Review

The Role and Diagnostic Accuracy of Artificial Intelligence in Pulmonary Function Tests: A Systematic Review

THOMAS ORPWOOD[1], IRINA-LAVINIA SOICA[2]

[1]East Surrey Hospital, Surrey and Sussex Healthcare NHS Trust, Redhill, UK
[2]University College London Faculty of Medical Sciences, 74 Huntley St, London, UK

ABSTRACT: The rapid advancement of artificial intelligence (AI) has highlighted its potential as a supportive tool in pulmonary function test (PFT) interpretation given the inherent biological variation, inter-rater variability, lack of confidence in result interpretation and restricted access within resource-constrained settings. This study aims to systematically review the published literature on the diagnostic accuracy of AI-based interpretation of PFTs and evaluate its implications for current and future clinical practice in healthcare. After screening, forty-seven publications met the inclusion criteria and were analysed to create a narrative summary. Four main over-arching themes were identified from the available literature. AI appears to consistently outperform non-specialists in diagnostic accuracy of PFTs and showed a synergistic effect when used as an adjunct in both specialist and non-specialist settings. Using AI software also had greater diagnostic accuracy than clinicians when presented with suboptimal or limited clinical information and investigations. It was also observed that the implementation of AI can address the issue of inter-rater variability by giving more consultant interpretations. Chronic obstructive pulmonary disease diagnosis saw the greatest accuracy with other conditions such as obstructive sleep apnoea showing limited evidence for the introduction of AI as a diagnostic tool. The evidence suggests that AI has the potential to play a significant role in the future of healthcare but should be used as an adjunctive tool as opposed to an independent diagnostic decision maker. One such way it could be utilised is as a triage/screening tool to help bridge the gap between primary and secondary care.

KEYWORDS: *Artificial Intelligence, machine learning, pulmonary function tests, diagnostic accuracy.*

Introduction

Chronic respiratory conditions exhibit a high world-wide prevalence and are among the most common non-communicable diseases globally. The pervasiveness of noxious occupational, environmental and behavioural inhalation exposures plays a significant role in this disease burden [1].

Pulmonary function tests (PFTs) are investigations that support the clinical assessment and diagnosis of such conditions.

PFTs such as spirometry, diffusion capacity for carbon monoxide (DLCO) and lung volume assessments measure the airflow mechanics, gas transfer ability and other physiological capabilities of the lungs. However, PFTs exhibit an inherent biological variation, and their use must therefore be complemented with clinical expertise and supporting information [2].

As a result, a significant degree of inter-rater variability has been observed as a long-standing limitation of PFT interpretation [3].

Beyond interpretive variability, limited access to PFTs or additional investigations, procedural complexity of PFTs and low clinician confidence in PFT interpretation have been highlighted as barriers to a widespread implementation in healthcare [4].

The advancement of artificial intelligence (AI) in recent years has been highlighted as a possible supportive tool in PFT interpretation on account of its unique pattern recognition proficiency. Moreover, AI can analyse large clinical datasets, identify complex, non-linear patterns not readily apparent to clinicians and evolve with new data inputs following advancements in machine learning (ML) [5,6].

Experience and pattern recognition are instrumental to PFT interpretation thus indicating AI software with access to a large data set would be well suited to this function.

This in turn could lead to improved diagnostic accuracy, greater clinician confidence and enhance time efficiency for senior clinicians enabling focus on priority tasks [6].

To date no systematic review has been conducted to specifically evaluate the diagnostic accuracy of AI-based interpretation of PFTs. This study systematically reviews and synthesises the current literature pertaining to AI interpretation of PFTs, with a focus on diagnostic accuracy as opposed to quality control [7].

The current and potential future roles of AI are explored and further recommendations for studies to investigate remaining gaps in the available literature are given.

10.12865/CHSJ.52.02.01

Objective

To systematically review the published literature on the diagnostic accuracy of AI-based interpretation of PFTs and evaluate its implications for current and future clinical practice in healthcare.

Methods

Design

A systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [8] and JBI guidelines for systematic reviews [9].

The PICO framework was applied to structure the research question when formulating the review. The population of interest was adults undergoing pulmonary function testing (spirometry with or without full PFTs).

The intervention of interest was AI-based interpretation of PFTs with comparators including clinician interpretation or guideline-based assessment. The outcomes of interest were diagnostic accuracy, sensitivity / specificity, agreement (Cohen's kappa) and classification accuracy (normal/obstructive/restrictive/mixed).

Search strategy

A systematic search of PubMed, Cochrane Library and Embase databases was conducted using the following search term:

("Pulmonary Function Tests" OR pulmonary function test*[tiab] OR lung function test*[tiab] OR lung function analys*[tiab] OR spirometr*[tiab] OR PFT[tiab] OR PFTs[tiab])

AND ("Artificial Intelligence"[MeSH] OR "achine Learning"[MeSH] OR deep learning[tiab] OR artificial intelligence[tiab] OR AI[tiab] OR machine learning[tiab] OR algorithm*[tiab] OR neural network*[tiab] OR automated[tiab] OR automation[tiab])

To reaffirm the sensitivity of the search strategy, four key papers [10-13] were identified through Google Scholar and their inclusion checked in the retrieved records.

Screening and study selection

All retrieved records from the databases were deduplicated in Rayyan [14].

This software was also used to facilitate title and abstract screening.

Titles and abstracts were screened using the eligibility criteria in Table 1 (75% TO, 25% IS).

This was followed by full text screening (75% TO, 25% IS). Discrepancies were resolved through discussion between reviewers.

Table 1. Eligibility criteria used to screen papers for inclusion/exclusion in the review.

Eligibility criteria used to screen papers for inclusion/exclusion in the review

Inclusion criteria	Exclusion criteria
Studies using AI to interpret PFTs	AI for quality control only
Studies reporting diagnostic performance	Case reports / editorials
Human adult population	Reviews and meta-analyses
Original research (cohort, validation, diagnostic studies etc.)	Non-English publications
Publications from 2016-2026	Purely predictive model studies
	Paediatric study population

Data extraction

A structured data extraction form was developed in Microsoft Excel to retrieve relevant data from the included studies (Table 2).

Table 2. Data extraction table.

Title	Author	Study design	Objective	Methods	Results	Key findings
Application of Artificial Intelligence in the Interpretation of Pulmonary Function Tests [10]	Saad et al. (2025)	Diagnostic accuracy / validation study	To explore the accuracy of an AI algorithm / machine learning in the interpretation of pulmonary function tests with limited clinical data.	An AI programme was trained in retrospective PFT and clinical history. The AI was then compared to 8 senior pulmonologists in their diagnosis of patients based on PFT and limited clinical history. The diagnostic categories were normal, other (mixed or restrictive) or obstructive (further sub-classified into mild, moderate, moderately severe, severe or very severe). The diagnostic accuracy was compared against a diagnosis made by 2 senior pulmonologists who had access to a full, extensive catalogue of clinical information and investigations.	Diagnostic accuracy (true positive) of AI was 86.59%. Clinician was 65.82% (ranged from 54 to 79%). Fleiss's Kappa was 0.46 amongst the 8 pulmonologists.	AI demonstrated higher diagnostic accuracy in the interpretation of pulmonary function tests when limited clinical information was available, compared with senior pulmonologists. The study also reported moderate inter-rater variability among pulmonologists, highlighting inconsistency in clinician interpretation.
AI-Assisted Interpretation of Pulmonary Function Tests: Enhancing Diagnostic Precision in Respiratory Medicine [11]	Singh et al. (2025)	Diagnostic accuracy / validation study	To evaluate the diagnostic accuracy of an AI-based interpretation system for PFTs in comparison with expert pulmonologist consensus and to assess its applicability in clinical settings.	An AI programme was trained on 10,000 labelled PFTs. Trained to classify as normal, restrictive, obstructive or mixed. 3 experienced pulmonologists interpreted 200 PFTs with clinical details in adults using 2019 ATS/ERS guidelines and their consensus was used as the "gold-standard" diagnosis. The AI then	Diagnostic accuracy (true positive) of the AI was 95%, 95%, 88% and 90% respectively for normal, obstructive, restrictive and mixed diagnoses. The AI system demonstrated a positive predicted value (PPV) of 92.1%, negative predicted value (NPV) 92.6% and a sensitivity and	AI demonstrated a high level of diagnostic accuracy and strong concordance with expert pulmonologists in the interpretation of pulmonary function tests. However, as the reference standard was based on pulmonologist consensus, the AI's performance was inherently constrained to being equivalent to or lower than this

				interpreted the same data and yielded an interpretation.	specificity of 91.5% and 93.2% respectively. Cohen's kappa 0.86 indicating strong agreement with the pulmonologists.	benchmark. Notably, reduced sensitivity was observed in the detection of restrictive patterns (88%). The absence of comparison with a broader range of clinicians may limit the generalisability of these findings.
Validation of artificial intelligence spirometric diagnostic support software in primary care: a blinded diagnostic accuracy study [12]	Sunjaya et al (2025)	Diagnostic accuracy / validation study	To assess the diagnostic performance of an AI diagnostic support software in the identification of COPD and other chronic respiratory diseases when provided with primary care spirometry data against a clinical reference diagnosis.	A previously validated AI (ArtiQ.Spiro) was used to interpret PFTs carried out in primary care clinics with limited clinical information and its diagnostic accuracy compared to that of the "gold-standard" diagnosis determined by a panel of two (three for diagnostic disagreements) pulmonologists with access to extensive clinical notes and secondary care investigations.	For COPD diagnoses, the AI had a sensitivity of 84.0%, specificity of 86.8%, NPV of 85.1% and PPV of 85.9%. AUC was 0.914 and accuracy was 85.4%. For Asthma diagnoses, the AI had a sensitivity of 55.1%, specificity of 86.9%, NPV of 94.8% and PPV of 30.9%. AUC was 0.814 and accuracy was 83.8%. For ILD diagnoses, the AI had a sensitivity of 75.1%, specificity of 85.9%, NPV of 92.3% and PPV of 60.5%. AUC was 0.900 and accuracy was 83.8%. For "normal" diagnoses, the AI had a sensitivity of 33.3%, specificity of 96.0%, NPV of 98.1% and PPV of 18.9%. AUC was 0.871 and accuracy was 94.4%.	AI demonstrated a high degree of usability in supporting COPD diagnosis in community settings, where it represented the most prevalent condition within the study population. Diagnostic performance was more variable for other conditions, such as interstitial lung disease (ILD) and asthma, with lower positive predictive values and sensitivity observed. This likely reflects the greater reliance on pulmonary function tests in COPD diagnosis, whereas PFTs play a more supportive role in other respiratory conditions. Notably, AI performance was highest in the detection of more advanced disease, particularly GOLD stage 3 and 4 COPD.
Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests [13]	Topalovic et al (2019)	Diagnostic accuracy / validation study	To explore the accuracy and interrater variability of pulmonologists when interpreting PFTs compared with artificial intelligence (AI)-based software.	120 pulmonologists from 16 European hospitals assessed 50 cases with PFT and clinical summary, resulting in 6000 independent interpretations. The AI software was exposed to the same data. American Thoracic Society/European Respiratory Society guidelines were used as the gold standard for PFT pattern interpretation. The gold standard for diagnosis was derived from clinical history, PFT as well as additional tests.	Pulmonologists made correct diagnoses in 44.6+/-8.7% of the cases with a large interrater variability (kappa=0.35). The AI-based software perfectly matched the PFT pattern interpretations (100%) and assigned a correct diagnosis in 82% of all cases.	AI was associated with reduced inter-rater variability in the interpretation of pulmonary function tests. It also outperformed clinicians when only spirometry and limited clinical information were available, compared to scenarios where full diagnostic data and investigations were accessible.
Integrating AI-assisted spirometry: Validation and user experience in a COPD diagnostic pathway [51]	Giffen et al. (2025)	Diagnostic accuracy / validation study	To compare the performance of ArtiQ.Spiro with clinician interpretation as part of efforts to integrate AI into a COPD diagnostic pathway.	2 retrospective cohorts in a direct-access COPD diagnostic pathway had their spirometry data analysed by ArtiQ.Spiro (AI) and compared to the interpretations of ATRP-registered clinicians and final diagnostic pathway outcomes. A prospective validation was also done in which the AI's interpretations were obtained by the assistant physiologists during data acquisition for 125 consecutive tests.	Near complete concordance between AI diagnosis of "normal" and clinician diagnosis of "normal" (NPV 0.942) and good agreement of AI alone in COPD identification (PPV 0.69). In the prospective validation, 90% of tests were correctly interpreted by both assistant physiologists and ArtiQ.Spiro. 10% of initial interpretations by assistant physiologists were incorrect but correctly identified by the AI.	AI demonstrated high accuracy in identifying normal pulmonary function test results, but comparatively lower accuracy in diagnosing COPD. The findings suggest a potential synergistic role for AI when used alongside assistant physiologists, with combined interpretation improving overall diagnostic performance.
Machine learning based detection of true ventilatory restriction [52]	Saha et al (2025)	Diagnostic accuracy / validation study	To determine if a diagnostic platform using machine learning to integrate spirometric data with demographics will enable accurate detection of true ventilatory restriction.	21,062 patients undergoing spirometry and lung volume testing at a quaternary referral hospital in the United states of America were selected as the study population. A LightGBM machine learning framework was trained on 80% of the data before using the remaining 20% as test data and comparing the accuracy against restrictive spirometry patterns alone to detect true ventilatory restriction. The FEV1, FVC, FEV1/FVC, FEV1 % predicted and FVC % predicted as well as age, sex and BMI were used by the machine learning framework to detect patients with true ventilatory restrictions.	The accuracy of RSP in detecting true ventilatory restriction was 0.61 with a sensitivity of 0.42, specificity of 0.91, PPV of 0.88, NPV of 0.51. The accuracy of the machine learning framework was 0.78 with a sensitivity of 0.74, specificity of 0.86, PPV of 0.89 and NPV 0.68.	Machine learning frameworks demonstrated greater accuracy in identifying true ventilatory restriction compared with the use of isolated spirometric parameters.
Artificial Intelligence- Based Deep learning Model for Detecting Central and Upper Airway Obstruction [44]	Prevot et al (2025)	Diagnostic accuracy / validation study	To address the gap in standardised diagnosis of intra- and extrathoracic airway obstruction using flow-volume loop patterns and improve the sensitivity of diagnostic decision making through the use of an AI deep learning model.	1094 flow-volume loops were analysed (766 healthy controls, 256 with fixed airway obstruction, 72 from COPD patients). 90:10 training: test split.	The AI demonstrated an overall PPV of 0.0.886, a sensitivity of 0.915 and an F1 score of 0.893 showing good concordance previously assigned diagnosis.	The AI demonstrated a high level of accuracy in detecting pathological patterns; however, its ability to differentiate between similar disease entities was limited. Interpretation of these findings is constrained by the availability of only abstract-level data for this study.
Machine Learning for Enhanced COPD Diagnosis: A Comparative Analyses of Classification Algorithms [38]	Elashmawi et al (2024)	Observational study with diagnostic accuracy / validation component	To determine the most effective machine learning algorithm in COPD diagnosis /prediction based on clinical information and spirometry data.	Various machine learning algorithms were compared against each other in detecting COPD based on spirometry and readily accessible patient data. 1522 data points were allocated	Random forest classifier outperformed models based on the following principles: logistic regression, support vector machine, gradient boosting classifier, Gaussian naive bayes, K-nearest	Random forest classifiers were identified as among the most effective machine learning algorithms for COPD diagnosis. This approach has been increasingly incorporated

