



AI-Enhanced Approaches to Interstitial Lung Disease: A Review of Machine Learning Advances

Editor's Pick

I have selected this review as my Editor's Pick for this issue of *EMJ Respiratory* because there is a growing interest in the integration of AI into healthcare, with machine learning playing a pivotal role in reshaping the landscape of disease management. Interstitial lung diseases represent a heterogeneous group for which an accurate diagnosis requires a multidisciplinary approach, combining clinical, radiological, and pathological assessments. This article aims to explore and discuss the transformative potential of machine learning in the management of interstitial lung disease.

Antonio Rossi
IQVIA, Milan, Italy



Authors:	*Geran Maule, ^{1,2} Joseph Ayim Zamora, ³ Aseed Mestarihi, ^{1,2} Akil Augustus, ³ Kyle Da Silva, ³ John Rickards, ⁴ Mohammad Khraisat, ^{1,2} Jon Beacher ²
	1. University of Central Florida College of Medicine, Orlando, USA 2. HCA Florida North Florida Hospital, Graduate Medical Education Internal Medicine Residency Program, Gainesville, USA 3. Trinity Medical Sciences University School of Medicine, Ribishi, Saint Vincent and the Grenadines 4. Department of Internal Medicine, Mercer University School of Medicine, Macon, Georgia, USA *Correspondence to geran.maule@hcahealthcare.com
Disclosure:	The authors have declared no conflicts of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Disclaimer:	The work was supported by HCA Healthcare and/or an HCA Healthcare-affiliated entity. The views expressed in this publication represent those of the author(s) and do not necessarily represent the official views of HCA Healthcare or any of its affiliated entities.
Received:	12.03.25
Accepted:	17.07.25
Keywords:	AI, interstitial lung disease (ILD), machine learning (ML).
Citation:	EMJ Respir. 2025; 109-122. https://doi.org/10.33590/emjrespir/MDK08232

Abstract

The integration of AI into healthcare has marked a transformative era, with machine learning (ML) playing a pivotal role in reshaping the landscape of interstitial lung disease (ILD) management. ML models excel in analysing complex datasets, such as medical imaging and electronic health records, offering unprecedented advancements in the diagnosis, prognosis, and treatment of ILDs. These models have demonstrated superior accuracy compared to traditional methods, particularly in diagnosing idiopathic pulmonary fibrosis, where delays

in diagnosis significantly impact patient outcomes. Early and precise diagnosis through ML-driven tools allows for the timely initiation of therapy, which is crucial for improving prognosis and extending patients' quality of life. Despite the challenges in data quality and model interpretability, the future of ML in pulmonary healthcare is promising, with continued advancements poised to enhance patient management and outcomes. This article aims to examine the transformative potential of ML in the management of ILD.

Key Points

1. Interstitial lung diseases (ILD) cause significant morbidity, yet diagnosis is often delayed. AI offers potential to improve early detection, prognostication, and treatment selection, addressing major unmet needs in ILD care.
2. This narrative review synthesised 26 primary peer-reviewed studies applying machine learning to ILD, covering diagnostic imaging, biomarker discovery, and prognostic modelling, with comparisons to human readers and evaluation of emerging AI tools.
3. Machine learning can match or surpass expert performance in ILD diagnosis, predict progression, and identify novel biomarkers, but widespread clinical adoption requires prospective validation, interpretability, and integration into real-world workflows.

INTRODUCTION

In recent years, the convergence of AI and healthcare has ushered in a transformative era marked by unprecedented advancements.¹ At the heart of this revolution lies machine learning (ML), a subset of AI that empowers systems to learn from data, identify patterns, and make decisions with minimal human intervention.¹ ML, broadly defined, involves algorithms that enable computers to learn from and analyse large volumes of data.¹ These algorithms improve their performance through iterative processes, making predictions or decisions based on historical data.¹ In healthcare, ML models offer the potential to enhance diagnostic accuracy, prognostic insights, and treatment planning.

Traditional diagnostic methods often rely on subjective interpretation of symptoms and diagnostic tests, which can be influenced by human error and variability. ML algorithms, however, can analyse complex datasets, including medical imaging, genetic information, and electronic health records (EHR), to identify subtle patterns that may elude human practitioners. This capability can lead to earlier and more accurate diagnoses, ultimately improving patient outcomes.

Additionally, ML enables personalised prognostic modelling by integrating diverse data such as patient demographics, lifestyle factors, and clinical history. This allows for individualised risk assessments and predictions about disease progression and treatment response, helping to optimise clinical decision-making and potentially prevent adverse outcomes.²

Interstitial lung diseases (ILD) represent a heterogeneous group of approximately 200 pulmonary disorders, characterised by varying degrees of inflammation and fibrosis affecting the lung interstitium.³ Accurate diagnosis often requires a multidisciplinary approach, combining clinical, radiological, and pathological assessments, given overlapping imaging patterns and heterogeneous presentations.⁴ The traditional diagnostic process for ILD is often lengthy and invasive. Patients may undergo multiple evaluations, and in complex cases, a surgical lung biopsy is sometimes required to establish a definitive diagnosis. Even with multidisciplinary discussions, misdiagnoses or significant diagnostic delays are common. Notably, emerging AI systems may facilitate the detection of pulmonary fibrosis even before overt clinical or radiological manifestations appear. For instance, a new

AI-driven screening tool was able to predict pulmonary fibrosis up to 4 years before a conventional diagnosis (area under the receiver operating characteristic curve [AUROC]: ~0.84 at 4 years),⁵ underscoring the potential of ML in identifying ILD at a pre-fibrotic stage, when early intervention could be most beneficial.

Idiopathic pulmonary fibrosis (IPF) is often diagnosed late, with a median delay of 2.1 years, largely due to misdiagnosis and delays at multiple healthcare levels, including general practitioners and community hospitals.⁶ This delay is particularly concerning, given that early initiation of therapy in IPF is associated with better outcomes and slower disease progression.⁷

In summary, ML holds transformative potential in the realm of ILDs, offering advancements in diagnosis, prognosis, and treatment.⁸ As the technology continues to evolve, it is poised to enhance patient management and outcomes, marking a significant leap forward in the quest for more effective and personalised healthcare. As highlighted by Barnes et al.,⁹ ML represents a 'new frontier' in radiology for ILD, offering a shift from subjective pattern recognition toward reproducible, high-throughput analysis.⁹

METHODOLOGY

Literature Search Strategy

To explore the impact of ML techniques in ILD, a comprehensive search of relevant literature was conducted. The authors searched the following databases from 1999–20 June 2025:

- PubMed;
- IEEE Xplore;
- Google Scholar; and
- Cochrane Library.

