



## OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>1</sup>Department of ECE, R.M.D Engineering College, Chennai, Tamil Nadu, India

<sup>2</sup>Department of ECE, Dr. M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India

<sup>3</sup>Department of ECE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Correspondence to: Jagadeesh Purushothaman, jagadeeshnarpavi@gmail.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Purushothaman J, Premkumar S, Balachander B and Amudha V. FibNet: A Deep Learning-Based Interpretable System for Automated Diagnosis of Pulmonary Fibrosis. Premier Journal of Science 2026;16:100256  
DOI: <https://doi.org/10.70389/PJS.100256>

### Peer Review

Received: 31 October 2025

Last revised: 1 January 2026

Accepted: 2 January 2026

Version accepted: 3

Published: 31 January 2026

Ethical approval: N/a

Consent: N/a

Funding: No funding

Conflicts of interest: No conflict of interest

# FibNet: A Deep Learning-Based Interpretable System for Automated Diagnosis of Pulmonary Fibrosis

Jagadeesh Purushothaman<sup>1,2,3</sup> , Bhuvaneshwari Balachander<sup>2</sup>, Premkumar Selvam<sup>3</sup> and Amudha Veerapan<sup>2</sup>

## ABSTRACT

Pulmonary fibrosis (PF) is a progressive interstitial lung disease (ILD) that severely impairs respiratory function and can lead to fatal outcomes. Accurate and early diagnosis from chest computed tomography (CT) scans is essential for timely treatment but remains challenging due to visual similarities with other lung conditions. Conventional manual assessment is time-consuming and prone to inter-observer variability, while traditional deep learning models like VGG16, ResNet50, and DenseNet121 often struggle with optimal accuracy and clinical interpretability. To address these challenges, we propose FibNet, an EfficientNet-B3-based deep learning framework enhanced with Squeeze-and-Excitation (SE) attention and Grad-CAM++ for improved accuracy and explainability. The model incorporates preprocessing steps including lung segmentation and Hounsfield unit normalization to enhance feature representation. Validation was conducted on two distinct publicly available datasets: the Open Source Imaging Consortium (OSIC) PF Progression dataset comprising 176 patients with more than 24,000 CT slices, and the ILD dataset from the Lung Tissue Research Consortium containing 128 patients with approximately 18,000 CT slices. Each dataset was split 70:30 for training and testing. The implementation was carried out using TensorFlow 2.12 with the Keras API on an NVIDIA RTX 3090 GPU (24 GB VRAM). Experimental results show that FibNet achieved an accuracy of 0.978, Dice coefficient of 0.97, IoU of 0.95, and Matthews Correlation Coefficient (MCC) of 0.96 on the OSIC dataset, and 0.968, 0.96, 0.93, and 0.95 respectively on the ILD dataset, outperforming all baseline models. In conclusion, FibNet provides a computationally efficient and interpretable solution for PF diagnosis, with robust validation across multiple datasets, offering strong potential for real-world clinical integration.

**Keywords:** Pulmonary fibrosis CT diagnosis, EfficientNet-B3 architecture, Squeeze-and-excitation attention, Grad-CAM++ interpretability, Hounsfield unit normalization

## Introduction

Pulmonary fibrosis (PF) diagnosis remains a highly challenging task despite continuous improvements in medical imaging technology, primarily due to subtle visual similarities with other pulmonary disorders.<sup>1</sup> For instance, parenchymal destruction and emphysematous alterations observed in chronic obstructive pulmonary disease (COPD) may resemble fibrotic patterns, while non-specific interstitial pneumonia often exhibits reticulation and ground-glass opacities that closely mimic PF manifestations.<sup>2</sup> Such overlapping

radiological characteristics significantly complicate differential diagnosis. Moreover, manual interpretation of computed tomography (CT) scans is time-consuming and susceptible to inter-observer variability, frequently leading to inconsistent clinical assessments.<sup>3</sup>

Although conventional convolutional neural networks (CNNs) such as VGG16, ResNet50, and DenseNet121 have demonstrated improved diagnostic performance, they typically involve a large number of parameters, lack domain-specific optimization, and offer limited interpretability, thereby restricting their clinical applicability. In addition, many existing deep learning models for PF diagnosis are often opaque, exhibit poor generalization across datasets and scanners, and impose substantial computational overheads. These approaches also rarely integrate preprocessing and attention mechanisms within a unified end-to-end explainable framework. To address these limitations, this article proposes *FibNet*, a deep-learning architecture built upon the EfficientNet-B3 backbone, explicitly designed to achieve robust performance with enhanced interpretability and computational efficiency.

Recent survey studies highlight the growing adoption of deep learning techniques for PF diagnosis.<sup>4</sup> However, the present work advances the state of the art by integrating EfficientNet-B3 with squeeze-and-excitation (SE) attention and Grad-CAM++ visualization within a single end-to-end framework. This unified design simultaneously improves classification accuracy, visual interpretability, and cross-dataset generalization, a combination that has not been comprehensively addressed in prior PF studies. This study introduces FibNet, a deep-learning architecture that employs EfficientNet-B3 as the backbone for effective feature extraction, incorporates SE blocks for adaptive channel attention, and utilizes Grad-CAM++ for enhanced visual interpretability. The model is designed to perform both classification and lesion localization, ensuring that the decision-making process is transparent to clinicians. The inclusion of preprocessing steps such as lung segmentation, Hounsfield unit normalization, and patch slicing further enhances feature consistency and reduces background noise.

To guide the development and evaluation of the proposed framework, this study seeks to answer the following research questions:

- RQ1: Can an EfficientNet-B3 backbone augmented with SE attention outperform traditional CNN architectures in PF classification?
- RQ2: Can Grad-CAM++ provide clinically interpretable localization of fibrotic regions in CT scans?

Author contribution:  
Jagadeesh Purushothaman –  
Writing –

original draft, Visualization,  
Validation; S. Premkumar –  
Funding acquisition,  
Formal analysis, Data  
curation, Conceptualization;  
Bhuvanewari Balachander –  
Software, Resources, Project  
administration; V. Amudha –  
Writing –

review & editing, Validation,  
Investigation, Data curation

Guarantor: Jagadeesh  
Purushothaman

Provenance and peer-review:  
Unsolicited and externally peer-  
reviewed

Data availability statement:  
N/a

- RQ3: Can the proposed FibNet framework generalize effectively across different datasets and imaging modalities?
- RQ4: Is the integrated end-to-end pipeline computationally efficient enough for practical clinical deployment.

The remainder of the article is organized as follows: Section “Related Work” reviews existing literature on PF diagnosis and interpretable deep-learning methods. Section “System Methodology” presents the proposed FibNet architecture and preprocessing pipeline. Section “Experimental Results” describes the datasets, experimental setup, and evaluation metrics. Section “Discussion” discusses the performance results, interpretability analysis, and ablation studies. Section “Conclusion” concludes the paper and outlines potential future directions.

