CLINICAL VALIDATION OF SWAASA® ARTIFICIAL INTELLIGENCE PLATFORM USING COUGH SOUNDS FOR SCREENING AND DIAGNOSIS OF RESPIRATORY DISEASES

Authors: Balamugesh Thangakunam*1, Sujith Thomas Chandy *1, Gowrisree Rudraraju*2, Narayana Rao Sripada*2, Jayanthy Govindaraj*2, Charishma Gottipulla*2, Baswaraj Mamidgi*2, Shubha Deepti Palreddy*2, Nikhil kumar Reddy Bhoge*2, Harsha Vardhan Reddy Narreddy*2, Prasanna Samuel P*1, Devasahayam Jesudas Christopher*1, Manmohan Jain*2, Venkat Yechuri*2

*1 Christian Medical College, Vellore,Tamil Nadu, 632004, India
*2 Salcit Technologies, Kothaguda, Hyderabad, 500084, India

Abstract

Introduction- Analysis of cough sounds have the potential to give a clue regarding the underlying respiratory disease. The SaaS® platform using artificial intelligence technology can analyze the cough sounds.

Methods- Patients (age >= 18 years) presenting to our departments with respiratory symptoms were prospectively recruited into the study. Patients cough sound was recorded by a mobile device in which the Swaasa software was installed. Spirometry was done and patients' clinical diagnosis is taken as reference standard for comparison. Polychotomous variables were analyzed in binary form as normal and abnormal. Multi-class outcomes were defined as obstructive, restrictive, or normal spirometry, mild-moderate-severe impairment of spirometry values or disease conditions like asthma, COPD, ILD, Bronchiectasis.

Results- We recruited 2179 patients with respiratory diseases and 827 normal individuals. Sixty-two percent of them were males. The Swaasa platform has 90% accuracy in distinguishing cough from normal individuals and patients as well as distinguishing normal from abnormal spirometry. The performance was consistent across demographic parameters such as age, gender, and BMI. Its accuracy in differentiating into various severity subtypes of spirometry as well as disease types can be improved further.

Conclusion- Swaasa® commends an elevated level of precision in classifying normal and abnormal conditions across demographic subgroups indicating its promise as a screening tool. While the accuracy measures emphasize further refinement to ensure consistent high performance across disease scenarios and severity levels, the study shows Swaasa®'s reasonable potential in identification of specific respiratory diseases and their severity.

INTRODUCTION

Respiratory sounds, produced during coughing and breathing, offer valuable insights for diagnosing respiratory diseases. Advances in technology now allow the acoustic characteristics of these sounds to be easily and non-invasively recorded, enhancing their effectiveness as diagnostic tools (1).

Recent advances in digital devices have led to the development of sophisticated technologies for recording and analyzing cough sounds. These technologies have been evaluated for monitoring, diagnosing, and evaluating treatment responses in patients with airway disorders like asthma and COPD, as well as in those who had Covid-19 infection and presenting with respiratory symptoms (2-5). Cough sound devices can indicate and analyze cough intensity, frequency, and waveform characteristics. Automatic cough frequency measurements are often used as endpoints and surrogates for primary endpoints. In 2007, the European Respiratory Society guidelines recognized cough frequency measurements using cough sounds as valid markers of respiratory status (6).

Respiratory diseases encompass any issues in the respiratory system that may hinder normal lung function. In clinical practice, pulmonary function testing (PFT) provides crucial feedback on patient's lung status based on their pulmonary function parameters, aiding doctors in diagnosing respiratory diseases (7, 8). Spirometry is one of the most readily available and useful tests for pulmonary function. They are classified into three patterns based on spirometry: obstructive, restrictive, and combined (features of both obstructive and restrictive).

