International Journal of Medical Informatics 176 (2023) 105093

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

Review article

Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review

Roneel V. Sharan *, Hania Rahimi-Ardabili

Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia

ARTICLE INFO

Keywords:
Acute respiratory diseases
Cough sound
Feature extraction
Machine learning
Pediatric

ABSTRACT

Background: Acute respiratory diseases are a leading cause of morbidity and mortality in children. Cough is a common symptom of acute respiratory diseases and the sound of cough can be indicative of the respiratory disease. However, cough sound assessment in routine clinical practice is limited to human perception and the skills of the clinician. Objective cough sound evaluation has the potential to aid clinicians in acute respiratory disease diagnosis. In this systematic review, we assess and summarize the predictive ability of machine learning algorithms in analyzing cough sounds of acute respiratory diseases in the pediatric population.
Method: Our systematic search of the Scopus, Medline, and Embase databases on 25 January 2023 identified six articles meeting the inclusion criteria. Quality assessment of the included studies was performed using the checklist for the assessment of medical artificial intelligence.
Results: Our analysis shows variability in the input to the machine learning algorithms, such as the use of various cough sound features and combining cough sound features with clinical features. The use of the machine learning algorithms also varies from conventional algorithms, such as logistic regression and support vector machine, to deep learning techniques, such as convolutional neural networks. The classification accuracy for the detection of bronchiolitis, croup, pertussis, and pneumonia across five articles is in the range of 82–96%. However, a significant drop is observed in the detection accuracy for bronchiolitis and pneumonia in the remaining article.
Conclusion: The number of articles is limited but, in general, the predictive ability of cough sound classification algorithms in childhood acute respiratory diseases shows promise.

1. Introduction

Acute respiratory diseases are one of the leading causes of morbidity and mortality in children worldwide [1]. Acute respiratory diseases include both upper and lower respiratory diseases, such as croup, bronchiolitis, pertussis, and pneumonia [2]. Cough is a common symptom of acute respiratory diseases and one of the most common presenting conditions in primary care globally [3]. The respiratory infections causing the acute respiratory disease can affect the airways differently, thereby producing variation in cough type, such as the distinctive barking cough of croup [4] and paroxysmal coughing ending in the characteristic "whoop" in pertussis [5]. Cough sound assessment is, therefore, useful in assessing the condition of the respiratory system. However, assessing cough sounds in clinical practice can be subjective, dependent on the training and skills of the clinician [6]. In addition, diagnostic errors are common in acute pediatric respiratory diseases, such as in emergency departments [7]. Objective artificial intelligence (AI) driven cough sound evaluation has the potential to aid clinicians in respiratory disease diagnosis. Despite the potential significance of objective cough sound evaluation in clinical decision making of acute respiratory diseases, no evidence syntheses have been completed on this topic. Therefore, we conducted a systematic review to determine the ability of machine learning methods to predict acute respiratory diseases in the pediatric population using cough sound.

2. Methods

The review was completed in accordance with the preferred reporting items for systematic reviews and meta-analysis (PRISMA) guidelines [8].

2.1. Search strategy

We performed a systematic literature search in Scopus, Medline

* Corresponding author.
E-mail address: roneel.sharan@mq.edu.au (R.V. Sharan).

https://doi.org/10.1016/j.ijmedinf.2023.105093
Received 21 March 2023; Received in revised form 21 April 2023; Accepted 7 May 2023
Available online 18 May 2023

(Ovid), and Embase (Ovid) databases. The search was performed on 25 January 2023. The following search terms were used in all three databases: ("pediatric" OR "paediatric" OR "child" OR "children" OR "childhood") AND ("cough") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "feature extraction" OR "accuracy"). Title, abstract, and keyword search was performed in all three databases in English language. The search terms were kept broad and no restriction was placed on the publication date and study location to maximize the search results. We did not search for any specific acute respiratory disease or include the term "acute respiratory disease" in our search terms. This ensured we were not restricting our search to a group of respiratory diseases or missing out on relevant articles that did not include "acute respiratory disease" in their title, abstract, or keyword, maximizing our search results. Using this approach allowed us to capture a broad range of articles to be screened manually for their relevancy. Gray literature was not considered in this systematic review.

2.2. Inclusion and exclusion criteria

This systematic review included articles that used cough sound features and AI algorithms to assess the diagnosis of acute respiratory diseases in the pediatric population. As a minimum criterion, cross-validation or training and testing or testing sets were required for a study to be included in this systematic review. This is because models lacking resampling procedures are less likely to appropriately generalize on independent datasets. In addition, studies with small sample sizes ($\leq$30) were excluded since cross-validation with small sample sizes, without training and testing sets, can cause model overfitting and lead to inflated and highly variable predictive results [9].