### Related Work

Several recent studies have explored deep learning and machine learning approaches for the detection and classification of pulmonary and other medical diseases using radiological images, electronic health data, and multimodal frameworks. PF/interstitial lung disease (ILD)-specific CT recent studies have integrated deep learning for predictive modeling. An example is Chantzi et al.<sup>5</sup> reviewed the application of radiomics and AI in PF, and pointed out the need of having explainable systems at these cross-regional collaborations level. Longitudinal multi-center studies, such as Zhang et al.<sup>6</sup> for combined PF and emphysema cohort studies, have demonstrated the need for dependable and versatile models to avoid overfitting to specific scanner types or acquisition protocols. Our work adds to these developments through the incorporation of explainability and attention mechanisms in a compact and efficient pipeline, segmentation-aware and empirically validated on two public datasets. Souid et al.<sup>7</sup> developed a fast-staged CNN for pulmonary disease and lung cancer detection, achieving efficient classification with high accuracy. The method, however, was dataset-specific, limiting broader applicability. Oltu et al.<sup>8</sup> proposed a deep learning-based chest X-ray classifier that effectively distinguished disease categories, though its accuracy was resolution-dependent and lacked multi-center validation. Nanthini et al.<sup>9</sup> introduced a DL framework for lung disease prediction integrating preprocessing and classification, but evaluation was limited to a single dataset without cross-domain testing. Borate et al.<sup>10</sup> compared multiple machine learning algorithms for lung disease prediction, highlighting accuracy-complexity trade-offs, yet relying solely on structured clinical data without medical imaging. Sun et al.<sup>11</sup> applied deep learning to bronchoscopy images for respiratory disease prediction, achieving high accuracy but without interpretability assessment. Cai et al.<sup>12</sup> proposed a local-to-global framework for COPD diagnosis using CT scans, effectively capturing fine and global features, but requiring high computational resources. Nguyen and Vo<sup>13</sup> developed a deep learning pipeline for

detecting lung diseases from X-rays, showing promising benchmark results but without testing under noisy or degraded images. Zhou et al.<sup>14</sup> proposed an adaptive multiscale feature fusion approach for chest radiography, achieving accurate recognition of pulmonary diseases through enhanced feature integration. Aljuaid et al.<sup>15</sup> developed RADAI, a deep learning framework for chest X-ray classification, which demonstrated strong performance in detecting diverse lung abnormalities. Khaled et al.<sup>16</sup> evaluated ML models for liver disease detection with big data, but did not consider applicability to lung diseases. Overall, the literature shows notable progress in disease detection using deep learning, though challenges remain in generalization, interpretability, and low-resource optimization. In addition to PF-specific models, most recent studies have investigated new architectural and multimodal approaches to the analysis of lung diseases with methodological implications applicable to our study. An example is a hybrid deep learning model with convolutional and recurrent layers to detect COVID-19 based on CT scans that is suggested by Khan et al.<sup>17</sup> and underlines the effectiveness of sequential feature learning in respiratory diagnostics. Their work, though focused on the classification of infectious diseases, highlights the severity of the hybrid architecture in terms of progressive parenchymal changes, which applies to the entire process of fibrosis progression monitoring. Going in the opposite direction, Patel et al.<sup>18</sup> presented a multimodal fusion method that combines CT images with clinical metadata to classify ILDs. Their system indicated that patient-specific clinical variables (i.e., age, smoking history) when used with imaging characteristics enhanced the accuracy of diagnosis when compared to image-only models. Although our present work is oriented to the analysis with the help of the CT technique, their results point to the significant direction of the future: the combination of multimodal data in order to provide it with strength and clinical significance. These articles demonstrate the current trends of hybrid modelling and multimodal fusion in pulmonary image analysis. Our FibNet framework is in line with these directions as it uses attention mechanisms and CT-specific preprocessing pipelines but also prepares the way to the further extension to multimodal and temporal modeling to monitor fibrosis (Table 1).

### Problem Statement and Justification

PF and related ILDs present significant diagnostic challenges due to overlapping visual patterns, subtle texture variations, and heterogeneous disease manifestations in high-resolution computed tomography (HRCT) scans. Conventional deep learning models such as VGG16, ResNet50, and DenseNet121, while effective for generic medical image classification, often lack domain-specific optimizations to capture fine-grained pathological features, leading to suboptimal accuracy and limited interpretability. Additionally, variability in image acquisition protocols and the need for robust cross-dataset generalization further

**Table 1 | Summary of related work**

Author [Ref]	Method	Findings	Limitations
Souid et al. <sup>7</sup>	Fast-staged CNN	Improved classification; competitive accuracy	Dataset-specific tuning limits generalization
Oltu et al. <sup>8</sup>	DL chest X-ray	High accuracy in disease differentiation	Dependent on resolution; limited multi-center tests
Nanthini et al. <sup>9</sup>	DL with preprocessing	Effective lung disease prediction	Single dataset; no cross-domain test
Borate et al. <sup>10</sup>	ML algorithm comparison	Accuracy–computation trade-offs	Structured data only; no imaging
Sun et al. <sup>11</sup>	DL on bronchoscopy	High diagnostic accuracy	No interpretability assessment
Cai et al. <sup>12</sup>	Local-to-global DL	Captures fine + holistic CT features	High computational cost
Nguyen and Vo <sup>13</sup>	DL on X-ray	Promising benchmark results	No noisy/low-quality analysis
Zhou et al. <sup>14</sup>	Workflow + cognition	Better diagnosis workflow	Limited technical details
Aljuaid et al. <sup>15</sup>	DL for plant images	Adaptable to biomedical imaging	No pulmonary-specific focus
Khaled et al. <sup>16</sup>	ML for liver disease	Predictive analytics potential	No lung-domain shift study

complicate automated diagnosis. To address these issues, the proposed *FibNet* framework leverages an EfficientNet-B3 backbone enhanced with SE attention for adaptive feature recalibration, Grad-CAM++ for visual interpretability, and domain-specific preprocessing steps including lung segmentation and Hounsfield unit normalization. This integrated design not only improves classification accuracy and Dice/Intersection over Union (IoU) metrics but also enhances clinical trust through explainable AI outputs, making it more suitable for real-world deployment in multi-center healthcare environments.

**System Methodology**

The proposed *FibNet* framework for automated PF diagnosis consists of four main stages: data preprocessing, feature extraction, attention-enhanced classification, and interpretability analysis. The workflow, as illustrated in Figure 1, is designed to optimize predictive accuracy while ensuring clinical transparency.

**Data Preprocessing**

The raw high-resolution CT scans are first subjected to lung region segmentation using a U-Net-based model to remove irrelevant anatomical structures. The U-Net follows a standard encoder–decoder design with four downsampling and upsampling blocks, each consisting of convolutional layers with batch normalization and ReLU activation. Skip connections are incorporated to preserve spatial detail during reconstruction. The network was trained on a publicly available lung segmentation dataset with expert-annotated CT slices, using the Dice loss function to optimize overlap between predictions and ground truth. Training was performed for 100 epochs with the Adam optimizer (learning rate  $1 \times 10^{-4}$ ) and a batch size of 8. After segmentation, voxel intensities are normalized to Hounsfield units (HU) in the range  $[-1000, 400]$  to preserve lung tissue contrast. The segmented volumes are then divided into non-overlapping patches of size  $300 \times 300$ , which focus on localized fibrosis patterns while enabling efficient batch processing. Such preprocessing steps have been shown to enhance feature consistency and improve fibrosis detection performance in prior studies.<sup>16</sup> The lung segmentation model used in the preprocessing stage was a U-Net architecture trained in-house on publicly available CT lung-mask datasets (LUNA16 and Vessel-12). The trained U-Net achieved a Dice coefficient of approximately 0.98 and an IoU of 0.97 on its