Spirometry measures the volume of air exhaled at specific time points during complete exhalation by force, which is preceded by a maximal inhalation. The test requires patients to place their mouth on the spirometer, inhale deeply, and exhale forcefully to expel all air quickly. This process must be completed with maximum effort and repeated until three consistent measurements are obtained (9). Patients are instructed and guided through the test procedure by trained Respiratory therapists. Reference values for 1-second forced expiratory flow (FEV1), forced vital capacity (FVC), and the FEV1/FVC ratio are estimated based on the patient's age, gender, height, and weight. The ratios of the measured values to the reference values, expressed as FEV1% and FVC%, indicate the severity of respiratory disease (10). FEV1/FVC and FVC% help distinguish between obstructive, restrictive, and normal respiratory patterns, while FEV1% determines the severity of obstructive diseases (11).

Although spirometry is a simple health screening test, it requires a patient's cooperation. Its sensitivity for diagnosing asthma related airway obstruction is only 29% (95% CI: 21-39), with a specificity of 90% (95% CI: 81-95) (12). Bonini et al. found that spirometry is poorly sensitive for detecting mild or intermittent asthma in athletes, as their spirometry values often remain close to 100% of predicted levels (13). Additionally, portable spirometers are inaccurate, leading to misclassification of obstructive patients (14). Only 60% of patients produced acceptable-quality tests, raising concerns about spirometry's utility in primary care without better quality assurance and training (15).

The reliability of spirometry tests is essential for their use as diagnostic instruments, both from the patient's perspective and for epidemiological purposes (16). Concerns about reliability have been raised, with Schermer et al. reporting a small but statistically significant difference in lung function values between spirometer performed in general practice and in a pulmonary function

laboratory (18). Additionally, Lusuardi et al. found that up to one-third of patients diagnosed with COPD using office-based spirometer in primary care had normal results when evaluated in a pulmonary function laboratory (19).

Another major limitation of spirometry is its reliance on forced expiratory maneuvers. The effort required often results in suboptimal performance in children and individuals with cognitive or physical impairments, complicating diagnosis, and management (20).

On the other hand, recording cough sounds requires minimal training unlike what is needed for spirometry, no physical contact, and is effective for disease assessment (21). Clinical research shows that cough severity indicates respiratory disease progression. Signal processing can extract diagnostic features from cough sounds, and studies have shown the importance of classifying these sounds to detect various respiratory illnesses (22). In this context, our previous studies have already proven strong correlation of cough sound characteristics with airflow characteristics including FEV1, FVC and their ratios, which are important in identifying the type of lung diseases through Swaasa AI (23). Following this, in the current study, we attempted to establish the high-level precision of the Swaasa AI in classifying normal and abnormal conditions across demographic subgroups, indicating it as a promise screening tool.

METHODS AND MATERIALS

Data collection: All patients included in the study were checked by a respiratory physician, followed by a spirometry and cough sound machine assessment sing mobile phones on the same day. This was done to ensure that there was no delay between the index and reference tests. Cough sounds were collected using the in-built mobile microphone. To record cough sounds, an in-built mobile microphone that is a part of the application was used. Five cough recordings were done per participant. Since there were noise corrupted coughs, invalid coughs and void recordings, an average of 3 recorded coughs from each participant were considered. A minimum gap of 10 seconds was maintained between consecutive cough samples.

Spirometry is done according to American Thoracic Society guidelines 2019 (24). Abnormal spirometry's were classified as obstruction or restriction based on FEV/FVC ratio. Those with isolated MMEF reductions were categorized separately. For calculation of accuracy measures, clinical diagnosis as concluded by the physician based on history, examination findings, spirometry and relevant radiological investigations was considered as the gold standard. The procedure was double blinded to avoid bias. i.e. neither the subjects, therapists nor the doctor(s), were provided with the Swaasa® report.

Sample size: For the sample size calculation, based on the assumption that the device could present a sensitivity and specificity of 80% with an error of 5% for 95% CI in the detection of respiratory conditions, and a prevalence of the same of 25%, a sample size of 990 was found to be adequate for validation. However, to ensure rigor and improve the accuracy of the estimates, a larger sample size of 1,300 was estimated for a sensitivity/specificity of 90%. Further, to account for patients with incomplete data and those unable to do spirometry, we aimed to include a total 1,800 participants in the study. Among the participants, half of them were designated normal controls while the remaining were planned to be evenly distributed among four disease

categories. (Asthma, COPD, ILD, Bronchiectasis). In the final analysis, a total of 1917 patients were available for validation.

