

Digital auscultation as a diagnostic aid to detect childhood pneumonia: A systematic review



Salahuddin Ahmed^{1,2}, Saima Sultana², Ahad M Khan^{1,2}, Mohammad S Islam^{1,3}, GM Monsur Habib^{1,4}, Ian M McLane⁵, Eric D McCollum^{6,7}, Abdullah H Baqui⁷, Steven Cunningham⁸, Harish Nair¹

¹Usher Institute, University of Edinburgh, Edinburgh, UK

²Projahnmo Research Foundation, Dhaka, Bangladesh

³Child Health Research Foundation, Dhaka, Bangladesh

⁴Bangladesh Primary Care Respiratory Society, Khulna, Bangladesh

⁵Sonavi Labs, Baltimore, Maryland, USA

⁶Global Program for Pediatric Respiratory Sciences, Eudowood Division of Paediatric Respiratory Sciences, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

⁷Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

⁸Department of Child Life and Health, Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK

Correspondence to:

Salahuddin Ahmed, MBBS
Centre for Global Health Research
Usher Institute
30 West Richmond Street
The University of Edinburgh
EH8 9DX
Edinburgh
United Kingdom
S.Ahmed-18@sms.ed.ac.uk

Background Frontline health care workers use World Health Organization Integrated Management of Childhood Illnesses (IMCI) guidelines for child pneumonia care in low-resource settings. IMCI guideline pneumonia diagnostic criterion performs with low specificity, resulting in antibiotic overtreatment. Digital auscultation with automated lung sound analysis may improve the diagnostic performance of IMCI pneumonia guidelines. This systematic review aims to summarize the evidence on detecting adventitious lung sounds by digital auscultation with automated analysis compared to reference physician acoustic analysis for child pneumonia diagnosis.

Methods In this review, articles were searched from MEDLINE, Embase, CINAHL Plus, Web of Science, Global Health, IEEEExplore database, Scopus, and the ClinicalTrials.gov databases from the inception of each database to October 27, 2021, and reference lists of selected studies and relevant review articles were searched manually. Studies reporting diagnostic performance of digital auscultation and/or computerized lung sound analysis compared against physicians' acoustic analysis for pneumonia diagnosis in children under the age of 5 were eligible for this systematic review. Retrieved citations were screened and eligible studies were included for extraction. Risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. All these steps were independently performed by two authors and disagreements between the reviewers were resolved through discussion with an arbiter. Narrative data synthesis was performed.

Results A total of 3801 citations were screened and 46 full-text articles were assessed. 10 studies met the inclusion criteria. Half of the studies used a publicly available respiratory sound database to evaluate their proposed work. Reported methodologies/approaches and performance metrics for classifying adventitious lung sounds varied widely across the included studies. All included studies except one reported overall diagnostic performance of the digital auscultation/computerized sound analysis to distinguish adventitious lung sounds, irrespective of the disease condition or age of the participants. The reported accuracies for classifying adventitious lung sounds in the included studies varied from 66.3% to 100%. However, it remained unclear to what extent these results would be applicable for classifying adventitious lung sounds in children with pneumonia.

Conclusions This systematic review found very limited evidence on the diagnostic performance of digital auscultation to diagnose pneumonia in children. Well-designed studies and robust reporting are required to evaluate the accuracy of digital auscultation in the paediatric population.

Pneumonia is the leading cause of mortality among infectious diseases in children under the age of five globally [1], accounting for an estimated 800 000 deaths per year, more than half of which occur in just five low- and middle-income countries (LMIC) [1,2]. In 2015, the estimated annual incidence of pneumonia in children under five years of age in developing countries was 231 episodes per 1000 children, resulting in about 138 million episodes of clinical pneumonia in this age group [3]. Prompt recognition of illness and care-seeking is critical to reducing pneumonia-related child deaths [4].

Currently, health care providers in LMICs use practical, standardized case management guidelines called the Integrated Management of Childhood Illnesses (IMCI) guidelines developed by the World Health Organization (WHO) for childhood pneumonia care [5,6]. IMCI guidelines have proven to be one of the most important childhood pneumonia interventions for LMICs to date [7,8].

IMCI guidelines prioritize sensitivity over specificity to ensure antibiotic treatment for children with an acute respiratory illness and suspected bacterial pneumonia [6]. Where successfully implemented, this algorithm has shown a 30%-40% reduction in case mortality [8]. Yet, IMCI has moderate sensitivity and a specificity largely contingent on disease severity. Specificity may range from 16% for children with an acute respiratory illness characterized by wheeze [9], to 49% for IMCI-defined non-severe pneumonia (ie, disease lacking clinical danger signs), to 95% for IMCI-defined very severe pneumonia (ie, disease with clinical danger signs) [10]. Limitations of IMCI specificity are thought to be related to the attribution of milder disease to viral pathogens, an epidemiological pattern catalysed by the expanded exposure of children in LMICs to *Haemophilus influenzae* type B and access to pneumococcal conjugate vaccines [11-13]. As a result, while the guidelines ensure few children with true bacterial pneumonia will be overlooked, most children, especially those with a milder illness, receive antibiotics inappropriately, resulting in antibiotic overuse.

