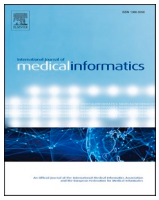
International Journal of Medical Informatics 199 (2025) 105912

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Review article

Wet and dry cough classification using cough sound characteristics and machine learning: A systematic review

Roneel V. Sharan [a,*], Hao Xiong [b]

[a] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom
[b] Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia

ARTICLE INFO

Keywords:
Cough sound
Feature extraction
Machine learning
Respiratory diseases
Wet cough

ABSTRACT

Background: Distinguishing between productive (wet) and non-productive (dry) cough types is important for evaluating respiratory health, assisting in differential diagnosis, and monitoring disease progression. However, assessing cough type through the perception of cough sounds in clinical settings poses challenges due to its subjectivity. Employing objective cough sound analysis holds promise for aiding diagnostic assessments and guiding the management of respiratory conditions. This systematic review aims to assess and summarize the predictive capabilities of machine learning algorithms in analyzing cough sounds to determine cough type.
Method: A systematic search of the Scopus, Medline, and Embase databases conducted on March 8, 2025, yielded three studies that met the inclusion criteria. The quality assessment of these studies was conducted using the checklist for the assessment of medical artificial intelligence (ChAMAI).
Results: The inter-rater agreement for annotating wet and dry coughs ranged from 0.22 to 0.81 across the three studies. Furthermore, these studies employed diverse inputs for their machine learning algorithms, including different cough sound features and time–frequency representations. The algorithms used ranged from conventional classifiers like logistic regression to neural networks. While the classification accuracy for identifying wet and dry coughs ranged from 78% to 87% across these studies, none of them assessed their algorithms through external validation.
Conclusion: The high variability in inter-rater agreement highlights the subjectivity in manually interpreting cough sounds and underscores the need for objective cough sound analysis methods. The predictive ability of cough-type classification algorithms shows promise in the small number of studies analyzed in this systematic review. However, more studies are needed, particularly those validating their models on independent and external datasets.

1. Introduction

Cough is a prevalent symptom of respiratory diseases and a common presenting condition in primary care settings globally [1]. It serves as a reflex mechanism to expel irritants from the respiratory system, with cough types broadly classified as productive (wet) or non-productive (dry). Wet coughs typically involve sputum production and may arise from infections, inflammation, or chronic conditions such as chronic obstructive pulmonary disease (COPD), whereas dry coughs can result from asthma or follow respiratory infections [2–4]. A dry cough has also been reported in the majority of COVID-19 patients [5]. In certain illnesses, the cough may initially present as dry but evolve into a phlegmy or wet cough as the disease progresses and airway secretions increase [6].

Differentiating between cough types is important for diagnosing and monitoring respiratory diseases, understanding disease progression, and guiding treatment decisions [7]. For instance, studies have shown that recognizing wet cough characteristics in COPD patients can help identify exacerbations early, potentially preventing hospitalization [8]. Similarly, distinguishing dry cough in asthma can aid in assessing disease progression and the efficacy of anti-inflammatory therapies [9]. However, subjective assessment methods in clinical practice, which rely on patient reporting or clinicians' perception of sounds associated with airway secretions [10], have inherent limitations. Although bronchoscopy [11] offers an alternative for evaluating airway secretions, it is invasive.

* Corresponding author.
E-mail address: roneel.sharan@essex.ac.uk (R.V. Sharan).

https://doi.org/10.1016/j.ijmedinf.2025.105912
Received 25 April 2024; Received in revised form 10 March 2025; Accepted 3 April 2025
Available online 6 April 2025

Developing objective assessment techniques is therefore important to overcome the limitations of subjective evaluations and to ensure consistent and reliable classification of cough types across various healthcare settings. These methods aim to categorize cough types based on sound characteristics, potentially enabling more accurate diagnosis and treatment planning. Despite the potential significance of objectively evaluating cough types in clinical decision-making, there is a lack of comprehensive evidence syntheses on this topic. Consequently, we undertook a systematic review to assess the capability of machine learning algorithms in predicting wet and dry cough types based on cough sound characteristics.

2. Methods

This systematic review adhered to the guidelines outlined by the preferred reporting items for systematic reviews and meta-analyses (PRISMA) [12].