Study design:

The study objective was to validate the Swaasa® AI platform in a real-world setting of patients diagnosed with respiratory conditions. This explorative study had a monocentric, statistical methodology framework. It was conducted on the patients of the Resiratory and Pulmonary medicine departments of the Christian Medical College (CMC), Vellore.

The main objectives of the statistical validation exercise were to
(1) estimate the accuracy of Swaasa® AI platform in detecting different respiratory diseases, (2) estimate the accuracy of Swaasa® AI platform in detecting underlying conditions, its pattern and severity,

The validation study included data from a prospective study, adopting a diagnostic accuracy study design. This process has been performed using 2 derivation plans. Plan 1 was employed to assess risk, pattern, severity, and spirometry values whereas plan 2 was applied for disease classification. These plans refer to the different ways in which the data was collected and organized for training and testing the AI model. In both the plans it was ensured that there was no overlap of subject data between the derivation and validation cohort.

Labelling of respiratory conditions from the collected samples were performed by clinicians and were used as gold standard or ground truth annotation to assess the accuracy of the AI platform. The accuracy of binary and multi-class disease outcomes was assessed using accuracy rate, kappa statistics and diagnostic accuracy measures (Sensitivity & Specificity). And intra-class correlation coefficients were used to measure the reliability of spirometry values.

The analysis is based on two distinct plans, Plan1 and Plan2, each designed to validate specific models.

Plan 1: The data from Plan 1 was utilized to validate models related to Risk, Spirometry Pattern, Pattern Severity. Among 1917 subjects, 450 were selected for derivation, and this dataset was incorporated into the existing training dataset. The remaining 1467 subjects are dedicated to the validation of the models mentioned above. (Plan 1 is used to predict Risk, Spirometry Pattern, Pattern Severity)

Plan 2: Plan 2 was focused on validating disease models. From the pool of 1917 subjects, 515 were chosen for derivation, and this dataset was integrated into the existing training dataset. The remaining 1402 subjects were employed for validating the performance of the disease models. (Plan 2 is used to predict Disease only)

RESULTS

Participants Information

From every participant three cough audio recordings were collected. In plan1 from 1467 subjects 3006 cough recordings were collected and in plan2 from 1402 subjects 2832 cough recordings were collected. Table 1 details the distribution of clinical and demographic variables of the subjects. In this study, demographic and clinical variables and other variables of interest were summarized using frequency tables.

Table 1: Distribution of clinical and demographic variables

	Data available for validation	
Variables	**n (%)** **N=3006**	**n (%)** **N=2832**
Age group		
18 – 47 years	1636 (54.4)	1559 (55.0)
48 – 63 years	934 (31.1)	877 (31.0)
64 – 89 years	436 (14.5)	396 (14.0)
Sex		
Male	1862 (61.9)	1730 (61.1)
Female	1144 (38.1)	1102 (38.9)
BMI		
Underweight (<18.5)	155 (5.2)	129 (4.6)
Normal (18.5 – 22.9)	424 (14.1)	385 (13.6)
Overweight (23 - 24.9)	228 (7.6)	218 (7.7)
Obese (>=25)	472 (15.7)	415 (14.6)
Missing	1727 (57.4)	1685 (59.5)

Comparison of swaasa results with the ground truths

To evaluate the diagnostic accuracy of Swaasa® as a screening tool, confusion matrix statistics were employed. These statistics included accuracy, Kappa statistic, sensitivity, and specificity. Accuracy, denoting the proportion of correct predictions out of all predictions made, serves as a fundamental measure of overall model performance. The Kappa statistics assess the agreement level between the predictive model and observed data. Sensitivity, or the true positive rate,

quantifies Swaasa®'s ability to correctly identify true positive cases. Specificity, or the true negative rate, measures the model's capacity to accurately identify true negative cases (controls). These metrics collectively provide a comprehensive understanding of the model's proficiency in classifying outcomes.