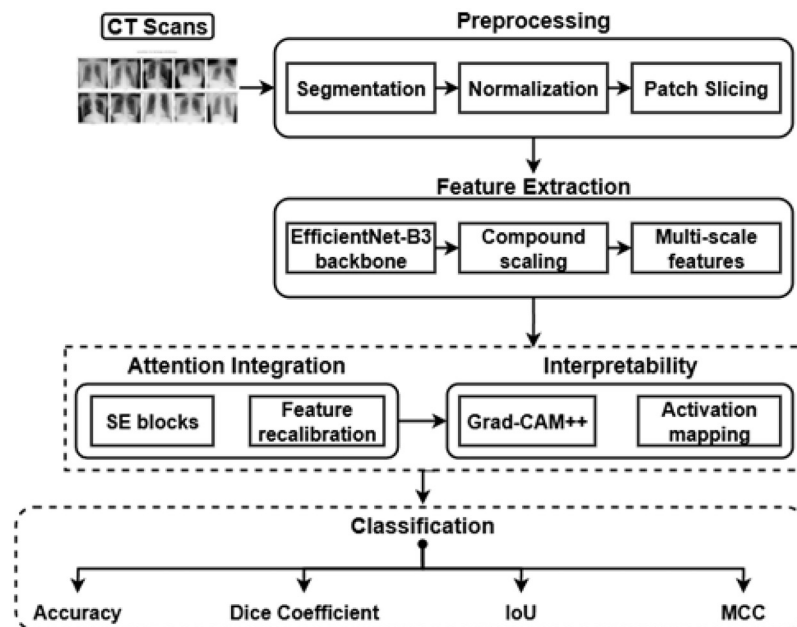


Fig 1 | Block diagram of the proposed FibNet framework

internal validation set, demonstrating reliable delineation of lung parenchyma for subsequent analysis. Typical failure cases included incomplete segmentation near the apices and occasional artefacts at basal slices affected by motion, which are acknowledged as known limitations of the preprocessing step. All splits were performed at the patient level to avoid any information leakage between training and testing sets.

**Feature Extraction**

Feature extraction is carried out using the EfficientNet-B3 backbone, which balances accuracy and computational cost through compound scaling. Given an input patch  $X \in \mathbb{R}^{300 \times 300 \times 3}$ , the convolutional transformation is represented in Equation (1): F. Training Objective.

The model is trained using the Binary Cross-Entropy (BCE) loss, given in (6):

$$\text{where } L_{\text{BCE}} = \frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad y_i$$

is the ground truth label and  $\hat{y}_i$  is the predicted probability for the  $i$ -th sample.

$$F = \phi_{\theta}(X) \tag{1}$$

where  $\phi_{\theta}(\cdot)$  denotes the convolutional feature mapping parameterized by weights  $\theta$ , and  $F$  represents the resulting feature tensor.

**Attention Mechanism**

To enhance the network’s focus on clinically significant areas, a SE block is applied. The squeeze operation aggregates global channel information using Equation (2):

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c(i, j) \tag{2}$$

where  $F_c$  is the  $c$ -th channel of the feature map. The excitation step adaptively reweights channels as shown in Equation (3):

$$s_c = \sigma(W_2 \delta(W_1 z_c)) \tag{3}$$

where  $W_1, W_2$  are learnable weights,  $\delta$  is the ReLU function, and  $\sigma$  is the sigmoid activation.

The SE block was preferred over more complex mechanisms such as CBAM, since it provides lightweight channel recalibration that effectively highlights subtle fibrosis-related texture patterns while avoiding additional spatial attention overhead.

**Classification Layer**

The refined features are flattened and passed through a fully connected layer with sigmoid activation for binary classification, computed using Equation (4):

$$\hat{y} = \sigma(W_f F' + b_f) \tag{4}$$

where  $F'$  is the attention-modulated feature vector,  $W_f$  and  $b_f$  are classifier parameters, and  $\hat{y}$  is the predicted probability of fibrosis.

**Interpretability Analysis**

Grad-CAM++ is employed to generate pixel-level heatmaps highlighting regions contributing most to the model’s decision. The class-specific importance weights are calculated using Equation (5):

$$\alpha_k^c = \frac{\partial^2 y^c}{\partial A_k^2} / \left( 2 \frac{\partial^2 y^c}{\partial A_k^2} + \sum_i \sum_j A_k^{ij} \frac{\partial^3 y^c}{\partial A_k^3} \right) \tag{5}$$

where  $A_k$  denotes the  $k$ -th feature map and  $y^c$  is the class score were extracted from this dataset, with expert-annotated labels for each class.

The complete pipeline is outlined in Algorithm 1, which details each stage from raw CT input to fibrosis prediction and interpretability generation.

Algorithm 1   Proposed FibNet workflow	
1:	Input: CT scan slices from OSIC and ILD datasets
2:	Perform lung segmentation to isolate lung regions
3:	Apply Hounsfield unit normalization for intensity standardization
4:	Slice CT scans into patches of 300 × 300 pixels
5:	Pass preprocessed images to EfficientNet-B3 backbone
6:	Integrate SE attention mechanism
7:	Generate deep feature maps from backbone output
8:	Perform classification using fully connected layers
9:	Apply Grad-CAM++ to visualize fibrosis-relevant regions
10:	Evaluate using Accuracy, Dice, IoU, and MCC
11:	Output final fibrosis prediction and interpretability maps

**Experimental Results**

This section presents a comprehensive evaluation of the proposed FibNet model across multiple publicly available datasets. Detailed analyses include performance metrics, interpretability assessment, ablation studies, computational efficiency, and training behavior.

**Dataset Summary**

In this study, two publicly available and widely recognized datasets were used to train and evaluate the proposed FibNet model for the automated diagnosis of PF (Table 2). The first dataset, OSIC Pulmonary Fibrosis Progression, contains HRCT scans of 176 patients diagnosed with PF. Along with imaging data, the dataset provides Forced Vital Capacity (FVC) values and risk of fibrosis progression, which are essential for correlating imaging features with clinical severity. A total of more than 24,000 CT slices were utilized, ensuring adequate variability in both disease presentation and scan quality. The second dataset, ILD

**Table 2 | Dataset summary**

Dataset	Patients	CT Slices	Labels
OSIC Pulmonary Fibrosis	176	24,000+	FVC values, fibrosis risk
ILD Dataset	128	18,000	Fibrosis, ground-glass, reticulation

Dataset (Lung Tissue Research Consortium), comprises CT scans from 128 patients with various ILD patterns. These include fibrosis, ground-glass opacities, and reticulation – key radiological manifestations relevant to differential diagnosis in fibrotic lung disorders. Approximately 18,000 CT slices.

For model development, a 70:30 split was applied to both datasets, allocating 70% of the data for training and 30% for testing. Although the datasets provide valuable diversity, the distribution of classes is imbalanced (e.g., fibrosis cases outnumber ground-glass and reticulation). To address this, we applied stratified splitting to preserve class ratios across training and testing, used class-weighting in the loss function to penalize underrepresented categories, and performed patch-level data augmentation (random rotations, flips, and scaling) to synthetically increase minority class samples. These strategies mitigated bias and supported balanced learning, ensuring that FibNet's results were not skewed toward dominant classes.

