

Performance of an artificial intelligence algorithm for interpreting lung sounds from children hospitalised with pneumonia in Malawi

Nadia E Hoekstra^{1,2}, Maganizo B Chagomerana^{3,4}, Zachary H Smith^{2,5}, Annapurna Kala⁶, Ian McLane⁶, Charl Verwey⁷, Daniel Olson⁸, W Chris Buck⁹, Justin Mulindwa¹⁰, Alex Gaudio⁶, Sunaina Kapoor², Holly B Schuh^{2,11}, Msandeni Chiume¹², Elizabeth Fitzgerald¹, Mounya Elhilali⁶, Tisungane Mvalo^{1,3}, Mina Hosseinipour^{3,4}, Eric D McCollum^{2,13}

¹Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

²Global Program in Pediatric Respiratory Sciences, Eudowood Division of Pediatric Respiratory Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

³University of North Carolina Project Malawi, Lilongwe, Malawi

⁴Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA

⁵Department of Pediatrics, Division of Pediatric Critical Care Medicine, Kaiser Permanente Northern California, Oakland, California, USA

⁶Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland, USA

⁷Department of Paediatrics and Child Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

⁸Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA

⁹University of California Los Angeles, David Geffen School of Medicine, Los Angeles, California, USA

¹⁰University Teaching Hospital, Lusaka, Zambia

¹¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

¹²Department of Paediatrics, Kamuzu Central Hospital, Lilongwe, Malawi

¹³Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Correspondence to:

Eric D McCollum, MD, MPH
Global Program in Pediatric Respiratory Sciences, Eudowood Division of Pediatric Respiratory Sciences, Johns Hopkins University School of Medicine
200 N. Wolfe St, Suite 3071, Baltimore, Maryland
USA
emccoll3@jhmi.edu

Background Pneumonia is a leading cause of death in under five year olds globally. World Health Organization (WHO) pneumonia diagnostic guidelines rely on non-specific clinical findings. Lung auscultation could improve pneumonia diagnosis, but conventional stethoscopes have implementation challenges. To address this, we developed an artificial intelligence (AI)-enabled digital auscultation system. We evaluated the system's AI lung sound analysis algorithm in children with severe pneumonia in Malawi.

Methods We enrolled children aged 2–59 months hospitalised with WHO-defined severe pneumonia. A study physician recorded lung sounds with a digital stethoscope at six chest positions. Recordings were de-identified, filtered, and interpreted by a trained and certified physician listening panel. Interpretable recordings were analysed by the AI algorithm. We evaluated the agreement of normal (absence of adventitial lung sounds) vs. abnormal (presence of adventitial lung sounds) classifications, by chest position and by patient, between the AI algorithm and the listening panel using raw percent agreement kappa statistics, both unadjusted and adjusted for chance agreement.

Results We enrolled 100 children and analysed 95 with interpretable recordings. The median age was 12.6 months (interquartile range (IQR)=5.4, 19.0) and 54% (51/95) were female. Among

interpretable recordings, 59.2% (294/497) of chest positions were abnormal per the listening panel compared to 52.7% (262/497) per the AI algorithm. The listening panel and AI algorithm agreed on classifications in 83.1% (413/497) of chest positions (unadjusted kappa 0.7; adjusted kappa 0.7) and 91.6% (87/95) of patients (unadjusted kappa 0.7; adjusted kappa 0.8). The AI algorithm's sensitivity and specificity for identifying abnormal lung sounds, compared to the listening panel, were 80.3% and 87.2% for chest positions and 96.3%, and 66.7% for patients.

Conclusions This AI lung sound classification algorithm accurately identified abnormal lung sounds in children with severe pneumonia. Next steps include training the algorithm to identify uninterpretable recordings and different abnormal sounds.

Pneumonia remains the leading infectious cause of mortality in children under five worldwide [1,2]. The World Health Organization (WHO) Integrated Management of Childhood Illnesses (IMCI) guidelines provide the current standard for managing paediatric pneumonia in low-income and middle-income countries (LMICs) [3]. Developed in the 1980s, the IMCI algorithm enables non-physician health care workers (HCWs) to identify children with likely bacterial pneumonia [3]. By increasing antibiotic treatment, IMCI has reduced paediatric pneumonia deaths over the past two decades [3,4]. In 2013, the IMCI approach was broadened to include hospitalised children and to strengthen diagnosis and treatment protocols [5]. Despite its success, recent studies have raised concerns about low diagnostic specificity due to reliance on clinical findings, such as chest wall retractions and elevated respiratory rates, which are seen in diseases beyond bacterial pneumonia [6–9]. Overdiagnosis exposes children to antibiotics, interventions, and avoidable costs, and delays treatment. This issue is critical given rising antimicrobial resistance and a shift towards viral respiratory diseases, largely due to effective vaccines against *Haemophilus influenzae* type B and *Streptococcus pneumoniae* [10–12]. These concerns highlight the need to develop new diagnostic technologies to improve IMCI's diagnostic performance.