Boolean operators were used to combine relevant terms, such as:

- 'interstitial lung disease' or 'idiopathic pulmonary fibrosis' or ILD or IPF; and

- 'artificial intelligence' or 'machine learning' or 'deep learning' or 'neural networks' or 'radiomics' or 'computer-assisted diagnosis'.

Filters were applied to restrict results to English-language, peer-reviewed, human-subject studies. Titles and abstracts were screened for relevance, followed by full-text review. Reference lists of key studies were also manually screened for additional sources. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-style flowchart (Figure 1) illustrates the selection process.

Studies included in this review met the following criteria:

- Involved human subjects diagnosed with any subtype of ILD;
- Applied AI or ML techniques for diagnostic, prognostic, or biomarker discovery purposes; and
- Reported specific outcomes, such as model performance metrics (e.g., AUROC, sensitivity, specificity), diagnostic accuracy, or clinical utility.

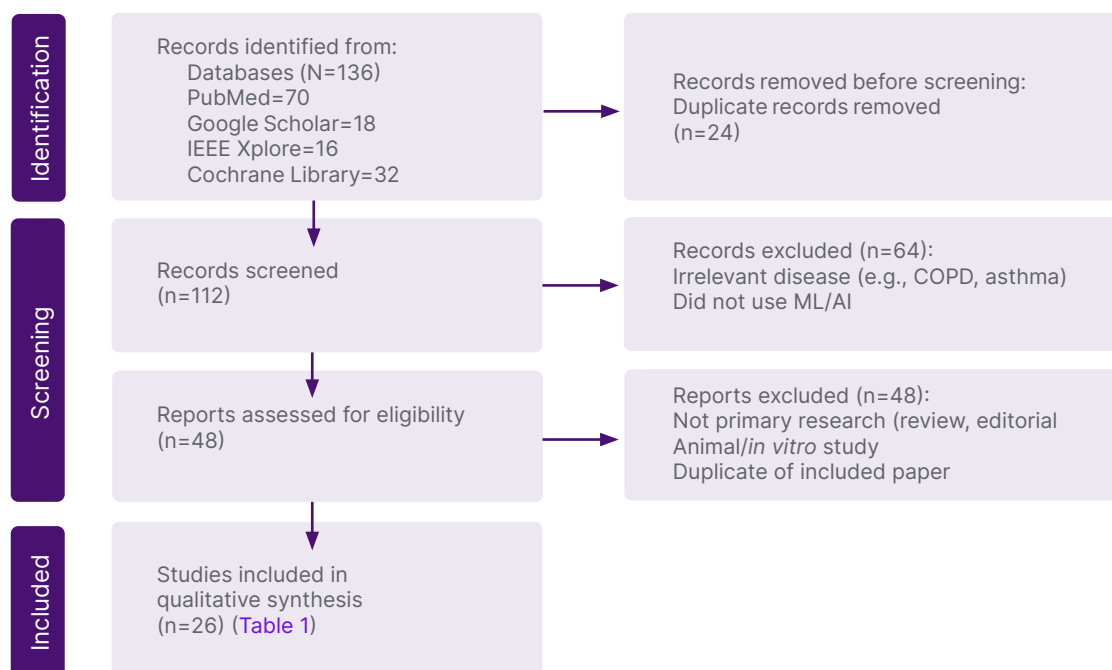
A total of 26 primary studies were included in the authors' results, each applying ML techniques to human subjects with ILD and reporting relevant diagnostic, prognostic, or biomarker outcomes. An additional set of systematic reviews and expert perspectives was referenced in the discussion to contextualise findings, but was not included in the formal study count.

Exclusion criteria included studies not directly related to ML applications in ILDs, studies focused solely on non-human subjects, studies lacking sufficient detail on methodology or results, and non-peer-reviewed material, including editorials, commentaries, conference abstracts, and technical white papers.

Data Synthesis and Analysis

A narrative synthesis approach was used due to the substantial heterogeneity among included studies in terms of ML models, data types, ILD subtypes, and outcome measures. Given the early-stage

Figure 1: PRISMA flow diagram of literature selection for a review of machine learning applications in interstitial lung disease.



ML: machine learning.

and exploratory nature of AI applications in ILD, with limited standardisation across methodologies, a narrative framework allowed for meaningful comparison, critical appraisal, and integration of diverse findings that would not be suitable for quantitative meta-analysis.

A total of 26 studies were included and analysed for this review, focusing on the application of ML in ILD. A summary of the results is shown below in [Table 1](#).

DISCUSSION

The application of ML in ILD represents a significant advancement in healthcare, offering promising improvements in diagnosis, prognosis, and treatment. As ML technologies continue to evolve, they provide increasingly sophisticated tools for analysing complex medical data, potentially addressing some of the longstanding challenges in ILD management. A pivotal aspect of these advancements is the comparison of ML models with

traditional diagnostic methods, including human readers, and the impact of these comparisons on clinical practice. Mekov et al.³⁴ offered an early overview, outlining how AI tools could bridge radiologic and clinical domains by supporting differential diagnosis and care planning in respiratory medicine, while Chan and Auffermann³⁵ emphasised the potential of AI to unify multimodal imaging in diffuse lung diseases.

Machine Learning in the Diagnosis of Interstitial Lung Disease

Radiological evaluation of ILD, particularly the identification and characterisation of pulmonary fibrosis, presents persistent challenges, even for experienced thoracic radiologists.³⁶ A key difficulty lies in detecting honeycombing, a defining feature of usual interstitial pneumonia (UIP), which is central to diagnosis but subject to high interobserver variability and diagnostic uncertainty.³⁶ This diagnostic ambiguity is especially pronounced in patients who do not meet criteria for a definitive UIP pattern.³⁷

Table 1: Summary of key studies on machine learning applications in interstitial lung disease. (Continued)

