BREATH: Building a Repository for Evaluation, Analysis, and Treatment for child respiratory Health

1. Background and Rationale

Acute respiratory infection (pneumonia) is a major cause of death and disability that disproportionately affect young children in low- and middle-income countries (LMICs).[1–3] The World Health Organization (WHO) estimates that pneumonia accounts for about 22% of child deaths with the highest burden in South Asia and Sub-Saharan Africa.[1,4,5] Diagnosis of child pneumonia poses a particular challenge related to non-specific clinical symptoms, since other respiratory illnesses such as upper respiratory infections, asthma exacerbations, and bronchiolitis can mimic the clinical features of pneumonia, are highly prevalent, and lack specific confirmatory tests.[6–11] These diagnostic challenges are compounded by the difficulty in distinguishing among bacterial, viral, and mycobacterial infections, and the high prevalence of co-infection with significant implications for individual patient management and population-level antimicrobial resistance.[12] For example, in a study in Uganda, 95% of children under age five with asthma or bronchiolitis were treated with antibiotics.[13] In view of these challenges, the Integrated Management of Childhood Illnesses (IMCI) guidelines aim to standardize the evaluation and treatment of pneumonia by front-line health workers.[14] However, this approach favors sensitivity over specificity, likely leading to over-treatment with antibiotics and contributing to antimicrobial resistance.[15–19]

Cough analysis has emerged as a potential adjunctive diagnostic tool in respiratory infections, as cough is a predominant and discernible symptom for most respiratory disorders. Artificial Intelligence (AI) and deep learning-based algorithms for audiometric analysis of cough signatures have shown promise in aiding the classification and triage of respiratory illnesses, including upper vs. lower respiratory infections, asthma exacerbations, bronchiolitis, pneumonia, croup, and tuberculosis.[20–26] When deployed on a smartphone, AI-based cough analysis provides a tool that is non-invasive, low-risk, and requires minimal technical expertise by the clinician.[27] In addition to analysis of cough sounds, AI-based tools can integrate multiple types of data, including physiologic parameters (including respiratory rate and SpO_2), patient history, local epidemiology of respiratory infections, as well as prediction of disease severity to guide triage and referral to appropriate facilities, often outperforming trained medical professionals.[28]

Achieving adequate accuracy with a deep learning algorithm requires substantial high-quality, well-phenotyped data for both training and validation of an AI model,[29] and while rapid progress is being made in the development of AI-based tools across many diseases, the data needed to inform AI-based tools for children in LMICs with respiratory infections are particularly lacking.[30–33] Our study aims to address this critical need by creating an open-source dataset of children in LMICs presenting with acute respiratory infections. Our overall hypothesis is that an AI-based cough analysis algorithm will provide greater diagnostic accuracy compared to existing tools used by front-line health workers in LMICs.

Preliminary Data

Our hypothesis is supported by preliminary data generated by our team's prior work. The PI of the proposed project (Dr. Moschovis, Massachusetts General Hospital/Harvard Medical School) previously led two large studies of ResAppDx, an algorithm using acoustic features of cough and parent-reported symptoms. The SmartCough-C2 study enrolled 1,468 children aged 1 month to 12 years with respiratory symptoms in three US hospitals, comparing the ResAppDx diagnosis to clinical and radiographic diagnoses provided by independent adjudicators. The algorithm demonstrated sensitivity between 63.0% and 76.1% and specificity between 59.6% and 86.0% for a range of diagnoses (upper vs. lower respiratory tract disease, asthma, bronchiolitis, pneumonia, and croup), diagnostic performance within the range of other common point-of-care tests.[7,34] Members of our group also conducted a pilot feasibility study among 137 adult patients in India (led by Dr. Philip, CMCH, Bihar, India) to measure the association between the acoustic pattern of a solicited cough using an AI model and chest radiograph abnormalities. An AI-enabled cough classifier demonstrated an AUC ranging from 0.70 to 0.78, suggesting that with further training of this algorithm, analysis of acoustic features of cough could reduce the use of chest radiography.[35]