				to the training set and 508 for testing.	neighbours classifiers, decision tree classifier, artificial neural networks.	into recent AI models for spirometric analysis.
Diagnosis and Severity if COPD Using a Novel Fast-Response Capnometer and Interpretable Machine Learning [36]	Talker et al (2024)	Diagnostic accuracy / validation study	To apply interpretable machine learning techniques to capnography data recorded by the N-Tidal device across five clinical studies and evaluate the performance of the diagnostic and severity classifiers.	115,053 capnograms (limited to 14 per patient in analysis) were collected from 1114 patients. Data was split into an 80:20 training: test split. Interpretations labelled patients as either COPD or non-COPD as well as the severity of COPD. Compared to gold-standard diagnosis of patients as per NICE guidelines.	The best performing machine learning model displayed n accuracy, AUROC, sensitivity, specificity, NPV and PPV of 0.811, 0.890, 0.771, 0.850, 0.792 and 0.834 respectively. Little variability between models; standard deviation of accuracy of the 3 models tested was <0.03. High performance seen with GOLD 2-4 diagnoses, less so with GOLD 1.	Machine learning algorithms demonstrated a high degree of concordance with reference standard diagnoses when incorporating capnography data alongside spirometry and clinical variables. Minimal variability was observed between different machine learning models. However, diagnostic performance was lower in mild COPD cases, which were the most challenging to detect.
Artificial Intelligence Applied to Forced Spirometry in Primary Care [16]	Mendez et al (2024)	Diagnostic accuracy / validation study	The goal is to create an AI model based on machine learning capable of predicting the presence or absence of an obstructive pattern using variables with the highest predictive power derived from an active search program for COPD in primary care.	Patients aged 30-80 with at least a 10 pack-year history were selected as the study population from six primary care centers in Velencia. 80:20 training: test split. The GOLD classification was used to assign diagnoses. The machine learning algorithm was given basic clinical information and spirometric data.	Machine learning was able to predict obstructive spirometry patterns with an accuracy and precision exceeding 90% using predicted FEV1 pre-bronchodilator therapy values.	The model demonstrated good precision and accuracy in detecting obstructive patterns using limited clinical information in a primary care setting.
Time-series Machine Learning Model Classifies Lung Function Using Spirometry Flow-Volume Loops Alone [53]	Mac et al (2024)	Diagnostic accuracy / validation study	To use flow-volume loops interpreted by AI to diagnose restrictive an obstructive airway patterns with a particular focus on small airway disease.	2871 pulmonary function tests including flow-volume loops from 895 patients were analysed and classified based on ATS/ERS guidelines. A machine learning model was trained on this data and a separate subsection used for testing that had not been seen before.	Accuracy of the AI was demonstrated at >90%.	AI demonstrated a high degree of diagnostic accuracy in identifying ventilatory defects, including small airway disease, when flow-volume loop data were incorporated into machine learning models, particularly in the absence of plethysmography.
Clinical Evaluation of Artificial Intelligence Supported Software in the Interpretation of Spirometry Results in a Primary Care Setting [17]	Maes et al (2024)	Diagnostic accuracy / validation study	To compare the performance of GP PFT interpretation with and without the use of AI.	A total of 6 primary care practices in Belgium used ArtiQ.Spiro, during a 3-month period alongside existing referral structures. The accuracy of the diagnostic hypothesis based on the clinical evaluation and interpretation of spirometry, supported by the AI software, was established by comparison to the diagnosis made by an expert panel of 3 pulmonologists. This panel reviewed the same data and all relevant information from the patient record 3 months after the spirometry session.	In 77% of cases, GPs agreed with the diagnosis proposed by ArtiQ.Spiro. Combined with spirometry and AI support, GPs labelled patients correctly in 67% of patients. When comparing the AI-proposed diagnosis with the pulmonologists' diagnosis, 82% of diagnoses were correct.	Non-specialists demonstrated lower concordance with pulmonologist consensus diagnoses compared to AI in the interpretation of pulmonary function tests. These findings suggest that AI may play a valuable role in improving spirometric interpretation in primary care settings, particularly when used by non-specialists prior to specialist referral.
Evaluating artificial intelligence-interpreted against pulmonologist-interpreted spirometry results: a comparative study [21]	Antonio, Montejo and Sy (2024)	Diagnostic accuracy / validation study	To compare the spirometry reading interpretation between AI and pulmonologists in terms of agreement and reliability.	197 patients were assessed using spirometry with bronchodilator studies to compare AI's accuracy to pulmonologists.	AI and pulmonologists had similar findings in normal (AI 50.76%, pulmonologists 55.1%-63.96%), restrictive (AI 19.29%, pulmonologists 15.74%-19.29%), and obstructive (AI 18.78%, pulmonologists 15.82%-20.30%) ventilatory defects, but differed in mixed defects and severity classification. AI's identification of significant bronchodilator response (15.23%) was within pulmonologists' range (5.58%-12.69%).	AI demonstrated comparable diagnostic performance and concordance with pulmonologists in the interpretation of pulmonary function tests.
Collaboration between explainable artificial intelligence and pulmonologists improves the accuracy of pulmonary function test interpretation [28]	Das et al (2023)	Diagnostic accuracy / validation study	To assess if the collaboration between a pulmonologist and AI with explanations (explainable AI (XAI)) is superior in diagnostic interpretation of pulmonary function tests (PFTs) than the pulmonologist without support.	The study was conducted in two phases, a monocentre study (phase 1) and a multicentre intervention study (phase 2). Each phase utilised two different sets of 24 PFT reports of patients with a clinically validated gold standard diagnosis. Each PFT was interpreted without (control) and with XAI's suggestions (intervention). Pulmonologists provided a differential diagnosis consisting of a preferential diagnosis and optionally up to three additional diagnoses.	The number of diagnoses in the differential diagnosis did not reduce, but diagnostic confidence and inter-rater agreement significantly increased during intervention. Pulmonologists updated their decisions with XAI's feedback and consistently improved their baseline performance if AI provided correct predictions.	The use of explainable AI was associated with improved diagnostic confidence and consistency in pulmonary function test interpretation compared with unaided clinician assessment, relative to a reference standard diagnosis.
AI models for respiratory disease from spirometry alone: validation based on UK Biobank [54]	Elmahy et al (2023)	Diagnostic accuracy / validation study	To evaluate the ability of AI to diagnose respiratory diseases in primary care on a representative dataset.	An AI model was trained on 1609 subjects who underwent PFTs in hospital with the addition of 500 healthy individuals from the UK Biobank. It was then tested on an independent dataset from the UK Biobank. This set consisted of physician-diagnosed asthma, bronchiectasis, COPD and ILD that have acceptable quality spirometry within a year of	The AI demonstrated an overall accuracy of 76%. Poor sensitivity (0.25) and PPV (0.06) for bronchiectasis.	The model demonstrated moderate discriminative ability between different diagnostic categories.

