–53.
- 15 Walsh SLF, Mackintosh JA, Calandriello L, *et al.* Deep learning-based outcome prediction in progressive fibrotic lung disease using high-resolution computed tomography. *Am J Respir Crit Care Med* 2022;206:883–91.
- 16 Cordelier P, Costa M, Fehrenbach J. Slow-fast model and therapy optimization for oncolytic treatment of tumors. *Bull Math Biol* 2022;84:64.
- 17 Feichtenhofer C, Fan H, Malik J, *et al.* Slowfast networks for Video recognition; 2019. 6202–11.
- 18 Jacob J, Hansell DM. HRCT of fibrosing lung disease. *Respirology* 2015;20:859–72.
- 19 Raghu G, Remy-Jardin M, Richeldi L, *et al.* 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* 2022;205:e18–47.
- 20 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:e44–68.
- 21 Russell BC, Torralba A, Murphy KP, *et al.* Labelme: a database and web-based tool for image annotation. *Int J Comput Vis* 2008;77:157–73.
- 22 Walsh SLF, Maher TM, Kolb M, *et al.* Diagnostic accuracy of a clinical diagnosis of idiopathic pulmonary fibrosis: an international case-cohort study. *Eur Respir J* 2017;50:1700936.
- 23 Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304:1491–4.
- 24 Kamiya H, Panlaqui OM. Systematic review and meta-analysis of prognostic factors of acute exacerbation of idiopathic pulmonary fibrosis. *BMJ Open* 2020;10:e035420.
- 25 Sakamoto S, Shimizu H, Isshiki T, *et al.* New risk scoring system for predicting 3-month mortality after acute exacerbation of idiopathic pulmonary fibrosis. *Sci Rep* 2022;12:1134.
- 26 Tzouveleki A, Kourtidou S, Bouros E, *et al.* Impact of depression on patients with idiopathic pulmonary fibrosis. *Front Med (Lausanne)* 2020;7:29.
- 27 Tzilas V, Tzouveleki A, Bouros E, *et al.* Clinical experience with antifibrotics in fibrotic hypersensitivity pneumonitis: a 3-year real-life observational study. *ERJ Open Res* 2020;6.
- 28 Kyriakopoulos C, Gogali A, Exarchos K, *et al.* Reduction in hospitalizations for respiratory diseases during the first COVID-19 wave in Greece. *Respiration* 2021;100:588–93.
- 29 Tzouveleki A, Karampitsakos T, Krompa A, *et al.* False positive COVID-19 antibody test in a case of granulomatosis with polyangiitis. *Front Med (Lausanne)* 2020;7:399.
- 30 Walsh SLF, Calandriello L, Sverzellati N, *et al.* Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016;71:45–51.
- 31 Karampitsakos T, Juan-Guardela BM, Tzouveleki A, *et al.* Precision medicine advances in idiopathic pulmonary fibrosis. *EBioMedicine* 2023;95:104766.
- 32 Ntatsoulis K, Karampitsakos T, Tsitoura E, *et al.* Commonalities between ARDS, pulmonary fibrosis and COVID-19: the potential of Autotaxin as a therapeutic target. *Front Immunol* 2021;12:687397.
- 33 Tzouveleki A, Antoniou K, Kreuter M, *et al.* The DIAMORFOSIS (diagnosis and management of lung cancer and fibrosis) survey: International survey and call for consensus. *ERJ Open Res* 2021;7.
- 34 Karampitsakos T, Spagnolo P, Mogulkoc N, *et al.* Lung cancer in patients with idiopathic pulmonary fibrosis: a retrospective Multicentre study in Europe. *Respirology* 2023;28:56–65.
- 35 Xylourgidis N, Min K, Ahangari F, *et al.* Role of dual-specificity protein phosphatase DUSP10/MKP-5 in pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2019;317:L678–89.
- 36 Song JW, Hong SB, Lim CM, *et al.* Acute exacerbation of idiopathic pulmonary fibrosis: incidence, risk factors and outcome. *Eur Respir J* 2011;37:356–63.
- 37 Bratt A, Williams JM, Liu G, *et al.* Predicting usual interstitial pneumonia histopathology from chest CT imaging with deep learning. *Chest* 2022;162:815–23.
- 38 Choe J, Hwang HJ, Seo JB, *et al.* Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology* 2022;302:187–97.
- 39 Karampitsakos T, Kalogeropoulou C, Tzilas V, *et al.* Safety and effectiveness of mycophenolate mofetil in interstitial lung diseases: insights from a machine learning radiographic model. *Respiration* 2022;101:262–71.
- 40 Agarwala S, Kale M, Kumar D, *et al.* Deep learning for screening of interstitial lung disease patterns in high-resolution CT images. *Clin Radiol* 2020;75:481.
- 41 Mei X, Liu Z, Singh A, *et al.* Interstitial lung disease diagnosis and prognosis using an AI system integrating longitudinal data. *Nat Commun* 2023;14:2272.
- 42 Biondini D, Balestro E, Sverzellati N, *et al.* Acute exacerbations of idiopathic pulmonary fibrosis (AE-IPF): an overview of current and future therapeutic strategies. *Expert Rev Respir Med* 2020;14:405–14.