Overview of Research Strategy

Needs assessment: To identify gaps in this field, we hosted a virtual discussion attended by approximately 60 pediatric respiratory research experts in September 2023. They provided input regarding an optimal study design, site selection, detailed data requirements (including radiography, lung ultrasonography, and

digital auscultation) for standardization of reference diagnoses, and advice regarding pathogen identification. We also met with investigators from the Pneumonia Etiology Research for Child Health (PERCH) study, which aimed to identify the most common pathogens in severe pneumonia across seven LMICs in Africa and Asia. The study found that viral pathogens, especially RSV, were predominant, along with several bacterial pathogens and TB.[36] The PERCH study also performed digital auscultation on a sample of participants, which demonstrated the feasibility of a centralized review of recordings, *but did not include cough sound recordings*.[37,38]

Proposed research framework: To respond to the needs raised by prior studies, we propose a three-phase research framework entitled "Building a Repository for Evaluation, Analysis, and Treatment for Pediatric Respiratory Health (BREATH)" (see Figure 1). The overall goal is to develop AI-based cough analysis tools for children with respiratory illnesses in LMICs. **Phase 1** of this study will build an open-source dataset that will be used to train and validate automated cough analysis algorithms (**Phases 2** and **3**) for pediatric respiratory disorders. Given the challenges of differentiating pneumonia from other disorders, this database with cough recordings, and clinical, radiology, and microbiology data will be a valuable resource for researchers to examine multiple questions relating to diagnosis and triage of children with a range of respiratory illnesses.

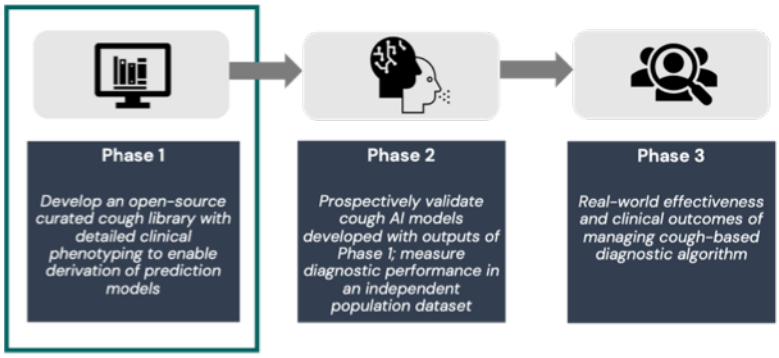

Figure 1. Overall research strategy

2. Study Objectives & Deliverables: Phase 1

Phase 1 of the BREATH study will establish a robust infrastructure for pediatric respiratory research by creating an open-source multidimensional database that includes cough sound recordings, clinical data, physical examination, pulse oximetry, digital imaging (ultrasonography and radiography), microbiologic testing, digital auscultation recordings, and adjudicated clinical and radiographic diagnoses. These data will lay the groundwork for the development of AI-based prediction models and real-world testing of point-of-care diagnostic tools for pediatric respiratory diseases, to be performed in Phases 2 and 3 of the research framework.

3. Study Design and Methodology

After obtaining institutional approvals, in the Phase 1 study we will enroll 7,000 symptomatic children aged 1 to 59 months across five tertiary care centers in Africa and South Asia over a period of 12 months (see sample size calculation below). Children presenting with respiratory symptoms will be recruited using the following criteria after obtaining informed consent from a parent or legal guardian.

Inclusion Criteria	***Exclusion Criteria***
• Age 1 to 59 months • Any cough or difficulty breathing within the past 24 hours[14,39]	• Mechanical ventilatory support or high-flow oxygen therapy • Tracheostomy, history of airway or vocal cord abnormalities • Medical instability per treating clinician (e.g., requiring immediate resuscitation or airway support) • Children with medical contraindications to voluntary cough (severe respiratory distress; recent pneumothorax, eye, chest, abdominal surgery, or recent hemoptysis) enrolled only if coughing spontaneously.