The study of respiratory diseases differs by the age of the participants with children more susceptible to certain respiratory diseases. Pediatric respiratory studies often focus on issues related to the development and growth of the respiratory system, as well as the diagnosis and management of respiratory conditions that are unique to children. Since this systematic review focuses on the pediatric population, only studies that included children with ages $\leq$ 18 years (216 months) were included.

Acute respiratory diseases are the leading cause of childhood illness and death and form the subject of this systematic review. As such, studies on chronic respiratory diseases, such as chronic asthma, were excluded. This systematic review focuses on detecting specific acute respiratory diseases against other respiratory diseases in children, therefore, studies on detecting groups of respiratory diseases, such as upper and lower respiratory diseases, were excluded. Similarly, studies that focused on the type of cough, such as productive (wet) and non-productive (dry) cough, were excluded. In addition, if the negative or non-disease class had only healthy subjects or subjects from only one non-target disease group then they were excluded as these can be considered relatively easier problems and of relatively less clinical value.

Studies that relied on non-cough sound features, such as demographic and clinical data, in addition to cough sound features were included. Abstracts and conference proceedings were included if sufficient data could be extracted from these publications or related publications. Some studies present their initial findings as a short paper, such as a conference abstract or paper, but followed by an extended publication. As such, where a preliminary study was followed by a more comprehensive study, only the latter was included.

2.3. Article selection

Search results were downloaded from the databases into EndNote X9. Duplicates were removed in EndNote and the remaining results were exported to Microsoft Excel. R.V.S. and H.R.A. independently screened the titles and abstracts of these articles for eligibility. Articles that were considered eligible for inclusion by either reviewer were reviewed for full-text screening and assessed for inclusion. Reasons for exclusion after full-text screening were entered in Excel and any disagreements were resolved by discussion, with a consensus reached for each study.

2.4. Data extraction and analysis

The methodological, demographic, and outcome data were extracted from the included articles by R.V.S. These were reviewed by H.R.A. and any disagreements were resolved through discussion. The extracted data included study characteristics (first author, year of publication, country where study was conducted, study design, sample size, feature extraction, feature selection, machine learning classifier, cough segmentation method); population characteristics (age, gender); and performance metrics (sensitivity, specificity, accuracy, and the 95% confidence interval for these metrics). If a study used multiple feature extraction, feature selection, and machine learning classifiers then all these data were extracted. In studies which did not report the accuracy and 95% confidence interval metrics, these were calculated using the relevant data available in these studies. The accuracy value was calculated as the total number of correct classifications divided by the overall sample size, where the number of correct classifications was calculated from the sensitivity and specificity values and the class sample size, and the 95% confidence interval was calculated using the method described by Newcombe [10]. To be consistent, the performance metrics are rounded to the nearest whole number in percentage for all the studies. Meta-analysis was not performed due to the small number of studies in each disease group and because of the overlap between the studies, which cannot be easily accounted for in meta-analysis [11].

2.5. Quality assessment

Quality assessment was performed by R.V.S using the checklist for the assessment of medical AI (ChAMAI) [12]. It was reviewed by H.R.A and any disagreements were resolved through discussion. The ChAMAI checklist is a tool proposed by the IJMEDI for assessing the quality of medical artificial intelligence studies. Its aim is to differentiate high-quality machine learning studies from basic medical data-mining studies. After literature selection, researchers can use this checklist to evaluate the quality of included articles based on the study purposes, inclusion criteria, and professional knowledge. The checklist comprises six dimensions: problem understanding, data understanding, data preparation, modeling, validation, and deployment, consisting of a total of 30 questions. Each question can be rated as OK (adequately addressed), mR (sufficient but improvable, minor revision needed), and MR (inadequately addressed, major revision needed). Based on [13], we linked scores to the responses for each item. Of the 30 questions in the ChAMAI checklist [12], 20 have been categorized as high-priority and 10 as low-priority. In high-priority items, scores of 2, 1, and 0 were assigned to OK, mR, and MR, respectively, while scores for low-priority items were halved. The maximum possible score is 50 points, and the study quality was classified as low (0–19.5), medium (20–34.5), or high (35–50) [13]. All eligible articles were included for analysis regardless of their quality assessment outcome.