2.1. Search strategy

We conducted a systematic literature search across the Scopus, Medline (Ovid), and Embase (Ovid) databases on March 8, 2025. The search utilized the following terms across all three databases: ("wet" OR "dry" OR "productive") AND ("cough") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "feature extraction" OR "accuracy" OR "classification"). Searches were performed in English, targeting titles, abstracts, and keywords. To maximize results, no restrictions were applied to publication date or study location, and the search terms were intentionally broad. Gray literature was not included in this systematic review.

2.2. Inclusion and exclusion criteria

This systematic review included studies that employed cough sound features and machine learning algorithms to differentiate between wet and dry cough types. To be eligible, studies were required to incorporate cross-validation, training and testing sets, or testing sets at a minimum. This criterion was established to ensure that models lacking resampling procedures, which may fail to generalize effectively to independent datasets, were excluded. Studies with small sample sizes ($\leq$30) were also excluded due to the risk of overfitting, which can result in inflated and highly variable predictive outcomes [13].

No restrictions were placed on the age or gender of the study populations; however, studies that failed to report demographic data were excluded. Additionally, studies without sufficient information to calculate sensitivity, specificity, or confidence intervals were excluded. Abstracts and conference proceedings were included only if they provided adequate data directly or through associated publications. If a preliminary study was succeeded by a more comprehensive version, only the latter was included in the review.

2.3. Study selection

Search results were imported into EndNote X9, and duplicate entries were removed. The remaining records were transferred to Microsoft Excel for screening. R.V.S. and H.X. independently reviewed titles and abstracts to assess eligibility. Studies deemed eligible by either reviewer were subjected to full-text screening. Reasons for excluding studies after full-text assessment were documented in Excel. Discrepancies between reviewers were resolved through discussion, with consensus achieved for each study.

2.4. Data extraction and analysis

R.V.S. extracted methodological, demographic, and outcome data from the included studies, which were then reviewed by H.X. Any discrepancies were resolved through discussion. Extracted information included study characteristics (e.g., first author, publication year, country, study design, sample size, feature extraction methods, feature selection techniques, machine learning classifiers, and cough segmentation methods), population demographics (age, gender), respiratory disease diagnosis, inter-rater agreement, and performance metrics (sensitivity, specificity, accuracy, and their corresponding 95% confidence intervals).

For studies employing multiple feature extraction, feature selection, or machine learning classifiers, all relevant data were extracted. When accuracy and confidence intervals were not explicitly reported, they were calculated using available data. Accuracy was derived by dividing the number of correct classifications by the overall sample size. Confidence intervals were calculated using Newcombe's method [14]. Performance metrics were rounded to the nearest whole percentage for consistency.

Meta-analysis was not conducted due to the limited number of studies within each population group.

2.5. Quality assessment

Quality assessments were conducted by R.V.S. using the checklist for the assessment of medical AI (ChAMAI) [15], which were subsequently reviewed by H.X. Any discrepancies were resolved through discussion. The ChAMAI checklist, developed by the IJMEDI, evaluates the quality of medical artificial intelligence studies, distinguishing between high-quality machine learning research and basic data-mining studies.

The checklist consists of six dimensions: problem understanding, data understanding, data preparation, modeling, validation, and deployment, encompassing a total of 30 questions. Each question was rated as OK (adequately addressed), mR (sufficient but improvable, minor revision needed), or MR (inadequately addressed, major revision needed). Scores were assigned as follows: 2, 1, and 0 for high-priority questions and halved for low-priority questions. The maximum possible score was 50, with quality classified as low (0–19.5), medium (20–34.5), or high (35–50) [16,17]. All eligible studies were included for analysis regardless of their quality assessment outcomes.

3. Results

3.1. Search results

The process of identifying eligible studies is depicted in the PRISMA flow diagram shown in Fig. 1. Searches conducted in Scopus, Medline (Ovid), and Embase (Ovid) yielded 639, 179, and 823 results, respectively, totaling 1641 studies. After removing 551 duplicate entries, 1090 studies underwent title and abstract screening. During this phase, 1058 studies were excluded, leaving 32 studies for full-text review. Subsequently, 29 studies were excluded after full-text screening, resulting in 3 studies meeting the inclusion criteria. These studies underwent quality assessment and were included in both qualitative synthesis and quantitative analysis.