				diagnosis that showed at least subtle signs of abnormal spirometry.		
ArtiQ.PFT: AI-powered decision support for the diagnosis of lung diseases: Does it help the pulmonologist to be more accurate? [29]	Autoren Desbordes et al (2023)	Diagnostic accuracy / validation study	To ascertain the accuracy and impact of AI on diagnosis and planned next steps in pulmonologists assessing PFTs.	60 cases are presented to 20 pulmonologists from 3 different countries. For each case, the full PFTs and a short clinical background are given. Pulmonologists provided a primary and up to 3 differential diagnoses as well as the next step they would take, once without and once with the support of ArtiQ.PFT.	Pulmonologists gave a similar number of differentials whether they used AI or not. It did seem to improve the diagnostic accuracy of clinicians when compared to interpretation without the aid of AI by 18.9%. It also seemed to influence the next steps pulmonologists wanted to take with an increase in suggestions for "immediate further investigations and "plan follow-up" as opposed to "back to GP" and "begin treatment".	AI demonstrated a synergistic effect when used alongside pulmonologists, improving diagnostic accuracy and influencing subsequent clinical management decisions.
Exploratory study of a risk prediction artificial intelligence model to diagnose asthma and COPD in Mexico [40]	Hernandez et al (2023)	Diagnostic accuracy / validation study	To develop an artificial intelligence (AI) model to standardize the diagnoses of asthma and COPD.	4223 subjects with asthma and 3830 with COPD were recruited between 2019 and 2021. Subject demographics were collected for all subjects. Using clinical variables in addition to spirometric variables, an artificial intelligence model was created and trained.	The AI identified 2 scorecards scaling from 0 to 250 and 300 to 620 points for asthma and COPD, respectively. A total of 14.95% subjects were diagnosed with asthma, and most patients (61.22%) had a score ranging from 176 to 220. COPD was diagnosed in 3.10% with 41.38% having a score of 540 to 620. The sensitivity, specificity, positive, and negative predictive values for asthma and COPD were 64%, 72%, 30%, 92%, and 83%, 77%, 12%, and 99%, respectively.	Interpretation of this study is limited by the lack of detailed methodological information. However, the reported low positive predictive values for asthma and COPD suggest a higher rate of false positives and potential overestimation of disease prevalence.
Development and evaluation of a computerized algorithm for the interpretation of pulmonary function tests [30]	Huang et al (2024)	Diagnostic accuracy / validation study	To compare a computerised algorithm to clinician interpretation of pulmonary function tests.	The accuracy of an AI generated model was compared to the clinician's interpretation of 6889 PFTs from patients under the care of Duke from 2018-2020 using ATS guidelines.	The agreement rates (Cohen's kappa) for obstruction, restriction and abnormal DLCO were 94.4% (0.868), 99.0% (0.979) and 87.9% (0.750) respectively. In 4,711 PFTs with FEV1/FVC≥0.7, the algorithm identified 190 tests with small airway disease index < lower limit of normal (LLN), suggesting small airway dysfunction. Of these, the clinicians (67.9%) noted 129 tests.	AI demonstrated a high level of concordance with clinician diagnoses, with fewer overt diagnostic errors compared to clinicians. Additionally, AI showed improved capability in identifying potential small airway disease.
Deep learning using multilayer perception improves the diagnostic acumen of spirometry: a single-centre Canadian study [23]	Mac et al (2022)	Diagnostic accuracy / validation study	To compare AI with clinician interpretation of spirometry alone in formulating a diagnosis when full PFTs including body plethysmography are not available.	The AI was trained using full sets of PFTs and clinical information and then compared with the ATS clinical decision tree and trained pulmonologists in terms of accuracy of clinical diagnosis made when all information is available.	The final model significantly outperformed (94.67%±2.63%, p<0.01 for both) interpretation of 82 spirometry tests by the decision tree (75.61%±0.00%) and pulmonologists (66.67%±14.63%).	AI outperformed both clinical decision trees and expert clinicians in scenarios where key clinical information was limited, which would otherwise be required to establish a reference standard diagnosis.
Automated interpretation of the pulmonary function test by a portable spirometer in Chinese adults [55]	Zhou et al (2022)	Diagnostic accuracy / validation study	To investigate the accuracy of automated interpretation of the PFT measured by a portable Yue Cloud spirometer in Chinese adults compared to standard care.	The portable Yue Cloud spirometer and AI diagnosis was compared against adults undergoing standard spirometry investigation with the conventional Jaeger MasterScreen and standard interpretation procedures.	Concordance verified by the Kappa statistic showed an excellent agreement in pulmonary dysfunction, small airway dysfunction, and severity evaluation between the Jaeger and Yue Cloud spirometers. The kappa values were all greater than 0.80 (pulmonary dysfunction, $\kappa = 0.890$; small airway dysfunction, $\kappa = 0.828$; and severity, $\kappa = 0.874$).	Portable or handheld spirometry combined with AI-based interpretation demonstrated high concordance with current standard diagnostic practice.
Deep Learning for Automatic Upper Airway Obstruction Detection by Analysis of Flow-Volume Curve [24]	Wang et al (2022)	Diagnostic accuracy / validation study	To develop a deep learning model to detect upper airway obstruction patterns and compare its performance to that of lung function clinicians.	Out of 45831 patient's spirometry records, 351 had upper airway obstruction confirmed through further assessment. These cases were used to train and test the AI and clinicians as well as 200 normal sets of PFTs. The two had their accuracy and agreement compared when viewing spirometry.	432 clinicians evaluated 80 spirometry cases. They assigned an accuracy of 41.2% (±15.4) with poor agreements ($\kappa = 0.12$). For the same cases, the model generated a correct detection of 81.3% ($p < 0.0001$).	AI demonstrated superior diagnostic performance compared to clinicians when limited data were available. It was also associated with reduced inter-rater variability among clinicians.
Explainable machine learning methods and respiratory Oscillometry for the diagnosis of respiratory abnormalities in sarcoidosis [56]	de Lima er al (2022)	Diagnostic accuracy / validation study	To assess the accuracy of explainable machine learning in detecting pulmonary abnormalities in sarcoid patients.	An explainable machine learning algorithm was developed and compared against other non-explainable algorithms/AI in interpreting forced Oscillometry to diagnose respiratory changes associated with sarcoidosis. 72 individuals were analysed with a mix of healthy controls, sarcoid diagnosis and spirometry abnormalities, sarcoid diagnosis and no spirometry.	Automatic classifiers could increase sarcoidosis diagnosis, particularly in those with normal spirometry. Using machine learning improved diagnostic accuracy compared to looking at isolated parameters. Explainable machine learning yielded an AUC >0.900 in both those with sarcoidosis and no spirometry changes as well as with abnormalities compared to controls.	Explainable machine learning approaches demonstrated accurate diagnostic performance and showed potential for application across a range of respiratory conditions, including sarcoidosis with pulmonary involvement.

Machine Learning for the identification of Restriction Without Lung Volume Measurements [25]	Moffett et al (2022)	Diagnostic accuracy / validation study	To identify a method of identifying lower limits of normal total lung volume without access to lung volume testing facilities.	ATS guidelines were used to interpret pulmonary function tests performed at the University of Pennsylvania Health System which included both spirometry and lung volume measurements. A 75-25 split sample was used with repeated 10-fold cross-validation to train AI models to identify restriction using the age, sex, standing height, percent predicted forced expiratory volume in 1 second (FEV1), percent predicted forced vital capacity (FVC), percent predicted FEV1/FVC, and the percent predicted forced expiratory flow (FEF25-75).	9,427 pulmonary function tests were used, of which 2,297 (24.3%) met ATS criteria for restriction. Performance was similar across each of the AI models, with the best performance observed with the random forest model. The c-statistic of this model was 0.980 (95% CI 0.976-0.984), while the area under the precision-recall curve was 0.934.	AI demonstrated a high degree of diagnostic accuracy when presented with limited clinical information.
Deep Learning-Based Analytic Models Based on Flow-Volume Curves for Identifying Ventilatory Patterns [18]	Wang et al (2022)	Diagnostic accuracy / validation study	To explore the accuracy of deep learning-based analytic models based on flow-volume curves in identifying the ventilatory patterns.	The gold standard for identifying ventilatory patterns was the ATS/ERS guidelines. One physician evaluated the ventilatory patterns according to the international guidelines. Ten deep learning models were developed to identify patterns from the flow-volume curves. The patterns obtained by the best-performing model were cross-checked with those obtained by the physicians.	90 physicians interpreted 100. The average accuracy achieved by the physicians was 76.9 +/- 18.4% with a moderate agreement (kappa = 0.46), physicians from primary care settings achieved a lower accuracy (56.2%), while the AI model accurately identified the ventilatory pattern in 92.0% of the 100 cases.	AI outperformed clinicians in the interpretation of flow-volume loops and was associated with greater consistency in diagnostic classification. This advantage was particularly evident when compared with general practitioners.
Machine learning associated with respiratory Oscillometry: a computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis [57]	Andrade et al (2021)	Diagnostic accuracy / validation study	To evaluate the performance of several ML algorithms associated with the respiratory Oscillometry analysis to aid in the diagnostic of respiratory changes in systematic sclerosis.	Oscillometry and spirometric exams were performed in 82 individuals, including controls (n = 30) and patients with systemic sclerosis with normal (n = 22) and abnormal (n = 30) spirometry. A machine learning algorithm was configured based on this. The best Oscillometry parameter (BOP) was sought in terms of accuracy in identifying patients who were controls or had systemic sclerosis with/without abnormal spirometry. This was then compared with the machine learning model.	The BOP was dynamic compliance which provided moderate accuracy (AUC = 0.77) in the scenario control group versus patients with sclerosis and normal spirometry. In the scenario control group versus patients with sclerosis and altered spirometry (CGvsPSAS), the BOP obtained high accuracy (AUC = 0.94). In control vs systemic sclerosis patients with normal spirometry, the machine learning model achieved the best result (AUC = 0.90), significantly improving the accuracy in comparison with the BOP ($p < 0.01$), while in control vs systemic sclerosis patients with abnormal spirometry, another machine learning model obtained the best results (AUC = 0.97), also significantly improving the diagnostic accuracy ($p < 0.05$).	AI enhanced the interpretation and diagnostic performance of oscillometry.
Machine learning for nocturnal diagnosis of chronic obstructive disease using digital oximetry biomarkers [34]	Levy et al (2021)	Diagnostic accuracy / validation study	To examine patterns and dynamics on COPD patients during overnight oximetry time series that may be unique to this condition.	The study used 350 overnight polysomnography recordings to extract oximetry-based digital biomarkers and demographic features, which were then used to train machine learning models to classify COPD vs non-COPD. The data were pre-processed, feature-engineered, and split into training/test sets, with models (e.g., random forest and logistic regression) evaluated using nested cross-validation and performance metrics such as AUROC and F1 score. Gold standard was achieved by performing spirometry on these patients.	A total of 350 patients' polysomnography (PSG) recordings were used. A random forest (RF) classifier was trained using these features. The RF classifier obtained F1 = 0.86 +/- 0.02 and AUROC = 0.93 +/- 0.02 on the test sets. A total of 8 COPD individuals out of 70 were misclassified. No severe cases (GOLD 3-4) were misdiagnosed. Including additional non-oximetry derived PSG biomarkers resulted in minimal performance increase.	AI demonstrated a high degree of diagnostic accuracy when assessing alternative or minimal data inputs. Performance was strongest at detecting more severe disease, whereas milder cases were more challenging to identify.
Deep learning algorithm for the classification of spirometries using flow-volume curves: proof of concept study [45]	Exarchos et al (2021)	Diagnostic accuracy / validation study	To evaluate a deep learning-based image classification algorithm that analyses solely the flow-volume loop curve and classifies it into obstructive, restrictive or normal.	The algorithm was trained and tested on a set of 108 flow-volume curves, almost equally distributed across the three patterns and accuracy assessed.	The algorithm yielded an overall accuracy of 96%; the class-specific performances for correctly identifying obstructive, restrictive and normal patterns were 94%, 97% and 98%, respectively.	AI demonstrated high accuracy in classifying spirometry results into broad categories, including normal and abnormal (obstructive and restrictive) patterns.
Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease [39]	Bodduluri et al (2020)	Diagnostic accuracy / validation study	To explore if it is possible that machine-learning algorithms can be trained on spirometry data to identify emphysema airway disease contributing to airway obstruction.	Participants enrolled in a large multicentre study (COPDGene) were included. The data points from expiratory flow-volume curves were trained using a deep-learning model to predict structural phenotypes of COPD based on CT findings, and results were compared with traditional spirometry metrics and an optimized random forest classifier. Area under the receiver operating characteristic curve (AUC) and weighted F-score were used to	The neural network was more accurate in discriminating predominant emphysema/airway phenotypes (AUC 0.80, 95%CI 0.79-0.81) compared with traditional measures of spirometry, FEV1/FVC (AUC 0.71, 95%CI 0.69-0.71), FEV1% predicted (AUC 0.70, 95%CI 0.68-0.71), and random forest classifier (AUC 0.78, 95%CI 0.77-0.79). The neural network was also more	AI demonstrated improved capability in interpreting spirometry to identify COPD phenotypes; however, it is not a substitute for reference standard diagnostic investigations.