Stethoscopes are a low-cost, non-invasive tool used in resource-rich settings to listen to and interpret lung sounds, and they have the potential to enhance IMCI's diagnostic accuracy. However, lung auscultation in children by HCWs remains subjective, with only moderate levels of agreement between examiners [13,14]. Lung auscultation in children is particularly challenging due to inconsistent patient cooperation, upper airway sound transmission, variable tidal volumes, and short respiratory cycles. Successful lung auscultation is challenging in clinical settings in LMICs where most paediatric patients are managed by non-physician HCWs [15]. In LMICs, there are limited numbers of HCWs who can effectively teach and perform this skill, and background noise in high-volume areas hinders auscultation.

Emerging technologies like digital stethoscopes and artificial intelligence (AI)-enabled lung sound analysis may overcome these barriers in paediatric auscultation. Digital stethoscopes, which transmit, filter, and amplify sounds, may help HCWs in LMICs better auscultate and evaluate lung sounds in children [16]. More importantly, AI lung sound analysis can classify lung sounds for HCWs with limited training, providing an automated solution to conventional lung auscultation challenges. Adventitious lung sounds, such as crackles and wheezes, are abnormal breath sounds that may indicate underlying respiratory pathology. These sounds have unique acoustic characteristics that can be identified by AI applications [17,18]. These tools could improve diagnostic accuracy and reduce the subjectivity associated with conventional stethoscopes. Although literature exists regarding AI classification of lung sounds, few studies focus on children in LMICs, where the pneumonia burden is highest [2].

To address this gap, in 2012 we embarked on a multi-disciplinary collaboration, involving paediatricians, paediatric pulmonologists, and sound engineers to develop a digital auscultation system designed for use in children in LMICs. The system consists of three elements: a prototype digital stethoscope, automated ambient noise filtering software, and an AI lung sound analysis algorithm. While we have previously reported on this system at various developmental stages,

we have yet to evaluate the AI lung sound analysis algorithm's performance on prospectively collected data from children in real-world LMIC contexts. Therefore, we aimed to evaluate the performance of the system's AI lung sound analysis algorithm, compared to a reference physician listening panel, in classifying lung sound recordings from hospitalised Malawian children aged 2–59 months with WHO-defined severe pneumonia.

METHODS

Study design and setting

This cross-sectional study was conducted at Kamuzu Central Hospital (KCH) in Lilongwe, Malawi. KCH is a government tertiary referral hospital for the Central Region of Malawi.

Enrollment and participant eligibility

Children aged two to 59 months admitted to the paediatric ward of KCH and diagnosed with severe pneumonia were eligible. A paediatrician and a nurse reviewed all patients admitted to the ward daily to determine study eligibility. Eligible children were hospitalised for less than 24 hours and met the WHO severe pneumonia definition. Severe pneumonia was defined as the presence of cough or difficulty breathing plus any of the following: oxygen saturation <90%, central cyanosis, severe respiratory distress with grunting or severe chest indrawing, or a general danger sign (inability to feed, lethargy, reduced level of consciousness, or convulsions) [4]. Exclusion criteria included wheezing that improved after bronchodilators, chronic lung disease other than asthma or reactive airways, medical instability, tracheostomy, or invasive or non-invasive ventilation.

Demographic, clinical and laboratory evaluation

At enrolment, demographic information, medical history, clinical data, and physical examination findings were obtained. All children received a malaria Rapid Diagnostic Test (RDT) and HIV counselling and testing as recommended by the Malawi Ministry of Health guidelines. Malaria RDT and HIV results (infected, uninfected, exposed, or unknown) were documented.

Lung sound auscultation and recording

An American Board of Pediatrics-eligible physician used a validated methodology to record lung sounds at six sequential anatomic chest positions, for at least 10 seconds per position (Figure S1 in the [Online Supplementary Document](#)), using a prototype digital stethoscope [19] connected to a Zoom H4n Pro Portable Recorder®. Each participant's recording, containing all six chest position recordings, was deidentified, uploaded to a secure computer server at the study site, and segmented by chest position using Audacity® audio editing software. Each chest position recording captured a minimum of three respiratory cycles. Participant factors during auscultation, such as motion and phonation, were documented by the study clinician. Following deidentification and segmentation, the recordings were securely transferred to Johns Hopkins University where ambient noise filtering was applied. This noise-cancellation software utilises digital stethoscope microphones to filter ambient signals from pulmonary sounds [20–22].