Study	ML method	Date type	Objective	Key findings	Performance
Mei et al. ⁴	DL	CT	Classification of ILD patterns	High accuracy in classifying ILD subtypes	AUROC: 0.91–0.94
Fanidis et al. ¹⁰	Radiomics + ML	CT	Radiomic feature analysis in ILD	Radiomics differentiated ILD types	Acc: 85%
Wu et al. ¹¹	CNN	CT	Predict progression	Predictive of disease worsening	AUROC: 0.88
Ahmad et al. ¹²	DL	CT	Differentiation of ILD types	Fibresolve model outperformed radiologists	AUROC: 0.92
Walsh et al. ¹³	CNN	CT	Diagnose UIP	DL matched expert radiologists	Sensitivity: 73%; Spec: 84%
Chutia et al. ¹⁴	SVM	CT	ILD classification	Effective feature-based classification	Acc 87%
Nishikiori et al. ¹⁵	CAD	CXR	Screening for fibrosing ILD	High sensitivity for fibrosing patterns	Sensitivity: 94.7%
Aoki et al. ¹⁶	CNN	CT	Detect SSc-ILD	Good correlation with clinical scores	AUROC: 0.89
Karampitsakos et al. ¹⁷	ML classifier	Biomarkers	ILD diagnosis via serum profiles	Identified key diagnostic markers	Acc: 90%
Huang et al. ¹⁸	Random forest	CT	Differentiate ILD types	Improved diagnostic accuracy	AUROC: 0.90
Kim et al. ¹⁹	Support vector machine	CT	UIP prediction	ML matched expert consensus	AUROC: 0.93
Qiang et al. ²⁰	CNN	CT	Automated pattern detection	Effective pattern classification	Acc: 88%
Oh et al. ²¹	Multimodal AI	CT+Clinical	IPF diagnosis and staging	Improved patient stratification	AUROC: 0.91
Maciukiewicz et al. ²²	Random forest	Clinical+Genetic	Predict SSc-ILD	Identified risk prediction features	C-index: 0.84
Devaraj et al. ²³	AI Scoring	CT	Disease burden quantification	Correlated with clinical outcomes	Not specified
Sun et al. ²⁴	AI quantitative model	CT	IPF survival prediction	Accurate 3-year survival estimation	AUROC: 0.85
Chen et al. ²⁵	AI prognostic model	CT	Predict IPF outcomes	Strong prognostic performance	AUROC: 0.87
Poynton et al. ²⁶	ML regression	CT+clinical	IPF progression modelling	Improved longitudinal predictions	RMSE improved
Choe et al. ²⁷	DL	CT	Content-based retrieval	Effective similarity matching	Top-1 accuracy: 80%
Qin et al. ²⁸	ML classifiers	Biomarkers	Diagnose RA-ILD	Biomarker model had high accuracy	AUROC 0.89
Agarwala et al. ²⁹	Deep learning	CT	ILD pattern screening	Efficient screening tool	AUROC 0.86
Moran-Mendoza et al. ³⁰	Deep learning classifier	CT	Predict mortality risk in ILD	Model output correlated with mortality across ILD subtypes	Not specified

Table 1: Summary of key studies on machine learning applications in interstitial lung disease. (Continued)

Study	ML method	Date type	Objective	Key findings	Performance
Ukita et al. ³¹	CAD	CXR	Detect fibrosing ILD	Promising CAD application	AUROC 0.87
Yu et al. ³²	Stress testing CNN	CT	Model robustness analysis	Performance declined with variability	Robustness- tested
Walsh et al. ²⁷	CNN	CT	Classify fibrotic lung disease	Accurate subtype classification	Accuracy 76%

Acc: accuracy; AUROC: area under the receiver operating characteristic curve; C-index: concordance index; CAD: computer-aided detection; CNN: convolutional neural network; CXR: chest X-ray; DL: deep learning; ILD: interstitial lung disease; IPF: idiopathic pulmonary fibrosis; ML: machine learning; RA-ILD: rheumatoid arthritis-associated interstitial lung disease; RMSE: root mean square error; Spec: specificity; SS-ILD: systemic sclerosis-associated interstitial lung disease; SVM: support vector machine; UIP: usual interstitial pneumonia.

Chang et al.³⁸ addressed this clinical gap by training an ML classifier on CT scans labelled with pathology- and clinical-supported diagnoses, intentionally excluding cases with clear UIP to focus on patients who are classified as ‘grey zone’, where radiologic interpretation alone may be insufficient.³⁸ Such models are especially valuable in real-world practice, where diagnostic confidence varies and multidisciplinary discussion is often required.

Building on this, Castillo-Saldana et al.³⁹ applied quantitative CT metrics to distinguish fibrotic ILD from emphysema, a common diagnostic dilemma. By leveraging densitometric and histogram-based features, their model captured subtle structural differences not readily appreciated by visual inspection, suggesting a role for quantitative imaging in phenotyping patients with overlapping clinical or radiographic features.³⁹ Complementing this, Ukita et al.³¹ developed a deep learning (DL)-based computer-aided detection system to identify fibrosing ILD on plain chest radiographs. Though the study did not directly compare ILD to emphysema, its use in a broad screening context highlights the potential of computer-aided detection tools to navigate diagnostic ambiguity and improve early identification.³¹

Convolutional neural networks (CNN) are particularly effective in medical imaging

due to their structure and functionality.⁴⁰ They start by representing an input image as a grid of numbers, with each number indicating the brightness of a pixel. CNNs use small squares called filters that slide across the image, performing a mathematical operation known as convolution to highlight specific features like edges or colours.⁴⁰ Following this, pooling reduces the image’s size by keeping only the most important parts, allowing the network to focus on key features while enhancing processing speed.⁴⁰ CNNs consist of multiple layers, with each layer recognising increasingly complex patterns, from basic edges in early layers to complete objects in later ones. Ultimately, the network classifies the image, identifying it as, for example, a cat or a dog, based on the features it has learned. This process enables CNNs to analyse and interpret a wide variety of images effectively. One of the key advantages of CNNs is their translation invariance, meaning they can recognise objects regardless of their position in the image. Beyond medical imaging, CNNs are widely used in real-world applications, including facial recognition and autonomous driving.⁴⁰

In a pivotal study by Mei et al.,⁴ CNN and vision transformer models were evaluated for both ILD subtype classification and survival prediction.⁴ Using CT scans and clinical data, the joint CNN-multilayer

perceptron model achieved an AUROC of 0.94, significantly outperforming a panel of seven human readers, whose combined AUROC was 0.88. The panel included radiologists and pulmonologists with varying experience levels, all of whom were provided identical CT scans and clinical metadata. The model also demonstrated higher sensitivity (90%) and specificity (87%) for diagnosing UIP compared to readers (sensitivity: 80%; specificity: 83%).⁴ These findings show the potential of ML to enhance diagnostic precision in ILD, particularly in complex or borderline cases. Still, the single-centre nature of this study warrants cautious interpretation until external validation is achieved.

In an early study from India, Agarwala et al.²⁹ developed a DL algorithm to detect ILD patterns on high-resolution CT, achieving an AUROC of 0.91.²⁹ This work is particularly notable for demonstrating the scalability and adaptability of AI models across diverse healthcare systems, including resource-limited environments.