Each participant will undergo timed cough audio event recordings for 15 minutes (either spontaneous or voluntary cough event recordings from older children able to cooperate). Research staff will record clinical observations and vital signs, and will perform ultrasonography and digital auscultation. Chest radiography will be performed if determined to be clinically indicated by local practice guidelines. Following the cough recording, nasopharyngeal swabs will also be collected for multiplex PCR tests to test for common viral pathogens (including RSV, influenza, and SARS-CoV-2; final list of pathogens to be determined after review of local epidemiology). Since these sites also have a high burden of TB, after recording cough sounds, we will obtain nasopharyngeal swabs and induced sputum samples for PCR testing (Xpert MTB/RIF Ultra) and mycobacterial culture. Results of any other additional microbiologic testing carried out for

clinical purposes (e.g., blood culture) will also be recorded. Follow-up will occur via phone or in-person two weeks after enrollment to assess clinical outcomes, including admission, recovery, or adverse events. All data will be collected via Qure.ai (an AI-based imaging solutions technology provider) application software and hosted on a secure platform at the Data Coordinating Center. The specific workflow of the proposed data collection study is depicted in Figure 2 along with the list of data variables.

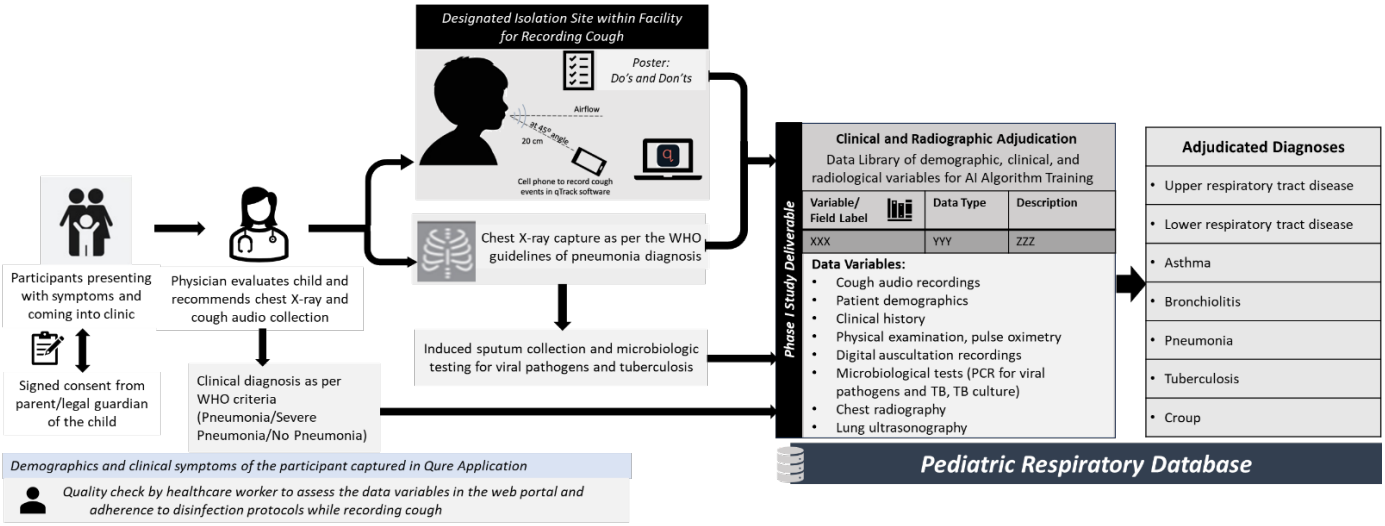

Figure 2. Proposed participant workflow for cough data collection

Adjudication Panels

A panel of pediatricians and a panel of radiologists with expertise in evaluation of pediatric respiratory disease in LMICs will be recruited to provide independent clinical and radiographic diagnoses using standardized case definitions.[27] Two adjudicators will review each case; a third adjudicator will review in case of disagreement among the initial reviewers, and any unresolved cases will be reviewed by the full committee. The adjudicated diagnoses will serve as reference standards for training of future algorithms.

4. Sample Size Estimation

We propose a sample size of 7,000 children recruited among five sites in Sub-Saharan Africa and South Asia to ensure adequate representation of each of the major clinical syndromes within a diverse population. Based on a review of historical data at participating sites, we estimate approximately 15% will have lower respiratory tract disease (including pneumonia, asthma, bronchiolitis, and TB), and 85% will have isolated upper respiratory tract disease. Using prior estimates of sensitivity and specificity (approximately 65% for each) obtained in the Smartcough-C-2 study,[27] a sample size of 7,000 provides the following precision for confidence intervals (CI) across a range of prevalence for various diagnoses.