				measure the discriminative accuracy of a fully convolutional neural network, random forest, and traditional spirometry metrics to phenotype CT as normal, emphysema-predominant (>5% emphysema), airway-predominant (Pi10>median), and mixed phenotypes. Comparisons were also made for the detection of functional small airway disease phenotype (>20% on parametric response mapping).	accurate in discriminating predominant emphysema/small airway phenotypes (AUC 0.91, 95%CI 0.90-0.92) compared with FEV1/FVC (AUC 0.80, 95%CI 0.78-0.82), FEV1% predicted (AUC 0.83, 95%CI 0.80-0.84).	
Application of machine learning to predict obstructive sleep apnoea syndrome severity [43]	Mencar et al (2020)	Diagnostic accuracy / validation study	To test the efficacy and clinical applicability of different machine learning methods based on demographic information and questionnaire data to predict obstructive sleep apnoea (OSA) syndrome severity.	Data surrounding demographic characteristics, spirometry values, gas exchange (PaO2, PaCO2) and symptoms of 313 patients with a previous diagnosis of obstructive sleep apnoea syndrome were collected. After principal component analysis, 19 variables were selected which were used for further preprocessing and to eventually train seven types of classification models and five types of regression models to evaluate the prediction ability of obstructive sleep apnoea syndrome severity, represented either by class or by apnoea-hypopnea index.	The best average classification accuracy on test sets is 44.7 percent, with the same average sensitivity (recall). In only 5.7 percent of cases, a severe obstructive sleep apnoea syndrome (class 4) is misclassified as mild (class 2). Regression results show a minimum achieved root mean squared error of 22.17.	AI demonstrated comparatively lower accuracy in diagnosing and classifying obstructive sleep apnoea (OSA) when using predominantly non-diagnostic variables, relative to its performance in other respiratory conditions.
Deep learning facilitates the diagnosis of adult asthma [58]	Tomita et al (2019)	Diagnostic accuracy / validation study	To explore whether the use of deep learning with combinations of symptom-physical signs and objective tests, such as lung function tests and the bronchial challenge test, can improve model performance in predicting the initial diagnosis of adult asthma compared to the conventional machine learning diagnostic method.	566 adult out-patients who visited Kindai University Hospital for the first time with complaints of non-specific respiratory symptoms had their data collected. Asthma was comprehensively diagnosed by specialists based on symptom-physical signs and objective tests. Model performance metrics were compared to logistic analysis, support vector machine (SVM) learning, and the new deep neural network (DNN) model.	For the diagnosis of adult asthma based on symptoms and physical signs alone, the accuracy of the DNN model was 0.68, whereas that for the SVM was 0.60 and for the logistic analysis was 0.65. When adult asthma was diagnosed based on symptoms and physical signs, biochemical findings, lung function tests, and the bronchial challenge test, the accuracy of the DNN model increased to 0.98 and was significantly higher than the 0.82 accuracy of the SVM and the 0.94 accuracy of the logistic analysis.	The AI demonstrated low diagnostic accuracy when relying solely on symptoms and physical signs; however, performance improved significantly with the inclusion of additional clinical and physiological data.
Artificial intelligence helps detecting lung disease with pulmonary function tests [26]	Topalovic et al (2018)	Diagnostic accuracy / validation study	To develop a smart software using AI which improves the clinical reading of a lung function and suggests a respiratory disease diagnosis if possible.	1430 subjects with respiratory symptoms had data extracted from 33 Belgian hospitals to develop the algorithm. The final diagnosis (healthy, asthma, COPD, ILD, neuromuscular disease, chest wall or pleural disease, pulmonary vascular disease, other obstructive disease) was obtained from clinical history, lung function and all additional tests, and confirmed by an expert panel. A cloud-based solution was incorporated into the clinical setting to validate the accuracy of the algorithm on a random sample of 136 new subjects. Finally, the software diagnoses were compared with the diagnostic opinions of 85 pulmonologists (from 11 different European hospitals) provided with PFT and clinical data of 50 new subjects.	At the external multicentric validation, the software-based automated diagnoses (82% accuracy) were superior to the suggested diagnoses of 85 pulmonologists (44.6 +/- 3.7 % mean accuracy). Great disagreement between readers is observed with a low kappa score of 0.34.	High levels of discordance were observed among expert clinicians, whereas AI systems demonstrated high diagnostic accuracy.
Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints [59]	Topalovic et al (2017)	Diagnostic accuracy / validation study	To assess whether an unbiased machine learning framework integrating lung function with clinical variables may provide alternative decision trees resulting in a more accurate diagnosis.	968 subjects admitted for the first time to a pulmonary practice had data collected. The final clinical diagnosis was based on the combination of complete pulmonary function with the investigations that were decided at the physician's discretion. Clinical diagnoses were separated into 10 different groups and validated by an expert pane.	ATS/ERS algorithm resulted in a correct diagnostic label in 38% of the subjects. Chronic obstructive pulmonary disease (COPD) was detected with an acceptable accuracy (74%), whereas all other diseases were poorly identified. The new data-based decision tree improved the general accuracy to 68% after 10-fold cross-validation when detecting the most common lung diseases, with a significantly higher positive predictive value and sensitivity for COPD, asthma, interstitial lung disease, and neuromuscular disorder (0.83/0.78, 0.66/0.82, 0.52/0.59, and 1.00/0.54, respectively).	Computer-based algorithms were associated with improved diagnostic accuracy.

Artificial intelligence detects lung diseases using pulmonary function tests [22]	Topalovic et al (2017)	Diagnostic accuracy / validation study	To explore whether the use of artificial intelligence may result in a more reliable interpretation of pulmonary function tests leading to an accurate detection of different lung diseases.	Data of 1452 subjects from two different cohorts were collected: 1) Belgian Pulmonary Function Study (N=968), a prospective cohort study that enrolled all successive new subjects in 33 different Belgian hospitals. A final clinical diagnosis was based on the combination of complete pulmonary function with the investigations that were decided at the physician's discretion and lastly validated by an expert panel. 2) Leuven cohort (N=484), a retrospectively collected data at the outpatient clinic of the University Hospital of Leuven, consisting of patients with known disease presence. Algorithm for disease detection and discrimination was developed using neural network model.	Applying the algorithm on the dataset presented an accuracy of 81% with a high positive predictive values and sensitivity for most of the disease groups: COPD, asthma, interstitial lung disease, neuromuscular disorder, pulmonary vascular disease and thoracic deformity (89% and 92%, 79% and 89%, 83% and 84%, 66% and 57%, 77% and 51%, 75% and 29%, respectively).	AI demonstrated high sensitivity and positive predictive value in the detection of disease.
Late Breaking Abstract-Applying artificial intelligence on pulmonary function tests improves the diagnostic accuracy [27]	Topalovic et al (2017)	Diagnostic accuracy / validation study	To develop a smart software which improves the clinical reading of a lung function and suggests a respiratory disease diagnosis if possible.	1430 subjects with respiratory symptoms were taken from 33 Belgian hospitals to develop the algorithm. The final diagnosis (healthy, asthma, COPD, ILD, neuromuscular disease, chest wall or pleural disease, pulmonary vascular disease, other obstructive disease) was obtained from the clinical history, lung function and all additional tests, and confirmed by an expert panel. A cloud-based solution was incorporated into the clinical setting to validate the accuracy of the algorithm on a random sample of 86 new subjects. Finally, the software diagnoses were compared with the diagnostic opinions of 16 pulmonologists provided with PFT and clinical data.	AI software presented a high accuracy of 74% after 10-fold cross-validation when detecting lung diseases (8 possible disease categories). The high accuracy was maintained in a real clinical setting (76%). The software-based automated diagnoses (76% accuracy) were superior over the suggested diagnoses of pulmonologists (53.5 +/- 6 % mean accuracy).	AI outperformed pulmonologists in diagnostic interpretation when key clinical data were limited.
Applying machine learning and pulmonary function data to detect interstitial lung disease in systemic sclerosis [41]	Le-Dong et al (2017)	Diagnostic accuracy / validation study	To apply a Machine learning (ML) approach to detect the presence of Interstitial Lung disease (ILD) in patients with systemic sclerosis (SSc), using combined data from multiple pulmonary function tests.	PFT data (Spirometry, Plethysmography and DLNO-CO measurements, all converted into Z-score scale) were obtained in 273 patients with confirmed diagnosis of SSc and 50 healthy non-smokers. The main outcome is ILD, as determined by HRCT.	AI software presented a high accuracy of 74% after 10-fold cross-validation when detecting lung diseases (8 possible disease categories). The high accuracy was maintained in a real clinical setting (76%). The software-based automated diagnoses (76% accuracy) were superior to the suggested diagnoses of pulmonologists (53.5 +/- 6 % mean accuracy).	Machine learning approaches may improve the detection of early systemic sclerosis (SSc) patients at higher risk of interstitial lung disease (ILD).
Automated Screening Methodology for Asthma Diagnosis that Ensembles Clinical and Spirometric Information [60]	Das et al (2016)	Diagnostic accuracy / validation study	To evaluate an efficient machine learning methodology for the computerized detection of asthma that ensembles clinico-epidemiological and spirometric information to help pulmonologists in diagnostic decision-making.	Three different neural networks were created using different models and their accuracy of diagnosis compared based on clinical features and spirometry available.	Of the trained neural network models, the multilayer perceptron model has the highest accuracy (96.55 % sensitivity, 97.18 % specificity, and 96.86 % overall classification accuracy) for the automated classification of asthma-based statistically significant features.	This study provides insight into the machine learning models most suitable for respiratory diagnostic applications.
Detection of interstitial lung disease in systemic sclerosis using a machine learning approach based on pulmonary function tests [61]	Le-Dong et al (2017)	Diagnostic accuracy / validation study	To develop and evaluate a screening tool for detecting ILD in SSc patients that based on Machine-learning algorithms and combined data from pulmonary function tests.	Clinical and PFT data were obtained in 244 patients with confirmed diagnosis of SSc, 29 patients with ILD-unrelated to SSc and 50 healthy non-smokers. The main outcome is ILD, as determined by High-resolution CT. The attributes for classification are SSc type, NYHA dyspnoea-score, cigarette smoking status and 12 PFT parameters including spirometry, DLNO-DLCO and Plethysmography.	Data exploration and Cluster analysis showed a clear dissimilarity in NYHA dyspnoea score, FVC, TLC, FEV1, DLNO, DLCO and Vc between patients with and without ILD. These PFT indices are implicated in most of the ML based decision rules for the screening of ILD. These classifiers showed a good to excellent prediction ability, with the mean values of ROC-AUC, Accuracy, Sensitivity, Specificity, PPV and NPV of 0.86, 0.81, 0.69, 0.88, 0.76 and 0.85, respectively. The best performing decision rule was based on support vector machine algorithm and Z-score, with a balanced accuracy of 0.84, a sensitivity of 0.60 and a specificity of 0.96.	Machine learning algorithms demonstrated the ability to identify interstitial lung disease (ILD) in patients with systemic sclerosis.
Deep Learning Using Multi-Layer Perceptron	Mac et al (2022)	Diagnostic accuracy /	To assess if a deep learning model using multi-layer perceptron	2582 tests including spirometry and plethysmography data for 478 patients were collected	Of the 2582 tests, 780 were normal, 335 obstructive, 1130 restrictive and 337	Machine learning algorithms demonstrated high diagnostic

Improves the Diagnostic Acumen of Spirometry [23]		validation study	(MLP) can differentiate normal, obstructive, restrictive, and mixed obstructive-restrictive defects based on spirometry.	from June 29th, 2018, to March 20th, 2020, at the Toronto General Hospital Pulmonary Function Laboratory. Tests were classified as obstructive, restrictive, or mixed lung disease using ATS guidelines. MLP was built. MLP randomly selected one test per patient for each experiment, and the 478 selected tests were randomly partitioned into three sets: training (80%), validation (10%), test (10%). MLP classified lung conditions based on different combinations of input parameters: (1) spirometry (FVC, FEV1, FEF25-75), plethysmography (TLC, RV) and biometrics (sex, age, height); (2) spirometry and biometrics; (3) spirometry and plethysmography; (4) spirometry.	mixed lung disease. MLP produced comparable accuracies for classifying the major lung conditions based on biometrics with spirometry and plethysmography (94.79±2.56%), and biometrics with spirometry (89.81±2.06%).	accuracy when applied to both complete and partial datasets.
Detection of small airway obstruction patterns in forced Oscillometry (FO) and spirometry using artificial neural network [62]	Xu et al (2018)	Diagnostic accuracy / validation study	To construct and evaluate an effective MLP model for detecting and classifying small airway obstruction patterns in FO and spirometry results.	The preliminary experiment involves the first 222 tests of 82 subjects. Each test is labelled as airflow obstruction (AO) or not according to ATS guidelines, and randomly divided into training, validation, and test datasets at 70:15:15. MLP models with different numbers of inputs, hidden layers, and hidden units are evaluated regarding their performance in classifying AO (positive/negative).	The best result is obtained by the MLP model with 44 input features, including 4 biometrics (age, gender, height, weight), 36 FO parameters (R, X, Rin-ex, Xin-ex, each at 9 different frequencies), 4 spirometric parameters (FEV1, FVC, FEF50, FEF75), 1 hidden layers with 140 units and a logistic activation function. By training 200 epochs at 0.001 learning rate, the average validation accuracy is 83% for unseen patterns.	Machine learning approaches demonstrated effectiveness in the interpretation of forced oscillometry data.
AI-powered evaluation of lung function for diagnosis of interstitial lung disease [31]	Gompelmann et al (2025)	Diagnostic accuracy / validation study	This study aims to explore the potential of AI software in assisting pulmonologists with PFT interpretation for ILD diagnosis.	In study phase 1, a cohort of 60 patients, 30 of whom had ILD, were retrospectively diagnosed by 25 pulmonologists by evaluating a PFT (body plethysmography and diffusion capacity) and a short medical history. The experts screened the cohort twice, without and with the aid of AI software and provided a primary diagnosis and up to three differential diagnoses for each case. In study phase 2, 19 pulmonologists repeated the protocol after using ArtiQ.PFT for 4-6 months.	AI improved the detection of lung fibrosis as the primary diagnosis from 42.8% without AI to 72.1% with AI. Phase 2 yielded a similar outcome: using AI increased ILD diagnosis based on primary diagnosis from 53.2% to 75.1%. ILD detections without AI support significantly increased between phase 1 and phase 2 but not with AI.	AI was associated with improved early detection of interstitial lung disease (ILD).
Machine learning for accurate detection of small airway dysfunction-related respiratory changes: an observational study [35]	Xu et al (2024)	Diagnostic accuracy / validation study	To evaluate the performance of several machine learning algorithms in diagnosing small airway disease (SAD) in a population with preserved pulmonary function (PPF).	280 datasets were used from healthy controls and those with respiratory symptoms (PPF patients without SAD, n=158, PPF patients with SAD, n=44, healthy control group, n=78). Data split 70:30 training: testing. SAD was diagnosed using spirometry measurements of FEF25-75%, FEF50% and FEF75%. Both impulse Oscillometry (IOS) alone and with machine learning aid was compared.	IOS showed respiratory resistance at 5Hz (R5) was the best determinant of distinguishing between healthy controls compared to those with "preserved pulmonary function" both without and with SAD (AUC 0.642, AUC 0.769) with moderate success but when combined with machine learning algorithms this dramatically increased to AUC 0.915 and 0.971 respectively.	Impulse oscillometry (IOS) may overcome certain limitations of spirometry, such as the inability to perform forced vital capacity (FVC) manoeuvres. The integration of machine learning further enhances its diagnostic performance, enabling more accurate identification of respiratory abnormalities.
SPIRO-AID: a randomized controlled trial of AI-assisted spirometry interpretation in primary care [19]	Doe et al (2025)	Randomised Control Trial	To evaluate whether an AI decision support software (ArtiQ. Spiro) improves the diagnostic prediction of primary care clinicians.	A parallel, two-group, randomised controlled trial of primary care clinicians in the UK who refer for, or interpret, spirometry. Clinicians were randomised 1:1 to independently interpret fifty de-identified, real-world patient spirometry sessions through an on-line platform either with (AI+) or without (AI-) an AI decision support software report. The primary outcome was diagnostic prediction performance (proportion of the fifty spirometry sessions where clinicians' preferred diagnosis matched the reference diagnosis, derived by consensus from independent review of primary and secondary care notes and investigations by 3 pulmonologists).	234 participants were randomised with 133 (57%) completing (Intervention n=67, Control n=66): 73% female, 42% were general practitioners, 50% nurses. The addition of AI decision support software improved the mean 'preferred diagnosis' prediction performance from 49.7% (Control) to 58.7% (Intervention) in all cases.	AI-based software was associated with improved diagnostic interpretation of pulmonary function tests in primary care settings.
Pulmonologists collaborate with explainable artificial intelligence for superior interpretation of	Das et al (2021)	Quasi-experimental	To explore the diagnostic accuracy difference between clinicians interpreting PFTs both with and without explainable AI.	A previously reported AI model was used to aid physicians. 16 pulmonologists interpreted 24 PFT reports; on each they provided a preferential and up to 4 differential diagnoses, first without (control) and then with	Pulmonologists updated their decision roughly 50% of the time with the use of AI. Pulmonologists were unlikely to follow XAI's incorrect suggestion indicating low automation bias. Preferential and	AI demonstrated a synergistic effect when used alongside pulmonologists, improving diagnostic accuracy.

