International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

Application Of Artificial Intelligence to Pulmonary Function Test Interpretation: A Scoping Review

Saranya J[1], Chirsty Varghese[2], Mohammed Nihal Cp[3]
[1]Faculty Respiratory Therapy, YSHACP, Bangalore,560064 Karnataka
[2]PG- Respiratory Therapy, YSHACP, Bangalore,560064 Karnataka
[3]UG- Respiratory Therapy, YSHACP, Bangalore,560064 Karnataka
Emails: *saranya.vnb@gmail.com[1],christyvarghesef1@gmail.com[2], mdnihalhaneefa@gmail.com[3]*

Abstract
Pulmonary medicine is the field where artificial intelligence (AI) is actively used, and respiratory data can be processed objectively and in a standardized way (3). The observance of complex trends among spirometric and volumetric values that are frequently ignored by clinicians can be identified by machine learning algorithms, and AI-aided Pulmonary Function Test (PFT) interpretation is a logical development (4). Nevertheless, the interpretation of PFT is still susceptible to the high interobserver variability, and the accuracy rate of pulmonologist varies in a range of 44.6% to 65.8 percent in contrast to expert consensus (4). Although AI has the potential to perform at the same level as experts or even better, the scoping study has not been performed in a systematic manner to map the new field. This review covered studies that used adults (over 18 years) that used AI, machine learning or deep learning algorithms to interpret, classify or diagnose respiratory disease using PFT data; those using AI on non-PFT signals were not included. PubMed, Scopus, Web of Sciences, and ProQuest were searched electronically since its inception and only English-language sources were included. The review was based on the JBI scoping review methodology and PRISMA-ScR. A standardized form was used to chart the data and NVivo version 15 was used to synthesize them thematically. Out of 312 records examined, five studies were included. AI identified 40 to 100 percent, versus 44.6 percent to 74.4 percent pulmonologists (4). Explainable AI enhanced the diagnostic accuracy of clinicians by 5-10 percent, and human-AI teams had higher diagnostic accuracy compared to individual human or AI (5). Nonetheless, AI was less sensitive to restrictive patterns and rare conditions (4), no researchers conducted external validation of heterogeneous populations (4, 5), and automation bias was found. AI demonstrates a high potential in standardizing the interpretation of PFT, though without external validation, low levels of algorithmic transparency, automation bias, and absence of prospective outcome trials are still major limitations to clinical translation (3, 4, 5).

Keywords*: Artificial intelligence; diagnostic accuracy; machine learning; Pulmonary Function Tests; spirometry.*

1. Introduction

Pulmonary Function Tests (PFTs) are fundamental diagnostic tools in respiratory medicine, measuring airflow dynamics, lung volumes, and gas exchange efficiency. They are essential for diagnosing, monitoring, and prognosticating conditions such as chronic obstructive pulmonary disease (COPD), asthma, interstitial lung disease, and neuromuscular respiratory disorders (1). However, the clinical value of PFTs is contingent upon the accuracy of their interpretation, which remains a complex and error-prone process. Recent advances in artificial intelligence (AI) offer a promising avenue to enhance diagnostic consistency and reduce interobserver variability. Despite encouraging early findings, no scoping review has systematically mapped the evidence on AI applications in PFT interpretation. This review aims to chart the scope and depth of existing research on the use of AI in interpreting, classifying, and diagnosing respiratory disease based on PFT data.

1.1.The Challenge of PFT Interpretation

Accurate PFT interpretation requires integrating

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

spirometric indices, lung volumes, diffusion capacity, and flow-volume loop morphology within the patient's clinical context (4). This task has long been subject to significant variability: in a multicentre study across 16 European hospitals, 120 pulmonologists matched ATS/ERS guideline-based patterns in only 74.4% of cases (nκ = 0.67), and diagnostic category accuracy averaged just 44.6%, ranging from 24% to 62% irrespective of clinical experience or institutional type (4). A cross-sectional study of 200 patients confirmed persistent misclassification and interobserver variance in respiratory diagnostics (1). These errors carry substantial clinical consequences, including misdiagnosis, inappropriate therapy, delayed referrals, and unnecessary investigations (1). Earlier algorithmic approaches operationalising ATS/ERS guidelines achieved only 38% diagnostic accuracy, improving to 68% with demographic data but remaining inferior to the 77% achieved by expert panels (4).