This study considers the detection of PF as a binary classification problem at the level of data patches (300 × 300 pixel areas taken from CT slices). Starting with the OSIC dataset, the ground-truth labels were based on clinically documented declines in Forced Vital Capacity (FVC) (10% over a 6-month period), and patches from the patients with marked declines in FVC were labeled as fibrotic, while the others were labeled as non-fibrotic. Regarding the ILD dataset, expert radiologist annotations at the slice level were used to label patches with fibrosis (honeycombing and reticular) as positive and those without such patterns as negative. This binary mapping prioritizes clinically significant fibrotic areas and ensures consistency across the datasets, and the validity of the mapping was confirmed with Grad-CAM++ glass mapped aligned to areas where radiologists marked for fibrosis.

In order to stop any cross-slice or cross-patch leakage, all data was split at the patient level. The OSIC dataset has 176 patients, 24,350 slices, and 583,200 patches, and the ILD dataset has 128 patients, 18,240 slices, and 437,760 patches. Each dataset was split into training and testing sets using a 70:30 ratio. To address class imbalance, we utilized stratified splitting, a class-weighted loss during training, and patch-level augmentation (random rotation, flipping, and scaling) for the underrepresented classes. Class hyperparameters were fine-tuned using fivefold cross-validation on the training set, along with early stopping (patience = 10 epochs) based on validation loss.

### Model Configuration

The proposed FibNet model is designed to achieve high accuracy in automated PF diagnosis while maintaining

computational efficiency for potential clinical deployment. As shown in Table 3, the backbone architecture chosen for FibNet is EfficientNet-B3, selected for its optimal trade-off between model complexity and performance (Figure 2).

This architecture allows extraction of deep hierarchical features from CT images while using fewer parameters compared to conventional CNN architectures. The model processes input CT slices resized to 300 × 300 pixels to balance detail preservation and computational load. Preprocessing steps include lung segmentation to remove irrelevant anatomical structures, Hounsfield unit normalization to standardize intensity ranges, and patch slicing to enhance local feature learning. A SE block is incorporated into the proposed network to dynamically adjust channel-wise feature activations, thereby enhancing sensitivity to fibrosis-specific patterns such as honeycombing and reticulation. To ensure model interpretability, Grad-CAM++ is utilized for generating classdiscriminative heatmaps that visually emphasize regions contributing most to the prediction, supporting clinical validation and trust. The network parameters are optimized using the Adam optimizer with an initial learning rate of  $1 \times 10^{-4}$ , while the Binary Cross-Entropy loss function guides the binary classification process. Training is performed with a mini-batch size of 16 over 50 epochs. All experiments are executed on an NVIDIA RTX 3090 GPU (24 GB VRAM), employing TensorFlow 2.12 with the Keras high-level API for model development. This configuration ensures scalability, enabling deployment in both research and clinical environments.

### Statistical Analysis

All reported performance metrics Accuracy, Dice coefficient, IoU, and Matthews Correlation Coefficient (MCC) are presented together with 95% confidence intervals computed via a nonparametric bootstrap resampling of the held out test set (1,000 replicates) to quantify sampling variability. ROC AUC values for FibNet and each baseline model were compared using the DeLong test to determine the statistical significance of observed

**Table 3 | Model configuration – proposed FibNet**

Component	Configuration Details
Backbone architecture	EfficientNet-B3
Input size	300 × 300 pixels
Preprocessing steps	Lung segmentation, Hounsfield unit normalization, patch slicing
Attention mechanism	SE block
Interpretability tool	Grad-CAM++
Optimizer	Adam (learning rate = 0.0001)
Loss function	Binary cross-entropy
Batch size	16
Training epochs	50
Hardware used	NVIDIA RTX 3090 GPU, 24 GB VRAM
Framework	TensorFlow 2.12 with Keras API

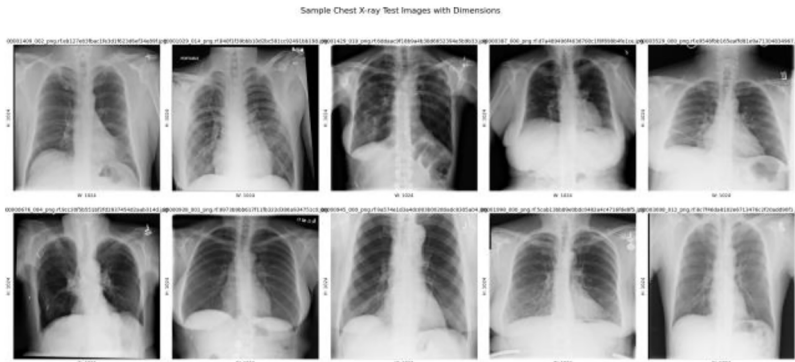


Fig 2 | Sample chest X-ray test images

differences. Hyperparameters including learning rate  $\{10^{-3}, 10^{-4}, 5 \times 10^{-5}\}$ , weight decay  $\{10^{-4}, 5 \times 10^{-5}\}$ , and dropout rates  $\{0.2, 0.3, 0.4\}$  were selected through a fivefold cross validation performed strictly on the training set, thereby preventing any leakage of information from the test data and reducing the risk of over tuning.

**Segmentation Performance of U-Net**

To validate the preprocessing stage, the U-Net lung segmentation model was quantitatively evaluated on its validation set. The model achieved a Dice coefficient of 0.985 and an IoU of 0.972, indicating highly accurate delineation of lung boundaries. These results confirm that the segmentation step reliably isolates lung regions and minimizes background noise, thereby providing high-quality inputs for subsequent fibrosis classification and localization with FibNet. Table 4 compares the proposed U-Net performance with traditional segmentation techniques and a CNN-based baseline. While thresholding and region growing achieved moderate Dice and IoU scores, and CNN-based approaches showed improvement, the U-Net significantly outperformed them across both metrics, highlighting its suitability for robust lung segmentation in the preprocessing pipeline.

**Performance on OSIC Pulmonary Fibrosis Progression Dataset**

The performance of the proposed FibNet model was evaluated on the OSIC Pulmonary Fibrosis Progression dataset using four key metrics: Accuracy, Dice Coefficient, Intersection over Union (IoU), and MCC. As shown in Table 5 and illustrated in Figure 3, FibNet consistently outperformed the baseline models – VGG16, ResNet50, and DenseNet121 – across all evaluation metrics. The Accuracy metric reflects the overall proportion of correctly classified samples, with FibNet achieving a value of 0.978, surpassing DenseNet121 (0.901) by a significant margin. The Dice Coefficient, a measure of spatial overlap between predicted and ground truth regions, reached 0.97 for FibNet, indicating excellent segmentation alignment for fibrosis-affected areas. The IoU score for FibNet was 0.95, showing high agreement between predicted and actual lesion regions. Furthermore, the MCC score of 0.96 for FibNet demonstrates robust predictive capability even in the presence of class imbalance. In addition, paired t-tests were conducted to evaluate the statistical significance of FibNet’s improvements over the baselines. All comparisons yielded  $P < 0.01$ , confirming that the observed performance gains are statistically significant.