pulmonary function tests [33]				XAI's feedback (intervention). Primary endpoint compared accuracy of preferential and differential diagnosis between control and intervention.	differential diagnostic accuracy increased with the use of explainable AI.	
Late Breaking Abstract-AI-powered decision support helps pulmonologists to diagnose ILD more accurately [32]	Desbordes et al (2023)	Quasi-experimental	This study aims to explore the potential of AI software in assisting pulmonologists with ILD diagnosis.	60 patients, 30 of whom had ILD, were retrospectively diagnosed by 25 pulmonologists by evaluating a full Pulmonary Function Test (body plethysmography, diffusion) and a short clinical background. The experts screened the cohort twice, first without the aid of AI, and then later with the assistance of ArtiQ.PFT. The pulmonologists provided a primary diagnosis and up to three differential diagnoses for each case. In Phase 2 of the study, half of the experts repeated the protocol after using ArtiQ.PFT for six months.	In Phase 1, AI improved the diagnosis of ILD from 42.8% without AI to 72.1% with AI (+68.5%). Phase 2displayed similar results. Using AI increased ILD diagnosis (52.5% to 77.9%). It did not affect the number of differentials the clinician gave.	AI demonstrated a synergistic effect when used alongside expert clinicians, improving diagnostic accuracy.
Artificial Intelligence Powered Spirometry Enables Early Detection of Interstitial Lung Disease [42]	Ray et al (2022)	retrospective cohort study	To evaluate if AI can detect ILD from spirometry earlier than clinicians.	500,000 volunteers in the UK Biobank dataset were used and data from participants had to meet the following criteria: 1. had ILD as a cause of death, 2. had performed an acceptable spirometry measurement up to 7 years prior to their date of death, 3. had not received an ILD diagnosis on the date of their spirometry measurement. A total of 109 subjects were selected. Spirometry data and subject demographic information were used as inputs to an AI clinical decision support software, which produced a probability for each subject of having normal lung function, asthma, COPD, ILD, another obstructive disease, or another unidentifiable respiratory disease.	In 27% of subjects (N=29), the AI software detected ILD may already be present (i.e. ILD was highest probability), with 66% (N=19) of these subjects exhibiting normal lung function using standard interpretation guidelines. AI software detected possible ILD up to 6.8 years prior to diagnosis by a clinician through standard care in 27% of patients who died because of ILD in the UK Biobank.	AI demonstrated superior ability to detect interstitial lung disease (ILD) earlier in the disease trajectory based on spirometry, compared with clinicians.
Insights into spirometry usage and the potential for AI-based decision support software in Belgian primary care [20]	De Vos et al (2023)	mixed methods observational study	To understand the delivery of and barriers to performing spirometry in primary care and investigate the added value of AI-based software for the quality assessment and interpretation of spirometry in helping primary care clinicians diagnose COPD.	Structured interviews were conducted in 18 Belgian GP practices at baseline and at the end of the study. GPs were asked to include patients with the following criteria to have PFTs: patients older than 35 years old, with a smoking history ≥ 10 pack-years and at least one respiratory complaint. ArtiQ.Spiro was used to assess the quality of spirometry and give a diagnosis. After every use, the added value of the software was rated on a Likert scale.	28.5% of patients at risk of COPD were flagged as having the disease. The primary barriers identified were time pressure and lack of good training in spirometry. ArtiQ.Spiro was effective for both quality assessment (4.13/5) and diagnostic aid (4.01/5).	AI provided a valuable adjunct to general practitioners in the interpretation of spirometry, particularly in settings where time constraints and limited confidence may impact diagnostic accuracy.
Big data insights in the ILD population through AI-driven analytics [63]	Elmahy et al (2024)	observational study	To utilize AI in interpreting pulmonary function tests in a large population to leverage interstitial lung disease related data insights.	ArtiQ.PFT was applied to anonymized PFT data from 87205 patients from a Belgian tertiary centre / reference centre and 7 referring centres. Compared the percentages of ILD-predicted patients, patients with suggestive restrictive spirometry, patients with restrictive volume, and patients with normal spirometry but reduced DLCO.	AI predicted 12366 patients (14%) to have ILD overall with an overall higher patient percentage at the reference centre (26%) compared to the referring centres (8%,SD 3%). The difference was driven by a high percentage of patients with restrictive volume (22% vs. 12%,SD 2%) and a large percentage of patients with normal spirometry but reduced DLCO (27% vs. 14%, SD 2%). Surprisingly, the percentage of patients at the reference centre exhibiting suggestive restrictive spirometry was similar to the referring centres (15% and 14% respectively).	AI demonstrated potential utility in epidemiological applications, enabling the analysis and interpretation of large-scale datasets.

Data appraisal

Studies meeting the inclusion criteria were critically appraised using the JBI critical appraisal tools appropriate to each study designs [9].

One researcher assessed the quality and scored all the publications, and a second researcher independently reviewed 25% of the studies included (100% TO, 25% IS).

Any discrepancies were brought to review and discussed to reach a consensus.

Data synthesis

The Popay et al. guideline for narrative synthesis was used to analyse and integrate the data, enabling the development of coherent

10.12865/CHSJ.52.02.01

conclusions from the included studies in the large dataset [15].

Results

Study inclusion

The database search yielded 4279 publications. After de-duplication, 2906 publications were reviewed for title screening, with 2727 publications excluded thereafter.

Abstracts of the remaining articles were screened and a further 118 were excluded. Sixty-one full-text articles were assessed for eligibility, of which forty-seven studies met the inclusion criteria and were included in the final analysis (Figure 1).

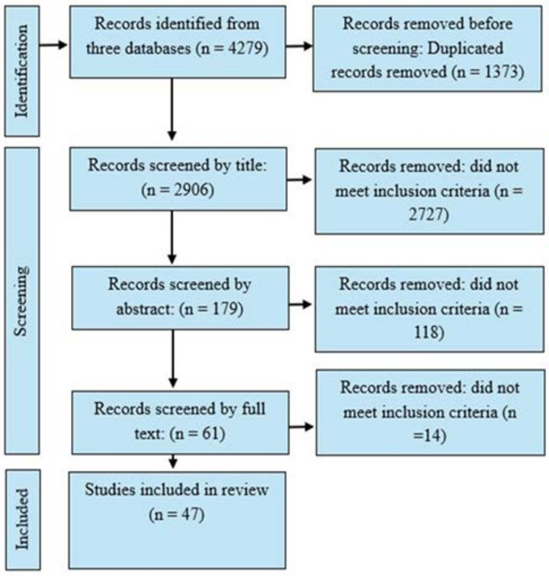

Figure 1. Schematic illustration of the study selection.

Characteristics and quality of included studies

This review encompassed forty-one diagnostic accuracy studies (87%), one randomised controlled trial (2%), two quasi-experimental studies (4%), two observational studies (4%) and one cohort study (2%). The included studies in the review exhibited an array of methodological quality following analysis with the JBI checklist (Tables 3-7); some being considered robust and others lacking rigor.

The majority of studies focused on COPD, while fewer investigated interstitial lung disease, asthma, upper airway obstruction and obstructive sleep apnoea. Commonly reported metrics include diagnostic accuracy, sensitivity, specificity, AUC and Cohen's kappa. The included studies were published between 2016 and 2025, with a median publication year of 2022 (IQR 2020-2024) and a modal year of 2024.

Although the outcome metrics reported differed significantly between studies, diagnostic performance parameters were analysed in all 47 included studies. The most commonly used metric was overall diagnosis accuracy; this was provided in 28 publications (59.6%). Thirty (63.8%) studies reported on sensitivity of AI diagnosis, whilst twenty-nine (61.7%) calculated specificity-this enables assessment of both illness detection and exclusion. Measures of inter-rater agreement, such as Cohen's or Fleiss' κ, were reported in 14 publications (29.8%).

Additionally, the area under the receiver operating characteristic curve (AUC) was reported in 18 studies (38.3%). Less frequently reported were positive and negative predictive values, which were generally used in conjunction with sensitivity and specificity, rather than as independent metrics.

The variety of outcome metrics analysed in the included studies highlighted the methodological heterogeneity of the data. There was no standard method for providing performance measures, despite the majority of studies aiming to assess AI diagnostic performance. Some studies provided disease-specific performance results or comparison with pre-defined reference standards without analysing overall accuracy. In contrast, others reported a single overall diagnostic accuracy.

As a result, even though all of the studies evaluated the same clinical question, there was limited direct quantitative comparison between them.

Table 3. JBI critical appraisal score for diagnostic accuracy studies.

Title, author and date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Application of Artificial Intelligence in the Interpretation of Pulmonary Function Tests-Saad et al. (2025)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
AI-Assisted Interpretation of Pulmonary Function Tests: Enhancing Diagnostic Precision in Respiratory Medicine-Singh et al. (2025)	N	Y	N	Y	N/A	Y	Y	Y	Y	Y
Validation of artificial intelligence spirometric diagnostic support software in primary care: a blinded diagnostic accuracy study-Sunjaya et al (2025)	Y	Y	Y	Y	N/A	Y	Y	Y	Y	N
Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests-Topalovic et al (2019)	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Integrating AI-assisted spirometry: Validation and user experience in a COPD diagnostic pathway-Giffen et al. (2025)	Y	Y	N	N	N/A	Y	N	Y	Y	N