1.2.Artificial Intelligence in PFT Interpretation Promise and Gaps

AI, including machine learning and deep learning, has demonstrated expert-level diagnostic performance across multiple medical fields (4). The standardised, numerical nature of PFT data makes it particularly amenable to AI-based analysis (4). Early results are encouraging: AI software achieved 100% pattern classification accuracy and 82% diagnostic accuracy in a European multicentre trial, significantly outperforming individual pulmonologists (4). A Decision Tree algorithm achieved 86.59% accuracy versus 65.82% by pulmonologists using identical limited data (2), while another study reported 92% overall AI accuracy with strong expert agreement (κ = 0.86) (1). One European university hospital has already integrated AI into routine PFT workflows, processing over 5,000 assessments in clinical practice (3). Beyond direct comparisons, Das et al. (2023) demonstrated that explainable AI (XAI) using Shapley values improved clinician accuracy by 5–10%, with human-AI collaboration outperforming either alone (5). However, critical barriers persist: training data composition may limit generalisability to diverse populations (3), algorithmic transparency and reproducibility remain essential for clinical trust (3), automation bias has been empirically documented (5), and no prospective trial has yet assessed whether AI integration improves tangible patient outcomes such as diagnostic timeliness or healthcare costs (3, 4, 5). Jump to latest

2. Method

This scoping review set out to chart the breadth and character of existing research on the use of artificial intelligence (AI) for interpreting, classifying, and diagnosing respiratory conditions from Pulmonary Function Test (PFT) data in adult populations (Arksey & O'Malley, 2005). Since it is a new field and the published literature is diverse in terms of quantitative, experimental, observational, and cross-sectional designs, a scoping review was deemed more suitable than a conventional systematic review, which generally combines effectiveness information on a single well-specific intervention. This reasoning followed advice from the Joanna Briggs Institute (JBI) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Page MJ et al., 2021; Peters et al., 2020). The review thus followed the five-step systems approach of conducting research as introduced by Arksey and O'Malley, and further refined by JBI: (1) framing the research question; (2) locating relevant studies; (3) selecting which studies to include; (4) extracting and tabulating the data; and (5) bringing together, summarizing and reporting the findings (Arksey and O'Malley, 2005).

Table 1. Proposed Search Strategy in PubMed, Web of Science and Scopus.

S.No	Data base	Search Terms	Result
1.	PubMed	(("artificial intelligence" OR "machine learning" OR "deep learning") AND ("Pulmonary Function Tests" OR spirometry OR "Pulmonary Function Test*") AND (interpretation OR diagnosis)	310
2.	Web Of Science	(("artificial intelligence" OR "machine learning" OR"deep learning") AND ("Pulmonary Function Tests" OR Spirometry OR "Pulmonary Function Test*") AND (Interpretation OR diagnosis))	1
3.	Scopus	(("artificial intelligence" OR "machine learning" OR "deep learning") AND ("Pulmonary Function Tests" OR Spirometry OR "Pulmonary Function Test*") AND (Interpretation OR Diagnosis))	1

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

3. Results And Discussion

3.1.Results

A total of 312 records were found in the search of the databases of PubMed, Scopus and Web of Science up to 2026. After the elimination of duplication and sifting of the records, five studies were retained and were put to the review. Synthesis and coding of the information was done in a narrative format after plotting the data according to Joanna Briggs Institute guidelines. The results of the studies included were used to produce descriptive codes which were gradually pooled to come up with the various thematic areas. NVivo software (version 15) has been used to support this. After this cycle of coding, 10 subthemes and three overarching themes were produced.