Chest position recording classification by a physician listening panel

De-identified, segmented, and denoised chest position recordings were randomised and distributed to a trained physician listening panel. The panel consisted of three primary panelists and two arbitrators, all trained to interpret lung sounds using our validated methodology (Figure 1) [23]. All panelists were physicians with advanced training in either Paediatrics or Paediatric Pulmonology and had passed a certification test prior to interpreting recorded lung sounds. Each recording was randomly assigned to two primary panelists for lung sound classification. Panelists reviewed lung sounds using Audacity software, classifying recordings as interpretable

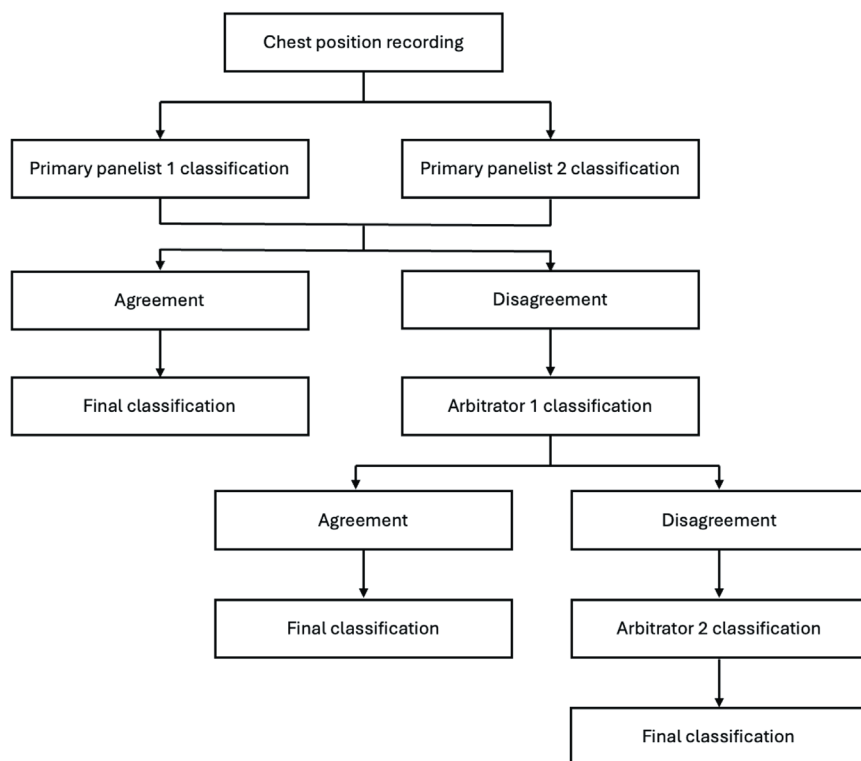


Figure 1. Chest position lung sound recording interpretation schema by the physician listening panel.

trator and both primary panelists. The secondary arbitrator's classification served as the final classification, even if it disagreed with the other panelists.

Chest position recording classification by the AI algorithm

The AI algorithm generated a classification for each denoised, interpretable chest position recording. This algorithm consists of two main components: a convolutional neural network (CNN) encoder, followed by a recurrent classifier. The CNN encoder applies multiple successive layers of filters at varying scales. Each layer captures local and increasingly abstract spectral and temporal acoustic features while reducing irrelevant noise or artifacts. The resultant condensed representation of the lung sound is passed to the recurrent classifier that examines the temporal dynamics of these extracted features, effectively capturing how the identified patterns evolve over time. The model's robustness was ensured through cross-validation, where the model was trained on various subsets of physician listening panel annotations. During this process, the data were divided into multiple testing sets to ensure robust evaluation of the model's performance. The algorithm was designed to classify interpretable lung sounds with at least one audible breath cycle as normal or abnormal. As training annotations were available for interpretable sounds only, the algorithm has not been trained or tested on uninterpretable samples. We combined the classifications from five algorithm iterations, using the majority output to make a final decision about each recording.

Patient lung sound classification

Chest positions classified as interpretable by the physician listening panel were regrouped to generate patient-level classifications of normal or abnormal. Specifically, for a given patient, if all chest positions were classified as uninterpretable, then the patient classification was uninterpretable. Alternatively, if all chest positions were classified as normal, the patient classification was normal. If a patient had one or more abnormal chest positions, then the patient classification was abnormal.

or uninterpretable. Interpretable recordings were then classified as normal or abnormal, with abnormal recordings receiving an additional classification of wheezes, crackles, or both wheezes and crackles. Classification agreement between the two primary panelists resulted in a final classification label. Panelists were masked to participant clinical data and other panelists' classifications. Recordings where primary panelists disagreed were interpreted by an initial arbitrator, who was also masked to all clinical data and other panelists' classifications. If the arbitrator agreed with one of the primary panelists, the classification was finalised. A second arbitrator, a pediatric pulmonologist with extensive experience in digital lung auscultation, reviewed recordings where disagreement persisted between the initial arbitrator and both primary panelists. The secondary arbitrator's classification served as the final classification, even if it disagreed with the other panelists.

Conventional lung sound auscultation

The study clinician performed conventional lung auscultation using a Littman Classic II Infant Stethoscope® at the same six chest positions immediately prior to the recording procedure. Following the same lung sound interpretation methodology, chest positions were classified in real-time by the clinician as interpretable or uninterpretable, with interpretable lung sounds further classified as normal or abnormal. Abnormal sounds included wheezes, crackles, or wheezes and crackles.