Machine learning based detection of true ventilatory restriction-Saha et al (2025)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Artificial Intelligence- Based Deep learning Model for Detecting Central and Upper Airway Obstruction-Prevot et al (2025)	Y	Y	N	N	N/A	Y	N	Y	Y	N
Machine Learning for Enhanced COPD Diagnosis: A Comparative Analysis of Classification Algorithms-Elashmawi et al (2024)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Diagnosis and Severity if COPD Using a Novel Fast-Response Capnometer and Interpretable Machine Learning-Talker et al (2024)	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
Artificial Intelligence Applied to Forced Spirometry in Primary Care-Moreno Mendez et al (2024)	N	Y	N	N	Y	Y	N	Y	Y	Y
Time-series Machine Learning Model Classifies Lung Function Using Spirometry Flow-Volume Loops Alone-Mac et al (2024)	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Clinical Evaluation of Artificial Intelligence Supported Software in the Interpretation of Spirometry Results in a Primary Care Setting-Maes et al (2024)	Y	Y	Y	Y	N/A	Y	Y	Y	Y	Y
Evaluating artifical intelligence-interpreted against pulmonologist-interpreted spirometry results: a comparative study-Antonio, Montejo and Sy (2024)	N	Y	Y	N	N/A	Y	N	Y	Y	Y
Collaboration between explainable artificial intelligence and pulmonologists improves the accuracy of pulmonary function test interpretation-Das et al (2023)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
AI models for respiratory disease from spirometry alone: validation based on UK Biobank-Elmahy et al (2023)	N	N	N	N	Y	Y	Y	Y	Y	Y
ArtiQ.PFT: AI-powered decision support for the diagnosis of lung diseases: Does it help the pulmonologist to be more accurate?-Autoren Desbordes et al (2023)	N	Y	N	Y	N/A	N	Y	Y	Y	Y
Exploratory study of a risk prediction artificial intelligence model to diagnose asthma and COPD in Mexico-Hernandez et al (2023)	N	Y	N	N	N/A	N	Y	Y	Y	Y
Development and evaluation of a computerized algorithm for the interpretation of pulmonary function tests-Huang et al (2024)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
Deep learning using multilayer perception improves the diagnostic acumen of spirometry: a single-centre Canadian study-Mac et al (2022)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Automated interpretation of the pulmonary function test by a portable spirometer in Chinese adults-Zhou et al (2022)	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Deep Learning for Automatic Upper Airway Obstruction Detection by Analysis of Flow-Volume Curve-Wang et al (2022)	N	Y	Y	Y	N/A	Y	Y	Y	Y	Y
Explainable machine learning methods and respiratory oscillometry for the diagnosis of respiratory abnormalities in sarcoidosis-de Lima er al (2022)	N	N	N	N	N	Y	Y	Y	N	Y
Machine Learning for the identification of Restriction Without Lung Volume Measurements-Moffeett et al (2022)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Deep Learning-Based Analytic Models Based on Flow-Volume Curves for Identifying Ventilatory Patterns-Wang et al (2022)	Y	Y	Y	N	Y	Y	N	Y	Y	Y
Machine learning associated with respiratory oscillometry: a computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis-Andrade et al (2021)	N	N	Y	N	Y	Y	Y	Y	Y	Y
Machine learning for nocturnal diagnosis of chronic obstructive disease using digital oximetry biomarkers-Levy et al (2021)	N	N	Y	N	Y	Y	Y	Y	N	Y
Deep learning algorithm doe the classification of spirometries using flow-volume curves: proof of concept study-Exarchos et al (2021)	N	Y	N	N	N	N	N	N	Y	N
Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease-Bodduluri et al (2020)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Application of machine learning to predict obstructive sleep apnea syndrome severity-Mencar et al (2020)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
Deep learning facilitates the diagnosis of adult asthma-Tomita et al (2019)	N	Y	N	N	N/A	Y	Y	Y	Y	Y
Artificial intelligence helps detecting lung disease with pulmonary function tests-Topalovic et al (2018)	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints-Topalovic et al (2017)	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Artificial intelligence detects lung diseases using pulmonary function tests-Topalovic et al (2017)	N	Y	Y	Y	Y	N	Y	N	N	Y
Late Breaking Abstract-Applying artificial intelligence on pulmonary function tests improves the diagnostic accuracy-Topalovic et al (2017)	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Applying machine learning and pulmonary function data to detect interstitial lung disease in systemic sclerosis-Le-Dong et al (2017)	N	N	N	N	N	Y	Y	Y	Y	Y
Automated Screening Methodology for Asthma Diagnosis that Ensembles Clinical and Spirometric Information-Das et al (2016)	N	Y	N	Y	Y	Y	Y	Y	Y	Y
Detection of interstitial lung disease in systemic sclerosis using a machine learning approach based on pulmonary function tests-Le-Dong et al (2017)	N	N	Y	N	N	Y	Y	Y	Y	Y
Deep Learning Using Multi-Layer Perceptron Improves the Diagnostic Acumen of Spirometry-Mac et al (2022)	Y	Y	N	Y	N	Y	Y	Y	Y	Y
Detection of small airway obstruction patterns in forced oscillometry (FO) and spirometry using artificial neural networks-Xu et al (2018)	Y	Y	N	N	N	Y	Y	Y	Y	Y
AI-powered evaluation of lung function for diagnosis of interstitial lung disease-Gompelmann et al (2025)	N	N	Y	N	N/A	Y	Y	N	Y	Y
Machine learning for accurate detection of small airway dysfunction-related respiratory changes: an observational study-Xu et al (2024)	N	Y	Y	N	Y	N	Y	Y	Y	Y

Note. N = No, Y = Yes, N/A = not applicable
Q1-Was a consecutive or random sample of patients enrolled?
Q2-Was a case control design avoided?
Q3-Did the study avoid inappropriate exclusions?
Q4-Were the index test results interpreted without knowledge of the results of the reference standard?
Q5-If a threshold was used, was it pre-specified?
Q6-Is the reference standard likely to correctly classify the target condition?
Q7-Were the reference standard results interpreted without knowledge of the results of the index test?
Q8-Was there an appropriate interval between index test and reference standard?
Q9-Did all patients receive the same reference standard?
Q10-Were all patients included in the analysis?

Table 4. JBI critical appraisal score for randomised control studies.

Title, author and date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
SPIRO-AID: a randomized controlled trial of AI-assisted spirometry interpretation in primary care-Doe et al (2025)	Y	N	Y	N	N	Y	Y	Y	Y	N	N	Y	Y

Note. N = No, Y = Yes, N/A = not applicable
Q1-Was true randomization used for assignment of participants to treatment groups?
Q2-Was allocation to treatment groups concealed?
Q3-Were treatment groups similar at the baseline?
Q4-Were participants blind to treatment assignment?
Q5-Were those delivering the treatment blind to treatment assignment?
Q6-Were treatment groups treated identically other than the intervention of interest?
Q7-Were outcome assessors blind to treatment assignment?
Q8-Were outcomes measured in the same way for treatment groups?
Q9-Were outcomes measured in a reliable way?
Q10-Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed?
Q11 Were participants analysed in the groups to which they were randomised?
Q12 Was appropriate statistical analysis used?
Q13 Was the trial design appropriate and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

10.12865/CHSJ.52.02.01

Table 5. JBI critical appraisal score for Quasi-experimental studies.

Title, author and date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Pulmonologists collaborate with explainable artificial intelligence for superior interpretation of pulmonary function tests-Das et al (2021)	Y	N	Y	N	N	Y	N	Y	Y
Late Breaking Abstract-AI-powered decision support helps pulmonologists to diagnose ILD more accurately-Desbordes et al (2023)	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note. N = No, Y = Yes, N/A = not applicable
Q1-Is it clear in the study what is the "cause" and what is the "effect" (i.e. there is no confusion about which variable comes first)?
Q2-Was there a control group?
Q3-Were participants included in any comparisons similar?
Q4-Were the participants included in any comparisons receiving similar treatment/care, other than the exposure or intervention of interest?
Q5-Were there multiple measurements of the outcome, both pre and post the intervention/exposure?
Q6-Were the outcomes of participants included in any comparisons measured in the same way?
Q7-Were outcomes measured in a reliable way?
Q8-Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analysed?
Q9-Was appropriate statistical analysis used?

Table 6. JBI critical appraisal score for studies.

Title, author and date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
Artificial Intelligence Powered Spirometry Enables Early Detection of Interstitial Lung Disease-Ray et al (2022)	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y

Note. N = No, Y = Yes, N/A = not applicable
Q1-Were the two groups similar and recruited from the same population?
Q2-Were the exposures measured similarly to assign people to both exposed and unexposed groups?
Q3-Was the exposure measured in a valid and reliable way?
Q4-Were confounding factors identified?
Q5-Were strategies to deal with confounding factors stated?
Q6-Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?
Q7-Were the outcomes measured in a valid and reliable way?
Q8-Was the follow up time reported and sufficient to be long enough for outcomes to occur?
Q9-Was appropriate statistical analysis used?
Q10-Were strategies to address incomplete follow up utilized?
Q11-Was appropriate statistical analysis used?

Table 7. JBI critical appraisal score for observational studies.

Title, author and date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Insights into spirometry usage and the potential for AI-based decision support software in Belgian primary care-De Vos et al (2023)	Y	Y	Y	Y	N	N	Y	Y
Big data insights in the ILD population through AI-driven analytics-Elmahy et al (2024)	N	N	Y	Y	N	N	Y	Y

Note. N = No, Y = Yes, N/A = not applicable
Q1-Were the criteria for inclusion in the sample clearly defined?
Q2-Were objective, standard criteria used for measurement of the condition?
Q3-Was the exposure measured in a valid and reliable way?
Q4-Were the outcomes measured in a valid and reliable way?
Q5-Were confounding factors identified?
Q6-Were strategies to deal with confounding factors stated?
Q7-Was appropriate statistical analysis used?
Q8-Were the study subjects and the setting described in detail?

A common limitation was the lack of blinding to the reference standard when deriving the index test results. Whilst this can prove challenging in the context of training AI models with representative datasets, some studies addressed this by having specific subsets of clinical data exclusively assigned for testing.

Quality appraisal

The majority of included studies (87%) were diagnostic accuracy/validation studies. The mean JBI critical appraisal score across these studies was 7.08 (IQR 6-9). Question 5 of the JBI diagnostic accuracy checklist was often not suitable for the included publications, as the reference standard used in the studies was based on expert consensus rather than an objective diagnostic threshold. The range of quality appraisal scores demonstrates methodological heterogeneity across the data.

Review findings

The four overarching themes developed from the included publications are:

1) The diagnostic accuracy of AI in non-specialist and specialist settings

2) Application with suboptimal data or alternative investigations

3) Improved consistency of diagnostic interpretations with AI

4) Clinical applicability and accuracy of AI in specific respiratory diseases.

Diagnostic accuracy in non-specialist and specialist settings

A consistent finding demonstrated throughout the publications included in this review was the superior diagnostic accuracy of AI compared to non-specialist interpretations.

In most of the studies, the non-specialists were primary care practitioners [12,16-20].

Studies found that AI had the best diagnostic accuracy when presented with COPD patient data but consistently outperformed inexperienced or unconfident clinicians [12].

The use of AI in clinical practice demonstrated a higher degree of concordance with expert interpretation than that achieved by primary care practitioners [17].

Moreover, AI was shown to have a positive synergistic effect when combined with general practitioner (GP) interpretation of PFTs [19,20].

A decision support tool with AI software was shown to improve the diagnostic accuracy of spirometry interpretation compared to clinicians who did not have access to such software [19].

AI regularly performed well across the studies in this review compared to the pulmonologists' interpretation and a very high degree of concordance was demonstrated between AI and expert clinicians in their interpretation of PFTs [10,21,22].

AI demonstrated superior diagnostic accuracy compared to senior pulmonologists when significant clinical data and investigations were unavailable [23-27].

Several studies alternatively investigated the synergistic effect of AI in diagnostic accuracy and consistency among specialists. It was demonstrated that incorporating AI or explainable AI into the tool set available to specialists reviewing PFTs improved their diagnostic confidence and accuracy as well as altering their initial management plans [28,29] PT-Conference Abstract, [30-32].

There was a lower chance of spurious errors occurring and the addition of AI influenced and altered their decisions up to 50% of the time [30,33].

Application with sub-optimal or alternative data

Multiple studies included in this review demonstrate that AI can maintain high diagnostic accuracy when comprehensive investigations are unavailable and suboptimal data is presented. AI often outperforms clinician diagnostic accuracy of spirometry in these settings [10,13,16,23-25,27,34].

One study included in this review demonstrated the ability to apply machine learning to alternative investigations such as impedance Oscillometry (IOS) in patients with preserved pulmonary function (FEV1/FVC≥0.70) with spirometry otherwise indicating small airway disease with excellent diagnostic concurrence (AUC>0.9) [35].

Capnography data was also analysed by machine learning algorithms to diagnose COPD with similarly successful results [36].

Improved consistency of diagnostic interpretation with AI

Interpretation of PFTs has long been limited by inter-rater variability [3].

Significant subjectivity was reaffirmed in the present review as multiple studies presented significant discrepancies amongst clinicians in PFT interpretation [10,13,24,26].

Reported kappa values ranged from 0.12-0.46 indicating only slight to moderate agreement in PFT interpretation amongst clinicians according to the classification proposed by Landis and Koch [37].

Variability between different AI-based models was substantially lower than that observed among clinicians [36].

Clinical applicability to specific diseases

Numerous studies focused on heterogenous PFTs amongst a population with varying pathology and clinical presentations. Some studies, however, focused on one specific respiratory disease. COPD was most commonly studied for the diagnostic accuracy of PFT interpretation by AI. The evidence consistently demonstrated high diagnostic accuracy of AI in COPD from PFT interpretation in both traditional and non-traditional means, such as capnography [12,14,34,36,38,39].

However, one study suggested that there was a high degree of false positives when using AI to diagnose COPD and asthma [40].