Auscultation, the process of listening to the human body's internal sounds by using a stethoscope [14], has been an effective tool for diagnosing pulmonary diseases for more than two centuries. This requires highly trained health professional, limiting its utility at first-level facilities in LMICs staffed by non-physician health workers. IMCI guidelines do not include lung auscultation in their pneumonia definition for frontline health care workers [6]. The guidelines' exclusion of auscultation likely results from its high inter-observer variability and subjectivity, regardless of health care providers' training level [15-19]. Furthermore, traditional stethoscopes are functionally limited since they attenuate higher frequency sounds, like wheezing and crackles, yet transmit ambient noises and tubular resonance effects [17,20]. Digital auscultation may overcome these limitations [21]. Digital auscultation has the potential advantages of signal amplification and ambient noise reduction [22-24], reducing inter-observer reporting variation [25], and not requiring auscultation training for health care providers. Digital stethoscopes can record a patient's respiratory sounds for automated computerised lung sound analysis, which has been the recent focus of a significant amount of research, with some commercial systems already available [26,27]. These advantages may be important in LMICs as they rely on varying cadres of non-physician clinicians, and as the burden of lower respiratory infections among children in LMICs is high.

To the best of our knowledge, there is currently no systematic review reporting the diagnostic accuracy of digital auscultation to detect adventitious lung sounds in childhood pneumonia. Therefore, we aimed to conduct a systematic review to summarize the current evidence on the diagnostic performance of digital auscultation to identify adventitious lung sounds compared to the physician's acoustic analysis to diagnose pneumonia in children.

METHODS

This systematic review was reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy (PRISMA-DTA) criteria [28]. The review protocol was registered with the PROSPERO database (registration number CRD42020180821).

Information sources and search strategy

Search terms ("pneumonia", "lower respiratory infection", "auscultation", "respiratory sound", "lung sound", "digital", "electronic", "automatic", "computerized", "crackles", "wheeze", "child") were used to generate comprehensive search strategies for the following electronic databases: MEDLINE, Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL) Plus, Web of Science, Global Health, IEEEExplore database, Scopus, and ClinicalTrials.gov (<http://clinicaltrials.gov/>). Search strategies for all databases are shown in Appendix S1 in the **Online Supplementary Document**. The initial search was conducted in February 2019 and was updated on October 27, 2021, with no language restrictions. To avoid missing relevant studies, reference lists of identified studies and relevant reviews were also screened.

Eligibility criteria

Studies were selected based on the following inclusion criteria:

1. Participants: Children under five years of age.
2. Index test: Digital auscultation/computerized analysis of lung sounds.
3. Reference standards: Physicians' diagnosis of adventitious lung sounds (crackles and/or wheeze) by conventional auscultation/acoustic analysis of recorded lung sounds.
4. Target condition: Pneumonia.
5. Outcome: Reported diagnostic accuracy measures of the digital/computerized analysis of adventitious lung sounds.
6. Types of studies: Observational and experimental studies.

Studies were excluded if 1) digital auscultation or computerized analysis of lung sounds were not used in the study, 2) reference standard was not a human classification of lung sounds, 3) the reports were reviews, conference proceedings, abstracts, case reports, editorials, and commentaries.

Screening and selection of studies

Search results were imported and merged into Endnote X9, and duplicates were removed. Two review authors (SA and MSI/SS/GMMH) independently screened the titles and abstracts of all included citations as per the predefined eligibility criteria, followed by a full-text review of all selected articles (SA and SS/AMK). Any discrepancies were resolved through discussion, or an arbiter (HN) was consulted, where necessary. Reasons for exclusion of studies were documented (Appendix S2 in the [Online Supplementary Document](#)).

Data extraction and quality assessment

Data extractions were independently carried out by two review authors (SA and AMK) using a standardized pretested data collection template (Appendix S3 in the [Online Supplementary Document](#)). Any discrepant judgments were evaluated by an arbiter (HN).

Two reviewers (SA and AMK) independently assessed the risk of bias and the applicability of the included studies using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool [29]. Non-consensus between the reviewers were resolved through consultation with an arbiter (HN). This tool includes four domains to judge bias in the included study: patient selection, index test, reference standard, and flow and timing. A study would have an overall judgment of “low risk of bias” if it was judged as “low” on all domains. In contrast, it would be judged as a “high risk of bias” if it was judged “high” in one or more domains. The “unclear risk of bias” was categorised only when insufficient data were reported. To judge the concerns regarding the applicability of the included study to the review question, three domains (ie, patient selection, index test, and reference standard) were used.

Data synthesis

A descriptive synthesis was performed following the predefined review objectives and outcome measures. Meta-analyses could not be performed due to insufficient data and the heterogeneity in included studies in terms of methodology and outcome measures.

RESULTS

The review process is summarised in [Figure 1](#) using the PRISMA flow diagram [30]. A total of 3798 citations were identified through the database search. After duplicates were removed, 3324 unique citations remained. In total, 3281 citations were excluded during the title and abstract screening, and 46 full-text articles (including three from citation searching) were reviewed. Of these, 10 articles were eligible for inclusion [21,31-39].