Statistical analysis

The sample size was determined based on the minimum acceptable agreement between the physician listening panel and the AI algorithm. Assuming an expected agreement rate of 70%, an intra-class correlation coefficient of 0.7, and a cluster size of 6 recordings per participant, a total of 91 participants was calculated to be sufficient to achieve a precision of 20% with 95% confidence. To accommodate potential data loss due to uninterpretable recordings or technical issues, the sample size was increased to 100 participants.

The analytic data set included all children with interpretable lung sound recordings, as classified by the physician listening panel. Descriptive statistics were used to summarise demographic information, medical history, and clinical data. Variables with a normal distribution were described using the mean and standard deviation (SD), while non-normally distributed variables were summarised using the median and interquartile range (IQR).

As a sensitivity analysis, we compared the demographic information, medical history, and clinical data of participants with no interpretable lung sound recordings to those with interpretable lung sound recordings. We used the χ^2 test to compare categorical variables and the Wilcoxon rank sum test for continuous variables.

The primary outcome was pairwise agreement of normal vs. abnormal chest position classifications between the AI algorithm and the physician listening panel. The secondary outcome was pairwise agreement of normal and abnormal patient classifications between the AI algorithm and physician listening panel. As a sensitivity analysis, we evaluated pair-wise agreement of patient classifications between the study clinician's conventional stethoscope interpretations and the physician listening panel, and between the study clinician and the AI algorithm.

Agreement was measured by raw percentage, Cohen's kappa statistic, and the Brennan and Prediger statistic, which adjusts for chance agreement [24]. The strength of agreement was defined as poor (≤ 0), slight (0.01, 0.19) fair (0.20, 0.39), moderate (0.40, 0.59), substantial (0.60, 0.79), or almost perfect (0.80, 1.0) [23,25]. We assessed the AI algorithm's performance in detecting abnormal lung sounds, considering the physician listening panel as the reference standard, for both chest position and patient classifications, by calculating sensitivity, specificity, positive and negative predictive values, positive and negative likelihood ratios, and diagnostic odds ratio (OR). The diagnostic OR measures a test's diagnostic performance by calculating the ratio of the odds of a positive test among diseased participants to the odds of a positive test among healthy participants [26]. We similarly assessed the AI algorithm's performance when considering the clinician as the reference standard.

We used generalised estimating equations with an exchangeable correlation matrix and binomial distribution with logit link function models to evaluate associations between participant characteristics and raw percentage agreement vs. disagreement in chest position classifications between the AI algorithm and the physician listening panel. Logistic regression was used to evaluate associations between participant characteristics and patient classification agreement between the AI algorithm and the physician listening panel.

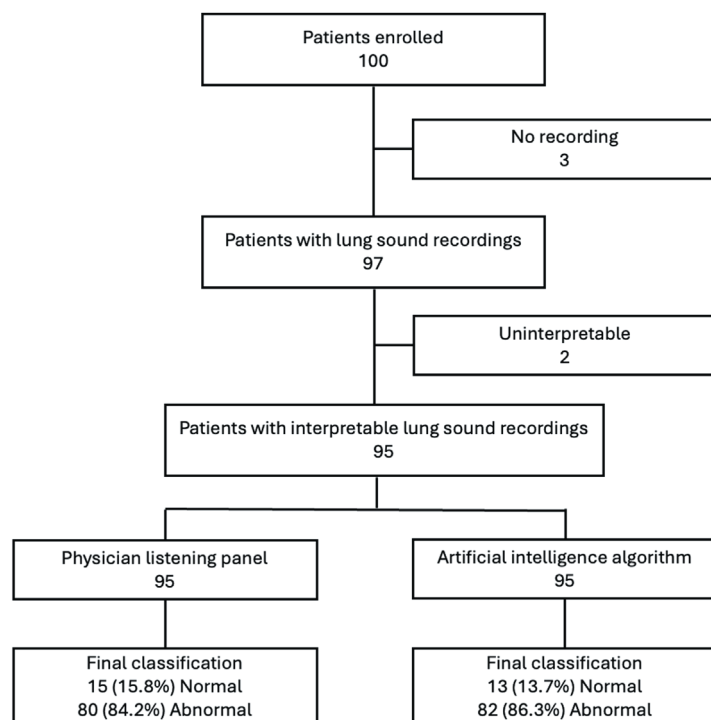


Figure 2. Artificial intelligence and physician listening panel lung sound classifications by patient.

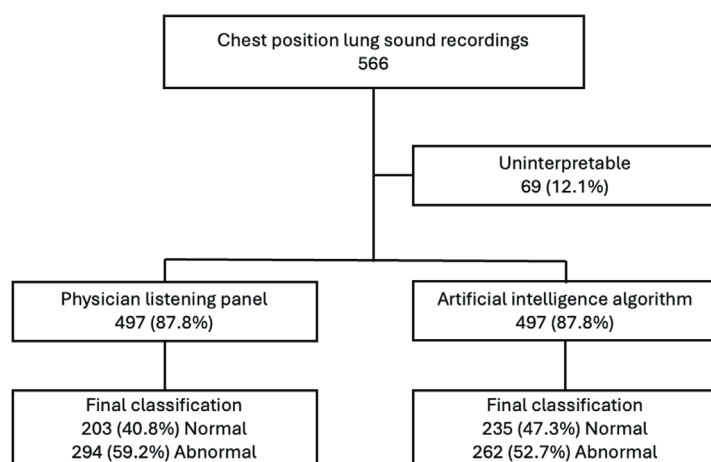


Figure 3. Artificial intelligence and physician listening panel lung sound classifications by chest position. Note: 4/570 (0.7%) recordings from the 95 participants with interpretable lung sound recordings were missing.

