TPM Vol. 32, No. S5, 2025
ISSN: 1972-6325
https://www.tpmap.org/

ENHANCING PULMONARY FUNCTION TEST REPORTING WITH ARTIFICIAL INTELLIGENCE: A RETROSPECTIVE OBSERVATIONAL STUDY

DR. SAKTHI BHALAN PANDIARAJAN[1], DR. ANBUMARAN PARIVAKKAN MANI[2], DR. GANGADHARAN VADIVELU[3]
[1,2,3]SAVEETHA MEDICAL COLLEGE, SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES.

Abstracts

Background: Pulmonary Function Tests (PFTs) are essential for diagnosing and monitoring respiratory disorders such as asthma, chronic obstructive pulmonary disease (COPD), and interstitial lung disease. Despite standardized protocols, manual interpretation of PFTs is prone to interobserver variability and delays in reporting. Artificial Intelligence (AI) offers a promising solution to improve diagnostic accuracy, consistency, and efficiency in functional respiratory diagnostics. This study aims to evaluate the accuracy and reliability of AI-assisted interpretation of PFTs compared with conventional assessment by pulmonologists.

Methods: A retrospective observational study was conducted at the Pulmonary Function Laboratory, Saveetha Medical College, Chennai, including 200 patients who underwent spirometry. Data were processed using a machine learning based algorithm trained on a subset of cases to classify ventilatory patterns into normal, obstructive, restrictive, and mixed categories. The diagnostic performance of AI interpretation was validated against pulmonologist reports. Key outcome measures included sensitivity, specificity, predictive values, and concordance with expert assessment.

Results: Among the study participants, the majority were male (86%), with most falling within the 20–39-year age group (51.5%). The AI model demonstrated strong diagnostic accuracy across all categories. For mixed obstructive patterns, the area under the curve (AUC) was 0.901; for obstructive, 0.930; for restrictive, 0.930; and for normal spirometry, 0.959. Sensitivity and specificity consistently exceeded 84% and 95%, respectively, with high positive and negative predictive values. The AI system reduced interobserver variability and produced consistent, reproducible outputs comparable to pulmonologist interpretations.

Conclusion: AI-based interpretation of PFTs achieves high diagnostic accuracy and reliability, offering a practical alternative to manual reporting. Its integration into clinical workflows and electronic health record systems can improve efficiency, reduce reporting delays, and enhance diagnostic consistency, particularly in resource-limited healthcare settings. Further validation across diverse populations is warranted before widespread adoption.

Keywords: Pulmonary function tests, Artificial intelligence, Spirometry, Diagnostic accuracy, Machine learning, Respiratory medicine

INTRODUCTION

Pulmonary Function Tests (PFTs) are widely recognized as fundamental diagnostic and monitoring tools in respiratory medicine. They are routinely employed to evaluate diseases such as asthma, chronic obstructive pulmonary disease (COPD), and interstitial lung disease, all of which contribute significantly to global morbidity and mortality. Parameters including spirometry indices, lung volumes, and diffusion capacities allow clinicians to quantify functional impairments, assess disease severity, and monitor therapeutic response (1). While standardized guidelines ensure technical accuracy, the interpretation of PFTs remains complex, as it is influenced by patient effort, reference standards, and clinician expertise. Manual interpretation is often time-consuming and may be affected by interobserver variability, resulting in inconsistent diagnostic outcomes and potential delays in initiating treatment (2). These challenges are particularly problematic in high-volume clinical environments and resource-limited healthcare systems.

In recent years, Artificial Intelligence (AI) has emerged as a promising tool to address such gaps in medical diagnostics. By leveraging machine learning and deep learning algorithms, AI has demonstrated superior performance in fields such as radiology, pathology, and cardiology, where it has equaled or surpassed human expertise in pattern recognition and disease classification (3). Within pulmonary medicine, AI has predominantly been applied to imaging studies such as chest radiographs and computed tomography. However, its application in functional diagnostics, particularly in PFT interpretation, remains relatively underexplored. Evidence suggests that AI-based models trained on large, diverse datasets can accurately classify obstructive, restrictive, and mixed ventilatory patterns while minimizing subjectivity and ensuring reproducibility (3). Furthermore, automated AI interpretation can provide rapid, real-time diagnostic support, substantially reducing turnaround time and integrating seamlessly into electronic health record (EHR) systems. Such advancements hold particular importance in resource-constrained healthcare environments, where specialist availability may be limited, and timely reporting is critical (4). The objective of this study is to assess the accuracy and reliability of AI-based pulmonary function test interpretation compared with conventional manual assessment methods.