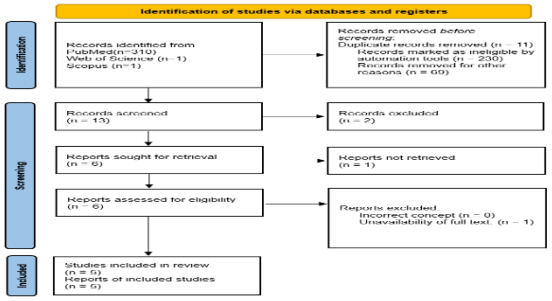

FIGURE 1. PRISMA- ScR Flow Chart PRISMA Flow Diagram for the scoping review process

3.2.Theme 1: Artificial Intelligence Diagnostic Performance.

In all the studies accounted, AI algorithms had a higher diagnostic accuracy than individual pulmonologists in the interpretation of PFT data. An AI system that was trained on more than 10,000 labelled PFTs in a cross-sectional study of 200 patients in a tertiary-care hospital demonstrated a total accuracy of 92, a sensitivity of 91.5, specificity of 93.2, and Cohen z03ba of 0.86, compared to the consensus of expert pulmonologists, identifying 38 of 40 normal, 76 of 80 obstructive, 44 of 50 restrictive, and 27 A Decision Tree algorithm in a retrospective-prospective study of 440 patients and eight senior pulmonologists in India has shown an accuracy of 86.59 which is considerably higher than the accuracy of the collective of pulmonologists of 65.82. The interrater agreement of the pulmonologists was at the best moderate in this study (Fleiss z3ba = 0.46), and individual pulmonologist predication accuracy ranged between 54 and 79 percent [2]. In the biggest study, a non-interventional, multicenter study with 120 pulmonologists in 16 hospitals in five European countries, the AI software accurately identified the disease category with an 82 percent accuracy, which is significantly higher than the mean accuracy of the individual pulmonologists (44.6) with no statistically significant effect of clinical experience or hospital type on the outcome [4].

3.3.Coherence in a variety of study designs and methods of AI.

The better-than-clinician performance of AI was identified regardless of the study design (cross-sectional, retrospective-prospective, and multicenter non-interventional), geographical location (India and Europe), and AI approaches z2014 between basic Decision Tree classifiers that used spirometry data only [2] and machine-learning models that used complete PFT suites comprising of plethysmography and diffusing capacity only [4]. This uniformity indicates that the benefit here lies in the innate ability of AI to manipulate dozens of numerical factors at once and identify multi-dimensional patterns of the disease that is habitually ignored by human readers [2,4]. AI accuracy on pattern-classification varied between 86.59% and 100% with pulmonologists scoring between 44.6% and 74.4% on the same tasks [1,2,4].

3.4.Demerits of rule-based algorithms.

The flaws of the rule-based software highlighted the usefulness of data-driven AI. Direct coding of interpretative guidelines in the format of the ATS/ERS created a tool that predicted diseases with only 38% accuracy; considering a basic level of demographic variables, including age and sex, could increase accuracy to 68 percent, which, however, was significantly lower than that of the expert panel and the data-based AI algorithm on the identical patient population [4]. This result substantiates the idea that the benefit of AI does not lie in the codification of the existing clinical rules, but in the learning patterns that cross the rules

3.5.Theme 2: Human and AI Collaboration.

The article revealed that explainable AI (XAI) can

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

have a positive impact on the performance of single clinicians in a two-phase repeated-measures study of 78 pulmonologists (16 in Phase 1, 62 in Phase 2) in a group of Europe-based institutions. The diagnostic prediction of the AI was presented to the pulmonologists in Phase 1 and Phase 2, which led to an increase in the preferential diagnostic performance of the pulmonologists of 10.4% ($p < 0.001$) in Phase 1 and 5.4% ($p < 0.001$) in Phase 2 (z2014) and the elaboration of the logic z2014 upon which the AI has reached using Shapley-value-based mathematical attribution z2014. The confidence of the intervention arm in diagnosing also improved [5].