Characteristics of the included studies

Table 1 summarises the characteristics of the included studies. Of the selected articles, five studies evaluated lung sound recordings from primary studies, among which one study was conducted across multiple centres in Africa (The Gambia, Kenya, South Africa, Zambia) and Asia (Bangladesh, Thailand) [21], one in Australia [36], one in India [37], and the remaining two were from the same study conducted in Egypt [31,35]. These five studies recruited children from different age groups and varied from one another in the number of study

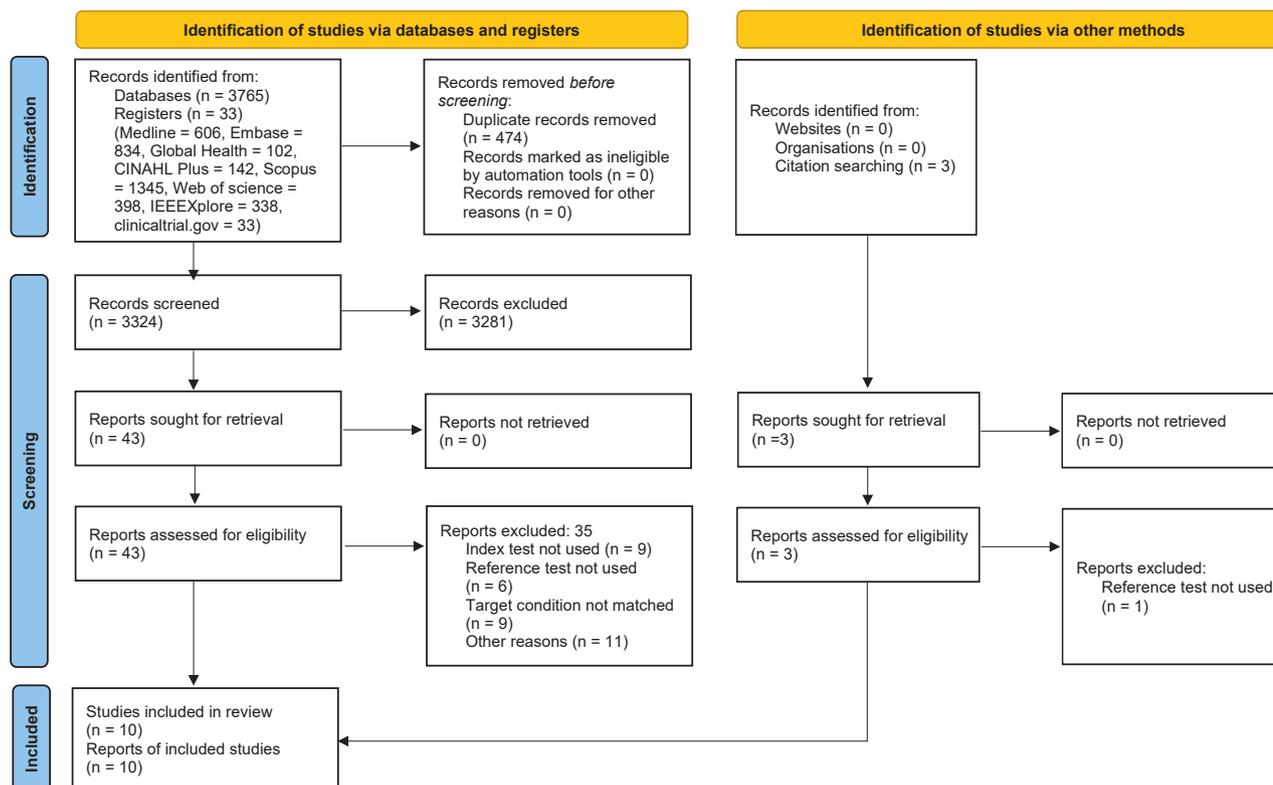


Figure 1. PRISMA flow diagram.

subjects. For instance, one study enrolled 1157 children aged 1-59 months [21], one study enrolled 600 children aged 0-12 years [31,35], while another study had only 20 children aged 4.6-17.1 years [36]. However, one study recruited 256 children but did not report any information on the participants' age [37]. Of these five studies, one study obtained lung sound recordings in outpatient and busy clinical settings [21], while another study recorded the lung sounds in a quiet room in the hospital [36]. Three studies also recorded lung sounds in a hospital setting [31,35,37], but is unclear whether they were recorded in a controlled environment or a noisy clinical setting. The other five studies [32-34,38,39] evaluated their proposed work using the publicly available International Conference on Biomedical and Health Informatics (ICBHI) scientific challenge respiratory sound database containing 920 annotated lung sound recordings from 126 subjects (49 children and 77 adults) [40,41].

We found only one study that specifically recruited children with IMCI-defined severe or very severe pneumonia and age-matched controls without clinical pneumonia [21], while seven studies included subjects with a variety of respiratory diseases, including pneumonia [32-34,36-39]. The other two studies did not report on the disease condition of the study subjects [31,35,37].

Stethoscopes used

For studies using the ICBHI data set, lung sounds were recorded using different devices (3M Littmann Classic II SE stethoscope, 3M Littmann 3200 electronic stethoscope, WelchAllyn Meditron Master Elite electronic stethoscope, and AKG C417 L Microphone) [32-34,38,39] from both clinical and home settings [41]. In the other four studies, a 3M Littmann 3200 electronic stethoscope was used [31,35-37], while the Clinicloud DS was also used in one of the studies [36]. The ThinkLabs ds32a digital stethoscope with a microphone (Sony-ICD-UX71-81) affixed to the stethoscope to record ambient noises was used in one study [21].