3. Results

3.1. Search results

The process for identifying eligible articles is illustrated in the PRISMA flow diagram of Fig. 1. Search in Scopus, Medline (Ovid) and Embase (Ovid) produced 464, 164 and 509 results, respectively, for a total of 1,137 articles. Removing 504 duplicate articles left 633 articles for the title and abstract screening. After title and abstract screening, 622 articles were excluded leaving 11 articles for full-text screening. Finally, 5 articles were excluded after the full-text screening with the remaining 6 articles meeting the inclusion criteria. These 6 articles were included for quality assessment. All 6 articles were included for

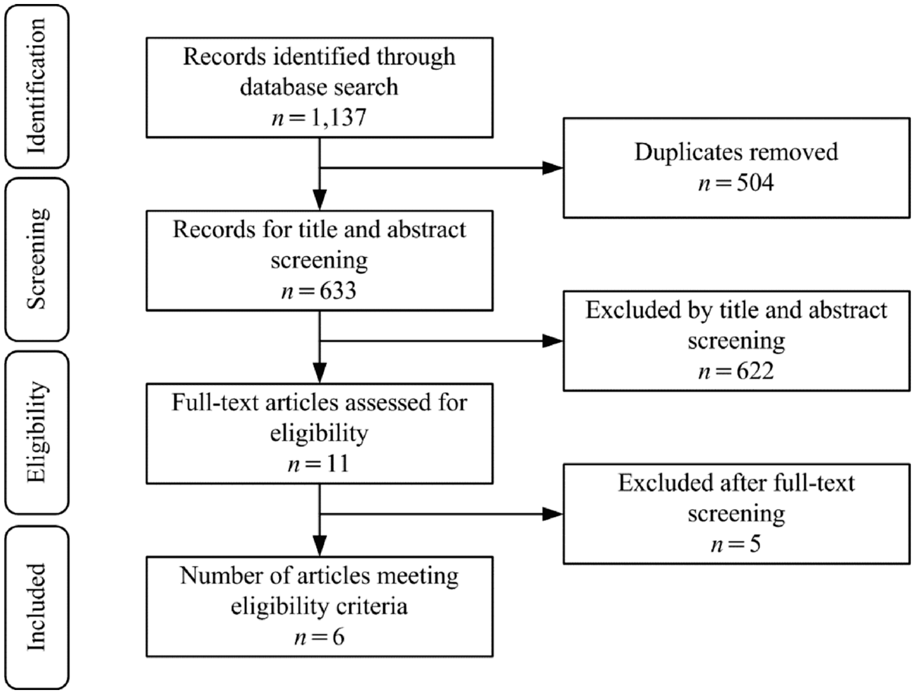

Fig. 1. PRISMA flow diagram of included articles which use cough sound features and machine learning to detect acute respiratory diseases in the pediatric population.

qualitative synthesis and quantitative analysis.

3.2. Study characteristics

A list of the included articles [14–19] and their relevant characteristics are given in Table 1. One article [16] focused on detecting croup from other respiratory diseases, one article [19] focused on detecting pertussis from other respiratory diseases, two articles [14,15] focused on detecting pneumonia from other respiratory diseases, and the remaining two articles [17,18] focused on detecting multiple acute respiratory diseases, such as bronchiolitis, croup, and pneumonia, from other respiratory diseases. The studies in two articles [16,17] were conducted in Australia, the studies reported in articles [14,15] use the same data which was collected in Indonesia, and the studies reported in article [18] was conducted in the United States. The data for the study in the remaining article [19] was collated from YouTube video postings, therefore, likely coming from many different countries. The age range of the subjects was 0–7 years in the study reported in article [19], 0–12

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

Study Reference	Study Disease (s)	Study Design	Country	Sample	Age Range (months)	Method	Cough Segmentation
Abeyratne et al. [14]	Pneumonia	Prospective	Indonesia	91	1–180	Cough features: BS, NGS, FF, LogE, ZCR, Kurt, MFCC Feature selection: *p*-value Classifier: LR	Manual
Kosasih et al. [15]	Pneumonia	Prospective	Indonesia	91	1–180	Cough features: BS, NGS, FF, LogE, ZCR, Kurt, MFCC, WF Feature selection: *p*-value Classifier: LR	Manual
Sharan et al. [16]	Croup	Prospective	Australia	479	0–192	Cough features: MFCC, CIF Feature selection: *p*-value, SFS Classifier: LR, SVM (RBF)	Automatic
Porter et al. [17]	Bronchiolitis Croup Pneumonia	Prospective	Australia	585	1–144	Cough features: BS, NGS, FF, LogE, ZCR, Kurt, MFCC, WF Clinical features: fever, rhinorrhea, wheeze, hoarse voice, maximum days of symptoms Feature selection: *p*-value Classifier: LR	Automatic
Moschovis et al. [18]	Bronchiolitis Pneumonia	Prospective	United States	1251	1–144	Cough features: BS, NGS, FF, LogE, ZCR, Kurt, MFCC, WF Clinical features: fever, rhinorrhea, wheeze, age, duration of symptoms Feature selection: *p*-value Classifier: LR	Automatic
Sharan et al. [19]	Pertussis	Retrospective	Worldwide	42	1–84	Cough features: MFCC, WF, CIF, mel-spectrogram, wavelet scalogram, cochleagram Feature selection: ANOVA, *t*-test Classifier: LR, NB, SVM (RBF), CNN	Manual