AI also demonstrated a high degree of diagnostic accuracy in the detection of interstitial lung disease (ILD), including the identification of subtle changes. In some cases, the AI detected such changes much earlier than experienced clinicians [31,32,41,42].

This includes patients with systemic pathology affecting the lungs such as systemic sclerosis and sarcoidosis which can present or be associated with respiratory symptoms and pathology. The AI was able to thoroughly and accurately detect those systemic sclerosis patients at higher risk of ILD and those with subtle lung changes [41].

Other conditions investigated include upper airway obstructions and obstructive sleep apnoea (OSA). Overall, the evidence for these conditions was mixed with limited research. Whilst AI interpretation of PFTs to identify upper airway obstruction was seen to be very successful and accurate, using AI to diagnose OSA had a very poor accuracy and did not seem to provide clinical benefit in this specific condition. The evidence on this is limited in quantity [18,43].

 10.12865/CHSJ.52.02.01

Discussion

The findings in this review emphasise the potential role of AI as a diagnostic support tool in the interpretation of PFTs, particularly in non-specialist and resource limited settings.

AI exhibited high concordance with expert interpretation throughout the evidence assessed and consistently outperformed non-specialists, which may reflect the known limitations in primary care clinician's confidence to interpret PFTs [12,16-20].

Thus, suggests that AI could be a useful tool in bridging the gap between primary and secondary care by improving diagnostic accuracy in community setting [20].

When patients only have preliminary studies complete, it appears AI may serve as the most accurate tool to diagnose patients or screen those who need further investigation to confirm a diagnosis through the standard pathways [10,13,16,23-25,27,34].

Using AI in this manner may thereby reduce the referral burden to secondary care and again could serve as an intermediary triage tool, although this remains to be validated in prospective studies.

However, a key limitation of AI seen in the literature relevant to the above propositions is the reduced ability of AI software to differentiate between mild impairment, moderate impairment and normal results when compared to its performance in identifying severe or clearly normal results [34].

Additionally, some studies reported challenges in accurately differentiating between specific conditions despite being successful in spotting abnormal or normal cases [44,45].

This reinforces the idea AI should function as a tool for clinicians, perhaps in screening or triaging as opposed to an independent decision-making model.

Studies not only compared AI to clinician interpretation, rather also investigated the pulmonologists' interpretation with and without the use of AI as an adjunctive tool. AI showed a consistent synergistic effect which improved clinicians' diagnostic accuracy and confidence [28-32] thus illustrating the potentiated effect combining the clinical experience of a pulmonologist and processing ability of AI can have. Accordingly, the perspective of AI going forwards needs to be more centered around it being used as a clinical adjunct, not clinician replacement. Radiology is one such specialty that has seen a synergistic effect of incorporating AI into clinical practice to improve efficiency, suggesting broader applicability across other diagnostic disciplines such as PFTs [46].

Again, this requires prospective validation.

In resource constrained environments, where access to gold standard investigations -such as body plethysmography and advanced imaging- for the diagnosis of chronic respiratory conditions is frequently limited (such as low- and middle-income countries), AI may offer a pragmatic solution. The limitation of resource scarcity has been widely recognized as a significant barrier to delivering high-quality healthcare in such environments. As a result, rates of misdiagnosis and underdiagnosis within resource limited environments of chronic respiratory diseases remain unacceptably high [47].

On account of the success seen with AI-based diagnostic accuracy with suboptimal clinical data, the implementation of AI in these settings may be a model that could aid in overcoming the limits of resource scarcity and insufficient healthcare access. Whilst this delivery style of healthcare would still not meet the criteria for "gold-standard" treatment, it may improve the current unacceptable levels of underdiagnosis [47].

Resource limitations are not the only reason for exploring the improvement of alternative or traditionally suboptimal investigations for diagnostic purposes in chronic respiratory conditions. Spirometry remains the most widely used tool to assess small-airway function; however, this requires good cooperation from patients and substantial effort. Certain patient groups, such as the elderly, can struggle with the forceful exercises required for good quality data thus alternatives may be more appropriate.

Unlike spirometry, IOS relies on quiet tidal breathing and is therefore more feasible for certain patients to carry out and could thus possibly lead to more efficacious results with machine learning incorporation [35].

These findings highlight the need to assess the potential utility of previously underutilized investigations as they may gain clinical relevance with the advancement of AI.

The implementation of AI-based interpretive software, particularly in settings where they demonstrated superior performance compared to clinicians, may facilitate a more standardized approach to care given the reduced inter-rater variability displayed [10,13,24,26].

Such systems have the potential to deliver consistent, scalable interpretation across large patient populations, thereby supporting equitable care without requiring proportional expansion of the clinical workforce.

The results were particularly robust for COPD specifically, the most frequently investigated condition, likely reflecting the central role spirometry plays in COPD diagnosis compared to the more supportive role it plays in other respiratory conditions [48,49].

In contrast, AI seemed to demonstrate a more limited diagnostic utility in obstructive sleep apnoea diagnosis where performance was inconsistent and generally lower [24,43].

Limitations

One of the main limitations of this study was the heterogeneity of the studies included. The aim of the study was to comprehensively appraise the current literature that met the study's inclusion criteria.

However, the resulting study designs were diverse with varying methodologies which limited quantitative analysis of the available literature.

The studies also exhibited a heavy reliance on non-standardised, clinician-led reference standards and diagnostic criteria thus leading to a degree of subjectivity and bias.

Moreover, the studies included in this review often lacked clear blinding to the reference standard when the index test results were interpreted which, in turn, reduced internal validity and may have overestimated diagnostic performance.

As mentioned, the reference standard was frequently based on a pulmonologist's interpretation or a group of pulmonologists making a consensus decision with a full clinical dataset of necessary investigations. Consequently, the reported diagnostic accuracy is inherently constrained by this clinician-defined standard.

There is also limited confirmed generalization of the data as the majority of studies were conducted in high income countries.

Furthermore, most studies focused on the binary presence or absence of disease and therefore diagnostic thresholds were not used and comparison was made with expert opinion as the reference standard. This can however create a "black box" effect in which there is a lack of transparency in decision making amongst AI tools. One such way to mitigate this is to use explainable AI as seen in the study by Das et al (2023) as this displays the factors which influence the AI's decision-making process [28].

However, one of the largest barriers to implementing AI in healthcare is perhaps the introduction of a "responsibility gap".

AI systems do not satisfy concepts such as agency, intent and accountability which help form legal responsibility and thus cannot bear responsibility for clinical decisions [50].

Responsibility must therefore lie with the clinician and as a result, AI should be used as an adjunctive tool, not an independent clinical decision maker.

Directions for future research

Future research should focus on the implementation of AI systems in low- and middle-income countries to assess its benefit in diagnostic accuracy with limited clinical data in such settings.

Furthermore, studies should incorporate clinical outcomes as opposed to diagnostic accuracy in isolation to assess if there is a tangible improvement in health outcomes.

Conclusions

Overall, AI demonstrates high diagnostic accuracy in the interpretation of PFTs when compared to both non-specialist and specialist physicians.

This advantage becomes particularly apparent when significant clinical data is unavailable.

AI systems also work synergistically with clinicians, improving diagnostic confidence and accuracy, thus should be used as an adjunctive clinical tool rather than independent decision-maker.

AI appears to have a particularly important role in primary care settings where confidence in PFT interpretation is generally lower, as well as in resource constrained environments where access to specialist investigations may be more limited.

Acknowledgement

None to declare.

Author Contributions

Conceptualisation: I.S. and T.O.; Methodology: I.S.; Literature search: I.S. and T.O.; Screening and study selection: I.S. and T.O.; Data extraction: T.O.; Quality appraisal: I.S. and T.O.; Data synthesis and manuscript writing: T.O.; Manuscript review and editing: I.S. and T.O.; All authors read and reviewed the final manuscript.

10.12865/CHSJ.52.02.01

Funding

There was no financial funding for this systematic review.

Conflicts of interest

The authors of this publication have no conflicts of interest to disclose.

Institutional Review Board

The study was conducted according to the guidelines of the Declaration of Helsinki.

Consent Statement

Not applicable.

Data availability

All data presented in the manuscript are available from the authors upon request.

References

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018, 392(10159):1789-1858.
2. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, Cooper BG, Culver B, Derom E, Hall GL, Hallstrand TS, Leuppi JD, MacIntyre N, McCormack M, McCormack M, Rosenfeld M, Swenson ER. ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J, 2022, 60(1):2101499.
3. Quadrelli SA, Roncoroni AJ, Porcel G. Analysis of Variability in Interpretation of Spirometric Tests. Respiration, 1996, 63(3):131-136.
4. Kim SH, Han MLK. Challenges and the Future of Pulmonary Function Testing in Chronic Obstructive Pulmonary Disease (COPD): Toward Earlier Diagnosis of COPD. Tuberc Respir Dis (Seoul), 2025, 88(3):413.
5. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ, 2020, 98(4):251-256.
6. Al-Anazi S, Al-Omari A, Alanazi S, Marar A, Asad M, Alawaji F, Alwateid S. Artificial intelligence in respiratory care: Current scenario and future perspective. Ann Thorac Med, 2024, 19(2):117-130.
7. López-Canay J, Casal-Guisande M, Represas-Represas C, Cerqueiro-Pequeño J, Bouza-Rodríguez JB, Comesaña-Campos A, Fernández-Villar A. Artificial Intelligence for Spirometry Quality Evaluation: A Systematic Review. Bioengineering, 2025, 12(12):1286.
8. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 2021, 372:n71.
9. Munn Z, Aromataris E, Tufanaru C, Stern C, Porritt K, Farrow J, Lockwood C, Stephenson M, Moola S, Lizarondo L, McArthur A, Peters M, Pearson A, Jordan Z. The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information, JBI SUMARI. JBI Evidence Implementation, 2019, 17(1):36-43.
10. Saad T, Pandey R, Padarya S, Singh P, Singh S, Mishra N. Application of Artificial Intelligence in the Interpretation of Pulmonary Function Tests. Cureus, 2025, 17(4):1-15.
11. Singh A, Sahu M, Padarya SK, Saad T, Mishra S, Singh MP. AI-Assisted Interpretation of Pulmonary Function Tests: Enhancing Diagnostic Precision in Respiratory Medicine. J Pharm Bioallied Sci, 2025, 17(Suppl 4):S3072.
12. Sunjaya A, Edwards GD, Harvey J, Sylvester K, Purvis J, Rutter M, Shakespeare J, Moore V, El-Emir E, Doe G, Van Orshoven K, Patel S, de Vos M, Elmahy A, Cuyvers B, Desbordes P, Sehdev S, Evans RA, Morgan MD, Russell R, Jarrold I, Spain N, Taylor S, Scott DA, Prevost AT, Hopkinson NS, Kon S, Topalovic M, Man WDC. Validation of artificial intelligence spirometry diagnostic support software in primary care: a blinded diagnostic accuracy study. ERJ Open Res, 2025, 11(5):00116-2025.
13. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, Janssen R, Kerstjens HAM, Liistro G, Louis R, Ninane V, Pison C, Schlesser M, Vercauter P, Vogelmeier CF, Wouters E, Wynants J, Janssens W. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. Eur Respir J, 2019, 53(4):1801660.
14. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev, 2016, 5(1):210.
15. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K, Duffy S. Developing guidance on the conduct of narrative synthesis in systematic reviews. A Prod from ESRC Methods Program Lancaster Inst Heal Res, 2005, 59(Suppl 1):A7.
16. Moreno Mendez R, Marín A, Ferrando JR, Rissi Castro G, Cepeda Madrigal S, Agostini G, Catalan Serra P. Artificial Intelligence Applied to Forced Spirometry in Primary Care. Open Respir Arch, 2024, 6:100313.
17. Maes J, Willaerts S, De Vos M, Topalovic M, Verbakel JY. Clinical Evaluation of Artificial Intelligence Supported Software in the Interpretation of Spirometry Results in a Primary Care Setting. Am J Respir Crit Care Med, 2024, 209(Supplement 1):A1461-A1461.
18. Wang Y, Li Q, Chen W, Jian W, Liang J, Gao Y, Zhong N, Zheng J. Deep Learning-Based Analytic Models Based on Flow-Volume Curves for Identifying Ventilatory Patterns. Front Physiol, 2022, 13:824000.