Ahmad et al.¹² developed Fibresolve, an ML tool designed to identify IPF from other ILDs using thin-slice CT imaging.¹² Notably, the algorithm outperformed clinical panels in cases with atypical UIP patterns that often require surgical biopsy for definitive diagnosis. Among patients who did not meet imaging criteria for IPF but had ≤ 3 mm CT slices, Fibresolve achieved a diagnostic yield of 53.1% and a specificity of 85.9%. These figures are particularly meaningful, considering that traditional diagnostic pathways for such cases are often prolonged and invasive. By reducing the median time to diagnosis (213 days), Fibresolve could meaningfully expedite care and reduce the need for invasive procedures. The system has since received FDA approval, further supporting its potential utility in clinical practice.

Walsh et al.³³ developed a DL algorithm using 1,157 anonymised, high-resolution CT scans to classify fibrotic lung disease.³³ The algorithm achieved an accuracy of 76.4%, surpassing 66% of 91 thoracic radiologists, whose median accuracy was 70.7%. Additionally, the algorithm showed

good interobserver agreement (weighted kappa [κ_w]=0.69), exceeding 62% of the radiologists (κ_w =0.67), and offered near-instantaneous diagnoses, taking only 2.31 seconds to evaluate 150 four-slice montages. This rapid and reproducible performance highlights the efficiency and reliability of ML algorithms compared to human readers,³³ but we must bear in mind that real-world performance may vary, and such systems would require rigorous external validation. As Yu et al.³² point out, real-world performance may diverge from training data benchmarks. Their study retrospectively evaluated DL models for IPF diagnosis and found variability in performance when applied to different institutions and CT acquisition protocols, emphasising the importance of cross-site robustness testing before clinical deployment.³²

A notable development in ML for ILD diagnosis was achieved by researchers at Sapporo Medical University Hospital, Japan, who created a DL model for detecting chronic fibrosing ILDs using chest radiographs.¹⁵ This model, which is the first to employ chest radiographs instead of CT scans, achieved an impressive area under the curve (AUC) of 0.979, with a sensitivity of 0.896 and specificity of 1.000. This performance is comparable to that of experienced radiologists and pulmonologists, demonstrating the model's potential as a valuable diagnostic tool.¹⁵ In the realm of histopathology, Fukuoka et al.⁴¹ conducted a large international study demonstrating that AI could help standardise histopathologic diagnoses of UIP by reducing interobserver variability among expert pathologists, establishing a potential reference framework for future diagnostic tools.⁴¹ Complementing this, Chung et al.⁴² evaluated a genomic classifier capable of identifying UIP even in patients lacking classic high-resolution CT patterns, reinforcing the value of AI-driven molecular diagnostics in complex or ambiguous cases.⁴²

It is also worth mentioning that, although most current AI applications in ILD are geared toward recognition of fibrotic disease patterns, there is increasing recognition that identifying pre-fibrotic

interstitial abnormalities, such as interstitial lung abnormalities or early non-specific interstitial pneumonia, can improve clinical outcomes by enabling earlier intervention.¹³ Incorporating AI into radiologic and histopathologic pipelines may enhance pattern recognition of subtle pre-fibrotic changes and support multidisciplinary team decision-making before irreversible damage occurs. AI's application in identifying pre-fibrotic conditions remains an underexplored but crucial frontier. Early detection, even before irreversible fibrosis sets in, could substantially improve long-term outcomes and reduce the need for invasive diagnostics.

Finally, content-based image retrieval systems are emerging as novel AI tools with both diagnostic and educational value. Choe et al.²⁷ developed a DL-based content-based image retrieval system that retrieves visually similar annotated CT scans to assist with ILD subtype recognition, achieving an AUROC of 0.922 for distinguishing UIP from nonspecific interstitial pneumonia.²⁷ These tools may enhance radiologists' confidence, reduce ambiguity in borderline cases, and promote standardisation across institutions.

Machine Learning in Biomarker Discovery for Idiopathic Pulmonary Fibrosis

While imaging remains the cornerstone of ILD diagnosis, biomarker discovery through ML is an increasingly active and promising frontier. These approaches aim to augment diagnostic accuracy, stratify risk, and ultimately tailor treatment by extracting patterns from high-dimensional molecular data, spanning transcriptomics, proteomics, and gene expression profiling.

A notable example is the work by Kim et al.,¹⁹ who applied ML to high-dimensional transcriptional data to classify UIP versus non-UIP patterns in ILD.¹⁹ Their model demonstrated high diagnostic accuracy, supporting the potential of molecular classifiers as adjuncts to radiologic and histopathologic assessment. This early, yet pivotal, study laid the groundwork for multi-omic ML models, bridging the gap between molecular pathology and clinical

phenotyping in ILD. Building on this, Huang et al.¹⁸ extended the scope to plasma proteomics, applying ML to quantify over 1,300 proteins from patients with ILD and controls.¹⁸ Their model achieved near-perfect discrimination (AUROC: 0.99 for ILD versus control; AUROC: 0.90 for IPF versus non-IPF), suggesting that proteomic signatures may soon complement imaging in classifying ILD subtypes. Importantly, this study highlights how proteomics could enable earlier and less invasive diagnosis if validated in external cohorts.

Fanidis et al.¹⁰ employed the eXtreme gradient boosting ML algorithm on gene expression data to explore potential molecular signatures associated with pulmonary fibrosis.¹⁰ The model achieved an encouraging accuracy (range: 0.85–0.95) and identified several candidate genes, including *IL13Rα2* and *PAPSS2*, with possible roles in fibrotic pathways. *IL13Rα2* is a key receptor that IL-13 uses to induce fibrosis, and its signalling is crucial for the production of TGF-β,⁴³ a major contributor to fibrotic processes in chronic inflammatory diseases.^{44,45} To interpret the model's predictions, Shapley additive explanation analysis was utilised, quantifying the contribution of each feature (gene) to the overall prediction. This methodology offers insight into model decision-making and helped to identify 76 candidate genes potentially associated with fibrosis. While these findings highlight promising avenues for further investigation, it is important to note that these biomarkers remain exploratory and have yet to undergo validation in large, prospective cohorts.

Wu et al.¹¹ conducted a parallel study focusing on differentially expressed genes and identified four critical biomarkers: *FHL2*, *HPCAL1*, *RNF182*, and *SLAIN1*.¹¹ These genes have demonstrated validated predictive value, particularly highlighting *SLAIN1* for its potential role in informing future therapeutic strategies. Notably, *FHL2* has been associated with tissue remodelling and fibrosis, further emphasising its significance within the context of IPF. Although these genes demonstrate predictive potential within retrospective datasets, their utility as diagnostic or

therapeutic biomarkers also requires further clinical validation, including reproducibility across diverse populations.