Lung sounds classification used

Variations were observed in classifying the lung sounds. For instance, three studies specifically performed the computerized analysis of lung sounds to classify normal and adventitious/pathological sounds [21,33,37], two studies performed a four-class classification (normal, crackle, wheeze, both wheeze and crackle) [32,38], one study performed a three-class classification (normal, crackle, and wheeze) [34], one study classified wheeze

Table 1. Study characteristics

AUTHOR, YEAR	COUNTRY	STUDY TYPE	POPULATION	NUMBER OF SUBJECTS	CLINICAL CONDITION OF THE SUBJECTS	SOUND/ PATHOLOGY ANALYZED	NUMBER OF RECORDINGS STUDIED	LUNG SOUND RECORDING DEVICE	FEATURE EXTRACTION METHOD	SOUND CLASSIFICATION METHOD	OUTCOME/RESULT
Fasseeh et al. 2015	Egypt	Case-control	Infants and children (0-12 y)	Case=500 Control=100	Not reported	Wheeze, stridor, rattle, normal	592	3M Littmann Electronic Stethoscope 3200	Short-time Fourier Transform (STFT)	Dynamic time warping (DTW) algorithm	Reported accuracy of validation for all wheezes was 81.93% (<12 mo=82.81% & ≥ 12 months=89.15%). Reported accuracy of validation for all normal sounds was 89% (<12 mo=96.15% & ≥ 12 months=90.54%).
Khan et al. 2017	India		Children	254	Not reported	Normal and pathological	254	Littmann 3200 electronic stethoscope	Short time Fourier transform (STFT)	k- Nearest Neighbour (k-NN); Support Vector Machine (SVM)	k-NN obtained sensitivity, specificity, and accuracy of 90.9%, 92.2% and 91.6%, respectively. SVM obtained sensitivity, specificity, and accuracy of 92.2%.
Kevat et al. 2017	Australia		Children (median age=6.7 y)	20	Cystic fibrosis, lower respiratory tract infection, asthma, preschool wheeze	Wheeze, crackles, normal	156	Littmann 3200 Electronic Stethoscope; Clinicloud DS	Not reported	Audio spectrographic analysis	Concordance between the Littmann electronic stethoscope and standard auscultation was found to be moderate for wheeze ($\kappa=0.44$) and almost perfect for crackles ($\kappa=1.0$). Concordance between the Clinicloud DS and standard auscultation was found to be moderate for wheeze ($\kappa=0.55$) and almost perfect for crackles ($\kappa=1.0$).
Abougabal et al. 2018	Egypt	Analysis of recorder lung sounds	Infants and Children (<13 y)	600	Not reported	Stridor, rattle and wheeze, normal	592	3M Littmann Electronic Stethoscope 3200	Wavelet Transform (WT) coefficients	Dynamic time warping (DTW) algorithm	Reported recognition accuracy of 88.2% for wheeze and 86% for normal breath sounds.
Emmanouilidou et al. 2018	Gambia, Kenya, South Africa, Zambia, Bangladesh, Thailand	Case-control	Children (median age 7 ± 11.4 mo)	1157	Pneumonia, Normal	Normal, abnormal (wheeze and/or crackle)	1095	ThinkLabs ds32a	Rich spectro-temporal feature space	Support-Vector Machine (SVM) classifier	The classification system achieved an accuracy of 86.7%, sensitivity of 86.9%, and specificity of 86.6%.
Chen et al. 2019		Analysis of respiratory sound database*	Not mentioned	Not mentioned	Not mentioned	Wheeze, crackle, normal	489	3M Littmann 3200 Electronic Stethoscope; Welch Allyn Elite Meditron	Optimized-S-Transform (OST)	Deep Residual Networks (ResNets)	Classification accuracy of 98.79% with sensitivity of 96.27% and specificity of 100% was obtained to classify wheeze, crackle, and normal sounds.
Perna et al. 2019		Analysis of respiratory sound database*	Children, adults	126	Pneumonia, Bronchiectasis, Bronchiolitis, COPD, Healthy, URTI	Normal, wheeze, crackle, both	920	AKG C417 L Microphone; 3M Littmann Classic II SE. Stethoscope; 3M Littmann 3200 Electronic Stethoscope; Welch Allyn Meditron Master Elite Electronic Stethoscope	Mel-Frequency Cepstral Coefficients (MFCC)	Recurrent Neural Networks (RNN) models: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)	A sensitivity of 0.62 and specificity of 0.85 were reported for LSTM based model on four class anomaly driven prediction. (ie, normal, presence of crackles, presence of wheezes, presence of both)

Table 1. continued

AUTHOR, YEAR	COUNTRY	STUDY TYPE	POPULATION	NUMBER OF SUBJECTS	CLINICAL CONDITION OF THE SUBJECTS	SOUND/ PATHOLOGY ANALYZED	NUMBER OF RECORDINGS STUDIED	LUNG SOUND RECORDING DEVICE	FEATURE EXTRACTION METHOD	SOUND CLASSIFICATION METHOD	OUTCOME/RESULT
Acharya et al. 2020		Analysis of respiratory sound database*	Children, adults	126	Pneumonia, Bronchiectasis, Bronchiolitis, COPD, Healthy, URTI	Normal, wheeze, crackle, both	920	AKG C417 L Microphone; 3M Littmann Classic II SE. Stethoscope; 3M Littmann 3200 Electronic Stethoscope; Welch Allyn Meditron Master Elite Electronic Stethoscope	Mel-spectrograms	Hybrid CNN-RNN model	A score of 66.31% was obtained on four class respiratory cycle classification and a score of a 71.81% was obtained for leave-one-out validation.
Alva Alicia et al. 2021		Analysis of respiratory sound database*	Children, adults	126	Pneumonia, Bronchiectasis, Bronchiolitis, COPD, Healthy, URTI	Normal, abnormal	920	AKG C417L Microphone; 3M Littmann Classic II SE. Stethoscope; Littmann 3200 Electronic Stethoscope; Welch Allyn Meditron Master Elite Electronic Stethoscope	Mel-spectrogram; Short time Fourier transform (STFT); Mel-Frequency Cepstral Coefficients (MFCC)	Convolutional Neural (CNN) Network Models	Accuracy values of 0.998 and 1 were obtained for normal sounds and abnormal sounds respectively. Accuracy values of 0.9959 and 0.9885 were reported for classification of pneumonia and other diseases.
Shuvo et al. 2021		Analysis of respiratory sound database*	Children, adults	87	Pneumonia, Bronchiectasis, Bronchiolitis, COPD, Healthy, URTI	Chronic classification (healthy, chronic diseases, non-chronic diseases) Pathological classification (Healthy, Bronchiectasis, Bronchiolitis, COPD, Pneumonia, URTI)	917	AKG C417L Microphone; 3M Littmann Classic II SE. Stethoscope; Littmann 3200 Electronic Stethoscope; Welch Allyn Meditron Master Elite Electronic Stethoscope	Hybrid scalogram using empirical mode decomposition (EMD) & continuous wavelet transform (CWT)	Lightweight Convolutional neural network (CNN) model	Weighted accuracy scores of 98.92% for three-class chronic classification and 98.70% for six-class pathological classification were obtained.