3.2. Study characteristics

Table 1 provides an overview of the included studies [18–20] and their characteristics. Data collection for one study occurred in Indonesia [18], another in the United States [19], while the third study [20] involved crowdsourced data from multiple countries. One study [18] focused on the pediatric population (ages 0–15 years), while the others targeted adolescents and adults (ages 18–68 years [19] and 14–60 years [20]). The sample sizes were 78 [18], 131 [19], and 88 [20] participants, respectively.

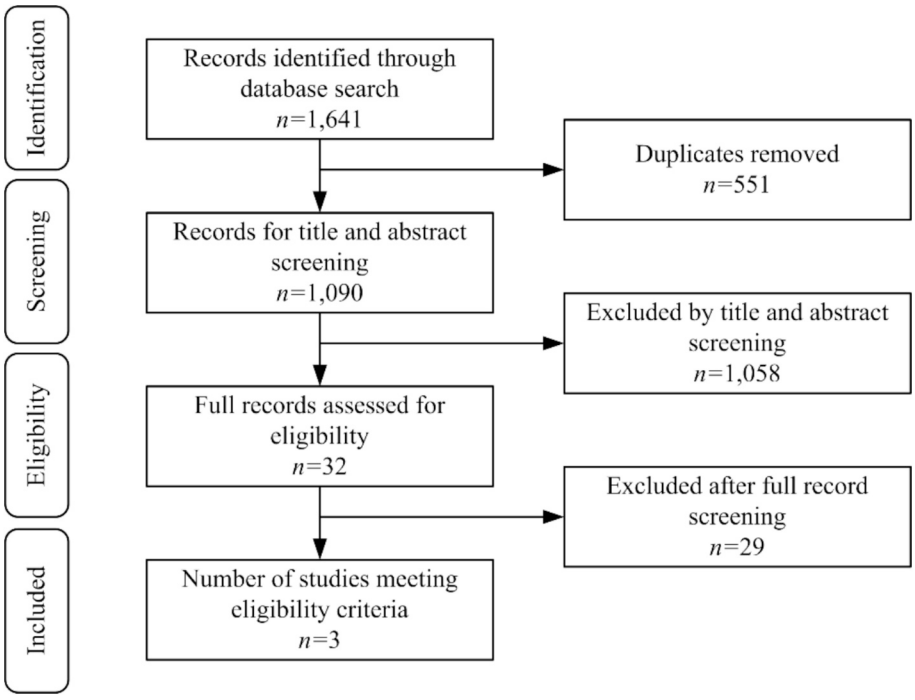


Fig. 1. PRISMA flow diagram of included studies using cough sound features and machine learning to differentiate between wet and dry cough.

Table 1
Overview of the studies included in the systematic review.

Study Reference	Country	No. of subjects (coughs)	Age Range (years)	Respiratory Disease	Method	Cough Segmentation
Swarnkar (2013) [18]	Indonesia	78 (536)	0–15	Pneumonia Bronchitis Asthma Rhinopharyngitis Others	Number of annotators: 2 Cough features: BGS, NGS, FF, P, LogE, ZCR, Kurt, MFCC Feature selection: *p*-value Classifier: LR	Manual
Nemati (2020) [19]	United States	131 (5971)	18–68	Asthma COPD Healthy	Number of annotators: 2 Cough features: OpenSmile toolbox, custom features Feature selection: correlation Classifier: RF	Manual
Sharan (2022) [20]	Multiple (crowdsourced)	88 (396)	14–60	COVID-19 Symptomatic Healthy Unknown	Number of annotators: 4 Cough features: cochleagram Classifier: CNN	Automatic

BGS – bispectrum score, CNN – convolutional neural network, FF – formant frequencies, Kurt – kurtosis, LogE – log energy, LR – logistic regression, MFCC – mel-frequency cepstral coefficients, NGS – non-Gaussianity score, P – pitch, RF – random forest, ZCR – zero-crossing rate.

3.3. Data recording

One study [18] recorded cough sounds using bedside non-contact microphones (Rode NT3) in a hospital setting. In contrast, the remaining studies used smartphones for data collection. For one study [20], the data was crowdsourced, implying varied recording environments. The recording environment for the second smartphone-based study [19] was not specified.

3.4. Respiratory disease

In the pediatric study [18], cough types were associated with conditions such as pneumonia, bronchitis, asthma, rhinopharyngitis, and other respiratory diseases. In the adolescent and adult studies [19,20], coughs were linked to asthma, COPD, COVID-19, symptomatic COVID-19, and healthy cases.