BS – bispectrum score, CIF – cochleagram image feature, CNN – convolutional neural network, FF – formant frequencies, Kurt – kurtosis, LogE – log energy, LR – logistic regression, MFCC – mel-frequency cepstral coefficients, NB – Naïve Bayes, NGS – non-Gaussianity score, RBF – radial basis function, SFS – sequential feature selection, SVM – support vector machine, WF – wavelet features, ZCR – zero-crossing rate.

years in studies reported in articles [17,18], 0–15 years in studies reported in articles [14,15], and 0–16 years in the study reported in the remaining article [16]. The number of subjects in the studies in three articles [14,15,19] was less than 100. The studies in the remaining three articles had more than 100 subjects, but the studies in two articles [17,18] had a different number of subjects for different disease studies. The number of subjects for the target acute respiratory disease and the study design were not reported in article [18] but they were obtained from other articles [20,21] reporting the same studies. The diagnosis of the subjects in all cases was done by clinicians, except one study [19] where the diagnosis was attributed by the information provided in the title and/or description of the online videos and later checked by a clinician to assess its plausibility. One study [19] was retrospective while the remaining studies [14–18] report results on prospective datasets.

3.3. Data recording

In the included articles, cough sound data was recorded using a smartphone [16–18] or using bedside microphones (Rode NT3) [14,15]. For the remaining article [19], the recordings were likely made using different devices but most, if not all, were believed to be made using smartphones. The recordings in all cases were made in a hospital environment, except for one article [19].

3.4. Cough data

The cough segmentation in studies reported in three articles [14,15,19] was performed manually, while the studies reported in the remaining three articles [16–18] used automatic cough segmentation. The auto segmentation algorithm in these articles [16–18] used various handcrafted features and time delay neural network. It is described in a separate study [16] and is built on an earlier work [22].

All studies used multiple coughs from each subject. In one article [16], the maximum number of coughs per subject is restricted to 10 with an average of 8.96 coughs per subject on the training dataset and 9.50 coughs per subjects on the test dataset. The maximum number of coughs was restricted to 20 in the study reported in one article [19], for a total of 542 coughs from the 42 subjects. In two articles [14,15], the training dataset had 2–15 coughs per subject for a total of 440 coughs from the 66 subjects and the test dataset had 15 coughs per subject for a total of 375 coughs from the 25 subjects. In the studies in the remaining two articles [17,18], the first 5 coughs were used from all the subjects. The coughing was spontaneous or voluntary in the studies reported in articles [16–18]. Spontaneous coughs were particularly needed when subjects, such as infants and young children, could not cooperate in providing voluntary coughs. Based on the description of the remaining three articles [17–19], we believe they mostly used spontaneous coughing.

3.5. Cough and clinical features

The studies reported in three articles [15,16,19] used only cough sound features in their classification models, studies in two articles [17,18] combined cough sound features and clinical features, and the study reported in one article [14] experimented with both, cough sound features only and combined cough sound and clinical features. Various cough features and clinical features were experimented with across the included studies. Mel-frequency cepstral coefficients (MFCCs) were the most widely used features, experimented with in all the studies. Three articles [15,17,18] used a similar feature set which includes MFCCs, bispectrum score (BS), non-Gaussianity score (NGS), formant frequencies (FF), log energy (LogE), zero-crossing rate (ZCR), kurtosis (Kurt), and wavelet features (WF). A similar feature set was also used in another article [14], but without WF. MFCCs were used together with cochleagram image features (CIF) in article [16], where cochleagram is a type of time–frequency representation utilizing gammatone filters. MFCCs, CIF, and WF were experimented with in study [19] together with three time–frequency representations: mel-spectrogram, wavelet scalogram, and cochleagram.

Different articles used slightly different clinical features. Breathing index (derived as the breathing rate for different age groups), cough, fever, and age were used in the study in article [23]; fever, rhinorrhea, wheeze, hoarse voice, and maximum days of symptoms were used in the studies in article [17]; and fever, rhinorrhea, wheeze, age, and duration of symptoms were used in the studies in article [18].

3.6. Classifiers

Classification using logistic regression (LR) was experimented with in all the studies. In the studies reported in articles [17,18], the output of multiple binary classification disease models built using LR were input to a softmax classifier for final prediction. In addition to the LR classifier, the study reported in article [16] utilized the support vector machine (SVM) classifier with a radial basis function (RBF) kernel. SVM with RBF kernel was also used in the study reported in article [19], in addition to the Naïve Bayes (NB) classifier, and convolutional neural network (CNN). As such, study [19] was the only study that made use of a deep learning classification technique. Due to their relatively small dataset, they used a relatively shallow CNN with data augmentation using *mixup* [24] during training to prevent model overfitting. Two-dimensional CNN was used and the inputs to the CNN were the three time–frequency image-like representations. One CNN was trained on each time–frequency image and the outputs of the CNNs were pooled together for classification using SVM.