Lastly, Qin et al.²⁸ focused on rheumatoid arthritis-associated ILD, developing ML classifiers using support vector machines and random forests to detect transcriptomic signatures specific to this subset.²⁸ Their model yielded strong diagnostic performance (AUROC: 0.89), reinforcing the idea that ML can help surface disease-specific molecular fingerprints in clinically overlapping ILD phenotypes.

Machine Learning in Interstitial Lung Disease Prognosis

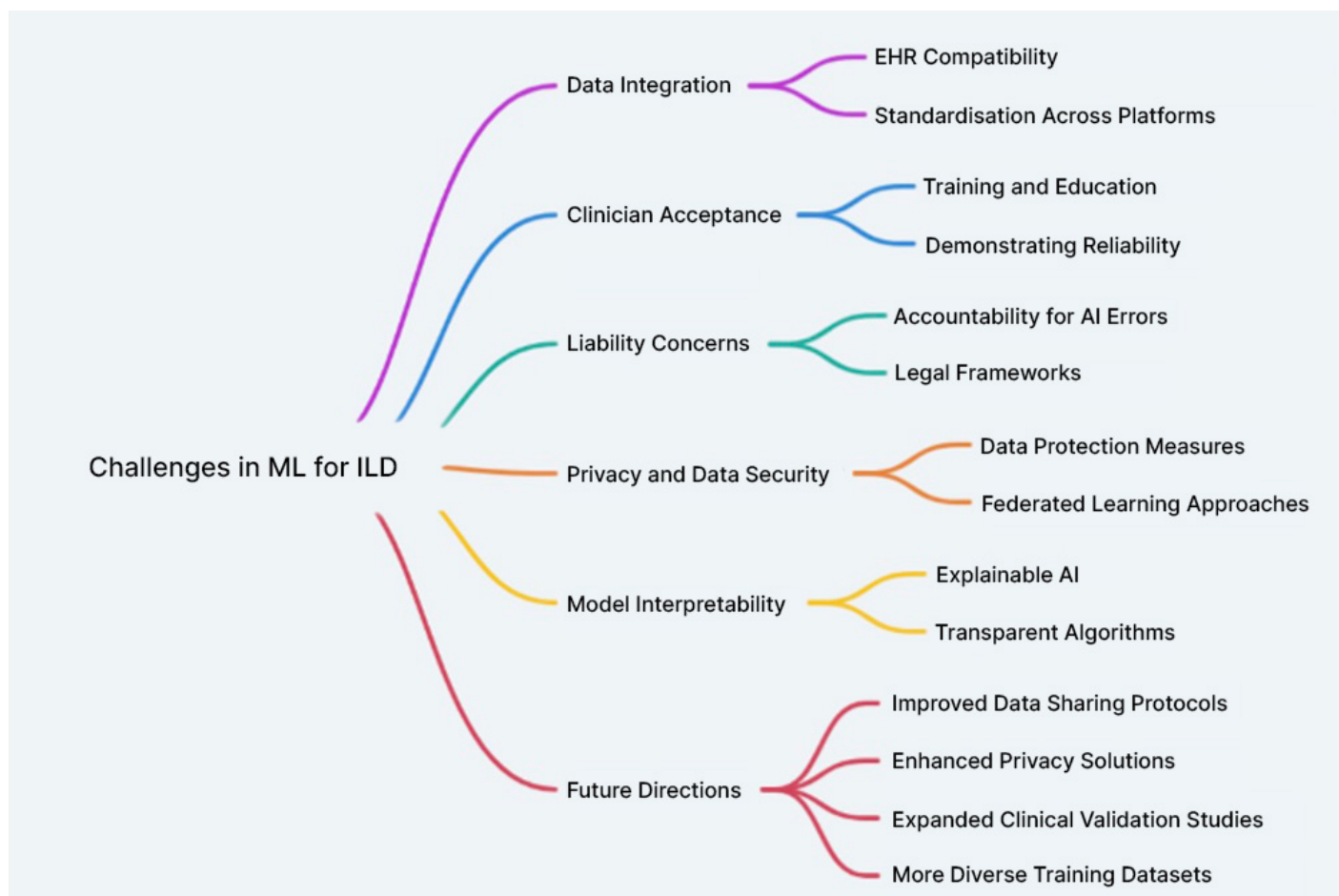
In their expert review, Bendstrup et al.⁴⁵ emphasised the importance of structured ILD monitoring using symptoms, spirometry, and imaging. These routine clinical touchpoints offer a natural opportunity for ML to augment care, whether by automating change detection on chest CTs or flagging subtle declines in pulmonary function tests before they cross conventional thresholds.⁴⁵ Within this prognostic domain, Chutia et al.¹⁴ developed a model to predict lung function decline in IPF by analysing 1,554 forced vital capacity (FVC) records from 176 patients, along with demographic data, smoking status, and CT scans.¹⁴ Using quantile regression combined with CNNs, the model achieved a striking 92% accuracy in forecasting lung function decline, supporting ML's potential to inform timely intervention and improve disease monitoring.

Imaging-based models have also shown significant promise. Chen et al.²⁵ trained a DL algorithm to predict mortality in IPF using chest CT features, achieving high predictive accuracy and reinforcing the role of imaging biomarkers in prognosis.²⁵ Similarly, Aoki et al.¹⁶ demonstrated that a DL-based quantification tool correlated strongly with FVC and diffusing capacity of the lungs for carbon monoxide, and achieved an AUROC of 0.78 for predicting ILD progression.¹⁶ These findings highlight the utility of quantitative CT metrics as surrogates for physiologic decline when automated via deep learning pipelines.