3.5. Annotation

Subjects contributed one or more cough samples. Two studies [18,20] used clinical experts to annotate cough types based on auditory perception. In [18], each cough was labeled as wet or dry by two scorers, whereas in [20], each recording was labeled by up to four scorers.

The third study [19] utilized crowdsourced annotations in a two-stage process: initial labeling by two scorers, followed by resolution of disagreements and wet cough confirmations by an additional four scorers. The scorers' clinical expertise was unclear.

3.6. Cough data

Only one study [20] employed automated segmentation for cough sounds, using the signal envelope to define start and end points [21], and classifying events via cochleagram representations with convolutional neural networks (CNNs). The other two studies relied on manual

segmentation. The cough sample sizes were 536 [18], 5971 [19], and 396 [20].

3.7. Cough features

Mel-frequency cepstral coefficients (MFCCs) were experimented with in all studies. In [18], 66 features, including MFCCs, bispectrum scores, non-Gaussianity scores, formant frequencies, log energy, zero-crossing rate, and kurtosis, were analyzed. Study [19] extracted 1597 features using the OpenSmile toolkit [22] and custom features. Study [20] experimented with multiple feature sets but primarily utilized cochleagram analysis.

3.8. Classifiers

Logistic regression (LR) was experimented with in all studies. While [18] exclusively relied on LR, the other studies [19,20] also employed random forests and support vector machines. Study [20] further utilized shallow CNNs for analyzing time–frequency characteristics, mitigating overfitting due to the small dataset.

3.9. Feature selection

Each study applied distinct feature selection methods. Study [18] used *p*-values of LR coefficient estimates, while [19] applied correlation analysis. Study [20] directly used cochleagram representations with CNNs, bypassing feature selection, although *t*-tests were used for baseline methods.

3.10. Data balancing

To address class imbalance, study [19] used undersampling to equalize class sizes. Study [20] employed the synthetic minority oversampling technique (SMOTE) [23] to augment the minority class.

3.11. Cough and subject classification

All studies performed binary classification, distinguishing between wet and dry coughs. Study [20] labeled entire recordings, assigning all coughs within a recording the same classification. The prediction scores of individual coughs were averaged to classify recordings.

3.12. Inter-rater agreement and diagnostic performance

Table 2 summarizes the annotation and evaluation results. Inter-rater agreement, measured using the kappa statistic, ranged from 0.26 to 0.81 across two scorers and 0.22–0.38 for three scorers [20]. Agreement among four scorers was 0.37 [19] and 0.24 [20].

Validation methods varied: two studies [19,20] used k-fold cross-validation, while [18] split the dataset into training, validation, and test sets. Sensitivity ranged from 84–100%, specificity from 76–86%, and accuracy from 78–87%.

3.13. Study quality

Detailed quality assessment results are in the Supplementary File. Table 3 summarizes individual study scores: one study [18] scored 39, another [19] scored 35.5, and the third [20] scored 34.5. On average, the studies achieved a score of 36.3, indicating two were of high quality and one of medium quality. Fig. 2 illustrates the distribution of responses across high- and low-priority items.

4. Discussion

This paper outlines the methodology used to conduct a systematic review of research investigating the detection of wet and dry coughs through analysis of cough sound characteristics and machine learning techniques. It also highlights important trends observed during the analysis of the included studies. Early investigations utilized traditional sound recording setups, while recent studies have adopted smartphones for recording, coupled with crowdsourced data collection and annotation methods. Furthermore, there has been a shift from conventional feature engineering and machine learning approaches in earlier studies to the use of neural networks in the most recent study, which demonstrated comparable performance even with relatively small datasets.

Table 2
Summary of the results for the included studies.

Study Reference	No. of subjects (coughs)		Inter-rater agreement (Kappa)	Performance [95% CI] (%)
	Training/ Validation	Test		
Swarnkar (2013) [18]	60 (310)	18 (117)	0.55	Sensitivity = 84 [67–93] Specificity = 76 [66–83] Accuracy = 78 [69–84]
Nemati (2020) [19]	131 (5971)	–	0.81 (First) 0.37 (Second)	Sensitivity = 88 [86–90] Specificity = 86 [85–87] Accuracy = 87 [86–87]
Sharan (2022) [20]	88 (396)	–	0.26–0.59 (Two) 0.22–0.38 (Three) 0.24 (Four)	Sensitivity = 100 [68–100] Specificity = 83 [73–89] Accuracy = 84 [75–90]

Across all included studies, we observed significant variance in inter-rater agreement for the manual annotation of wet and dry cough types. This highlights the subjectivity in interpreting wet and dry cough sounds and emphasizes the need for objective methods of cough type analysis.