Of 497 interpretable chest position recordings, 294 (59.2%) were classified as abnormal by the physician listening panel compared to 262 (52.7%) by the AI algorithm (**Figure 2**). The physician listening panel and the AI algorithm agreed on 83.1% (413/497) of chest position classifications (**Table 2**). The agreement beyond what is expected by chance, measured by both Cohen's kappa and adjusted kappa statistics, was substantial for interpretable chest position recordings with or without abnormal lung sounds (**Table 2**).

RESULTS

Of the 100 patients enrolled, 95 (95%) had lung sound recordings available and interpretable (**Figure 2**). Of 566 chest position recordings, 497 (88%) were interpretable (**Figure 3**). The median age of analysed participants was 12.6 months (IQR=5.4, 19.0), with seven (7%) born prematurely. Seven (7%) patients were HIV-exposed, and 11 (12%) tested malaria positive (**Table 1**). The average WHO weight-for-height Z-score was -0.3 with an SD of 1.2. Sensitivity analysis comparing participants with no interpretable recordings to those with interpretable recordings showed no differences in demographic information, medical history, or clinical data (**Table S1** in the **Online Supplementary Document**).

Table 1. Characteristics of children with World Health Organization-defined severe pneumonia

Characteristics	Participants (n = 95)
Demographics	
Age in months, median (IQR)	12.6 (5.4, 19.0)
Age categories in months, n (%)	
2–11	46 (48)
12–23	32 (34)
24–59	17 (18)
Females, n (%)	51 (54)
Past medical history	
Prematurity*, n (%)	7 (7)
Tuberculosis contacts, n (%)	6 (6)
History of tuberculosis, n (%)	1 (1)
Vaccinations up to date†, n (%)	86 (90)
Hospitalisation characteristics	
HIV status, n (%)	
Infected	0 (0)
Uninfected	77 (81)
Exposed	7 (7)
Unknown	11 (12)
Malaria positive‡, n (%)	11 (12)§
Corticosteroid treatment, n (%)	8 (8)
Bronchodilator treatment	21 (22)
Clinical features	
WHO Weight-for-height Z score, mean (SD)	-0.3 (1.2)
Axillary temperature in degrees Celsius, median (IQR)	36.8 (36.4, 37.5)
Respiratory rate in breaths/min, mean (SD)	54.4 (14.7)
SpO ₂ in room air, median (IQR)	95.0 (92.0, 97.0)

IQR – indicates interquartile range, SD – standard deviation, SpO₂ – peripheral arterial oxyhemoglobin saturation, WHO – World Health Organization,

*By mother's verbal report.

†Up-to-date if documented in the child's health passport or verbally confirmed by guardian.

‡Positive rapid diagnostic test.

§27 (28%) with missing data.

Table 2. Lung sound classification agreement on abnormal vs. normal lung sounds between the artificial intelligence algorithm and physician listening panel for chest positions and patients

Normal or abnormal	Chest position classification, n = 497	Patient classification, n = 95
Agreement, n (%)	413 (83.1)	87 (91.6)
Kappa Statistic (95% CI)	0.659 (0.592, 0.725)	0.665 (0.447, 0.884)
Adjusted Kappa Statistic* (95% CI)	0.662 (0.596, 0.728)	0.832 (0.718, 0.945)

CI – confidence interval

*Brennan and Prediger statistic.

At the patient level, 80/95 patients (84.2%) had an abnormal classification by the listening panel compared to 82/95 (86.3%) classified by the AI algorithm (**Figure 3**), resulting in an agreement of 91.6% (87/95) (**Table 2**). Agreement by the adjusted kappa statistic was almost perfect for interpretable patient classifications with or without abnormal lung sounds (**Table 2**).

The AI algorithm demonstrated a sensitivity of 80.3% (95% CI = 75.3, 84.7) and a specificity of 87.2% (95% CI = 81.8, 91.5) in detecting abnormal lung sounds in chest position recordings, compared to the physician listening panel. The diagnostic OR of the AI algorithm indicates it was 27.7 times more likely to classify chest position recordings as abnormal and 52.3 times more likely to classify patients as abnormal when the physician listening panel also classified them as abnormal, compared to when they classified them as normal (**Table 3**).