*The study analysed International Conference on Biomedical Health Informatics (ICBHI) 2017 data set.

and crackles [36] and another two analysed wheeze, rattle, stridor, and normal lung sounds [31,35]. One study performed a three-class chronic classification (healthy, non-chronic diseases, chronic diseases) and six-class pathological classification (healthy, bronchiectasis, bronchiolitis, COPD, Pneumonia, URTI) based on the respiratory sound signal processing and computerised analysis [39].

Algorithm used to classify lung sounds

A substantial variation was observed in feature extraction methods (ie, the process of identifying distinctive features of respiratory sound signals) and sound classification algorithms/models used in the included studies. For feature extraction, three studies employed Short-time Fourier Transform (STFT) [33,35,37], two studies employed Melspectrogram [32,33], two used Mel Frequency Cepstral Coefficients (MFCC) [33,38], while the other four studies employed enhanced distinctive feature extraction methods –Wavelet Transform (WT) coefficients [31], rich spectro-temporal feature space [21], Optimized-S-Transform (OST) [34], and hybrid scalogram using empirical mode decomposition (EMD) and continuous wavelet transform (CWT) [39]. Wide-ranging sound classification models/algorithms used across the studies are: Dynamic time warping (DTW) algorithm [31,35], Support Vector Machine (SVM) [21,37], k-Nearest Neighbour (k-NN) based classifier [37], audio spectrographic analysis [36], Deep Residual Networks (ResNets) [34], Recurrent Neural Networks (RNN) models with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [38], hybrid CNN-RNN model [32], Convolutional Neural Network (CNN) models [33], and lightweight convolutional neural network (CNN) model [39].

Diagnostic performance of used algorithms/models

Regarding diagnostic performance, diverse methodologies and diagnostic accuracy measures were used. Furthermore, in studies that included participants with different target conditions and age groups (ie, children and adults), the reported performance measures were only limited to overall accuracy on adventitious lung sounds classification [32-34,38] or by pathological conditions classification [39]. For these studies, reported accuracies varied from 66.3% to 100%. Of the five studies that recruited children [21,31,35-37], only one study [21] reported on the accuracy of the computerized sound classification system in children aged 1-59 months in pneumonia. In that study, the accuracy of the classification system was 86.7% (sensitivity=86.8%; specificity=86.6%) to differentiate between normal and adventitious lung sounds. Further, the researchers suggested that the system performance varied with the different window analyses of breath cycles ranging

from shorter to longer duration. For example, setting the analysis window at 0.5s yielded an accuracy of 84.1% (sensitivity=87.2%; specificity=81%) while it was about 77% in longer time windows (>1s); thus, a short window size was recommended [21].

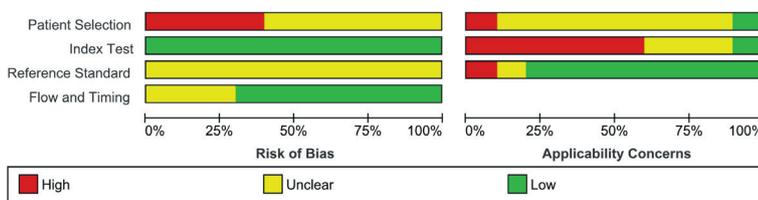


Figure 2. Risk of bias and applicability concerns graph: review authors’ judgements about each domain presented as percentages across included studies.

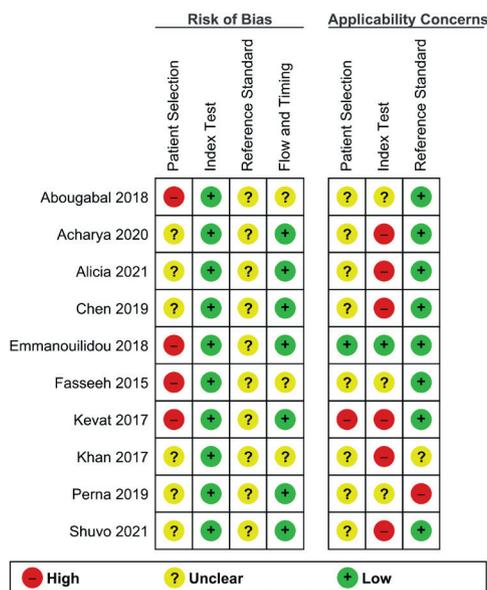


Figure 3. Risk of bias and applicability concerns summary: review authors’ judgements about each domain for each included study.