The limited number of studies included in this systematic review poses challenges in discerning long-term trends in quality assessment. However, the overall quality scores of the studies mostly fell within the upper bounds of medium quality or were rated as high quality. Factors contributing to reduced scores included cases where index test results were reported only during cross-validation, without a distinct test set. While all studies adequately addressed (OK) 70% or more of the high-priority requirements, only 40% or fewer of the low-priority requirements were adequately addressed, with 40% of these requiring major revisions (MR). Additionally, deployment performance was notably deficient across all studies, with scores not exceeding 31.25% of a maximum possible score of 8. Moreover, none of the studies underwent external validation, such as testing on data from different healthcare settings.

The choice of feature selection methods impacted the performance of machine learning models. For the LR classifier, selecting features based on *p*-values yielded moderate performance (sensitivity = 84%, specificity = 76%) [18], reflecting the method's effectiveness in identifying statistically significant predictors. In contrast, using feature correlation improved performance in another study (sensitivity = 88%, specificity = 86%) [19], suggesting that prioritizing independent, less correlated features enhances model accuracy. Meanwhile, the cochleagram input fed directly into the CNN achieved the highest sensitivity (100%) but slightly lower specificity (83%) [20], demonstrating CNNs' ability to handle raw, complex data without explicit feature selection. This comparison underscores the importance of aligning feature selection techniques with model architecture and data characteristics to optimize outcomes.

Among the included studies, only one [20] utilized neural network classification, despite having the smallest subject sample size. This study

Table 3
Quality assessment scores based on the ChAMAI checklist.

Study Reference	Problem Understanding (10)	Data Understanding (6)	Data Preparation (8)	Modeling (6)	Validation (12)	Deployment (8)	Total (50)
Swarnkar (2013) [18]	10	6	7	6	7.5	2.5	39
Nemati (2020) [19]	5.5	6	8	6	7.5	2.5	35.5
Sharan (2023) [20]	5.5	5	8	6	7.5	2.5	34.5

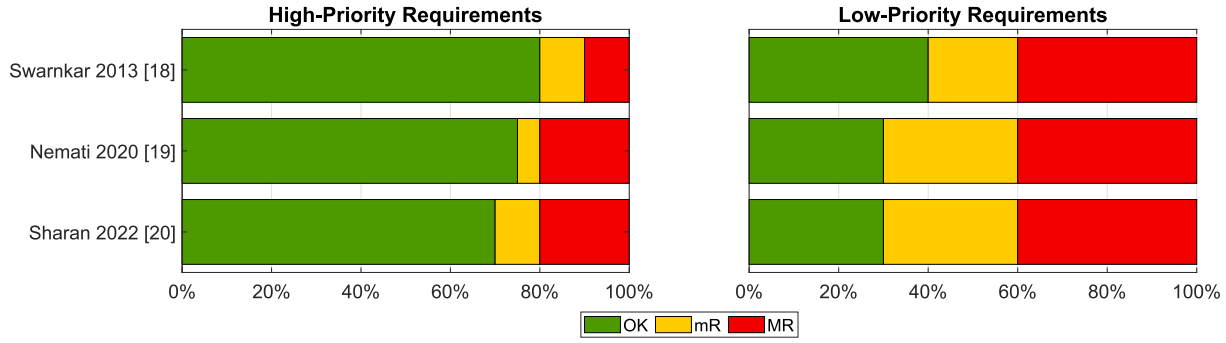


Fig. 2. Percentage of OK (adequately addressed), mR (sufficient but improvable, minor revision needed), and MR (inadequately addressed, major revision needed) within both high-priority and low-priority requirements.

relied on data augmentation and employed a shallow CNN to mitigate overfitting. Advanced pretrained CNN architectures designed for audio classification, such as YAMNet and VGGish [24], show promise for improved performance. Moreover, recent advancements in learning directly from one-dimensional raw audio signals, such as SincNet [25], may further enhance performance. Although neural networks introduce complexity, they provide the potential for effective feature engineering and learning without extensive signal transformation, enabling the capture of nonlinear characteristics.