Swaasa AI platform predicts Risk, Spirometry Pattern, Pattern Severity and Disease type from cough audio recording.

Risk: It refers to the detection of the presence of an underlying respiratory problem. If any abnormalities are detected in a subject, Swaasa classifies the Risk as "Abnormal"; otherwise, it categorizes them as "Normal."

Spirometry Pattern: The term pattern refers to the detected respiratory disease pattern. Swaasa classifies patterns as "Obstructive," "Restrictive," or "Normal" based on the features extracted from the input data. Those with MMEF reduction alone were categorized separately.

Pattern Severity: Pattern severity refers to the assessment of the severity of respiratory disease patterns mentioned above. The severity categories include "Normal", "Mild", "Moderate," and "Severe." The spirometry report values are used as the reference to generate ground truth data and to build the severity model. When the spirometry report says normal, the ground truth of the patient is considered to be Normal. The severity categories are mapped to the FEV1% predicted values as: Mild (>70), Moderate (70-50), Severe (<50) (25).

Disease Model: The Disease model is focused on detecting various respiratory conditions. Swaasa outputs diseases as "Asthma," "COPD," "ILD," "Bronchiectasis," and "Normal." In the case of Abnormal vs. Normal detection, all diseases, including multi-labelled values from the ground truth, are considered abnormal.

Table 2 and table 3 illustrate the outcomes of binary and multi-class classification respectively

Table 2: Frequencies for the Ground Truth and Swaasa® risk, Spirometry pattern variables with binary classification

Outcomes	Ground Truth n (%)	Swaasa® n (%)
Risk		
Normal	827 (27.5)	835 (27.8)
Abnormal	2179 (72.5)	2171 (72.2)
Total	3006 (100.0)	3006 (100.0)
Spirometry Pattern		
Normal	1486 (49.4)	1377 (45.8)
Abnormal	1214 (40.4)	1629 (54.2)
Missing	306 (10.2)	NA
Total	3006 (100.0)	3006 (100.0)

Table 3: Frequencies for the Ground Truth and Swaasa® risk, pattern, severity and disease variables with multiple classes

Outcomes	Ground Truth n (%)	Swaasa® n (%)
Spirometry Pattern		
Normal	746 (24.8)	1377 (45.8)
Obstructive	681 (22.7)	1098 (36.5)
Restrictive	533 (17.7)	531 (17.7)
MMEF	740 (24.6)	NA
Missing	306 (10.2)	NA
Total	3006 (100.0)	3006 (100.0)
Pattern Severity		
Normal	786 (26.2)	846 (28.1)
Mild	1058 (35.2)	704 (23.4)
Moderate	482 (16.0)	630 (21.0)
Severe	374 (12.4)	826 (27.5)
Missing	306 (10.2)	NA
Total	3006 (100.0)	3006 (100.0)
Diagnosis Disease Type		
Normal	852 (30.1)	857 (30.3)
Asthma	510 (18.0)	746 (26.3)
Bronchiectasis	80 (2.8)	286 (10.1)
COPD	252 (9.0)	577 (20.4)
ILD	156 (5.5)	365 (12.9)
Multi-labelled	845 (29.8)	NA
Missing	137 (4.8)	1 (0.0)
Total	2832 (100.0)	2832 (100.0)

Performance Metrics of the Swaasa AI Models

Confusion matrix statistics for risk, pattern, severity and disease with binary classification for complete data is in Table 4, at gender level is in table 5, and at sub age level is in table 6.