3.7. Feature selection

In the studies reported in articles [14–18], feature selection for the LR classifier was performed using the *p*-value of the coefficient estimates of the LR. In addition, sequential feature selection (SFS) was experimented with in the study in article [16] while one-way analysis of variance and *t*-test methods were used for feature selection in the study in article [19].

3.8. Cough and subject classification

All the included studies used binary classification and the first stage of classification was cough classification, where the coughs in each recording were assigned the same disease label as the subject. In the study in article [16], the subject was determined to have the target disease if at least one cough was classified positive for the target disease. In the studies reported in the other five articles [14,15,17–19], cough index was used in determining whether or not a subject had the target disease, where cough index is the number of coughs classified as the target disease divided by the number of coughs from the subject. A cough index threshold was then computed to classify a subject into two classes, positive (disease) and negative (non-disease), where the non-disease class was comprised of all non-target respiratory diseases.

3.9. Diagnostic performance

In this systematic review, we focus on the accuracy metric to analyze the classification performance. However, some studies had highly imbalanced class distributions because of which we also report the sensitivity and the specificity metrics. We also include the 95% confidence interval for these metrics. The evaluation results are summarized in Table 2.

The accuracy in detecting croup from other respiratory diseases was 82% or more using cough sound features alone and when combining cough and clinical features. Our search found only one study [19] on detecting pertussis using cough sound which achieved an accuracy of 90%. The two studies [17,18] on bronchiolitis used a combination of

Table 2
Summary of the results for the included studies.

Study Disease	Study Reference	Sample (Training/ Validation)	Sample (Test)	Input to ML classifier	Performance [95% CI] (%)
Bronchiolitis	Porter et al. [17]	–	157	Cough + clinical features	Sensitivity = 84 [77–90] Specificity = 81 [61–93] Accuracy = 83 [77–89]
	Moschovis et al. [18]	–	131	Cough + clinical features	Sensitivity = 76 [61–88] Specificity = 60 [49–70] Accuracy = 65 [56–73]
Croup	Sharan et al. [16]	364	115	Cough features only	Sensitivity = 92 [78–100] Specificity = 85 [78–92] Accuracy = 86 [80–92]
	Porter et al. [17]	–	568	Cough + clinical features	Sensitivity = 85 [75–93] Specificity = 82 [78–85] Accuracy = 82 [79–85]
Pertussis	Sharan et al. [19]	42	–	Cough features only	Sensitivity = 95 [77–99] Specificity = 86 [65–95] Accuracy = 90 [78–96]
Pneumonia	Abeyratne et al. [14]	66	25	Cough features only	Sensitivity = 94 [73–99] Specificity = 75 [41–93] Accuracy = 88 [70–96]
				Cough + clinical features	Sensitivity = 94 [73–99] Specificity = 100 [68–100] Accuracy = 96 [80–99]
	Kosasih et al. [15]	66	25	Cough features only	Sensitivity = 94 [73–99] Specificity = 88 [53–98] Accuracy = 92 [75–98]
	Porter et al. [17]	–	569	Cough + clinical features	Sensitivity = 87 [75–94] Specificity = 85 [82–88] Accuracy = 85 [82–88]
	Moschovis et al. [18]	–	1250	Cough + clinical features	Sensitivity = 63 [53–72] Specificity = 62 [59–65] Accuracy = 62 [59–65]

cough and clinical features. While one study [17] achieved an accuracy of 83%, the accuracy dropped to 65% in the second study [18]. Two studies [14,15] on pneumonia reported results using cough sound features only and three studies [14,17,18] reported results using cough and clinical features. The accuracy in detecting pneumonia using cough sound features only was 88% or higher and, in one study [14], the accuracy improved from 88% to 96% when cough features were combined with clinical features. However, the accuracy dropped to 85% [17] and 62% [18] in two studies using a similar approach.

3.10. Study quality

Details of the results of the quality assessment are provided in the Supplementary File. The scores of each requirement and the total score in each study are summarized in Table 3. The average score of the included studies was 31.2 with one study [19] having a score of 24 and the remaining five studies [14–18] having a score in the range 32–33. As such, all the articles were of a medium quality. Fig. 2 shows the proportion of the different answers in the high- and low-priority items.