This systematic review has several limitations. Only three studies met the inclusion criteria, precluding the possibility of conducting a meta-analysis due to the limited number of studies and the diversity among population groups. Meta-analysis is generally discouraged for such small datasets because it increases the risk of statistical inaccuracies and reduces robustness. Small sample sizes tend to produce wide confidence intervals, diminishing the reliability of effect estimates [26]. Furthermore, the inclusion of only a few studies increases the potential for publication or selective reporting bias, potentially skewing results toward individual study outcomes rather than providing an accurate summary of the evidence [27].

Additionally, the heterogeneity in study populations and methodologies further complicates aggregation, making it inappropriate to synthesize results quantitatively through meta-analysis [28]. The three studies represent a mixture of pediatric (1 study) and adult (2 studies) populations, highlighting challenges in combining these datasets. In clinical practice, respiratory diseases in children and adults are often studied and managed separately due to key physiological and clinical differences. For example, children are more prone to conditions like bronchiolitis, while adults are more likely to develop chronic diseases such as COPD. Differences in airway structure, immune response, and disease progression necessitate distinct diagnostic and treatment approaches for these age groups [29]. Combining pediatric and adult data into a single meta-analysis could obscure these differences and potentially lead to misleading conclusions.

The studies analyzed primarily focus on cough types associated with a limited range of respiratory diseases, such as pneumonia, asthma, COPD, and COVID-19. The scarcity of relevant studies highlights the nascent stage of research into cough type detection using sound analysis, emphasizing the need for further exploration in this area, particularly in light of the COVID-19 pandemic. Another limitation across the included studies is the lack of external validation, a challenge commonly encountered in AI applications in healthcare [30].

Despite these limitations, the studies provide important insights that can inform future research designs, emphasizing the importance of data standardization [31] and the reduction of subjectivity in ground truthing [32]. The insights derived from this systematic review, covering aspects such as data acquisition, annotation, cough sound features, time–frequency representation, and machine learning techniques, offer a valuable foundation for future studies.

Advancing the practical application of cough sound analysis algorithms is important for driving research in this area. The collection of cough audio data is relatively straightforward, as many devices, both in clinical and non-clinical settings, including personal smartphones, can be utilized for this purpose. Demonstrating the real-world effectiveness of these algorithms could stimulate further research and deeper exploration in this domain. For instance, studies have developed smartphone-based algorithms for continuous cough monitoring in hospital wards, showing high sensitivity and specificity in detecting coughs [33]. Additionally, smartphone-based cough detection algorithms have been validated in pediatric populations, demonstrating strong correlations between automated and manual cough counts [34]. Furthermore, smartphone-based algorithms have been developed to identify acute exacerbations of asthma by analyzing cough sounds and patient-reported symptoms, achieving high agreement with clinical diagnoses [35]. These examples emphasize the feasibility and potential of implementing machine learning algorithms in real-world settings, thereby encouraging further research and application in this field.

Also, the integration of AI in medical diagnostics necessitates adherence to regulatory frameworks that ensure safety and efficacy. Notably, the European In Vitro Diagnostic Regulation (IVDR) explicitly includes software within its scope, presenting challenges for in vitro diagnostic devices that utilize machine learning algorithms for data analysis and decision support. This inclusion highlights the need for explainable AI methods that empower biomedical professionals to take responsibility for their decisions.

Explainable AI techniques, such as layer-wise relevance propagation, can interpret the specific input features influencing a model's output, thereby enhancing transparency and trustworthiness [36]. The concept of causability extends this by providing metrics to assess the quality of explanations produced by AI systems [36]. Incorporating these

methodologies is important to demonstrate scientific validity, as well as analytical and clinical performance, for AI-based in vitro diagnostic devices. This aligns with the IVDR's requirements and supports the development of trustworthy AI in medical diagnostics.

In conclusion, the small number of studies included in this review underscores the early stage of research into cough sound analysis for determining cough types. Nevertheless, the promising diagnostic precision observed highlights its potential as a tool for assessing respiratory diseases, which affect millions worldwide. The findings from this systematic review provide valuable insights for future study designs and will benefit various stakeholders, including regulatory bodies, technology developers, engineers, data scientists, and clinicians.