MATERIALS AND METHODS

This study was designed as a retrospective observational analysis and was carried out at the Department of TB and Respiratory Medicine, Saveetha Medical College, Chennai, following approval from the Institutional Ethics Committee (IEC-Reference Number: 014/09/2024/IEC/SMCH). The primary aim was to evaluate the accuracy and reliability of an Artificial Intelligence (AI)-based framework for Pulmonary Function Test (PFT) interpretation compared with conventional manual assessment performed by pulmonologists.

The study population consisted of 200 patients who underwent spirometry during the study period. Eligibility criteria required that tests meet acceptability and reproducibility standards in line with the American Thoracic Society/European Respiratory Society (ATS/ERS) 2022 guidelines. Exclusion criteria included incomplete manoeuvres, artifacts, or suboptimal patient effort that compromised the reliability of the test. Demographic data such as age, sex, body mass index (BMI), and smoking history were recorded to describe the study participants.

For AI model development, labelled spirometry datasets were used to train a machine learning algorithm. The model was designed to classify ventilatory impairments into four categories: obstructive, restrictive, mixed, and normal patterns. Validation was performed against interpretations made by expert pulmonologists, who served as the reference standard. The AI algorithm was also configured to generate standardized reports for direct comparison with manually prepared clinical interpretations.

Outcome measures included diagnostic accuracy, sensitivity, specificity, and concordance between AI and pulmonologist classifications. These metrics were analyzed to determine the performance of the AI system relative to expert interpretation. Additionally, the study evaluated the extent to which AI reduced interobserver variability, a common limitation in manual reporting. Statistical analysis included descriptive statistics for baseline characteristics and performance metrics. A p-value <0.05 was considered statistically significant.

RESULTS

Table 1: Demographic variable of study participants (n=200)

S.no	Variable	Category	Number of Patients (n)	Percentage (%)
1	Age Group (years)	<20	9	4.5
		20–39	103	51.5
		40–59	45	22.5
		60–79	39	19.5
		≥80	4	2
2	Gender	Male	172	86
		Female	28	14

A total of 200 patients were included in the study. The patients were categorized based on age, and gender (Table 1). The majority (51.5%) were between 20–39 years of age, followed by 22.5% in the 40–59 age group, 19.5% in the 60-79 age group, 4.5% under 20 years, and only 2% aged 80 years or above. The study participants were predominantly male (86%), with females comprising 14% of the study population.

Table 2: Body Mass Index (BMI) categories of study participants (n=200)

BMI	Category	Number of patients (n)	Percentage (%)
< 18.5	**Underweight**	6	3.00
18.5 - 24.9	**Normal Weight**	117	58.50
25 - 29.9	**Overweight**	61	30.50
> 30	**Obese**	16	8.00

Table 3: Smoking status of study participants (n=200)

Group	Number of patients (n)	Percentage (%)
Smoker	41	20.50
Non- Smoker	159	79.50
Total	200	100.00

With respect to body mass index (BMI) (Table 2), 58.5% of individuals had a normal BMI, 30.5% were overweight, 8% were obese, and 3% were underweight. In terms of smoking history (Table 3) 20.5% were smokers, while 79.5% reported no smoking history.

Among the 200 patients included in the study, the first 50 patient records were used for training the AI software to predict spirometry outcomes. The remaining 150 patients were evaluated by both the AI model and pulmonologists to assess the accuracy of AI-based spirometry interpretation.

Figure 1: Diagnostic Accuracy of the AI software in detecting Mixed obstructive pattern: ROC Curve Analysis Against Pulmonologists diagnosis.

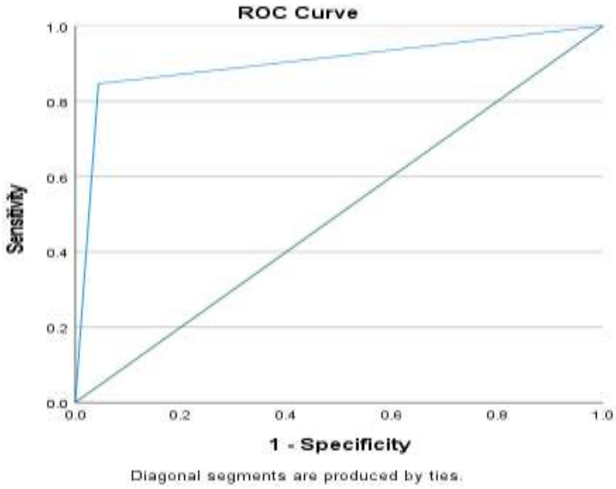

Table 4: Diagnostic Accuracy of the AI software in detecting Mixed obstructive pattern: Area Under the Curve