4. Discussion

In this paper, we describe the methodology in our systematic review of studies focused on detecting acute respiratory diseases using cough sound characteristics and machine learning. We also summarize key trends that emerged when analyzing the included studies. Earlier studies used a conventional sound recording setup while more recent studies used smartphones and all studies analyzed multiple cough sounds from each subject in predicting the respiratory disease of the subject. We noted that earlier works relied on conventional feature engineering and machine learning methods but a more recent study based on deep learning techniques outperformed such conventional methods, even on a relatively small dataset. Respiratory infections that are well-known to cause a distinctive cough sound, such as the infection of croup and pertussis (whooping cough), are seen to give high accuracy using cough sound features on its own or when combined with clinical features.

However, we observed a significant drop in classification accuracy for bronchiolitis and pneumonia from one prospective study to another, even though both studies used similar case definitions and data acquisition setup. This highlights the subjectiveness in clinical decision-making of respiratory diseases, especially for bronchiolitis and pneumonia [7], and the need for objective respiratory disease evaluation

Table 3
Quality assessment scores of the six included studies according to the ChAMAI checklist.

Study Reference	Problem Understanding (10)	Data Understanding (6)	Data Preparation (8)	Modeling (6)	Validation (12)	Deployment (8)	Total (50)
Abeyratne et al. [14]	10	3	3	6	8.5	2	32.5
Kosasih et al. [15]	10	3	3	6	8.5	1.5	32
Sharan et al. [16]	10	4	3	6	8	2	33
Porter et al. [17]	10	4	3	6	8.5	1.5	33
Moschovis et al. [18]	10	4	3	6	8	1.5	32.5
Sharan et al. [19]	4	1	4	6	7.5	1.5	24

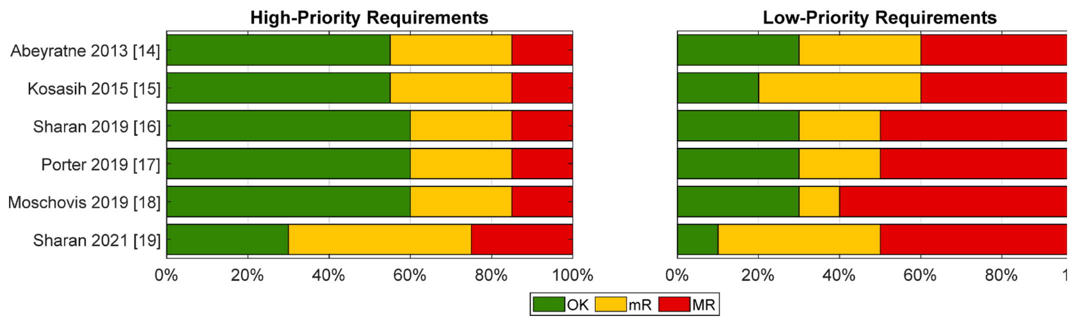

Fig. 2. Proportion of OK (adequately addressed), mR (sufficient but improvable, minor revision needed), and MR (inadequately addressed, major revision needed) in the high- and low-priority requirements.

tools. We also note that in detecting croup, cough sound features [16] yielded higher classification accuracy than the combined cough and clinical features [17]. This could be because subjects with croup have a distinctive barking cough making the cough sound features very informative. The inclusion of clinical features may not add value to the cough sound features in the croup classification model. However, the two studies [16,17] employed different sample size, cough sound features, feature selection method, and machine learning classifier which could also be the reason for this difference in performance. The contribution of cough features and clinical features to the pneumonia classification model were easy to determine in one article [14] as performance metrics using cough features and a combination of cough and clinical features are reported. However, articles [17,18] reported performance metrics using only the combined cough and clinical features which makes it difficult to gauge the contribution of each feature type to their disease classification models.

The number of articles included in this systematic review is small which makes it difficult to see long-term overall trends in quality assessment. However, the overall score of studies in articles [14–18] were close to the upper limit of medium quality while the study reported in one article [19] was close to the lower limit. This was due to several reasons, such as the reference diagnosis in this study [19] did not follow standard clinical guidelines and the index test results were reported in cross-validation only, without an separate test set. Articles [14–18] adequately addressed (OK) more than 50% of the high-priority requirements but this dropped to 30% or less for low-priority requirements, and 40–60% of the low-priority requirements require major revision (MR). In addition, all the articles performed poorly in deployment, with a total score of 25% or less from the maximum possible score of 8, and none of the studies had been externally validated on data from a different healthcare setting.