Table 3. Performance of artificial intelligence algorithm for detecting abnormal lung sounds in children with World Health Organization-defined severe pneumonia in Malawi using the physician listening panel as the reference

	AI algorithm performance							
	Abnormal classification, n/N (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	LR+ (95% CI)	LR- (95% CI)	Diagnostic OR (95% CI)
Chest position classification	262/497 (52.7)	80.3 (75.3, 84.7)	87.2 (81.8, 91.5)	90.1 (85.8–93.4)	75.3 (69.3, 80.7)	6.27 (4.36, 9.01)	0.23 (0.18, 0.29)	27.7 (16.8, 45.7)
Patient classification	82/95 (86.3)	96.3 (89.4, 99.2)	66.7 (38.4–88.2)	93.9 (86.3, 98.0)	76.9 (46.2, 65.0)	2.89 (1.41, 5.91)	0.06 (0.02, 0.18)	51.3 (11.2, 234)

AI – artificial intelligence, CI – confidence interval, diagnostic OR – diagnostic odds ratio, LR+ – positive likelihood ratio, LR- – negative likelihood ratio, NPV – negative predictive value, PPV – positive predictive value

Predictors of raw percentage agreement between the AI algorithm and the physician listening panel for chest position classifications and patient-level classifications are presented in Table S3–4 in the **Online Supplementary Document**. Children who were uncooperative during lung auscultation had 49% lower odds of agreement for chest position classifications compared to cooperative children (Table S3 in the **Online Supplementary Document**). No factors were significantly associated with agreement for patient-level classifications (Table S4 in the **Online Supplementary Document**).

DISCUSSION

We evaluated a novel digital auscultation system, which includes a prototype digital stethoscope, filters for ambient noise removal, and an AI algorithm developed to analyse lung sounds in acutely ill children in LMIC clinical contexts with dynamic noise environments. Our data demonstrate that the AI algorithm can accurately identify abnormal lung sounds in children with severe pneumonia in a paediatric ward in Malawi, showing high discriminatory power and substantial reliability compared to a physician listening panel's reference standard interpretations of the same recordings. These results represent an important step forward for the potential application of AI-powered digital auscultation systems in identifying abnormal lung sounds in acutely ill children in real-world clinical contexts where HCWs lack training to effectively perform lung auscultation.

The performance of our AI algorithm compares favourably with existing paediatric literature on automated lung sound analysis. A 2022 systematic review of 10 studies across multiple countries evaluated the effectiveness of digital auscultation with automated lung sound analysis compared to conventional physician lung auscultation for pneumonia diagnosis in ~3000 children [27]. The review found a wide range of accuracies for classifying adventitious (abnormal) lung sounds, from 66.3 to 100% [27]. Children in the reviewed studies had various clinical conditions, including pneumonia and other respiratory conditions. Our AI algorithm achieved an accuracy of 83.1% for chest positions and 91.6% for patients in classifying abnormal lung sounds in children with severe pneumonia in a challenging environment with high ambient noise levels. To our knowledge, few studies have shown that an AI algorithm can achieve high accuracy in classifying lung sounds in children with pulmonary disease in this type of real-world context [28], further supporting the potential usability for our diagnostic platform.

In its current version, the AI algorithm of our digital auscultation system has several limitations. First, the algorithm cannot yet classify a recording as uninterpretable, as it maps the signal onto one of two possible cases: normal or abnormal. Consequently, the HCW using this system must judge the quality of the measurement in real time. The algorithm requires further development to instruct the user to repeat the measurement when there are no audible breath sounds, which may occur due to background noise or movement artifacts, both common during chest auscultation in sick, agitated children. Second, the AI results are not generated in real-time, which is necessary for clinical application. Third, the algorithm cannot yet differentiate between adventitious (abnormal) sounds, such as high or low-pitched wheezes and fine or coarse crackles, which may represent different underlying disease processes and require different treatment. Fourth, the algorithm is also unable to distinguish upper respiratory sounds (like stridor, stertor, or vocalisations) from lower respiratory sounds, potentially leading to misclassification of recordings. Lastly, the algorithm may perform less accurately when exposed to new patient data sets. To overcome this limitation, it may be necessary to recalibrate the algorithm when introduced into new contexts, including other LMIC settings that have different patient demographics or ambient noise profiles. Feasible implementation strategies, such as algorithm ‘fine-tuning’ or retraining using new data, could be utilised.

The clinical use-case of this digital auscultation system deserves discussion, as its application in the management of children with non-severe pneumonia is clearer than in severe pneumonia. For patients with non-severe pneumonia, a ‘normal vs. abnormal’ AI algorithm may help determine which children require immediate antibiotic treatment and which can be safely observed without antibiotics. This could help save valuable resources and promote antibiotic stewardship in settings with resource limitations. Indeed, the randomised, double-blinded placebo trial BLAAAST (Bangladesh Lung Auscultation with Artificial Intelligence for Antibiotic Stewardship) seeks to address this important question for children aged 2–59 months in Bangladesh. However, in hospitalised patients with WHO-defined severe pneumonia, where the risk of poor outcomes is higher and clinicians are more likely to treat empirically with antibiotics, a binary classification may have limited impact on clinical decision-making. In such scenarios, an AI algorithm capable of further differentiating between adventitious sounds such as wheezes and crackles may be more beneficial for clinical management. For instance, crackles may be associated with bacterial pathogens that require antibiotics, whereas wheezing may indicate viral-induced airway inflammation that does not need antibiotics. We previously reported that crackles identified by a human listening panel are associated with higher odds of radiographic pneumonia and mortality, while wheezes are associated with lower odds of radiographic pneumonia and mortality, consistent with this clinical framework [29–31]. We acknowledge that there can be clinical overlap between wheezes and crackles, as some patients with wheezing may also have a secondary bacterial infection.