Summary Points:

- Reviews the current state of automated cough sound analysis for detecting wet and dry cough types, emphasizing its potential as a diagnostic tool.
- Examines important components including data collection methods, annotation processes, signal processing techniques, machine learning approaches, and performance evaluation metrics.
- Identifies significant variability in inter-rater agreement, underscoring the subjectivity inherent in manual cough sound interpretation and the need for objective analysis methods.
- Outlines key challenges and research gaps, providing actionable insights and directions for advancing this emerging field.

CRediT authorship contribution statement

Roneel V. Sharan: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Hao Xiong:** Writing – review & editing, Methodology.

Ethics approval and consent to participate

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2025.105912.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] C.R. Finley, et al., What are the most common conditions in primary care? Canadian Family Phys. 64 (11) (2018) 832–840. Available: http://www.cfp.ca/content/64/11/832.abstract.

[2] R.A. Haque, O.S. Usmani, P.J. Barnes, Chronic idiopathic cough: A discrete clinical entity? *Chest* 127 (5) (2005) 1710–1713, https://doi.org/10.1378/chest.127.5.1710.

[3] K. Lai, et al., Clinical practice guidelines for diagnosis and management of cough - Chinese Thoracic Society (CTS) Asthma Consortium, *J. Thorac. Dis.* 10 (11) (2018) 6314–6351, https://doi.org/10.21037/jtd.2018.09.153.

[4] P. Kardos, Management of cough in adults, *Breathe* 7 (2) (2010) 122, https://doi.org/10.1183/20734735.019610.

[5] World Health Organization, Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020.

[6] M.D. Shields, S. Thavagnanam, The difficult coughing child: Prolonged acute cough in children, *Cough* 9 (1) (2013) 11, https://doi.org/10.1186/1745-9974-9-11.

[7] A.H. Morice, L. McGarvey, I. Pavord, Recommendations for the management of cough in adults, *Thorax* 61 (Suppl 1) (2006) i1–i24, https://doi.org/10.1136/thx.2006.065144.

[8] I.D. Pavord, P.W. Jones, P.-R. Burgel, K.F. Rabe, Exacerbations of COPD, *Int. J. Chron. Obstruct. Pulmon. Dis.* 11 (2016) 21–30.

[9] A. Niimi, Cough and asthma, *Current Respiratory Med. Rev.* 7 (1) (2011) 47–54.

[10] M. Weinberger, M. Hurvitz, Diagnosis and management of chronic cough: Similarities and differences between children and adults, *F1000Research* 9 (757) (2020) 1–10, https://doi.org/10.12688/f1000research.25468.1.

[11] A.B. Chang, J.T. Gaffney, M.M. Eastburn, J. Faoagali, N.C. Cox, I.B. Masters, Cough quality in children: A comparison of subjective vs. bronchoscopic findings, *Respir. Res.* 6 (1) (2005) 3, https://doi.org/10.1186/1465-9921-6-3.

[12] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *Ann. Intern. Med.* 151 (4) (2009) 264–269, https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

[13] D. Watts, R.F. Pulice, J. Reilly, A.R. Brunoni, F. Kapczinski, I.C. Passos, Predicting treatment response using EEG in major depressive disorder: A machine-learning meta-analysis, *Transl. Psychiatry* 12 (1) (2022) 332, https://doi.org/10.1038/s41398-022-02064-z.

[14] R.G. Newcombe, Two-sided confidence intervals for the single proportion: Comparison of seven methods, *Stat. Med.* 17 (8) (1998) 857–872, https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E.

[15] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies, *Int. J. Med. Inf.* 153 (2021) 104510, https://doi.org/10.1016/j.ijmedinf.2021.104510.

[16] Y. Zhou, et al., Machine learning predictive models for acute pancreatitis: A systematic review, *Int. J. Med. Inf.* 157 (2022) 104641, https://doi.org/10.1016/j.ijmedinf.2021.104641.

[17] R.V. Sharan, H. Rahimi-Ardabili, Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review, *Int. J. Med. Inf.* 176 (2023) 105093, https://doi.org/10.1016/j.ijmedinf.2023.105093.

[18] V. Swarnkar, U.R. Abeyratne, A.B. Chang, Y.A. Amrulloh, A. Setyati, R. Triasih, Automatic identification of wet and dry cough in pediatric patients with respiratory diseases, *Ann. Biomed. Eng.* 41 (5) (2013) 1016–1028, https://doi.org/10.1007/s10439-013-0741-6.