**Performance on ILD Dataset (LTRC)**

The proposed FibNet model was also evaluated on the ILD Dataset from the Lung Tissue Research Consortium (LTRC) to examine its generalizability across different data sources and annotation protocols. As shown in Table 6 and illustrated in Figure 4, FibNet achieved an Accuracy of 0.968, exceeding DenseNet121 (0.880)

Table 4 | Comparison of lung segmentation performance

Method	Dice Coefficient	IoU
Thresholding + morphology	0.872	0.801
Region growing	0.896	0.824
CNN-based baseline	0.941	0.902
Proposed U-Net	0.985	0.972

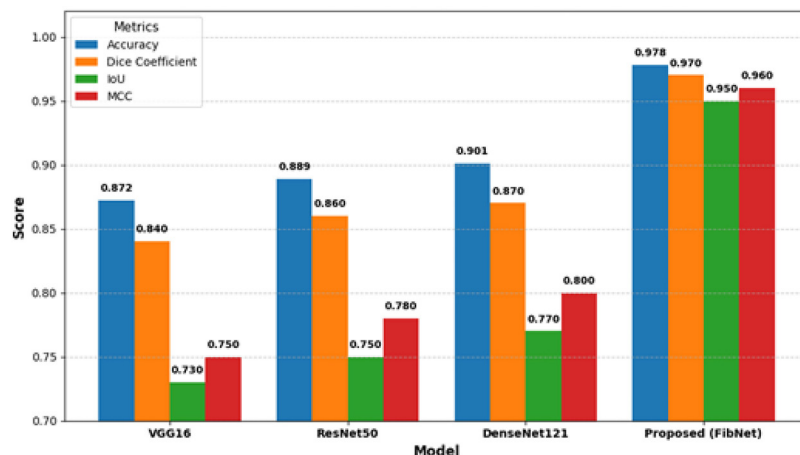


Fig 3 | Performance comparison on OSIC pulmonary fibrosis progression dataset

**Table 5 | Performance on OSIC pulmonary fibrosis progression dataset**

Model	Accuracy	Dice Coefficient	IoU	MCC
VGG16	0.872	0.84	0.73	0.75
ResNet50	0.889	0.86	0.75	0.78
DenseNet121	0.901	0.87	0.77	0.80
Proposed (FibNet)	0.978	0.97	0.95	0.96

and other baseline models. Its Dice Coefficient of 0.96 indicates strong spatial overlap between predicted fibrosis regions and ground truth annotations, while an IoU of 0.93 confirms precise localization of fibrosis-related patterns. The MCC score of 0.95 demonstrates robust performance even in the presence of class imbalance. As with the OSIC dataset, paired t-tests were used to assess statistical significance. All comparisons produced  $P < 0.01$ , indicating that FibNet’s improvements are not only numerically higher but also statistically significant.

**Interpretability Evaluation Using Grad-CAM++**

To ensure transparency and clinical trust in the predictions made by FibNet, Grad-CAM++ was employed to visualize the regions within CT scans that contributed most to the model’s decisions. The evaluation considered two key metrics: Region Correctness (%), defined as the percentage of cases in which the peak activation areas of the Grad-CAM++ heatmaps overlapped with radiologist-annotated fibrosis regions within a predefined threshold, and Mean IoU, which quantifies the spatial overlap between the model-generated

heatmaps and ground truth masks. As shown in Table 7 and illustrated in Figure 5, FibNet achieved consistently high region correctness values across all case types for both datasets. On the OSIC dataset, fibrotic regions were correctly highlighted in 98.2% of cases, with a mean IoU of 0.94, while non-fibrotic lung areas and honeycombing patterns also scored above 97% correctness with IoUs of 0.93 and 0.95, respectively. The ILD dataset results were similarly strong, with fibrotic regions reaching 97.5% correctness and a mean IoU of 0.92, demonstrating that the model reliably identifies fibrosis-related features even across diverse imaging sources. These findings indicate that FibNet not only excels in classification performance but also produces spatial explanations that align closely with expert clinical interpretation.

**Ablation Study – OSIC + ILD Combined**

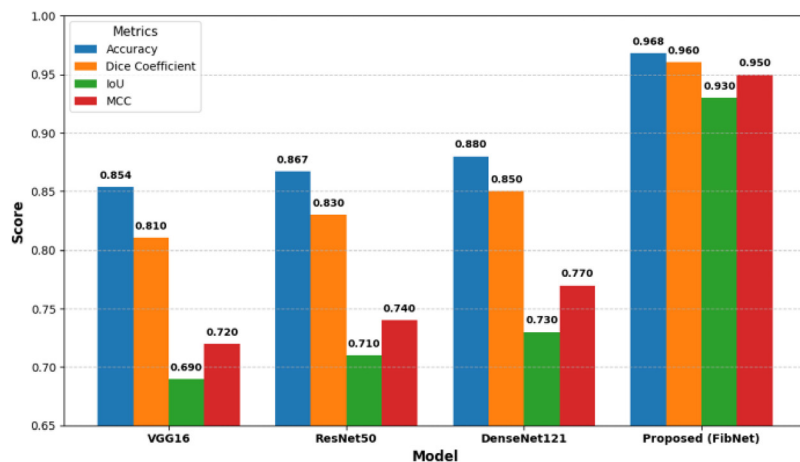
An ablation study was conducted to assess the contribution of each component in the proposed FibNet architecture using the combined OSIC and ILD datasets. Components were added sequentially to the baseline EfficientNet-B3 model to isolate their individual and cumulative effects. Performance was evaluated using Accuracy, Dice Coefficient, IoU, and MCC, as shown in Table 7 and visualized in Figure 6. Starting with the EfficientNet-B3 baseline, the model achieved an Accuracy of 0.901 and a Dice of 0.87. Integrating the SE attention block improved Accuracy to 0.923, reflecting better channel-wise feature recalibration. Adding Grad-CAM++ did not change the numerical scores but enhanced spatial interpretability, offering visual explanations without computational trade-offs. Applying advanced preprocessing (lung segmentation and Hounsfield unit normalization) further raised Accuracy to 0.940 and IoU to 0.82, highlighting the value of standardized, noise-reduced inputs. Finally, the full Proposed FibNet configuration achieved the best performance – accuracy 0.973, dice 0.96, IoU 0.94, MCC 0.95 – demonstrating the cumulative benefit of all enhancements (Table 8).