Expanding this work, Teramachi et al.⁴⁶ developed a longitudinal DL model that incorporated clinical data and environmental exposures to predict acute exacerbations and mortality in patients with ILD.⁴⁶ Walsh et al.,¹³ who had earlier applied DL to classify fibrotic lung disease, extended their model to predict mortality in progressive fibrotic ILD, demonstrating the broader applicability of AI-derived radiologic scores in outcome prediction.¹³ Likewise, Moran-Mendoza et al.³⁰ found that their ML-derived CT classifier score correlated significantly with mortality in a real-world ILD cohort, highlighting the prognostic potential of AI beyond simple subtype classification.³⁰

Other models have extended ML-based prognostic prediction to rare ILD phenotypes. For instance, Qiang et al.²⁰ trained a random forest model on CT and serum biomarkers to predict rapid progression in idiopathic inflammatory myopathy-associated ILD, achieving an AUC of 0.883.²⁰ Oh et al.²¹ similarly demonstrated that DL-derived fibrosis extent on CT predicted transplant-free survival independently of radiologist-assigned pattern.²¹ A related application of radiomics-based ML was explored by Karampitsakos et al.,¹⁷ who developed a random forest classifier trained on quantitative CT features to predict fibrotic ILD progression in survivors of COVID-19.¹⁷ Their model achieved robust predictive performance at 3 and 6 months (AUC: 0.827 and 0.851, respectively), demonstrating the adaptability of ML-based prognostic tools beyond idiopathic disease and into viral-induced ILD phenotypes. Other radiomics applications further support the utility of quantitative imaging. Chassagnon et al.³⁷ developed an automated DL system to assess ILD severity in systemic sclerosis using CT imaging.³⁷ Maciukiewicz et al.²² demonstrated that radiomic features from high-resolution CTs could predict FVC decline in systemic sclerosis-associated ILD using a random forest classifier.²² Sun et al.²⁴ applied ensemble learning in connective tissue disease-associated ILD, integrating demographics, radiographic data, and pulmonary function tests to predict long-term mortality.²⁴

Figure 2: Challenges and future directions of machine learning in interstitial lung disease.



EHR: electronic health record; ILD: interstitial lung disease; ML: machine learning.

Recent studies have highlighted the prognostic potential of imaging-based ML models. In a post hoc analysis of a Phase II trial, Devaraj et al.²³ used the e-Lung platform to derive the weighted reticulovascular score (WRVS), which outperformed traditional metrics such as diffusing capacity of the lungs for carbon monoxide in predicting disease progression in patients with IPF over 52 weeks.²³

In the study conducted by Mei et al.,⁴ they also developed a model for ILD prognosis.⁴ The study aimed to predict 3-year mortality in patients with ILD using advanced ML models: a Long Short-Term Memory model and a transformer model. Both models incorporated 165 features, 32 high-level CT features extracted from chest CT scans using a pretrained CNN model, and 18

clinical variables, such as medication history and treatment details. These features were longitudinally assessed to create dynamic models for survival prediction.⁴ The transformer model consistently outperformed the Long Short-Term Memory model, showing a 15.8% better performance.

By the third year, the transformer model's AUROC of 0.868 indicated strong predictive performance for 3-year mortality, signifying that the model could distinguish between patients who would survive and those who would not with high accuracy.⁴ The model's negative predictive values (ranging from 89.66–94.55%) suggest that it was particularly reliable at identifying patients who would survive, minimising

false negatives. The increase in sensitivity from 54.55% after 1 year to 72.73% by the third year further demonstrates that the model became more accurate in identifying patients at risk of death as more follow-up data was added. This improvement highlights the importance of continuous clinical monitoring, with the model gaining more predictive power as patient history and response to treatment accumulate over time. This suggests that longer follow-up periods allow for more accurate prognosis and could help clinicians make more informed decisions about patient management.⁴

Radiomics has also shown promise in detecting subclinical progression. Poynton et al.²⁶ applied radiomic analysis to serial chest CTs in high-risk individuals and successfully differentiated progressive interstitial lung abnormalities from stable cases, often before overt clinical or functional decline was apparent.²⁶ This highlights the role of radiomics in early surveillance strategies. Molecular markers may further augment prognostic models. Libra et al.⁴⁷ recently proposed candidate plasma biomarkers for IPF progression using ML-based analysis, offering a glimpse into how future prognostic tools might integrate multi-omic data to personalise risk stratification.⁴⁷

Finally, the Duke EMPOWER app (Duke University Health System, Durham, North Carolina, USA) exemplifies the integration of digital tools with AI to enhance patient engagement and research participation.⁴⁸ By offering ILD-specific education, enabling self-screening for research studies, and collecting longitudinal data on patient outcomes and biometric measures, the app illustrates a practical use of technology in managing rare diseases.⁴⁸ Its success in increasing study enrolment and promoting healthy behaviours underscores the potential for AI-driven tools to improve clinical research and patient management. This direct-to-patient approach presents a promising model for other conditions, bridging the gap between technology and personalised healthcare.

Challenges and Future Directions

Despite the promising advancements in ML for diagnosing and managing ILD, several challenges must be addressed to fully realise its potential (Figure 2). One significant challenge is the integration of ML systems with existing electronic health record systems. Seamless integration is crucial for ensuring that ML tools are easily accessible and effectively utilised in clinical practice. However, this integration often involves complex technical and regulatory hurdles, including interoperability issues and the need for robust data exchange protocols. Addressing these challenges will be essential for the widespread adoption of ML tools in healthcare.

The generalisability of current models remains a significant limitation.⁴⁹ Of the seven studies included in this review, four focus exclusively on IPF, which, though clinically significant, represents only a small subset of the more than 200 ILD subtypes. As a result, current ML models are often optimised for the recognition of fibrotic patterns associated with IPF, potentially limiting their diagnostic performance when applied to less common or non-fibrosing ILDs such as sarcoidosis, hypersensitivity pneumonitis, or connective tissue disease-associated ILD. Also, expanding model training and validation across the full ILD spectrum is essential to ensure broader clinical applicability and to prevent inequities in diagnosis and treatment. In parallel with these diagnostic innovations, Soffer et al.⁵⁰ conducted a comprehensive systematic review of AI applications in ILD, highlighting the growing body of work focused on chest CT analysis.⁵⁰ Their review emphasises the heterogeneity in model architectures, training datasets, and outcome definitions, which collectively pose barriers to reproducibility and clinical adoption. Importantly, they call for greater standardisation and transparency in AI development, a theme echoed throughout the authors' review. These findings reinforce the need for interpretable and externally validated models before widespread implementation in clinical workflows.

Additionally, a challenge lies in the availability and quality of training data.

ILDs are relatively rare and heterogeneous diseases, and the development of robust ML models requires large, well-annotated, and structured datasets, resources typically available only at large academic or tertiary care centres. This concentration of data introduces potential biases, as models may underperform in community settings or underserved populations where disease presentations, imaging protocols, and EHRs may differ. Furthermore, smaller centres often lack the infrastructure to collect high-resolution imaging data or comprehensive clinical annotations necessary for model development.⁴⁹ Addressing this limitation will require greater collaboration across institutions, federated learning approaches, and efforts to democratise access to high-quality ILD datasets.

Another obstacle is acceptance of AI among clinicians, hospitals, and patients. Many healthcare professionals may be hesitant to rely on AI due to concerns about its reliability and the potential for reduced human oversight.^{51,52} Building trust in AI systems will require demonstrating their efficacy through rigorous validation studies and providing adequate training for users.^{51,52} Additionally, patient acceptance of AI-driven diagnostics will depend on transparency about how these tools work and how patient data is handled.