3.6.Clinician and AI team synergistic performance.

It is also worth noting that clinician-plus-AI team was significantly better than only clinician and only AI: 12 z201313% ($p < 0.0001$) [5]. This finding could favour the result of having a decision-support model of AI as a partner to clinical expertise rather than a substitute. The synergistic performance that is observed is consistent with the existing trends in the other specialties, including automated ECG interpretation in cardiology, where the machine presupposes the initial search, which must be quick and predictable, but the medical specialist is the ultimate interpreter [3,4].

3.7.Practice-based clinical integration.

It is stated in the published letters of correspondence that an artificial intelligence system was already implemented into the workflow in the University Hospital of Leuven, and this system already has already have more than 5,000 fully completed PTs in the regular workflow with patients [3]. The system has been characterized as a real second opinion that saves the laboratory time on pulmonary functions and also improves the quality of interpretations. Besides being used as a decision-support system, the system was also proposed to be a possible educational tool that can be provided to trainees [3].

3.8.Theme 3: Clinical Translation barriers.

The largest gap, established in all the five studies was the total lack of external validation in ethnically/geographically diverse populations. All the reviewed algorithms were created and evaluated in the same institution or within a group of demographically similar hospitals [1,2,4,5]. It is not at all clear whether these models can be generalized across ethnicities, ranges of body habitus, distributions of disease prevalence, resource constrained environments. It was noted by commentators that the composition of training datasets has a fundamental impact on the output of algorithms and that datasets that are enriched with rare diagnoses are unlikely to recapitulate the spectrum of the disease that can be expected in clinical practice [3].

3.9. Disproportional disease type performance.

There was no constant performance of AI across ventilatory patterns and disease categories. The AI system, which was used in the 200-patient study, showed much lower sensitivity of 88% in restrictive patterns, with that of 95% in obstructive and normal categories [1]. The multicenter algorithm in Europe also experienced a challenge in the less common diagnoses like bronchiectasis and neuromuscular disease [4]. The healthcare professionals themselves also scored worst in intermediate severity levels [2]. In both human and machine performance, therefore, it became worsened along the extremes of the disease range where there were fewest training exemplars.

4. Algorithms in healthcare

Responsibility Algorithms transparency and automation bias DSS control and informed consent Automated decision-making Systems oriented to patients and care environments The algorithm transparency was reported as a long-standing obstacle to the trust of clinicians. Most AI systems are black boxes, and even explainable-AI method which helped clinicians improve their performance could not eliminate automation bias [3,5]. Whenever the algorithm provided the wrong suggestion, the accuracy of medical workers was reduced, but not significantly, which shows that not all professionals approached the machine results with seriousness. Though the rate of agreement increased significantly with the right predictors ($p < 0.001$), implying that the risk is not so high, any implementation in large-scale clinical practice where time constraints are more evident may increase this trend [5]. Absence of prospective patient-outcome evidence

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

4.1. Absence of prospective patient-outcome evidence.

Most fundamentally, all of the five studies did not quantify actual patient outcomes. The issue of AI-aided interpretation of PFTs causing faster diagnostics, reduced inappropriate referrals, decreased healthcare costs, and enhanced patient satisfaction is never assessed in any prospective trial [2,4,5]. Without such evidence, the even high level of accuracy does not have any certain clinical value, as even significant accuracy improvements remain under unknown clinical utility, and whether AI-based PFT interpretation can be translated into measurable clinical utility is a question of its own.