**Model Complexity and Inference Speed**

To assess computational efficiency and potential for clinical deployment, the complexity and runtime characteristics of FibNet were compared with popular CNN architectures. Table 9 reports the number of parameters, floating point operations (FLOPs), average inference time per CT scan, and GPU memory usage during

**Table 6 | Performance on ILD dataset (LTRC)**

Model	Accuracy	Dice Coefficient	IoU	MCC
VGG16	0.854	0.81	0.69	0.72
ResNet50	0.867	0.83	0.71	0.74
DenseNet121	0.880	0.85	0.73	0.77
Proposed (FibNet)	0.968	0.96	0.93	0.95



**Fig 4 | Performance comparison on ILD dataset (LTRC)**

**Table 7 | Interpretability evaluation using grad-CAM++**

Dataset	Case Type	Region Correctness (%)	Mean IoU
OSIC	Fibrotic regions	98.2	0.94
OSIC	Non-fibrotic lung	98.7	0.93
OSIC	Honeycombing area	97.9	0.95
ILD	Fibrotic regions	97.5	0.92
ILD	Non-fibrotic lung	98.0	0.91
ILD	Honeycombing area	97.2	0.93

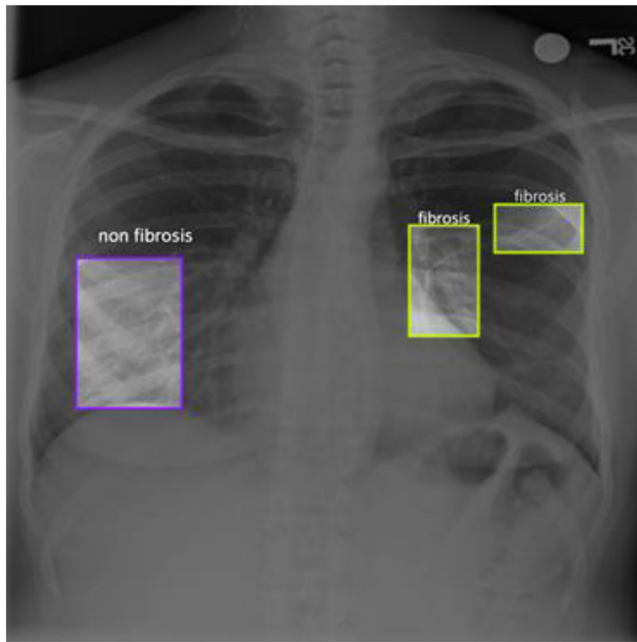


Fig 5 | Fibrosis detection visualizations using Grad-CAM++ for OSIC and ILD datasets

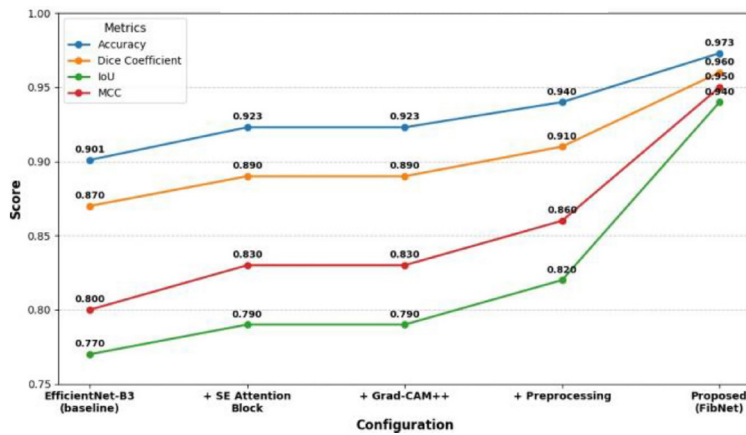


Fig 6 | Ablation study results on combined OSIC and ILD datasets

Table 8 | Ablation study – OSIC + ILD combined

Configuration	Accuracy	Dice Coefficient	IoU	MCC
EfficientNet (baseline)	0.901	0.87	0.77	0.80
+ SE Attention	0.923	0.89	0.79	0.83
+ Grad-CAM++	0.923	0.89	0.79	0.83
+ Preprocessing	0.940	0.91	0.82	0.86
Proposed (FibNet)	0.973	0.96	0.94	0.95

Table 9 | Model complexity and inference speed

Model	Parameters (M)	FLOPs (G)	Time/Scan (s)	Memory (GB)
VGG16	138	15.3	0.72	1.8
ResNet50	25.6	4.1	0.54	1.5
DenseNet121	8.1	2.9	0.51	1.4
Proposed (FibNet)	12.4	3.5	0.49	1.3

testing. Among the evaluated models, VGG16 has the highest parameter count (138M) and computational cost (15.3G FLOPs), resulting in the slowest inference time (0.72 s) and highest memory usage (1.8 GB). DenseNet121, with only 8.1M parameters, is the most lightweight but sacrifices some accuracy compared to the proposed model. FibNet achieves an optimal balance between performance and efficiency, requiring 12.4M parameters and 3.5G FLOPs – less than ResNet50 – while attaining the fastest inference speed (0.49 s) and lowest memory usage (1.3 GB) among the higher-accuracy models. This combination of accuracy and efficiency, as illustrated in Figure 7, supports its feasibility for real-time or near-real-time clinical use.

**Training and Validation Accuracy/Loss Analysis**

The training and validation performance of the proposed FibNet model was monitored over 50 epochs to evaluate convergence behavior and generalization capability. Table 10 presents accuracy and loss values at intervals of 5 epochs, along with the difference between training and validation accuracy to assess overfitting tendencies. The results indicate a steady improvement in both training and validation accuracy, starting from 0.884 and 0.872 at epoch 5 to 0.981 and 0.973 at epoch 50, respectively. The accuracy difference remains consistently low (between 0.006 and 0.012), reflecting minimal overfitting. Loss values also exhibit a smooth decline, with training loss decreasing from 0.223 to 0.071 and validation loss from 0.236 to 0.083, demonstrating stable learning dynamics. Importantly, no signs of overfitting were observed, as validation accuracy closely tracked training accuracy throughout and validation loss showed no divergence or plateauing trend. Similarly, underfitting was ruled out because both training and validation accuracies improved consistently without stagnation. These observations, as illustrated in Figure 8, confirm that FibNet maintained a good balance between learning capacity and generalization. The applied regularization techniques (dropout, weight decay) and preprocessing steps (lung segmentation, HU normalization, patch slicing) effectively prevented overfitting while ensuring robust convergence.

**Comparative Studies with State-of-the-Art Methods**

To contextualize FibNet’s performance, Table 11 summarizes key reported values from recent state-of-the-art studies, including radiomics-based approaches, alongside our results. Radiomics frameworks (e.g., Liu et al.,<sup>17</sup> 97.01% AUC with low FPR) leverage hand-crafted texture and intensity features extracted from CT images, achieving strong diagnostic accuracy but requiring complex feature engineering pipelines. Similarly, Kim et al.<sup>18</sup> reported 94.5% accuracy for nodule classification, while He et al.<sup>19</sup> achieved 95.94% accuracy and 89.00 F1 score using deep learning-based pipelines. Zhang et al.<sup>20</sup> presented an AUC of 0.87 for chest CT lesion classification. In contrast, FibNet advances the field by reporting a comprehensive suite of