Furthermore, only one study [19] used deep learning classification, even though it had the smallest number of subjects of all the studies included in this work. Due to the relatively small dataset, their method relied on data augmentation and used a relatively shallow CNN to prevent overfitting. However, more sophisticated pretrained CNN architectures, such as AlexNet [25] and GoogleNet [26], and variations of CNN, such as with residual connections, ResNet [27], have the potential to perform better if presented with enough training data and finding an effective image-like representation of the one-dimensional cough sound signals. While AlexNet, GoogleNet, and ResNet have been pretrained on images, pretrained CNNs for audio classification [28] are also available which take in two-dimensional mel-spectrogram representation of cough sounds as input. More recently, learning directly from one-dimensional raw audio signals is also possible, such as with SincNet [29,30]. While neural networks can be complex, it has the potential to perform effective feature engineering and learning without significant signal transformation and learn non-linear characteristics. Neural networks have been widely explored in medical imaging but their potential is yet to be realized in detecting childhood acute respiratory diseases using cough sounds.

This systematic review has limitations. We found only six articles in the literature that met our inclusion criteria. In addition, only one or two studies were available for each acute respiratory disease, except pneumonia, and there was overlap between some of these studies because of which meta-analysis was not performed. The study reported in article [19] had a relatively poor quality assessment as the reference diagnosis did not follow clinical guidelines and the index test was not evaluated on an independent test dataset. Our systematic review yielded studies on only four acute respiratory diseases but children are affected by several other types of acute respiratory diseases. The lack of studies shows the infancy of research in detecting acute respiratory diseases using cough sounds and the recent pandemic shows the need for more research in objective cough sound evaluation.

The findings from this systematic review, such as data acquisition, cough sound and clinical features, and machine learning techniques, would provide a useful starting point in the design of similar studies in the future. Due to the relatively small number of included studies in this systematic review, learnings from similar studies in the adult population, such as pneumonia [31], would also be useful. One shortcoming of the studies included in this systematic review is the lack of external validation. In detecting COVID-19 using cough sounds in adult subjects, a recent study [32] reported promising results but the predictive performance of the algorithm declined when evaluated on an independent dataset [33]. In the included studies, a similar drop in algorithmic performance is observed when the algorithms developed in one setting [17] are adopted and evaluated in another setting [18]. Similar problems have been reported in some other AI in healthcare applications as well, such as in sepsis risk prediction in adults using electronic health record data [34]. Such studies highlight the challenges of using AI in healthcare applications, especially in different settings, but they also offer valuable insights and lessons that can be used in future study designs, such as the need for data standardization at the initial development stage to minimize heterogeneity in health data [35] and the need to reduce subjectiveness in ground truthing [36].

In conclusion, the limited number of included studies shows the infancy of work in analyzing cough sounds using signal processing and machine learning techniques for detecting acute respiratory diseases in children. However, the promising diagnostic accuracy in most of the reviewed studies shows its potential as a respiratory disease assessment tool. Knowledge gained from this systematic review can be used in future study designs and also be useful to regulatory bodies, technology manufacturers, engineers and data scientists, and clinicians. Millions of children are affected by respiratory infections every year where cough is one of the common symptoms. Visiting a clinician with an infective respiratory disease can lead to other people getting infected. Virtual healthcare has seen an increased uptake during COVID-19 and this is expected to continue [37]. There is a potential for such objective cough sound assessment technology to be integrated into existing virtual healthcare systems to aid clinical diagnosis.

Ethics approval and consent to participate:

Not applicable.

CRediT authorship contribution statement

Roneel V. Sharan: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Hania Rahimi-Ardabili:** Methodology, Supervision, Writing – review & editing.

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 material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2023.105093.