5. Discussion

The main conclusion of the present scoping review is the fact that artificial intelligence may always be equal or even better than the performance of single healthcare professionals when it comes to interpreting Pulmonary Function Test results; however, the evidence base underlying this potential is still limited and confined to a number of still not directly resolved questions that lie between the promising results of laboratory results and the responsible adoption at the bedside. In the five studies utilized, AI algorithms classified patterns and classes with an accuracy of 86.59 to 100 percent, and single pulmonologists with an accuracy of 44.6 to 74.4 percent on the same tasks [1,2,4]. Such differences were not insignificant, and the performance disparity did not decrease regardless of the years of experience of the clinicians, hospital environment, or the country in which the clinicians practiced [2,4]. The consistency of the result in the literature, with the variation in study designs, geographical locations, and AI methodologies z2014 by simple Decision Tree algorithms using only spirometric data and the whole PFT suite using plethysmography and diffusing capacity z2014 implies that the benefit lies in the ability of AI to simultaneously handle dozens of numeric parameters and identify multi-dimensional disease fingerprints that the human eye will make regular failures to detect [2,4]. The inability of rule-based software that was built in direct correspondence with ATS/ERS guidelines to achieve 38% accuracy in disease prediction, and the data-driven AI algorithm to get 82% with the same sample of patients [4], once again proves the point that the power of AI lies not in the codification but the learning patterns that supersede it. Equally significant is the evidence that AI functions optimally as a collaborative partner rather than a replacement for human expertise. When pulmonologists were presented not only with the AI's predicted result but also with an explanation of the algorithm's reasoning z2014 employing the Shapley-value method z2014 their individual accuracy improved by 5 z201310%, and the combined human z2013AI team significantly outperformed both parties working independently [5]. This collaborative model mirrors the manner in which automated ECG interpretation already functions in cardiology: the machine provides a rapid, consistent first reading, while the healthcare professional retains final interpretive authority [3]. At least one AI system has already been integrated into routine clinical practice at a European university hospital, where it processes more than 5,000 patient PFTs, functioning as an "accurate second opinion" that enhances efficiency and elevates the overall quality of interpretations [3]. Even though these encouraging results were found, there were a number of critical limitations in which any suggestion to move into large-scale uses is obscured. The most significant one is the total lack of external validation within various groups of people. All the reviewed algorithms have been developed and tested either within the same institution or a network of demographically similar hospitals [1,2,4,5]. The extent to which these models can be generalized in other ethnicity, body habitus distributions, prevalence distributions of disease and resource settings is completely unclear z2014 a issue brought out explicit by commentators who observed that the composition of training-datasets has a significant influence on algorithm output and that datasets with rare diagnoses might be atypical of real-world clinical practice [3]. The issue of biased performance in disease categories is closely connected: AI was less sensitive in restrictive forms (88% versus 95% in obstructive) [1] and proved the most challenging to health practitioners themselves [4] and lower prevalence diseases (such as bronchiectasia or neuromuscular

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

disease) [2]. The performance of humans and machines, consequently, becomes worse at the periphery of the disease spectrum where training examples are the most limited. Another barrier is the algorithm transparency. Most AI systems are black boxes, and even the so-called explainable-AI algorithm that was proven to lead to improved clinician performance could not avoid the presence of automation bias: even when the algorithm provided a false recommendation, the accuracy of healthcare professionals decreased marginally, which suggests that some health practitioners blindly accepted the output provided by the machine [5]. Even though consensus was significantly stronger in the case of correct forecasts, it implied to propose the total risk as low ([5], which would be increased in the case of high-volume clinical environments where time pressure is higher. Most essentially, there was no study that evaluated practical patient outcomes. The question of whether AI-aided PFT interpretation results in expedited diagnostics, more effective use of healthcare resources, less expensive healthcare, or the increase in patient satisfaction has not been studied with any prospective-based trial [3,4,5]. Even a significant change in diagnostic accuracy may not be significant clinically in the absence of outcome assessment. The review has several limitations. Only 5 studies met the inclusion criteria which might be due to the youthfulness of the field and limits the level of synthesis that can be made. The search was limited to publications in the English language. Since a scoping review is a process that is consistent with the JBI methodology, the formal quality appraisal of individual studies was not carried out. Since the time the search was carried out, there could have been more work on the topic that is relevant to this study because of the high rate of AI research. Finally, AI-based PFT interpretation has already demonstrated proof of concept but is yet to demonstrate proof of clinical utility. The future requires multicenter external validation in an ethnically and geographically heterogeneous population, prospective clinical trials with patient-centered outcomes including time to diagnosis, cost-effectiveness, additional improvement of explainability approaches to reduce automation bias, and the creation of standardized reporting systems of AI-PFT trials. As long as these milestones are not reached, AI is to be viewed as a highly promising decision-support tool that is yet to obtain the necessary evidence to be implemented safely and transparently in clinical practice [1,5].