Area	95 % confidence interval	
	Lower range	Upper range
0.901***	0.785	1.000

Table 5: Diagnostic Accuracy of the AI software vs Pulmonologists in detecting Mixed obstructive pattern

Sensitivity	84.60%
Specificity	95.60%
PPV	64.70%
NPV	98.50%

The diagnostic accuracy of the AI software was assessed across four spirometry patterns - mixed obstructive, pure obstructive, restrictive, and normal. For the detection of mixed obstructive patterns, the area under the ROC curve (AUC) was 0.901 (95% CI: 0.785–1.000) (Figure 1) (Table 4). The algorithm achieved a sensitivity of 84.6%, specificity of 95.6%, positive predictive value (PPV) of 64.7%, and negative predictive value (NPV) of 98.5% (Table 5).

Figure 2: Diagnostic Accuracy of the AI software in detecting Pure obstructive pattern: ROC Curve Analysis Against Pulmonologists diagnosis.

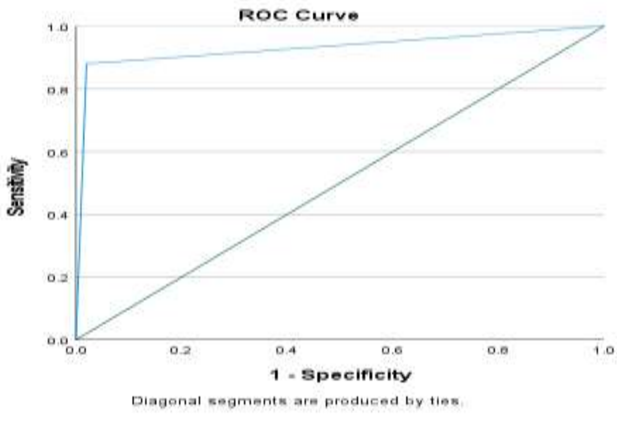

Table 6: Diagnostic Accuracy of the AI software in detecting Pure obstructive pattern: Area Under the Curve

Area	95 % confidence interval	
	Lower range	Upper range
0.930***	0.875	0.985

Table 7: Diagnostic Accuracy of the AI software vs Pulmonologists in detecting Pure obstructive pattern

Sensitivity	88.00%
Specificity	98.00%
PPV	95.70%
NPV	94.20%

In identifying pure obstructive patterns, the model demonstrated an AUC of 0.930 (95% CI: 0.875–0.985) (Figure 2) (Table 6). Sensitivity and specificity were 88.0% and 98.0%, respectively, with a PPV of 95.7% and NPV of 94.2% (Table 7).

Figure 3: Diagnostic Accuracy of the AI software in detecting restrictive pattern: ROC Curve Analysis Against Pulmonologists diagnosis

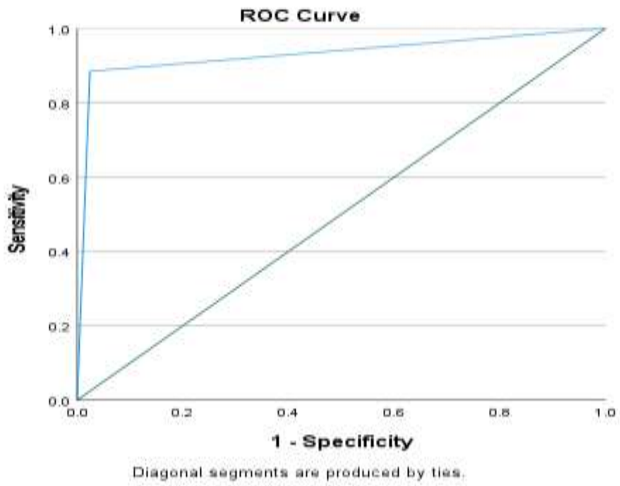

Table 8: Diagnostic Accuracy of the AI software in detecting restrictive pattern: Area Under the Curve

Area	95 % confidence interval	
	Lower range	Upper range
0.930***	0.857	1.000

Table 9: Diagnostic Accuracy of the AI software vs Pulmonologists in detecting restrictive pattern

Sensitivity	88.50%
Specificity	97.60%
PPV	88.50%
NPV	97.60%

For restrictive ventilatory patterns, the diagnostic performance was similarly robust, with an AUC of 0.930 (95% CI: 0.857–1.000) (Figure 3) (Table 8). Sensitivity and specificity were 88.5% and 97.6%, respectively, and both PPV and NPV were 88.5% and 97.6% (Table 9).