Assessment of risk of bias and applicability

The quality of the included studies according to the QUADAS-2 tool is summarized and displayed graphically in Figure 2 and Figure 3. In general, very few of the included studies met most of the quality criteria. Most of the studies were judged to be of unclear methodological quality because of insufficient reporting. For patient selection, four studies were evaluated as having a high risk of bias, and six were evaluated as having an unclear risk of bias due to inappropriate or poorly described sampling methods. For the index test, all 10 studies were judged as having low risk of bias because there was no chance of non-blinding, as index tests were the machine classifications. For the reference standard, all the studies were evaluated as having an unclear risk of bias due to poor reporting of target conditions and/or blinding status. In the flow and timing domain assessment, seven studies were deemed to have a low risk of bias as the recorded sounds were the second interpretation, and three were evaluated as an unclear risk because of underreporting. The overall concerns regarding applicability for this review were unclear, high, and low, for patient selection, index test, and reference standard, respectively.

DISCUSSION

This systematic review included studies evaluating the diagnostic performance of digital auscultation/computerized analysis of adventitious lung sounds in pneumonia in children aged under 5 years. The literature search identified 10 articles that met the inclusion criteria. Of these included articles, five of them evaluated lung sound recordings from primary studies, while the remaining five utilised a publicly available respiratory sound database to evaluate their proposed lung sound classification system. Substantial variation was observed in the included studies in terms of study subjects, sample size, and use of feature extraction methods and sound classification models. For instance, seven types of feature extraction methods and nine types of sound classification models/algorithms were used across the studies. Hence, we could not draw any conclusion from this systematic review. However, the reported accuracies for classifying adventitious lung sounds ranged from 66.3% to 100%. This review identified only one study [21] that involved children aged under 5 years with pneumonia to evaluate an integrated computerized lung sound classification framework to detect adventitious lung sounds at an accuracy of 86.7% (sensitivity=86.5%; specificity=86.6%). Although the rest of the studies demonstrated excellent accuracy, sensitivity, and specificity values, their results were primarily limited to overall diagnostic accuracy parameters in classifying lung sounds and did not provide disaggregated results by age and/by target condition/s. Therefore, the findings could not be directly applied to the paediatric population to diagnose childhood pneumonia. These shortfalls are important findings of this review and suggest a need for improved reporting of study findings. The authors should account for participants' contributing factors (such as age, disease/target condition) while analysing and reporting their data. Evidently, the methodological quality of the included studies was deemed to be unclear due to insufficient reporting in participant selection, sampling methods, and clinical/target conditions of the participants. The concerns regarding the applicability of the patient selection and index test domain were unclear and high, respectively.

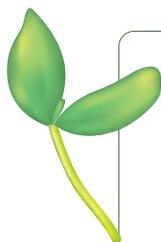
To the best of our knowledge, this is the first systematic review assembling evidence on the discriminatory power of digital auscultation for the detection of adventitious lung sound/s against conventional auscultation or acoustic analysis of recorded lung sounds by physicians in childhood pneumonia diagnosis. Prior reviews focused on summarizing existing evidence on artificial intelligence and algorithms to classify adventitious lung sounds [27,42]. Another systematic review and meta-analysis by Gurung et al. [20] evaluated the performance of computerized lung sounds analysis to detect adventitious lung sounds in respiratory diseases against chest radiography or clinical diagnosis.

In line with our findings, this review emphasised methodological and analytical standardization, including completeness and transparency in reporting by following standardized guidelines, and the need for conducting more studies on the paediatric population. To date, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 statement remains the most used tool for reporting of studies investigating diagnostic test accuracy and performance [43,44]. However, this tool has some shortcomings when reporting studies evaluating artificial intelligence (AI) driven interventions due to unclear methodological interpretation, lack of standardised nomenclature, use of unfamiliar outcome measures, and other issues, thereby limiting the comprehensive appraisal of these technologies [45]. Thus, our study findings further reiterate the need for developing an AI-specific STARD guideline to ensure complete and robust reporting of studies evaluating AI-driven technologies and interventions [45,46].

The heterogeneous nature of the included studies and insufficient data prohibited us from performing meta-analyses and drawing conclusions on the diagnostic accuracy of digital auscultation and computerized analysis of lung sounds in childhood pneumonia diagnosis.

CONCLUSIONS

Given the paucity of available data in this area, there is a need for additional well-designed studies to generate evidence on accuracy, sensitivity, and specificity of digital auscultation and/or computerized analysis of lung sounds involving the paediatric population to diagnose pneumonia. Future investigations should also consider conducting the studies in real-life, noisy clinical settings rather than in a controlled laboratory or clinical environment to broaden the usability of the automated applications. This is a rapidly evolving field, and new and advanced applications with robust reporting of methods and findings within studies will enable meta-analysis in future reviews.



Acknowledgements: We are grateful to Marshall Dozier, the Academic Support Librarian at the University of Edinburgh, UK for her assistance in developing the search strategies.