Liability concerns also pose a significant challenge.⁵³ Determining accountability when AI systems make mistakes is complex. If an AI system provides an incorrect diagnosis or treatment recommendation, it raises questions about who is responsible: the developers, the healthcare providers, or the institutions using the technology. Clear guidelines and legal frameworks will be necessary to address these issues and ensure that patients receive safe and effective care.⁵³

Interpretability remains a key challenge to AI adoption in ILD, as many high-performing models, such as CNNs and transformers, operate as 'black boxes' with limited transparency. This lack of explainability can hinder clinician trust and complicate integration. Clinicians are understandably hesitant to act on an AI's

prediction without understanding the basis, especially in high-stakes diagnoses like ILD subtyping or prognostication.

Regulatory Considerations

To advance clinical impact, future research should prioritise prospective validation and implementation studies that assess ML tools in real-world settings, measuring outcomes like diagnostic accuracy, workflow integration, and patient benefit. These studies are essential to move beyond retrospective analysis and ensure meaningful clinical adoption.

Privacy concerns are another critical issue. The use of AI in healthcare involves processing vast amounts of sensitive patient data and ensuring the protection of this data; addressing potential privacy breaches is paramount.⁵⁴ Strategies such as federated learning, which allows models to be trained on decentralised data without compromising privacy, offer promising solutions but require further development and validation.⁵⁵

From a regulatory and ethical perspective, compliance with evolving frameworks is critical. The FDA has released guidance on Software as a Medical Device (SaMD) and adaptive AI systems, emphasising transparency, clinical validation, and post-market monitoring.^{53,56} ML tools for ILD must align with these standards, particularly when offering diagnostic suggestions that could influence patient care. Likewise, adherence to data privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the USA, and the General Data Protection Regulation (GDPR) in the European Union, is paramount. These frameworks mandate strict governance over patient data, informed consent, and secure data storage and transfer, especially when training models on multi-institutional or international datasets.

Finally, model deployment should emphasise ethical use, explainability, and integration into clinical workflows. Collaboration with regulatory bodies, clinicians, and patients will be key to ensuring that AI tools are safe, effective, and trusted in practice.⁵⁷ AI should be viewed not as a replacement for expert

clinical judgment but as an assistive tool aimed at improving diagnostic consistency and efficiency in diverse care settings.⁵⁸

CONCLUSION

ML is revolutionising the diagnosis and management of ILD. From CNNs and vision transformers that have outperformed experienced human experts, to integrative models that combine imaging and clinical data, ML has demonstrated substantial potential for improving diagnostic accuracy and patient outcomes. Applications like the Fibresolve system and biomarker discovery pipelines also highlight the versatility and promise of ML in ILD care.

Although these advancements exist, real-world clinical adoption remains limited. To bridge this gap, future research must prioritise prospective validation studies across diverse clinical settings. Such studies should assess not only diagnostic and prognostic accuracy, but also the impact of ML tools on clinical

decision-making, patient outcomes, and workflow efficiency. Furthermore, model deployment studies are needed to evaluate integration with EHRs, clinician usability, and patient acceptability.

Future work should also prioritise the development of interpretable models to enhance clinician trust and transparency, external validation across multicentre and multinational cohorts to strengthen generalisability, the adoption of federated learning and secure data-sharing frameworks to address privacy concerns and expand access to high-quality ILD datasets, and early regulatory engagement to ensure alignment with standards set by the FDA, HIPAA, and GDPR.

Realising the full clinical potential of ML in ILD will require a coordinated approach that emphasises prospective evaluation, ethical implementation, and multidisciplinary collaboration. Progress in this space will depend on integration into real-world practice.

References

- Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94-8.
- Alowais SA et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23(1):689.
- Althobiani MA et al. Interstitial lung disease: a review of classification, etiology, epidemiology, clinical diagnosis, pharmacological and non-pharmacological treatment. *Front Med (Lausanne)*. 2024;11:1296890.
- Mei X et al. Interstitial lung disease diagnosis and prognosis using an AI system integrating longitudinal data. *Nat Commun*. 2023;14(1):2272.
- Onishchenko D et al. Screening for idiopathic pulmonary fibrosis using comorbidity signatures in electronic health records. *Nat Med*. 2022;28(10):2107-16.
- Hoyer N et al. Risk factors for diagnostic delay in idiopathic pulmonary fibrosis. *Respir Res*. 2019;20(1):103.
- Alsomali H et al. Early diagnosis and treatment of idiopathic pulmonary fibrosis: a narrative review. *Pulm Ther*. 2023;9(2):177-93.
- Mlodzinski E et al. Machine learning for pulmonary and critical care medicine: a narrative review. *Pulm Ther*. 2020;6(1):67-77.
- Barnes H et al. Machine learning in radiology: the new frontier in interstitial lung diseases. *Lancet Digit Health*. 2023;5(1):e41-50.
- Fanidis D et al. An explainable machine learning-driven proposal of pulmonary fibrosis biomarkers. *Comput Struct Biotechnol J*. 2023;21:2305-15.
- Wu Z et al. Identifying potential biomarkers of idiopathic pulmonary fibrosis through machine learning analysis. *Sci Rep*. 2023;13(1):16559.
- Ahmad Y et al. A machine learning system to indicate diagnosis of idiopathic pulmonary fibrosis non-invasively in challenging cases. *Diagnostics (Basel)*. 2024;14(8):830.
- Walsh SLF 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(7):883-91.
- Chutia U et al. Analysis of idiopathic pulmonary fibrosis through machine learning techniques. 2021;DOI:10.1109/ICSCC51209.2021.9528243.
- Nishikiori H et al. Deep-learning algorithm to detect fibrosing interstitial lung disease on chest radiographs. *Eur Respir J*. 2023;61(2):2102269.
- Aoki R et al. Effects of automatic deep-learning-based lung analysis on quantification of interstitial lung disease: correlation with pulmonary function test results and prognosis. *Diagnostics (Basel)*. 2022;12(12):3038.
- Karampitsakos T et al. Post-COVID-19 interstitial lung disease: insights from a machine learning radiographic model. *Frontiers in Medicine*. 2023;9:1083264.
- Huang Y et al. Machine learning of plasma proteomics classifies diagnosis of interstitial lung disease. *Am J Respir Crit Care Med*. 2024;210(4):444-54.
- Kim SY et al. Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med*. 2015;3(6):473-82.