Figure 4: Diagnostic Accuracy of the AI software in detecting normal spirometry: ROC Curve Analysis Against Pulmonologists diagnosis.

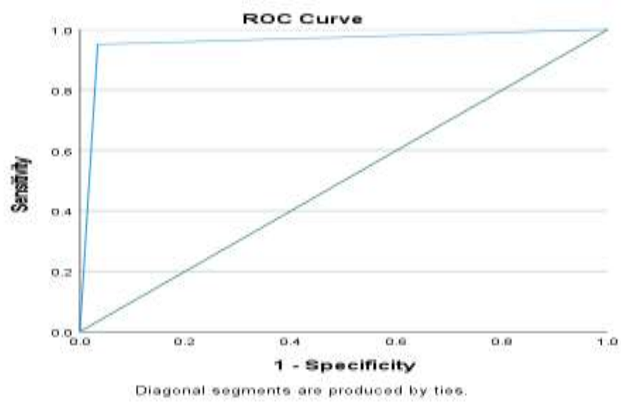

Table 10: Diagnostic Accuracy of the AI software in detecting normal spirometry: Area Under the Curve

Area	95 % confidence interval	
	Lower range	Upper range

0.959***	0.920	0.997

Table 11: Diagnostic Accuracy of the AI software vs Pulmonologists in detecting normal spirometry

Sensitivity	95.10%
Specificity	96.60%
PPV	95.10%
NPV	96.60%

When classifying normal spirometry, the AI tool achieved the highest performance metrics. The AUC was 0.959 (95% CI: 0.920–0.997) (Figure 4) (Table 10), with a sensitivity of 95.1%, specificity of 96.6%, PPV of 95.1%, and NPV of 96.6% (Table 11). Overall, the AI-based interpretation system showed high concordance with pulmonologist assessments across all categories, with particularly strong reliability in distinguishing normal and obstructive patterns.

DISCUSSION

This study highlights that artificial intelligence (AI)–based interpretation of pulmonary function tests (PFTs) provides high diagnostic accuracy and reliability when compared with conventional evaluation by pulmonologists. Automated systems effectively identified complex ventilatory patterns including obstructive, mixed obstructive, restrictive, and normal spirometry with strong sensitivity and specificity (3,5).

AI-supported interpretation demonstrated excellent ability to discriminate among spirometric categories. The area under the receiver operating characteristic (ROC) curve (AUC) ranged from 0.901 to 0.959 for mixed obstructive, pure obstructive, restrictive, and normal groups, reflecting strong predictive performance. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were consistently high. These outcomes are consistent with earlier reports, confirming that AI algorithms can achieve expert-level accuracy in clinical practice (3). Manual PFT reporting is prone to interobserver variability, which may compromise diagnostic confidence. In this study, AI reduced such variability by applying standardized algorithms and producing reproducible outputs. This consistency is particularly important in high-volume clinical practices and in settings with limited specialist availability, ensuring timely and uniform assessment (1,6).

Beyond diagnostic accuracy, AI enhances workflow efficiency. Automated interpretation can be integrated into electronic health record (EHR) systems, enabling rapid reporting and reducing turnaround time. Such integration provides broader access to specialist-level insights, especially in resource-constrained environments (7,8). Despite promising outcomes, limitations remain. The robustness of AI models depends on the quality and diversity of training datasets, which may introduce bias if certain populations are underrepresented. Broader multi-center validation and the adoption of explainable AI frameworks will further support clinician trust and safe integration of AI into healthcare systems. (9,10).

REFERENCES

1. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948–68.
2. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324–43.
3. Topalovic DB, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. Eur Respir J. 2019;53(4):1801660.
4. Attali D, Boussadi A, Bousquet C, Lamy JB. Explainable artificial intelligence in healthcare: state of the art and future perspectives. Artif Intell Med. 2021;121:102164
5. Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, et al. Standardization of spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. Am J Respir Crit Care Med. 2019;200(8):e70–88.
6. Pedersen JH, Madsen H, Jensen MT. Machine learning in respiratory medicine: new tricks for an old trade. Eur Clin Respir J. 2020;7(1):1763615.
7. Duong ML, Himes BE. Application of artificial intelligence to pulmonary function tests. Chest. 2021;159(5):2121–34.
8. Ruppel GL, Enright PL. Pulmonary function testing. Respir Care. 2012;57(1):165–75.
9. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31–8.
10. U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning in Software as a Medical Device. FDA; 2021.