Funding: This research was funded by the UK National Institute for Health Research (NIHR) (Global Health Research Unit on Respiratory Health (RESPIRE); 16/136/109) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Government. SA, AMK, MSI, GMMH are supported by PhD studentships from the NIHR Global Health Research Unit on Respiratory Health (RESPIRE), the University of Edinburgh. RESPIRE is funded by the National Institute of Health Research using Official Development Assistance (ODA) funding. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The RESPIRE collaboration comprises the UK Grant holders, Partners and research teams as listed on the RESPIRE website (www.ed.ac.uk/usher/respire) including Hilary Pinnock.

Authorship contributions: SA, HN, EDM, and SC conceptualised this systematic review. SA, MSI, SS, AMK, and GMMH obtained and appraised data with discrepancies resolved by HN. SA, SS, and AMK prepared tables and figures presented in the manuscript. SA prepared the first draft. All co-authors reviewed the manuscript and provided intellectual inputs and approved the final version.

Competing interests: The authors completed the ICMJE Unified Competing Interest form (available upon request from the corresponding author) and declare no conflicts of interest.

Additional material

Online Supplementary Document

REFERENCES

- 1 Every child's right to survive: A 2020 agenda to end pneumonia deaths. 2020. Available: <https://www.unicef.org/media/64701/file/Every-childs-right-to-survive-pneumonia-2020.pdf>. Accessed: 2 November 2021.
- 2 UNICEF Save the Children. Fighting for Breath Call to Action: End childhood pneumonia deaths. 2018. Available: <https://www.savethechildren.org.uk/content/dam/gb/reports/policy/fighting-for-breath-call-to-action.pdf>. Accessed: 2 November 2021.
- 3 McAllister DA, Liu L, Shi T, Chu Y, Reed C, Burrows J, et al. Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. *Lancet Glob Health*. 2019;7:e47-57. Medline:30497986 doi:10.1016/S2214-109X(18)30408-X
- 4 Wardlaw TM, Johansson EW, Hodge MJ. Pneumonia: the forgotten killer of children: New York: UNICEF; 2006.
- 5 World Health Organization. Pocket book of hospital care for children: guidelines for the management of common childhood illnesses: World Health Organization; 2013.
- 6 World Health Organization. WHO IMCI Chart booklet 2014. Geneva: World Health Organisation. 2014.
- 7 Theodoratou E, Al-Jilaihawi S, Woodward F, Ferguson J, Jhass A, Balliet M, et al. The effect of case management on childhood pneumonia mortality in developing countries. *Int J Epidemiol*. 2010;39:i155-71. Medline:20348118 doi:10.1093/ije/dyq032
- 8 Sazawal S, Black RE, Pneumonia Case Management Trials G. Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials. *Lancet Infect Dis*. 2003;3:547-56. Medline:12954560 doi:10.1016/S1473-3099(03)00737-0
- 9 Cardoso M-RA, Nascimento-Carvalho CM, Ferrero F, Alves FM, Cousens SN. Adding fever to WHO criteria for diagnosing pneumonia enhances the ability to identify pneumonia cases among wheezing children. *Arch Dis Child*. 2011;96:58-61. Medline:20870628 doi:10.1136/adc.2010.189894
- 10 Puumalainen T, Quiambao B, Abucejo-Ladesma E, Lupisan S, Heiskanen-Kosma T, Ruutu P, et al. Clinical case review: a method to improve identification of true clinical and radiographic pneumonia in children meeting the World Health Organization definition for pneumonia. *BMC Infect Dis*. 2008;8:95. Medline:18644109 doi:10.1186/1471-2334-8-95
- 11 Goodman D, Crocker ME, Pervaiz F, McCollum ED, Steenland K, Simkovich SM, et al. Challenges in the diagnosis of paediatric pneumonia in intervention field trials: recommendations from a pneumonia field trial working group. *Lancet Respir Med*. 2019;7:1068-83. Medline:31591066 doi:10.1016/S2213-2600(19)30249-8
- 12 Mackenzie G. The definition and classification of pneumonia. *Pneumonia*. 2016;8:14. Medline:28702293 doi:10.1186/s41479-016-0012-z
- 13 Mulholland K. Problems with the WHO guidelines for management of childhood pneumonia. *Lancet Glob Health*. 2018;6:e8-9. Medline:29241619 doi:10.1016/S2214-109X(17)30468-0
- 14 Earis J. Lung sounds. *Thorax*. 1992;47:671. Medline:1440457 doi:10.1136/thx.47.9.671
- 15 Margolis PA, Ferkol TW, Marsocci S, Super DM, Keyes LL, McNutt R, et al. Accuracy of the clinical examination in detecting hypoxemia in infants with respiratory illness. *J Pediatr*. 1994;124:552-60. Medline:8151469 doi:10.1016/S0022-3476(05)83133-6
- 16 Murphy RL, Vshedskiy A, Power-Charnitsky VA, Bana DS, Marinelli PM, Wong-Tse A, et al. Automated lung sound analysis in patients with pneumonia. *Respir Care*. 2004;49:1490-7. Medline:15571639
- 17 Grenier MC, Gagnon K, Genest J Jr, Durand J, Durand LG. Clinical comparison of acoustic and electronic stethoscopes and design of a new electronic stethoscope. *Am J Cardiol*. 1998;81:653-6. Medline:9514471 doi:10.1016/S0002-9149(97)00977-6
- 18 Brooks D, Thomas J. Interrater reliability of auscultation of breath sounds among physical therapists. *Phys Ther*. 1995;75:1082-8. Medline:7501711 doi:10.1093/ptj/75.12.1082