Conclusion

This scoping review of five studies demonstrates that AI consistently outperforms individual pulmonologists in PFT interpretation, achieving diagnostic accuracy of 82–100% compared to 44.6–74.4% for clinicians, with 100% concordance on ATS/ERS pattern classifications [2, 5]. This advantage held across different study designs, geographical settings, and AI methodologies [2]. Explainable AI improved individual clinician accuracy by 5–10%, and the human–AI collaborative team significantly outperformed either party working alone, reinforcing AI's role as a decision-support partner [3]. At least one AI system is already operational in routine clinical practice at a European university hospital, processing over 5,000 PFT assessments [1]. However, critical barriers to widespread adoption remain. AI showed reduced sensitivity for restrictive patterns (88% vs. 95% for obstructive) and struggled with rare diagnoses such as bronchiectasis and neuromuscular disease [2, 5]. No algorithm underwent external validation in ethnically or geographically diverse populations, raising concerns about generalisability [1, 2, 3, 5]. Algorithm transparency remains limited, and automation bias was observed even with explainable AI approaches [1, 3]. Most importantly, no prospective trial has evaluated whether AI-assisted PFT interpretation improves patient outcomes such as time to diagnosis, unnecessary testing, or healthcare costs [1, 2, 3]. AI-based PFT interpretation has demonstrated proof of concept but not yet proof of clinical utility. Responsible clinical adoption requires multicentre external validation across diverse populations, prospective outcome-based trials, enhanced algorithmic transparency, and standardised reporting frameworks for AI-PFT studies. Until then, AI should be regarded as a promising decision-support tool that complements, rather than replaces, clinical expertise [1, 2, 3, 5].