Table 4: Confusion matrix statistics for disease identification and spirometry abnormality with binary classification (differentiating normal and abnormal)

Parameters	Accuracy (%)	Kappa Statistic (%)	Sensitivity (%)	Specificity (%)
Risk	89.1	72.7	92.3	80.7
Spirometry Pattern	79.0	58.2	85.3	73.8
Pattern Severity	76.0	42.6	82.2	60.9
Disease Type	90.0	76.6	93.4	82.6

Table 5: Confusion matrix statistics for risk, pattern, severity and disease with binary classification at gender level

Parameters	Accuracy (%)		Kappa Statistic (%)	
	Male	Female	Male	Female
Risk	88.2	90.6	68.6	78.2
Spirometry Pattern	80.5	76.5	61.2	52.6

Table 6: Confusion matrix statistics for risk, pattern, severity and disease with binary classification at age level

Parameters	Accuracy (%)			Kappa Statistic (%)		
	18-47 yrs	48-63 yrs	64-89 yrs	18-47 yrs	48-63 yrs	64-89 yrs
Risk	85.6	92.2	95.6	69.5	73.2	73.0
Spirometry Pattern	77.2	79.8	84.5	52.3	59.2	61.2

Confusion matrix statistics for risk, pattern, severity and disease with multi-class variables for complete data is in Table 7.

Table 7: Confusion matrix statistics for spirometry pattern, spirometry severity and disease with multi-class variables

Parameters	Accuracy (%)	Kappa Statistic (%)	Sensitivity (%)	Specificity (%)
Spirometry Pattern				
Normal	79.0	58.2	73.8	85.3
Obstructive	79.9	53.3	81.2	79.5
Restrictive	88.4	61.2	63.4	94.5
Pattern Severity				
Normal	76.0	42.6	60.9	82.2
Mild	78.9	52.7	55.4	94.0
Moderate	84.5	48.1	58.7	90.1
Severe	83.5	49.1	83.2	83.5
Diagnosis Disease Type				
Normal	89.1	77.9	82.6	94.7
Asthma	83.0	57.2	68.2	88.6
Bronchiectasis	93.1	36.2	52.5	94.9
COPD	86.3	46.8	61.1	90.2
ILD	92.1	47.8	51.3	95.8

References

1. Kapetanidis P, Kalioras F, Tsakonas C, Tzamalis P, Kontogiannis G, Karamanidou T, Stavropoulos TG, Nikoletseas S. Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review. Sensors (Basel). 2024;24(4):1173.

2. Umayahara Y., Soh Z., Sekikawa K., Kawae T., Otsuka A., Tsuji T. Clinical Significance of Cough Peak Flow and Its Non-Contact Measurement via Cough Sounds: A Narrative Review. Appl. Sci. 2020;10:2782.

3. Mouawad P., Dubnov T., Dubnov S. Robust Detection of COVID-19 in Cough Sounds: Using Recurrence Dynamics and Variable Markov Model. SN Comput. Sci. 2021;2:34

4. Rohmetra H., Raghunath N., Narang P., Chamola V., Guizani M., Lakkaniga N.R. AI-enabled remote monitoring of vital signs for COVID-19: Methods, prospects and challenges. Computing. 2021:1–27.

5. Laguarta J., Hueto F., Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J. Eng. Med. Biol. 2020;1:275–281.

6. Morice A.H., Fontana G.A., Belvisi M.G., Birring S.S., Chung K.F., Dicpinigaitis P.V., Kastelik J.A., McGarvey L.P., Smith J.A., Tatar M., et al. ERS guidelines on the assessment of cough. Eur. Respir. J. 2007;29:1256–1276.

7. Ponce MC, Sankari A, Sharma S. Pulmonary Function Tests. [Updated 2023 Aug 28]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482339/

8. Dempsey TM, Scanlon PD. Pulmonary Function Tests for the Generalist: A Brief Review. Mayo Clin Proc. 2018;93(6):763-771.

9. Lamb K, Theodore D, Bhutta BS. Spirometry. 2023 Aug 17. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 32809361.