References

[1] A.A. Tazinya, G.E. Halle-Ekane, L.T. Mbuagbaw, M. Abanda, J. Atashili, M. T. Obama, Risk factors for acute respiratory infections in children under five years attending the Bamenda Regional Hospital in Cameroon, BMC Pulm. Med. 18 (1) (2018) 7.
[2] C.F. Lanata, R.E. Black, Acute lower respiratory infections, in: R.D. Semba, M. W. Bloem, P. Piot (Eds.), Nutrition and health in developing countries, Humana Press, Totowa, NJ, 2008, pp. 179–214.
[3] C.R. Finley, et al., What are the most common conditions in primary care? Can. Fam. Phys. 64 (11) (2018) 832–840.
[4] C.L. Bjornson, D.W. Johnson, Croup, Lancet 371 (9609) (2008) 329–339.
[5] World Health Organization, Pertussis vaccines: WHO position paper – August 2015, World Health Organization, 2015.
[6] M. Binnekamp, K.J. van Stralen, L. den Boer, M.A. van Houten, Typical RSV cough: myth or reality? A diagnostic accuracy study, Eur. J. Pediatr. 180 (1) (2021) 57–62.
[7] P. Porter, et al., Diagnostic errors are common in acute pediatric respiratory disease: A prospective, single-blinded multicenter diagnostic accuracy study in Australian emergency departments, Front. Pediatr. 9 (2021), 736018.
[8] 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.
[9] 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. Psych. 12 (1) (2022) 332.
[10] R.G. Newcombe, Two-sided confidence intervals for the single proportion: Comparison of seven methods, Stat. Med. 17 (8) (1998) 857–872.
[11] M. Bracher-Smith, K. Crawford, V. Escott-Price, Machine learning for genetic prediction of psychiatric disorders: A systematic review, Mol. Psychiatry 26 (1) (2021) 70–79.
[12] 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. Informatics 153 (2021), 104510.
[13] Y. Zhou, et al., Machine learning predictive models for acute pancreatitis: A systematic review, Int. J. Med. Informat. 157 (2022), 104641.
[14] U.R. Abeyratne, V. Swarnkar, A. Setyati, R. Triasih, Cough sound analysis can rapidly diagnose childhood pneumonia, Ann. Biomed. Eng. 41 (11) (2013) 2448–2462.
[15] K. Kosasih, U.R. Abeyratne, V. Swarnkar, R. Triasih, Wavelet augmented cough analysis for rapid childhood pneumonia diagnosis, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1185–1194.
[16] R.V. Sharan, U.R. Abeyratne, V.R. Swarnkar, P. Porter, Automatic croup diagnosis using cough sound recognition, IEEE Trans. Biomed. Eng. 66 (2) (2019) 485–495.
[17] P. Porter, et al., A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children, Respir. Res. 20 (1) (2019) 81.
[18] P.P. Moschovis, et al., A cough analysis smartphone application for diagnosis of acute respiratory illnesses in children, in: American thoracic society international conference, Dallas, Texas, 2019, p. A1181.
[19] R.V. Sharan, S. Berkovsky, D.F. Navarro, H. Xiong, A. Jaffe, Detecting pertussis in the pediatric population using respiratory sound events and CNN, Biomed. Signal Process. Control 68 (2021), 102722.
[20] ResApp Health Limited, "ResApp Reports Positive Preliminary Results from SMARTCOUGH-C-2 Study for Diagnosis of Childhood Respiratory Disease using Cough Sounds", Brisbane, Australia, 2018. https://www.resapphealth.com.au/wp-content/uploads/2018/10/1863471.pdf (Accessed 19 April 2023).
[21] P.P. Moschovis, et al., The diagnosis of respiratory disease in children using a phone-based cough and symptom analysis algorithm: The smartphone recordings of cough sounds 2 (SMARTCOUGH-C 2) trial design, Contemp. Clin. Trials 101 (2021), 106278.
[22] Y.A. Amrulloh, U.R. Abeyratne, V. Swarnkar, R. Triasih, A. Setyati, Automatic cough segmentation from non-contact sound recordings in pediatric wards, Biomed. Signal Process. Control 21 (2015) 126–136.
[23] U.R. Abeyratne, V. Swarnkar, A. Setyati, R. Triasih, "Cough sound analysis can rapidly diagnose childhood pneumonia", (in English), Ann Biomed Eng 41 (11) (2013) 2448–2462.
[24] H. Zhang, M. Cisse, Y.N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in: International conference on learning representations, Vancouver, Canada, 2018, pp. 1–13.
[25] A. Krizhevsky, I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, (eds.), Advances in neural information processing systems (NIPS), Nevada, USA, 2012, vol. 25, pp. 1097–1105.
[26] C. Szegedy et al., Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9.
[27] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA 2016, pp. 770–778.
[28] 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, 2017, pp. 131–135.
[29] M. Ravanelli, and Y. Bengio, Speaker recognition from raw waveform with SincNet, in: IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 2018, pp. 1021–1028.
[30] R.V. Sharan, Cough sound detection from raw waveform using SincNet and bidirectional GRU, Biomed. Signal Process. Control 82 (2023), 104580.
[31] P. Paul, et al., Diagnosing community-acquired pneumonia via a smartphone-based algorithm: a prospective cohort study in primary and acute-care consultations, BR. J. GEN. PRACT. 71 (705) (2021) e258.
[32] ResApp Health Limited, "ResApp announces positive results for a new novel smartphone-based COVID-19 screening test," Brisbane, Australia, 2022. https://www.resapphealth.com.au/wp-content/uploads/2022/03/2358427.pdf (Accessed 19 April 2023).
[33] ResApp Health Limited, "Results from data confirmation study," Brisbane, Australia, 2022. https://www.resapphealth.com.au/wp-content/uploads/2022/06/2396080.pdf (Accessed 19 April 2023).
[34] A. Wong, et al., External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients, JAMA Intern. Med. 181 (8) (2021) 1065–1070.
[35] 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.
[36] 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.
[37] P. Webster, Virtual health care in the era of COVID-19, Lancet 395 (10231) (2020) 1180–1181.