19. Doe G, Banya W, Edwards G, Topalovic M, El-Emir E, Evans R, Russell R, Smets E, Orshoven K Van, Sylvester K, Sunjaya A, Scott D, Vos M De, Elmahy A, Taylor S, Hutchinson A, Hopkinson N, Morgan M, Kon S, Patel S, Jarrold I, Kennington E, Sehdev S, Spain N, Prevost A, Man WC. S11 SPIRO-AID: a randomized controlled trial of AI-assisted spirometry interpretation in primary care. Thorax, 2025, 80(Suppl 2):A14-A14.
20. Vos M De, Maes J, Orshoven K Van, Desbordes P, Vanheule V, Quataert C, Fovel I, Topalovic M, Topalovic M. Insights into spirometry usage and the potential for AI-based decision support software in Belgian primary care. Eur Respir J, 2023, 62(suppl 67):OA764.
21. Antonio GL, Montejo Y, Sy R. Evaluating artificial intelligence-interpreted against pulmonologist-interpreted spirometry results: a comparative study. Eur Respir J, 2024, 64(suppl 68):PA4336.
22. Topalovic M, Aerts JM, Decramer M, Troosters T, Janssens W. Artificial Intelligence Detects Lung Diseases Using Pulmonary Function Tests. Am J Respir Crit Care Med, 2017, 195 (Supplement 1):A5678-A5678.
23. Mac A, Xu T, Wu JKY, Belousova N, Kitazawa H, Vozoris N, Rozenberg D, Ryan CM, Valaee S, Chow CW. Deep learning using multilayer perception improves the diagnostic acumen of spirometry: a single-centre Canadian study. BMJ Open Respir Res, 2022, 9(1):e001396.
24. Wang Y, Li Y, Chen W, Zhang C, Liang L, Huang R, Jian W, Liang J, Zhu S, Tu D, Gao Y, Zhong N, Zheng J. Deep Learning for Automatic Upper Airway Obstruction Detection by Analysis of Flow-Volume Curve. Respiration, 2022, 101(9):841-850.
25. Moffett AT, Kreider M, Halpern SD, Weissman GE. Machine Learning for the Identification of Restriction Without Lung Volume Measurements. Am J Respir Crit Care Med, 2022, 205(Supplement 1):A5021-A5021.
26. Topalovic M, Das N, Troosters T, Decramer M JW. Artificial intelligence helps detecting lung disease with pulmonary function tests. Respirology, 2018, 23(Suppl 1):21.
27. Topalovic M, Das N, Troosters T, Decramer M, Janssens W. Late Breaking Abstract-Applying artificial intelligence on pulmonary function tests improves the diagnostic accuracy. Eur Respir J, 2017, 50(suppl 61):OA3434.
28. Das N, Happaerts S, Gyselinck I, Staes M, Derom E, Brusselle G, Burgos F, Contoli M, Dinh-Xuan AT, Franssen FME, Gonem S, Greening N, Haenebalcke C, Man WDC, Moisés J, Peché R, Poberezhets V, Quint JK, Steiner MC, Vanderhelst E, Abdo M, Topalovic M, Janssens W. Collaboration between explainable artificial intelligence and pulmonologists improves the accuracy of pulmonary function test interpretation. Eur Respir J, 2023, 61(5):2201720.
29. Desbordes P, Steinwender M, Helfenstein E, Marginean C, Henzi N, Cerkl P, Spielmanns M, Keusch S, Stolz D, Simon M, Gompelmann D, Topalovic M. ArtiQ.PFT: AI-powered decision support for the diagnosis of lung diseases: Does it help the pulmonologist to be more accurate? Pneumologie, 2023, 77(S 01):S73-S74.
30. Huang YCT, Henriquez L, Chen H, Henriquez C. Development and evaluation of a computerized algorithm for the interpretation of pulmonary function tests. PLoS One, 2024, 19(1):e0297519.
31. Gompelmann D, Gysan MR, Desbordes P, Maes J, Van Orshoven K, De Vos M, Steinwender M, Helfenstein E, Marginean C, Henzi N, Cerkl P, Heeb P, Keusch S, Calderari G, von Boetticher P, Baumgartner B, Stolz D, Simon M, Prosch H, Janssens W, Topalovic M. AI-powered evaluation of lung function for diagnosis of interstitial lung disease. Thorax, 2025, 80(7):445-450.
32. Desbordes P, Maes J, Orshoven K Van, Vos M De, Steinwender M, Helfenstein E, Marginean C, Henzi N, Cerkl P, Heeb P, Keusch S, Calderari G, Boetticher P, Baumgartner B, Stolz D, Simon M, Gompelmann D, Janssens W, Topalovic M. Late Breaking Abstract-AI-powered decision support helps pulmonologists to diagnose ILD more accurately. Eur Respir J, 2023, 62(suppl 67):PA3483.
33. Das N, Happaerts S, Topalovic M, Janssens W. Pulmonologists collaborate with explainable artificial intelligence for superior interpretation of pulmonary function tests. Eur Respir J, 2021, 58(suppl 65):PA3443.
34. Levy J, Alvarez D, Del Campo F, Behar JA. Machine learning for nocturnal diagnosis of chronic obstructive pulmonary disease using digital oximetry biomarkers. Physiol Meas, 2021, 42(5):054001.
35. Xu WJ, Shang WY, Feng JM, Song XY, Li LY, Xie XP, Wang YM, Liang BM. Machine learning for accurate detection of small airway dysfunction-related respiratory changes: an observational study. Respir Res, 2024, 25(1):286.
36. Talker L, Dogan C, Neville D, Lim RH, Broomfield H, Lambert G, Selim A, Brown T, Wiffen L, Carter J, Ashdown HF, Hayward G, Vijaykumar E, Weiss ST, Chauhan A, Patel AX. Diagnosis and Severity Assessment of COPD Using a Novel Fast-Response Capnometer and Interpretable Machine Learning. COPD J Chronic Obstr Pulm Dis, 2024, 21(1):2321379.
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 1977, 33(1):159.
38. Elashmawi WH, Djellal A, Sheta A, Surani S, Aljahdali S. Machine Learning for Enhanced COPD Diagnosis: A Comparative Analysis of Classification Algorithms. Diagnostics, 2024, 14(24):2822.
39. Bodduluri S, Nakhmani A, Reinhardt JM, Wilson CG, McDonald ML, Rudraraju R, Jaeger BC, Bhakta NR, Castaldi PJ, Sciurba FC, Zhang C, Bangalore PV, Bhatt SP. Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease. JCI Insight, 2020, 5(13):e132781.
40. Hernandez R, Rodriguez R, Villalva O, Pimentel J, Miguel JL, Villasis MA. Exploratory study of a risk prediction artificial intelligence model to diagnose asthma and COPD in Mexico. Eur Respir J, 2023, 62(suppl 67):PA3784.
41. Le-Dong NN, Hua-Huy T, Nguyen-Ngoc HM, Dinh-Xuan AT. Applying machine learning and pulmonary function data to detect interstitial lung disease in systemic sclerosis. Eur Respir J, 2017, 50(suppl 61):OA3438.

10.12865/CHSJ.52.02.01

42. Ray K, Coenegrachts T, Van Steenbergen S, Halilovic A, Maes J, De Sadeleer L, Janssens WLH, Topalovic M. Artificial Intelligence Powered Spirometry Enables Early Detection of Interstitial Lung Disease. Am J Respir Crit Care Med, 2022, 205(Supplement 1):A4884-A4884.
43. Mencar C, Gallo C, Mantero M, Tarsia P, Carpagnano GE, Foschino Barbaro MP, Lacedonia D. Application of machine learning to predict obstructive sleep apnea syndrome severity. Health Informatics J, 2020, 26(1):298-317.
44. Prevot J, Mathi Mariappan N, Wickramaratne S, Sherman Y, Parekh A, Schotland H. Artificial Intelligence-Based Deep Learning Model for Detecting Central and Upper Airway Obstruction on Flow-Volume Loops. Am J Respir Crit Care Med, 2025, 211(Supplement 1):A7662-A7662.
45. Exarchos K, Potonos D, Aggelopoulou A, Sioutkou A, Kostikas K. Deep learning algorithm for the classification of spirometries using flow-volume curves: proof of concept study. Eur Respir J, 2021, 58(suppl 65):PA3442.
46. Mohanarajan M, Murugiah S, Hota P, Kumar S, Venkatesh SK. Advancements in Machine Learning and Artificial Intelligence in the Radiological Detection of Pulmonary Embolism. Cureus, 2025, 17(1):e77828.
47. Meghji J, Mortimer K, Agusti A, Allwood BW, Asher I, Bateman ED, Bissell K, Bolton CE, Bush A, Celli B, Chiang CY, Cruz AA, Dinh-Xuan AT, El Sony A, Fong KM, Fujiwara PI, Gaga M, Garcia-Marcos L, Halpin DMG, Hurst JR, Jayasooriya S, Kumar A, Lopez-Varela M V., Masekela R, Mbatchou Ngahane BH, Montes de Oca M, Pearce N, Reddel HK, Salvi S, Singh SJ, Varghese C, Vogelmeier CF, Walker P, Zar HJ, Marks GB. Improving lung health in low-income and middle-income countries: from challenges to solutions. Lancet, 2021, 397(10277):928-940.
48. Kostikas K, Hillas G, Gogali A. 2025 GOLD Report: What is New and What is Noteworthy for the Practicing Clinician. COPD J Chronic Obstr Pulm Dis, 2025, 22(1):2451613.
49. Richter DC, Joubert JR, Nell H, Schuurmans MM, Irusen EM. Diagnostic value of post-bronchodilator pulmonary function testing to distinguish between stable, moderate to severe COPD and asthma. Int J COPD, 2008, 3(4):693-699.
50. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ, 2020, 98(4):251.
51. Giffen E, Carroll D, Smets E, Orshoven K Van, Topalovic M, Walker E, Taylor A, Lowe D, Carlin C. P3 Integrating AI-assisted spirometry: validation and user experience in a COPD diagnostic pathway. Thorax, 2025, 80(Suppl 2):A125-A126.
52. Saha P, Chaudhary MFA, Rabby ASA, Bodduluri S, Nakhmani A, Zhang C, Bhatt SP. Machine learning based detection of true ventilatory restriction. Respir Med, 2025, 248:108314.
53. Mac A, Wu J, Ryan CM, Valaee S, Chow CW. Time-Series Machine Learning Model Classifies Lung Function Using Spirometry Flow-Volume Loops Alone. Am J Respir Crit Care Med, 2024, 209(Supplement 1):A2624-A2624.
54. Elmahy A, Maes J, Desbordes P, Vos M De, Topalovic M. AI models for respiratory diseases from spirometry alone: validation based on UK Biobank. Eur Respir J, 2023, 62(suppl 67):PA3508.
55. Zhou J, Wang P, Guo L, Cao J, Zhou M, Dai R. Automated interpretation of the pulmonary function test by a portable spirometer in Chinese adults. Clin Respir J, 2022, 16(8):555-561.
56. de Lima AD, Lopes AJ, do Amaral JLM, de Melo PL. Explainable machine learning methods and respiratory oscillometry for the diagnosis of respiratory abnormalities in sarcoidosis. BMC Med Inform Decis Mak, 2022, 22(1):274.
57. Andrade DSM, Ribeiro LM, Lopes AJ, Amaral JLM, Melo PL. Machine learning associated with respiratory oscillometry: a computer-aided diagnosis system for the detection of respiratory abnormalities in systemic sclerosis. Biomed Eng Online, 2021, 20(1):31.
58. Tomita K, Nagao R, Touge H, Ikeuchi T, Sano H, Yamasaki A, Tohda Y. Deep learning facilitates the diagnosis of adult asthma. Allergol Int, 2019, 68(4):456-461.
59. Topalovic M, Laval S, Aerts JM, Troosters T, Decramer M, Janssens W. Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints. Respiration, 2017, 93(3):170-178.
60. Das DK, Chakraborty C, Bhattacharya PS. Automated Screening Methodology for Asthma Diagnosis that Ensembles Clinical and Spirometric Information. J Med Biol Eng, 2016, 36(3):420-429.
61. Le-Dong NN, Hua-Huy T, Nguyen Ngoc HM, Martinot JB, Dinh Xuan AT. Detection of Interstitial Lung Disease in Systemic Sclerosis Using a Machine Learning Approach Based on Pulmonary Function Tests. Am J Respir Crit Care Med, 2017, 195(Supplement 1):A2531-A2531.
62. Xu T, Salehinejad H, Huang Q, Cho E, Valaee S, Chow CW. Late Breaking Abstract-Detection of small airway obstruction patterns in forced oscillometry(FO) and spirometry using artificial neural networks. Eur Respir J, 2018, 52(suppl 62):OA308.
63. Elmahy A, Maes J, Smets E, Vos M De, Janssens W, Topalovic M. Big data insights in the ILD population through AI-driven analytics. Eur Respir J, 2024, 64(suppl 68):PA2466.

Corresponding Authors: Thomas Orpwood, East Surrey Hospital, Redhill, UK,
e-mail: tom.orpwood1@nhs.net
Irina-Lavinia Soica, University College London Faculty of Medical Sciences,
e-mail: irina.soica.20@ucl.ac.uk