Prevalence	0.05	0.1	0.2	0.3	0.4	0.5	0.6
# of cases	350	700	1400	2100	2800	3500	4200
Sensitivity [95% CI]	0.650 [0.600, 0.700]	0.650 [0.615, 0.685]	0.650 [0.625, 0.675]	0.650 [0.630, 0.670]	0.650 [0.632, 0.668]	0.650 [0.634, 0.666]	0.650 [0.636, 0.664]
Specificity [95% CI]	0.650 [0.639, 0.661]	0.650 [0.638, 0.662]	0.650 [0.638, 0.662]	0.650 [0.637, 0.663]	0.650 [0.636, 0.664]	0.650 [0.634, 0.666]	0.650 [0.632, 0.668]

5. Site Selection

The following study sites were selected for this study based on the prevalence of pneumonia, volume of patients seen, operational feasibility, and prior experience and expertise in study implementation for pneumonia and other lung diseases (Table 1).

Site Name	**Country**	**Site Principal Investigator**
Christian Medical Centre and Hospital, Purnia, Bihar	India	Dr. Alex Phillip
University of Ibadan	Nigeria	Dr. Adegoke Falade

Aga Khan University	Pakistan	Dr. Zahra Hoodbhoy
SA-MRC Unit on Child & Adolescent Health, U. Cape Town	South Africa	Dr. Heather Zar
Makerere University/Mbarara University of Science and Technology	Uganda	Dr. Rebecca Nantanda Dr. Elias Kumbakumba

Table 1. Confirmed sites and investigators for Phase 1 of cough data collection

6. Roles and Responsibilities

Working closely with partner sites, the Massachusetts General Hospital (Principal Investigator/Data Coordinating Center) and Qure.ai propose to undertake the proposed study with guidance from Every Breath Counts Coalition, a public-private partnership focused on ending child pneumonia deaths using data-based control strategies. Two committees will also be constituted to support the study. The Steering Committee will include PIs from each site and the Data Coordinating Center PI, the chairs of the clinical and radiologic adjudication committees, a statistician, and the senior project and data managers. An Expert Advisory Group (EAG) will include experts in pneumonia research and data science, including representatives of the BMGF, academic researchers, and members of Every Breath Counts Coalition.

Stakeholder	**Responsibilities**
Every Breath Counts Coalition	• Steering and Facilitation Partner, Advocacy, and Impact Dissemination
Donor	• Approval of final protocol and workflow, overall guidance, and progress review
Massachusetts General Hospital: Data Coordinating Center	• Principal Investigatory Support to draft and review protocol, finalize workflow for data collection with site partners, IRB assistance, project oversight, and support during implementation. • Forming Adjudicator Panels for clinical and radiographic adjudication • Hosting the Open-Source Data Platform and Launching AI Challenge to collaborate with technology partners and researchers
Qure.ai	• Stakeholder Coordination and Project Management Support • Building the concept, quality control, and monitoring the SOP adherence at sites • Developing clinical and radiographic adjudication platforms
Partner Sites	• Input into design of protocol, implementation of final protocol, IRB approval, and project management • Training and orientation of data collectors and healthcare workers • Procurement, Quality Control, Protocol Adherence and Reporting
Steering Committee	• Oversee the direction and progress of the project, providing strategic guidance, decision-making, and ensuring alignment with project goals.
Expert Advisory Group	• Advisory support and guidance with respect to pediatric respiratory health, implications for policy change, and strategies for knowledge dissemination.
Technology Partners	• Lung ultrasound, digital auscultation, pulse oximetry, and microbiological testing

Table 2. Stakeholder roles and responsibilities in the project

7. Next Steps and Future Studies

The work performed in this Phase 1 study will foster collaboration among childhood respiratory disease researchers globally, develop partnerships, draw attention to the need for improved diagnostics, and build capacity for the subsequent phases of the research framework. Following completion of Phase 1, our team will issue a Request for Applications (RFA) for commercial, NGO, and academic teams to use the data collected in Phase 1 to develop AI-based diagnostic algorithms (similar to the CODA-TB Challenge).[40] We are aware of several groups conducting work in this space, including Kenya Medical Research Institute/ University of Washington,[41] ResApp/Pfizer,[42] Hyfe.ai,[43] Swaasa,[44] and Raisonance.[45]

By including a wide variety of predictor and outcome variables (including reference-grade adjudication), the

dataset will also be a rich resource for additional research questions. For example, further work could assess severity of disease and develop models for triaging and referral, monitor respiratory disease progression over time, assess response to treatment, and provide clinical decision support to front-line workers based on national/international treatment guidelines.