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

TABLE 2. Proposed Data Extraction ForScoping Review

S. No	Author/ year/ country	Objective	Participants' characteristic	Study design	Sampling techniques and Sample size	Data collection method/ type of interview/ group discussion	RESULTS	
							AI technologies used	Findings Related to AI implementation.
1.	Singh et al., 2025, India	To evaluate diagnostic accuracy of an AI-assisted system for interpreting PFTs compared with pulmonologist consensus.	200 adults (≥18 yrs) undergoing spirometry and full PFT at a tertiary hospital; excluded poor-quality spirometry and neuromuscular disease.	Cross-sectional diagnostic accuracy study.	Consecutive clinical sampling; n=200.	Clinical data and PFT indices collected; no interviews or group discussions.	Supervised learning model trained on >10,000 labelled PFTs	AI achieved 92% overall accuracy (κ = 0.86) vs. expert consensus across 200 patients; lower sensitivity for restrictive patterns (88%).
2.	Saad et al., 2025, India	To compare AI vs pulmonologists in interpreting PFT patterns using limited clinical data and spirometry.	3,230 retrospectives + 440 prospective adult cases; majority males; most aged 21–60; high smoking prevalence.	Retrospective-prospective observational analytical study.	Convenience dataset from tertiary canter; training n=3230; test n=440.	Digitalized anonymized records; pulmonologists reviewed via Google Forms; no interviews.	Decision Trees, SVM, KNN, Naïve Bayes, Neural Networks (Python-based); trained on 3,230 retrospective PFTs	AI (Decision Tree) achieved 86.59% accuracy vs. 65.82% by eight pulmonologists; Fleiss' κ = 0.46 indicating moderate inter-rater agreement; AI struggled with the "others" (restrictive/mixed) category.
3.	Gonem & Siddiqui, 2019, UK.	To discuss implications, challenges, and opportunities of AI for PFT interpretation	Not applicable (correspondence; no primary participants).	Commentary / correspondence.	Not applicable.	Conceptual discussion; no interviews.	Editorial/correspondence (no novel AI developed)	Highlighted need for AI transparency, reproducibility, and open-source code; emphasised AI's potential for educational use and knowledge transfer back to clinicians.
4.	Topalovic et al., 2019, Europe (multicentre)	To compare pulmonologists' interpretation accuracy and variability with AI-based software.	120 pulmonologists from 16 hospitals; 50 adult patient PFT cases across disease categories.	Multicentre non-interventional study.	Random selection of 50 clinical cases; 6,000 interpretations generated.	Structured independent evaluation sessions; no interviews/group discussion.	Machine learning model (R-based) trained on 1,430 cases using PFT indices and clinical data	AI achieved 100% PFT pattern accuracy and 82% diagnostic accuracy vs. 74.4% pattern and 44.6% diagnostic accuracy by 120 pulmonologists across 16 hospitals.
5.	Das et al., 2023, Europe (multicentre)	To assess whether collaboration between pulmonologists and explainable AI improves diagnostic accuracy.	Phase 1: 16 pulmonologists; Phase 2: 62 pulmonologists; adult anonymized PFT cases with clinical data.	Repeated-measures intervention study (monocentre + multicentre).	Purposive case selection; each interpreted with and without AI assistance.	Online structured interpretation platform; no interviews.	Explainable AI (XAI) using Shapley values for interpretation of a machine learning diagnostic model	XAI improved pulmonologists' preferential diagnostic accuracy by 5–10%; human-AI collaboration significantly outperformed both AI alone and clinicians alone.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0148
e ISSN: 2584-2854
Volume: 04 Issue: 04
April 2026
Page No: 973 - 980

Acknowledgements

The authors would like to acknowledge Yenepoya (Deemed to be University Bangalore) for providing the necessary support and resources to conduct this study. No external funding was received for this research.

References

[1]. Singh, S., et al. (2025). AI-assisted interpretation of Pulmonary Function Tests: Cross-sectional diagnostic accuracy study in a tertiary care teaching hospital. [Journal name, volume, page numbers, and DOI needed to complete this entry]

[2]. Saad, T., Pandey, R., Padarya, S., Singh, P., Singh, S., & Mishra, N. (2025). Application of artificial intelligence in the interpretation of Pulmonary Function Tests. Cureus, 17(4), e82056. https://doi.org/10.7759/cureus.82056

[3]. Gonem, S., & Siddiqui, S. (2019). Artificial intelligence for Pulmonary Function Testing. European Respiratory Journal, 53(4), 1900405. https://doi.org/10.1183/13993003.00405-2019

[4]. Topalovic, M., Das, N., Gianella, P. R., Gayan-Ramirez, G., Haddad, R., Leemans, N., Salonidis, T., & Decramer, M. (2019). Artificial intelligence outperforms pulmonologists in the interpretation of Pulmonary Function Tests. European Respiratory Journal, 53(4), 1801660. https://doi.org/10.1183/13993003.01660-2018

[5]. Das, N., Happaerts, S., Germonpré, P., Janssens, W., & Topalovic, M. (2023). Collaboration between explainable artificial intelligence and pulmonologists improves the accuracy of Pulmonary Function Test interpretation. European Respiratory Journal, 61(1), 2201720. https://doi.org/10.1183/13993003.01720-2022