10. Torén K, Schiöler L, Lindberg A, Andersson A, et al., The ratio FEV1 /FVC and its association to respiratory symptoms-A Swedish general population study. Clin Physiol Funct Imaging. 2021;41(2):181-191.

11. David S, Edwards CW. Forced Expiratory Volume. [Updated 2022 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK540970/

12. Meneghini AC, Paulino ACB, Pereira LP, Vianna EO. Accuracy of spirometry for detection of asthma: a cross-sectional study. Sao Paulo Med J. 2017;135(5):428-433.

13. Bonini M, Lapucci G, Petrelli G, et al. Predictive value of allergy and pulmonary function tests for the diagnosis of asthma in elite athletes. Allergy. 2007;62(10):1166-70.

14. Hegewald MJ, Gallo HM, Wilson EL. Accuracy and Quality of Spirometry in Primary Care Offices. Ann Am Thorac Soc. 2016;13(12):2119-2124.

15. Kaur R, Kumar A, Hadda V, Kalaivani M, Nongkynrih B, Kant S, Gupta SK, Vignesh D. Quality of spirometry testing in a community setting: A study among elderly persons in a rural area of Haryana. J Family Med Prim Care. 2019;8(11):3718-3725.

16. Soumagne T, Guillien A, Roux P, Laplante JJ, Botebol M, Laurent L, Roche N, Dalphin JC, Degano B. Quantitative and qualitative evaluation of spirometry for COPD screening in general practice. Respir Med Res. 2020;77:31-36.

17. van de Hei SJ, Flokstra-de Blok BMJ, Baretta HJ, Doornewaard NE, van der Molen T, Patberg KW, Ruberg ECM, Schermer TRJ, Steenbruggen I, van den Berg JWK, Kocks JWH. Quality of spirometry and related diagnosis in primary care with a focus on clinical use. NPJ Prim Care Respir Med. 2020;30(1):22

18. Schermer TR, Jacobs JE, Chavannes NH, Hartman J, Folgering HT, Bottema BJ, van Weel C. Validity of spirometric testing in a general practice population of patients with chronic obstructive pulmonary disease (COPD). Thorax. 2003; 58(10):861-866

19. Donner CF. A randomized controlled trial on office spirometry in asthma and COPD in standard general practice: data from spirometry in Asthma and COPD: a comparative evaluation Italian study. Chest. 2006;129(4):844-852.

20. Kouri A, Dandurand RJ, Usmani OS, Chow CW. Exploring the 175-year history of spirometry and the vital lessons it can teach us today. Eur Respir Rev. 2021 Oct 5;30(162):210081.

21. Kang HS, Lee EG, Kim CK, Jung A, Song C, Im S. Cough Sounds Recorded via Smart Devices as Useful Non-Invasive Digital Biomarkers of Aspiration Risk: A Case Report. Sensors (Basel). 2021;21(23):8056.

22. Alqudaihi KS, Aslam N, Khan IU, Almuhaideb AM, Alsunaidi SJ, Ibrahim NMAR, Alhaidari FA, Shaikh FS, Alsenbel YM, Alalharith DM, Alharthi HM, Alghamdi WM, Alshahrani MS. Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities. IEEE Access. 2021;9:102327-102344.

23. Gowrisree Rudraraju, ShubhaDeepti Palreddy, Baswaraj Mamidgi, Narayana Rao Sripada, Y. Padma Sai, Naveen Kumar Vodnala, Sai Praveen Haranath,

Cough sound analysis and objective correlation with spirometry and clinical diagnosis,

Informatics in Medicine Unlocked. 2020. 19:100319,

24. BrianL.Graham et al., Standardization of Spirometry 2019 Update, Am J Respir Crit Care Med Vol 200, Iss8, ppe70–e88 ,Oct15, 2019. DOI:10.1164/rccm.201908-1590ST.

25. R. Pellegrino et al., Interpretative strategies for lung function tests, European Respiratory Journal 2005 26(5): 948-968; DOI: https://doi.org/10.1183/09031936.05.00035205