CONCLUSIONS

In summary, this novel digital auscultation system, which includes a prototype digital stethoscope, ambient noise filtering, and an AI lung sound classification algorithm, shows potential for accurately identifying abnormal lung sounds in children with pneumonia. The next steps include further developing the algorithm to classify lung sounds as uninterpretable and to distinguish between wheezes, crackles, and upper respiratory sounds. Additionally, it will be important to evaluate how the AI algorithm's lung sound findings relate to the outcomes of children with pneumonia in LMICs.

Acknowledgements: We thank the children and their caregivers who participated, the Kamuzu Central Hospital and the Malawi Ministry of Health for their support of this research, and the dedicated study staff at the University of North Carolina (UNC) Project.

Ethics statement: This study was approved by the National Health Science Research Committee of Malawi (reference: 18/05/2052) and the University of North Carolina at Chapel Hill Institutional Review Board (reference: 18-1180). Caregivers of eligible children provided written informed consent to participate.

Data availability: The data that support the findings of this study are available upon reasonable request from the corresponding author.

Funding: Research reported in this publication was supported by the Fogarty International Center, National Heart Lung and Blood Institute, and National Institute of Neurological Disorders and Stroke of the National Institutes of Health under Award Number D43 TW009340. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authorship contributions: Funding acquisition: ZHS and EDM. Conceptualization and design: ZHS, TM, EDM. Data curation: NEH, MBC, ZHS, and EDM. Data collection: ZHS. Data analysis: NEH, MBC, ZHS, EDM. Data interpretation: NEH, ZHS, MBC, CV, DO, WCB, JM, and EDM. Writing – original draft: NEH. Writing – review & editing: NEH, MBC, ZHS, AK, IM, CV, DO, WCB, JM, AG, SK, HBS, MC, EF, ME, TM, MH, and EDM.

Disclosure of interest: The authors completed the ICMJE Disclosure of Interest Form (available upon request from the corresponding author) and disclose no relevant interests.

Additional material

Online Supplementary Document

REFERENCES

- 1 Perin J, Mulick A, Yeung D, Villavicencio F, Lopez G, Strong KL, et al. Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet Child Adolesc Health*. 2022;6:106–15. [Medline:34800370](#) [doi:10.1016/S2352-4642\(21\)00311-4](#)
- 2 GBD 2021 Lower Respiratory Infections and Antimicrobial Resistance Collaborators. Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021. *Lancet Infect Dis*. 2024;24:974–1002. [Medline:38636536](#) [doi:10.1016/S1473-3099\(24\)00176-2](#)
- 3 World Health Organization. WHO recommendations on the management of diarrhoea and pneumonia in HIV-infected infants and children: Integrated Management of Childhood Illness (IMCI). Geneva: World Health Organization; 2010.
- 4 World Health Organization. Revised WHO classification and treatment of pneumonia in children at health facilities: evidence summaries. Geneva: World Health Organization; 2014.
- 5 World Health Organization. Pocket book of hospital care for children: guidelines for the management of common childhood illnesses. 2nd ed. Geneva: World Health Organization; 2013.
- 6 Cardoso MRA, 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](#)
- 7 Hazir T, Nisar YB, Qazi SA, Khan AM, Raza M, Abbasi S, et al. Chest radiography in children aged 2–59 months diagnosed with non-severe pneumonia as defined by World Health Organization: descriptive multicentre study in Pakistan. *BMJ*. 2006;333:629. [Medline:16923771](#) [doi:10.1136/bmj.38915.673322.80](#)