- 19 Gjhrup T, Bugge PM, Jensen AM. Interobserver variation in assessment of respiratory signs. Physicians' guesses as to interobserver variation. *Acta Med Scand.* 1984;216:61-6. Medline:6485882 doi:10.1111/j.0954-6820.1984.tb03772.x
- 20 Gurung A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med.* 2011;105:1396-403. Medline:21676606 doi:10.1016/j.rmed.2011.05.007
- 21 Emmanouilidou D, McCollum ED, Park DE, Elhilali M. Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments. *IEEE Trans Biomed Eng.* 2018;65:1564-74. Medline:28641244 doi:10.1109/TBME.2017.2717280
- 22 Leng S, San Tan R, Chai KTC, Wang C, Ghista D, Zhong L. The electronic stethoscope. *Biomed Eng Online.* 2015;14:66. Medline:26159433 doi:10.1186/s12938-015-0056-y
- 23 McLane I, Emmanouilidou D, West JE, Elhilali M. Design and Comparative Performance of a Robust Lung Auscultation System for Noisy Clinical Settings. *IEEE J Biomed Health Inform.* 2021;25:2583-94. Medline:33534721 doi:10.1109/JBHI.2021.3056916
- 24 Nelson G, Rajamani R. Accelerometer-based acoustic control: Enabling auscultation on a black hawk helicopter. *IEEE/ASME Trans Mechatron.* 2016;22:994-1003. doi:10.1109/TMECH.2016.2603444
- 25 Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: advances beyond the stethoscope. *Am J Respir Crit Care Med.* 1997;156:974-87. Medline:9310022 doi:10.1164/ajrccm.156.3.9701115
- 26 Andrés E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. *J Med Life.* 2018;11:89. Medline:30140315
- 27 Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS One.* 2017;12:e0177926. Medline:28552969 doi:10.1371/journal.pone.0177926
- 28 Salameh J-P, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): explanation, elaboration, and checklist. *BMJ.* 2020;370:m2632. Medline:32816740 doi:10.1136/bmj.m2632
- 29 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-36. Medline:22007046 doi:10.7326/0003-4819-155-8-201110180-00009
- 30 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. Medline:33782057
- 31 Abougabal MM, Diao Moussa N. A novel technique for validating diagnosed respiratory noises in infants and children. *Alex Eng J.* 2018;57:3033-41. doi:10.1016/j.aej.2018.05.003
- 32 Acharya J, Basu A. Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. *IEEE Trans Biomed Circuits Syst.* 2020;14:535-44. Medline:32191898 doi:10.1109/TBCAS.2020.2981172
- 33 Alva A, Alexander A-G, William C, Cieza Terrones M, Victor H-A, Sebastian R-C. Abnormal Pulmonary Sounds Classification Algorithm using Convolutional Networks. *Int J Adv Comput Sci Appl.* 2021;12.
- 34 Chen H, Yuan X, Pei Z, Li M, Li J. Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks. *IEEE Access.* 2019;7:32845-52. doi:10.1109/ACCESS.2019.2903859
- 35 Fasseeh NA, Heiba DA, Abougabal MM, Meguid BSA, Rizk M. Towards Validating Diagnosed Respiratory Sounds Using Dynamic Time Warping at Alexandria University Children Hospital (AUCH)—Egypt. *Life Sci J.* 2015;12.
- 36 Kevat AC, Kalirajah A, Roseby R. Digital stethoscopes compared to standard auscultation for detecting abnormal paediatric breath sounds. *Eur J Pediatr.* 2017;176:989-92. Medline:28508991 doi:10.1007/s00431-017-2929-5
- 37 Khan SI, Ahmed V, Jawarkar NP, editors. Application of signal processing techniques for preliminary detection of adventitious lung sounds in paediatric population using electronic stethoscope. 2017 IEEE 2nd Int Conf Big Data Anal Comput Intel. 2017.
- 38 Perma D, Tagarelli A, editors. Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks. 2019 IEEE 32nd Int Symp Comput-Based Med Syst; 2019.
- 39 Shuvo SB, Ali SN, Swapnil SI, Hasan T, Bhuiyan MIH. A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram. *IEEE J Biomed Health Inform.* 2021;25:2595-603. Medline:33373309 doi:10.1109/JBHI.2020.3048006
- 40 Rocha BM, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, et al, editors. A Respiratory Sound Database for the Development of Automated Classification. *Precision Medicine Powered by pHealth and Connected Health*; 2018; Singapore: Springer Singapore.
- 41 Rocha BM, Filos D, Mendes L, Serbes G, Ulukaya S, Kahya YP, et al. An open access database for the evaluation of respiratory sound classification algorithms. *Physiol Meas.* 2019;40:035001. Medline:30708353 doi:10.1088/1361-6579/ab03ea
- 42 Palaniappan R, Sundaraj K, Sundaraj S. Artificial intelligence techniques used in respiratory sound analysis—a systematic review. *Biomed Tech (Berl).* 2014;59:7-18. Medline:24114889 doi:10.1515/bmt-2013-0074
- 43 Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hoof L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6:e012799. Medline:28137831 doi:10.1136/bmjopen-2016-012799
- 44 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clin Chem.* 2015;61:1446-52. Medline:26510957 doi:10.1373/clinchem.2015.246280
- 45 Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hoof L, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. *BMJ Open.* 2021;11:e047709. Medline:34183345 doi:10.1136/bmjopen-2020-047709
- 46 Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med.* 2020;26:807-8. Medline:32514173 doi:10.1038/s41591-020-0941-1