20. Qiang Y et al. Development of a machine learning model in prediction of the rapid progression of interstitial lung disease in patients with idiopathic inflammatory myopathy. *Quant Imaging Med Surg.* 2024;14(12):9258-75.
21. Oh AS et al. Deep learning-based fibrosis extent on computed tomography predicts outcome of fibrosing interstitial lung disease independent of visually assessed computed tomography pattern. *Ann Am Thorac Soc.* 2024;21(2):218-27.
22. Maciukiewicz M et al. OP0150: Machine learning approaches for risk modelling in interstitial lung disease associated with systemic sclerosis using high dimensional image analysis. *Ann Rheum Dis.* 2021;80(Suppl 1):90
23. Devaraj A et al. e-Lung computed tomography biomarker stratifies patients at risk of idiopathic pulmonary fibrosis progression in a 52-week clinical trial. *Am J Respir Crit Care Med.* 2024;209(9):1168-9.
24. Sun D et al. Prediction of long-term mortality by using machine learning models in Chinese patients with connective tissue disease-associated interstitial lung disease. *Respir Res.* 2022;23(1):4.
25. Chen Z et al. Applications of CT with artificial intelligence in the prognostic model of idiopathic pulmonary fibrosis. *Ther Adv Respir Dis.* 2024;18:17534666241282538.
26. Poynton C et al. Radiomics detects interstitial lung abnormalities progression in high-risk patients. *Am J Respir Crit Care Med.* 2021;203(9):A3521.
27. Choe J et al. Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology.* 2022;302(1):187-97.
28. Qin Y et al. Identification of biomarkers by machine learning classifiers to assist diagnose rheumatoid arthritis-associated interstitial lung disease. *Arthritis Res Ther.* 2022;24(1):115.
29. Agarwala S et al. Deep learning for screening of interstitial lung disease patterns in high-resolution CT images. *Clin Radiol.* 2020;75(6):481.e1-481.e8.
30. Moran-Mendoza O et al. Computed tomography machine learning classifier correlates with mortality in interstitial lung disease. *Respir Investig.* 2024;62(4):670-6.
31. Ukita J et al. Detection of fibrosing interstitial lung disease-suspected chest radiographs using a deep learning-based computer-aided detection system: a retrospective, observational study. *BMJ Open.* 2024;14(1):e078841.
32. Yu W et al. Evaluating the robustness of deep learning models trained to diagnose idiopathic pulmonary fibrosis using a retrospective study. *Medical Physics.* 2025;52(6):4239-49.
33. Walsh SLF et al. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med.* 2018;6(11):837-45.
34. Mekov E et al. Artificial intelligence and machine learning in respiratory medicine. *Expert Rev Respir Med.* 2020;14(6):559-64.
35. Chan J, Auffermann WF. Artificial intelligence in the imaging of diffuse lung disease. *Radiol Clin North Am.* 2022;60(6):1033-40.
36. Levin DL. Deep learning and the evaluation of pulmonary fibrosis. *Lancet Respir Med.* 2018;6(11):803-5.
37. Chassagnon G et al. Deep learning-based approach for automated assessment of interstitial lung disease in systemic sclerosis on CT images. *Radiol Artif Intell.* 2020;2(4):e190006.
38. Chang M et al. Analysis of validation performance of a machine learning classifier in interstitial lung disease cases without definite or probable usual interstitial pneumonia pattern on CT using clinical and pathology-supported diagnostic labels. *J Imaging Inform Med.* 2024;37(1):297-307.
39. Castillo-Saldana D et al. Using quantitative computed tomographic imaging to understand COPD and fibrotic interstitial lung disease. *J Thorac Imaging.* 2020;35(4):246-54.
40. Sharma A. A comprehensive guide to convolutional neural networks – the ELI5 way. Towards Data Science. 2018. Available at: <https://medium.com/data-science/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>. Last accessed: 8 August 2025.
41. Fukuoka J et al. Finding “ground truth” for the diagnosis of UIP with artificial intelligence: standardization of diagnoses by 16 expert pathologists from 9 countries. *Mod Pathol.* 2020;DOI:10.1038/s41379-020-0485-4.
42. Chung J et al. Evaluating clinical utility of a UIP genomic classifier in subjects with and without a HRCT pattern of UIP. *Chest.* 2019;156(4):A175-8.
43. Fichtner-Feigl S et al. IL-13 signaling through the IL-13alpha2 receptor is involved in induction of TGF-beta1 production and fibrosis. *Nat Med.* 2006;12(1):99-106.
44. Wynn TA. Fibrotic disease and the T(H)1/T(H)2 paradigm. *Nat Rev Immunol.* 2004;4(8):583-94.
45. Bendstrup E et al. Current best clinical practices for monitoring of interstitial lung disease. *Expert Rev Respir Med.* 2022;16(11-12):1153-66.
46. Teramachi R et al. Deep learning for predicting acute exacerbation and mortality of interstitial lung disease. *Ann Am Thorac Soc.* 2025;22(5):689-97.
47. Libra A et al. Highlights on future treatments of IPF: clues and pitfalls. *Int J Mol Sci.* 2024;25(15):8392.
48. Swaminathan AC et al. Development of Duke Empower: a platform to engage, inform, probe, promote wellness, and foster community among patients with rare diseases. *Am J Respir Crit Care Med.* 2023;207:A4962.
49. Trusculescu AA et al. Deep learning in interstitial lung disease-how long until daily practice. *Eur Radiol.* 2020;30(11):6285-92.
50. Soffer S et al. Artificial intelligence for interstitial lung disease analysis on chest computed tomography: a systematic review. *Acad Radiol.* 2022;29 Suppl 2:S226-35.
51. Chen M et al. Acceptance of clinical artificial intelligence among physicians and medical students: a systematic review with cross-sectional survey. *Front Med (Lausanne).* 2022;9:990604.
52. Lambert SI et al. Author correction: an integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digit Med.* 2023;6(1):125.
53. Cestonaro C et al. Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. *Front Med (Lausanne).* 2023;10:1305756.
54. Yadav N et al. Data privacy in healthcare: in the era of artificial intelligence. *Indian Dermatol Online J.* 2023;14(6):788-92.
55. Rahman A et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster Comput.* 2022;DOI:10.1007/s10586-022-03658-4.
56. Administration USFaD. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD). 2019. Available at: <https://www.fda.gov/media/122535/download?attachment>. Last accessed date. 8 August 2025.
57. Calandriello L, Walsh SLF. Artificial intelligence for thoracic radiology: from research tool to clinical practice. *Eur Respir J.* 2021;57(5):2100625.
58. Handa T. The potential role of artificial intelligence in the clinical practice of interstitial lung disease. *Respir Investig.* 2023;61(6):702-10.