- 8 Hazir T, Nisar YB, Abbasi S, Ashraf Y, Khurshid A, Tariq P, et al. Comparison of oral amoxicillin with placebo for the treatment of World Health Organization-defined nonsevere pneumonia in children aged 2–59 months: a multicenter, double-blind, randomized, placebo-controlled trial in Pakistan. *Clin Infect Dis*. 2011;52:293–300. [Medline:21189270](#) [doi:10.1093/cid/ciq142](#)
- 9 Puumalainen T, Quiambao B, Abucejo-Ladesma E, Lupisan S, Sombrero L, 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](#)
- 10 Laxminarayan R, Duse A, Wattal C, Zaidi AK, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance—the need for global solutions. *Lancet Infect Dis*. 2013;13:1057–98. [Medline:24252483](#) [doi:10.1016/S1473-3099\(13\)70318-9](#)
- 11 McCollum ED, Nambiar B, Deula R, Zadrozny S, Mvalo T, Lufesi N, et al. Impact of the 13-valent pneumococcal conjugate vaccine on clinical and hypoxemic childhood pneumonia over three years in Central Malawi: an observational study. *PLoS One*. 2017;12:e0168209. [Medline:28052071](#) [doi:10.1371/journal.pone.0168209](#)
- 12 Pneumonia Etiology Research for Child Health (PERCH) Study Group. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet*. 2019;394:757–79. [Medline:31257127](#) [doi:10.1016/S0140-6736\(19\)30721-4](#)
- 13 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](#)
- 14 Margolis P, Gadomski A. The rational clinical examination. Does this infant have pneumonia? *JAMA*. 1998;279:308–13. [Medline:9450716](#) [doi:10.1001/jama.279.4.308](#)
- 15 Harper BD, Nganga W, Armstrong R, Forsyth KD, Ham HP, Keenan WJ, et al. Where are the paediatricians? An international survey to understand the global paediatric workforce. *BMJ Paediatr Open*. 2019;3:e000397. [Medline:30815583](#) [doi:10.1136/bmjpo-2018-000397](#)
- 16 Grenier MC, Gagnon K, Genest J, 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](#)
- 17 Reichert S, Gass R, Brandt C, Andr  s E. Analysis of respiratory sounds: state of the art. *Clin Med Circ Respirat Pulm Med*. 2008;2:45–58. [Medline:21157521](#) [doi:10.4137/CCRP.M.S530](#)
- 18 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](#)
- 19 Elhilali M, West JE. The stethoscope gets smart: engineers from Johns Hopkins are giving the humble stethoscope an AI upgrade. *IEEE Spectr*. 2019;56:36–41. [Medline:34588704](#) [doi:10.1109/MSPEC.2019.8635815](#)
- 20 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](#)
- 21 Emmanouilidou D, Elhilali M. Characterization of noise contaminations in lung sound recordings. *Annu Int Conf IEEE Eng Med Biol Soc*. 2013;2013:2551–4. [Medline:24110247](#) [doi:10.1109/EMBC.2013.6610060](#)
- 22 Emmanouilidou D, McCollum ED, Park DE, Elhilali M. Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries. *IEEE Trans Biomed Eng*. 2015;62:2279–88. [Medline:25879837](#) [doi:10.1109/TBME.2015.2422698](#)
- 23 McCollum ED, Park DE, Watson NL, Buck WC, Bunthi C, Devendra A, et al. Listening panel agreement and characteristics of lung sounds digitally recorded from children aged 1–59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study. *BMJ Open Respir Res*. 2017;4:e000193. [Medline:28883927](#) [doi:10.1136/bmjresp-2017-000193](#)
- 24 Brennan RL, Prediger DJ. Coefficient kappa: some uses, misuses, and alternatives. *Educ Psychol Meas*. 1981;41:687–99. [doi:10.1177/001316448104100307](#)
- 25 Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85:257–68. [Medline:15733050](#) [doi:10.1093/ptj/85.3.257](#)
- 26 Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56:1129–35. [Medline:14615004](#) [doi:10.1016/S0895-4356\(03\)00177-X](#)
- 27 Ahmed S, Sultana S, Khan AM, Islam MS, Habib GM, McLane IM, et al. Digital auscultation as a diagnostic aid to detect childhood pneumonia: a systematic review. *J Glob Health*. 2022;12:04033. [Medline:35493777](#) [doi:10.7189/jogh.12.04033](#)

- 28 Zhang J, Wang HS, Zhou HY, Dong B, Zhang L, Zhang F, et al. Real-world verification of artificial intelligence algorithm-assisted auscultation of breath sounds in children. *Front Pediatr*. 2021;9:627337. **Medline:33834010** **doi:10.3389/fped.2021.627337**
- 29 McCollum ED, Park DE, Watson NL, Fancourt NSS, Focht C, Baggett HC, et al. Digital auscultation in PERCH: associations with chest radiography and pneumonia mortality in children. *Pediatr Pulmonol*. 2020;55:3197–208. **Medline:32852888** **doi:10.1002/ppul.25046**
- 30 Park DE, Watson NL, Focht C, Feikin D, Hammitt LL, Brooks WA, et al. Digitally recorded and remotely classified lung auscultation compared with conventional stethoscope classifications among children aged 1–59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study. *BMJ Open Respir Res*. 2022;9:e001144. **Medline:35577452** **doi:10.1136/bmjresp-2021-001144**
- 31 McCollum ED, Park DE, Watson NL, Buck WC, Bunthi C, Devendra A, et al. Listening panel agreement and characteristics of lung sounds digitally recorded from children aged 1–59 months enrolled in the Pneumonia Etiology Research for Child Health (PERCH) case-control study. *BMJ Open Respir Res*. 2017;4:e000193. **Medline:28883927** **doi:10.1136/bmjresp-2017-000193**