[19] E. Nemati, M.M. Rahman, V. Nathan, K. Vatanparvar, J. Kuang, A comprehensive approach for classification of the cough type, in: *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 20–24 Jul 2020, pp. 208–212, doi: 10.1109/EMBC44109.2020.9175345.

[20] R.V. Sharan, Productive and non-productive cough classification using biologically inspired techniques, *IEEE Access* 10 (2022) 133958–133968, https://doi.org/10.1109/ACCESS.2022.3231640.

[21] A.V. Oppenheim, R.W. Schafer, J.R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Prentice-Hall Inc, Upper Saddle River, New Jersey, 1998.

[22] F. Eyben, M. Wöllmer, B. Schuller, Opensmile: The Munich versatile and fast open-source audio feature extractor, in:Proceedings of the 18th ACM International Conference on Multimedia, Firenze, Italy, 2010: Association for Computing Machinery, pp. 1459–1462, doi: 10.1145/1873951.1874246.

[23] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357, https://doi.org/10.1613/jair.953.

[24] S. Hershey *et al.*, CNN architectures for large-scale audio classification, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, 5–9 Mar 2017, pp. 131–135, doi: 10.1109/ICASSP.2017.7952132.

[25] R.V. Sharan, Cough sound detection from raw waveform using SincNet and bidirectional GRU, *Biomed. Signal Process. Control* 82 (2023) 104580, https://doi.org/10.1016/j.bspc.2023.104580.

[26] J.C. Valentine, T.D. Pigott, H.R. Rothstein, How many studies do you need?: A primer on statistical power for meta-analysis, *J. Educ. Behav. Stat.* 35 (2) (2010) 215–247.

[27] M. Borenstein, L.V. Hedges, J.P.T. Higgins, H.R. Rothstein, *Introduction to Meta-Analysis*, John Wiley & Sons Ltd, UK, 2009.

[28] J.P.A. Ioannidis, T.A. Trikalinos, The appropriateness of asymmetry tests for publication bias in meta-analyses: A large survey, *Can. Med. Assoc. J.* 176 (8) (2007) 1091–1096.

[29] M. Trivedi, E. Denton, Asthma in children and adults—What are the differences and what can they tell us about asthma? *Front. Pediatr.* 7 (2019) 256.

[30] A.C. Yu, B. Mohajer, J. Eng, External validation of deep learning algorithms for radiologic diagnosis: A systematic review, *Radiol. Artif. Intell.* 4 (3) (2022), https://doi.org/10.1148/ryai.210064 e210064.

[31] J. He, S.L. Baxter, J. Xu, J. Xu, X. Zhou, K. Zhang, The practical implementation of artificial intelligence technologies in medicine, *Nat. Med.* 25 (1) (2019) 30–36, https://doi.org/10.1038/s41591-018-0307-0.

[32] P.-H.-C. Chen, C.H. Mermel, Y. Liu, Evaluation of artificial intelligence on a reference standard based on subjective interpretation, *Lancet Digital Health* 3 (11) (2021) e693–e695, https://doi.org/10.1016/S2589-7500(21)00216-8.

[33] F. Barata, et al., Nighttime continuous contactless smartphone-based cough monitoring for the ward: Validation study, *JMIR Formative Res.* 7 (2023), https://doi.org/10.2196/38439 e38439.

[34] M.D. Kruizinga, et al., Development and technical validation of a smartphone-based pediatric cough detection algorithm, *Pediatr. Pulmonol.* 57 (3) (2022) 761–767, https://doi.org/10.1002/ppul.25801.
[35] P. Porter, J. Brisbane, U. Abeyratne, N. Bear, S. Claxton, A smartphone-based algorithm comprising cough analysis and patient-reported symptoms identifies acute exacerbations of asthma: A prospective, double blind, diagnostic accuracy study, *J. Asthma* 60 (2) (2023) 368–376, https://doi.org/10.1080/02770903.2022.2051546.
[36] H. Müller, A. Holzinger, M. Plass, L. Brcic, C. Stumptner, K. Zatloukal, Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation, *N. Biotechnol.* 70 (2022) 67–72, https://doi.org/10.1016/j.nbt.2022.05.002.