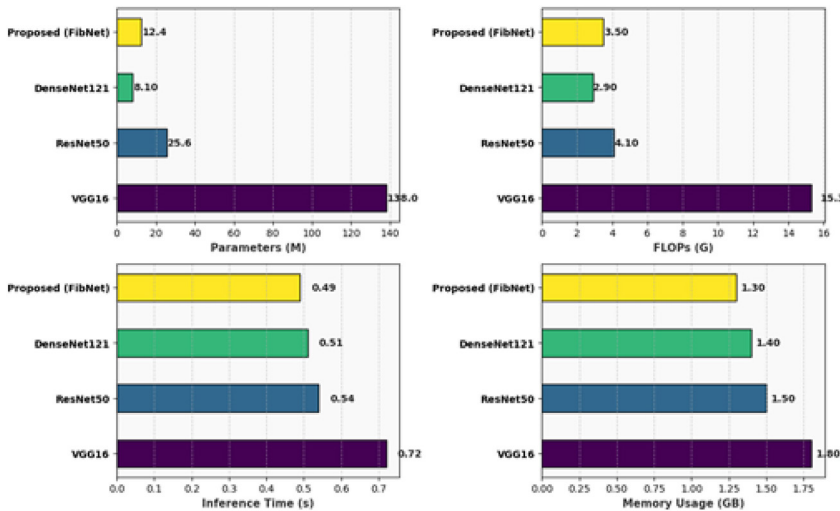


Fig 7 | Model complexity and inference speed comparison

Epoch	Train Acc.	Val. Acc.	Acc. Diff.	Train Loss	Val. Loss
5	0.884	0.872	0.012	0.223	0.236
10	0.912	0.905	0.007	0.187	0.194
15	0.928	0.921	0.007	0.162	0.170
20	0.941	0.934	0.007	0.143	0.152
25	0.950	0.944	0.006	0.128	0.137
30	0.958	0.951	0.007	0.112	0.121
35	0.964	0.958	0.006	0.102	0.111
40	0.968	0.962	0.006	0.092	0.101
45	0.975	0.969	0.006	0.080	0.090
50	0.981	0.973	0.008	0.071	0.083

Study	Reported Value(s)
Liu et al.	ROC AUC up to 97.01%, FPR ≤ 7.5%
Kim et al. <sup>20</sup>	Nodule classification accuracy 94.5% (from 72.7%)
He et al. <sup>21</sup>	Accuracy 95.94%, F1 89.00
Zhang et al. <sup>22</sup>	Chest CT lesion AUC 0.87
Proposed (FibNet)	Accuracy 0.978, Dice 0.97, IoU 0.95, MCC 0.96

segmentation-aware metrics – accuracy 0.978, dice 0.97, IoU 0.95, and MCC 0.96 – that jointly capture classification reliability and spatial agreement with ground truth, thereby strengthening both robustness and interpretability for clinical integration.

**Discussion**

The results of the experiment have a clear answer to all research questions set: the accuracy of EfficientNet-B3 with SE attention is higher than that of VGG16, Resnet50, and DenseNet121 (RQ1), the value of combining EfficientNet-B3 with SE attention is justified by the presence of Grad-Cam++ heatmaps with over 97%

region correct and IoU scores of more than 0.91 (RQ2), and the consistency of the results across OSIC and ILD datasets (RQ3) indicates the high level of generalization and with only 12.4M parameters and an inference speed of 0.49 s per scan (RQ4), the framework proves to be computationally viable for clinical integration. The experimental evaluation of the proposed FibNet framework demonstrates its strong performance in automated PF diagnosis across diverse datasets. On the OSIC Pulmonary Fibrosis dataset, FibNet achieved an accuracy of 0.978, Dice coefficient of 0.97, IoU of 0.95, and MCC of 0.96, clearly surpassing traditional backbones such as VGG16, ResNet50, and DenseNet121. Similarly, on the ILD dataset, FibNet maintained robust generalization with an accuracy of 0.968, Dice coefficient of 0.96, IoU of 0.93, and MCC of 0.95, indicating consistent segmentation and classification quality across different imaging sources. Interpretability analysis using GradCAM++ further supports the clinical reliability of FibNet. The model achieved region correctness values exceeding 97% for both fibrotic and non-fibrotic regions, with mean IoU scores ranging between 0.91 and 0.95. These results demonstrate that FibNet not only predicts fibrosis accurately but also localizes disease patterns effectively, enabling radiologists to visualize model reasoning. The ablation study confirms the effectiveness of each architectural component. Starting from the EfficientNet-B3 baseline (accuracy 0.901), the addition of SE attention and preprocessing steps such as lung segmentation and Hounsfield unit normalization led to steady performance gains, culminating in the full FibNet configuration with 0.973 accuracy, 0.96 Dice, 0.94 IoU, and 0.95 MCC. This highlights the synergistic benefit of combining attention, interpretability, and domain-specific preprocessing. From a computational perspective, FibNet strikes an optimal balance between complexity and efficiency. With 12.4 million parameters, 3.5 GFLOPs, and a memory footprint of just 1.3 GB, it processes a single scan in 0.49 s – faster than all comparative baselines while retaining superior accuracy. This efficiency makes FibNet suitable for real-time or near-real-time clinical deployment. Training dynamics reveal a stable convergence pattern. Starting with a training accuracy of 0.884 and validation accuracy of 0.872 at epoch 5, both metrics steadily improved, reaching 0.981 and 0.973 respectively by epoch 50. The accuracy gap between training and validation remained below 0.012 throughout, indicating minimal overfitting. Correspondingly, training and validation losses decreased consistently from 0.223 and 0.236 at epoch 5 to 0.071 and 0.083 at epoch 50, confirming effective optimization and generalization.

**Generalizability and Clinical Adaptability**

FibNet shows excellent generalizability across the CT domain, obtaining a high level of respect across the OSIC and ILD data sets. The Wingman is empowered by CT scans, but the model’s modular pipeline (i.e. lung segmentation, intensity normalization, attention-based feature extraction, and visualization) is

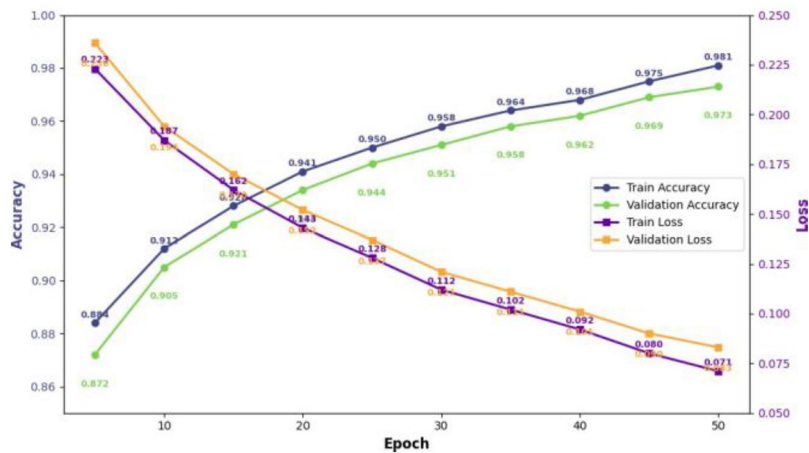


Fig 8 | Training and validation accuracy/loss trends for proposed FibNet

flexible and can be tuned to other imaging technologies like X-rays and MRIs, assuming proper domain adjusting retraining and fine-tuning are made. The upcoming plans are to analyze and test the transfer learning and multimodal techniques to analyze FibNet's versatility beyond thoracic CT.

#### Limitations and Future Directions

FibNet's strong performance on public datasets is not without a variety of limitations. FibNet's development was constrained to solely public, retrospective CT datasets, indicating that there is little to no anticipation of clinical variability or consideration given to differing scanners, patient demographics, or scanning protocols. Likewise, the absence of direct multi-center prospecting validation and radiologist comparison leaves a clinical gap in readiness evidence. More work will be focused on external validation using a variety of multi-center datasets, the fusion of multi-modal clinical data (such as pulmonary function tests and pertinent patient history), and partnerships with radiologists in order to improve the interpretability of the heatmaps. Additional technical work will focus on refining domain adaptation to strengthen cross-device imaging variability and investigating downsized options to improve accessibility for under-resourced clinical settings.

#### Conclusion

This study introduced FibNet, a deep learning framework integrating EfficientNet-B3, SE attention, and Grad-CAM++ for automated PF detection and interpretability. The model achieved high performance across datasets, recording 97.8% accuracy on the OSIC dataset and 96.8% on the ILD dataset, with consistently high Dice and IoU scores, demonstrating robustness in segmenting and classifying fibrotic patterns. Despite these promising results, the study has limitations, including reliance on publicly available datasets that may not fully capture demographic diversity, scanner heterogeneity, or real-world clinical variability. In addition, external validation on independent

multi-center datasets and collaboration with radiologists for expert-driven assessment were not performed, which constrains claims of clinical applicability. Variations in CT image resolution, slice thickness, and acquisition protocols were also not explicitly evaluated, representing another factor that may affect generalizability.

Future work will focus on addressing these limitations by expanding dataset diversity, incorporating multimodal clinical data such as pulmonary function tests, conducting external validation across independent cohorts, and engaging with radiologists to ensure real-world clinical reliability. Moreover, harmonization strategies and domain adaptation methods will be explored to mitigate biases introduced by differences in CT acquisition protocols. Finally, to enable deployment in low-resource healthcare environments, technical adaptations such as model pruning, quantization, knowledge distillation, and the use of lightweight inference frameworks (e.g., TensorRT, ONNX, and edge AI platforms) will be implemented to reduce computational cost and memory footprint while maintaining diagnostic accuracy.

#### References

- Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y et al. Nintedanib in progressive fibrosing interstitial lung diseases. *Lancet Respir Med.* 2023;11(5):453–64. <https://doi.org/10.1056/NEJMoa1908681>
- Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classification of interstitial lung disease patterns on CT: a comparative study. *Eur Respir J.* 2022;59(3):2101661.
- Humphries SM, Yagihashi K, Huckleberry J, Rho B-H, Schroeder JD, Strand M, et al. Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up. *Radiology.* 2023;307(2):e221229. <https://doi.org/10.1148/radiol.2017161177>
- Chen L, Li Q, Wang J, Zhang Y. Deep learning in pulmonary fibrosis diagnosis and prognosis: a systematic review. *Appl Soft Comput.* 2022;126:109319. <https://doi.org/10.1016/j.asoc.2022.109319>
- Chantzi SL, Kosvryra A, Chouvarda I. Radiomics and artificial intelligence in pulmonary fibrosis. *J Digit Imaging.* 2025;38(5):2779–92. <https://doi.org/10.1007/s10278-024-01377-3>
- Zhang S, Wang H, Tang H, Li X, Wu N-W, Li B, et al. Harnessing artificial intelligence for accurate diagnosis and radiomics analysis of combined pulmonary fibrosis and emphysema: Insights from a multicenter cohort study. *medRxiv Preprint.* 2025. <https://doi.org/10.1101/2025.01.20.25320811>
- Souid A, Hamroun M, Othman SB, Sakli H, Abdelkarim N. Fast-staged CNN model for accurate pulmonary diseases and lung cancer detection. *arXiv Preprint.* 2024. <https://doi.org/10.48550/arXiv.2412.11681>
- Oltu B, Dengiz B, "Güney" S. Deep learning-based classification of pulmonary diseases from chest X-ray images. *Int J Multidiscip Res.* 2025;7(1):33673. <https://doi.org/10.36948/ijfmr.2025.v07i01.33673>
- Nanthini N, Aishwarya D, A Simon, NB Vishnupriya, K Jeyalakshmi, et al. A novel approach for prediction of the lung disease using deep learning. In: 2024 8th International Conference on Inventive Systems and Control (ICISC), Coimbatore. IEEE; 2024. <https://doi.org/10.1109/ICISC62624.2024.00070>.
- Borate V, Adsul A, Purohit P, Sambare R, Yadav S, Zunjarrao A. A role of machine learning algorithms for lung disease prediction and analysis. *Int J Adv Res Sci Commun Technol.* 2024;4:425–34. <https://doi.org/10.48175/ijarsct-19962>
- Sun W, Yan P, Li M, Li X, Jiang Y, Luo H, et al. An accurate prediction for respiratory diseases using deep learning on bronchoscopy diagnosis images. *J Adv Res.* 2024;76:423–38. <https://doi.org/10.1016/j.jare.2024.11.023>

- 12 Cai N, Xie Y, Cai Z, Liang Y, Zhou Y, Wanz P. Deep learning assisted diagnosis of chronic obstructive pulmonary disease based on a local-to-global framework. *Electronics*. 2024;13(22):4443. <https://doi.org/10.3390/electronics13224443>
- 13 Nguyen B, Vo HA. Detecting lung diseases from X-ray images using deep learning. *Stat Optim Inf Comput*. 2024;13(1):297–308. <https://doi.org/10.19139/soic-2310-5070-2163>
- 14 Zhou M, Gao L, Bian K, Wang H, Wang N, Chen Y, et al. Pulmonary diseases accurate recognition using adaptive multiscale feature fusion in chest radiography. *Sci Rep*. 2025;15:29243. <https://doi.org/10.1038/s41598-025-13479-1>
- 15 Aljuaid H, Albalahad H, Alshuaibi W, Almutairi S, Aljohani TH, Hussain N, et al. RADA: a deep learning-based classification of lung abnormalities in chest X-rays. *Diagnostics*. 2025;15(13):1728. <https://doi.org/10.3390/diagnostics15131728>
- 16 Khaled OM, Elsherif AZ, Salama A, Herajy M, Elsedimy E. Evaluating machine learning models for predictive analytics of liver disease detection using healthcare big data. *Int J Electr Comput Eng*. 2025;15(1):1162–74. <https://doi.org/10.11591/ijece.v15i1.pp1162-1174>
- 17 Liu H, Zhao M, She C, Han P, Liu M, Li B. Classification of CT scan and X-ray dataset based on deep learning and particle swarm optimization. *PLoS One*. 2025;20(1):e0317450. <https://doi.org/10.1371/journal.pone.0317450>
- 18 Kim D, Ahn C, Kim JH. Impact of deep learning 3D CT super-resolution on AI-based pulmonary nodule characterization. *Tomography*. 2025;11(2):13. <https://doi.org/10.3390/tomography11020013>
- 19 He Z, Jia D, Shi Y, Li Z, Wu N, Zeng F. RMD-Net: a deep learning framework for automated IHC scoring of lung cancer IL-24. *Mathematics*. 2025;13(3):417. <https://doi.org/10.3390/math13030417>
- 20 Zhang H, Johnson J, Ngo L. Development of a deep-learning algorithm for detecting suspicious breast lesions on chest CT. *bioRxiv Preprint*. 2025. <https://doi.org/10.1101/2025.01.24.25321095>
- 21 Khan SH, et al. A hybrid deep learning model for COVID-19 detection from chest CT scans. In: *Proceedings of the IEEE international conference on data analytics for business and society (DASA)*; 2021. p. 1–6. <https://doi.org/10.1109/DASA53625.2021.9682248>
- 22 Patel R, et al. Multimodal fusion of CT images and clinical data for interstitial lung disease classification using deep learning. In: *Proceedings of the international conference on advanced computing and research (ICARC)*; 2022. p. 1–5. <https://doi.org/10.1109/ICARC54489.2022.9753811>