8. References

1. World Health Organization. Pneumonia in children [Internet]. [cited 2024 Feb 15];Available from: https://www.who.int/news-room/fact-sheets/detail/pneumonia

2. Marangu D, Zar HJ. Childhood pneumonia in low-and-middle-income countries: An update. Paediatr Respir Rev 2019;32:3–9.

3. Zar HJ, Ferkol TW. The global burden of respiratory disease-impact on child health. Pediatr Pulmonol 2014;49(5):430–4.

4. Deardorff KV, McCollum ED, Ginsburg AS. Pneumonia Risk Stratification Scores for Children in Low-Resource Settings: A Systematic Literature Review. Pediatr Infect Dis J 2018;37(8):743–8.

5. Gothankar J, Doke P, Dhumale G, et al. Reported incidence and risk factors of childhood pneumonia in India: a community-based cross-sectional study. BMC Public Health 2018;18(1):1111.

6. Goodman D, Crocker ME, Pervaiz F, 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(12):1068–83.

7. Gentilotti E, De Nardo P, Cremonini E, et al. Diagnostic accuracy of point-of-care tests in acute community-acquired lower respiratory tract infections. A systematic review and meta-analysis. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 2022;28(1):13–22.

8. Li Y, Wang X, Blau DM, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. Lancet Lond Engl 2022;399(10340):2047–64.

9. Troeger C, Blacker B, Khalil IA, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis 2018;18(11):1191–210.

10. Jin X, Ren J, Li R, et al. Global burden of upper respiratory infections in 204 countries and territories, from 1990 to 2019. eClinicalMedicine [Internet] 2021 [cited 2024 Jan 22];37. Available from: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(21)00266-2/fulltext

11. Wang Z, Li Y, Gao Y, et al. Global, regional, and national burden of asthma and its attributable risk factors from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. Respir Res 2023;24:169.

12. Wei M, Yongjie Zhao, Zhuoyu Qian, et al. Pneumonia caused by Mycobacterium tuberculosis. Microbes Infect 2020;22(6):278–84.

13. Nantanda R, Tumwine JK, Ndeezi G, Ostergaard MS. Asthma and pneumonia among children less than five years with acute respiratory symptoms in Mulago Hospital, Uganda: evidence of under-diagnosis of asthma. PloS One 2013;8(11):e81562.

14. Integrated management of childhood illness [Internet]. [cited 2024 Feb 21];Available from: https://www.who.int/teams/maternal-newborn-child-adolescent-health-and-ageing/child-health/integrated-management-of-childhood-illness

15. Allwell-Brown G, Hussain-Alkhateeb L, Sewe MO, et al. Determinants of trends in reported antibiotic use among sick children under five years of age across low-income and middle-income countries in 2005–17: A systematic analysis of user characteristics based on 132 national surveys from 73 countries. Int J Infect Dis 2021;108:473–82.

16. Kolstad PR, Burnham G, Kalter HD, Kenya-Mugisha N, Black RE. Potential implications of the integrated management of childhood illness (IMCI) for hospital referral and pharmaceutical usage in western Uganda. Trop Med Int Health 1998;3(9):691–9.

17. Sl W, Rg B, Mc M, Mi N. Application of the world health organization criteria to predict radiographic pneumonia in a US-based pediatric emergency department. Pediatr Infect Dis J [Internet] 2012 [cited 2024 Feb 28];31(6). Available from: https://pubmed.ncbi.nlm.nih.gov/22333702/

18. Wingerter SL, Bachur RG, Monuteaux MC, Neuman MI. Application of the World Health Organization criteria to predict radiographic pneumonia in a US-based pediatric emergency department. Pediatr Infect Dis J 2012;31(6):561–4.

19. House DR, Rijal S, Adhikari S, Cooper ML, Hohl CM. Prospective evaluation of World Health Organization guidelines for diagnosis of pneumonia in children presenting to an emergency department in a resource-limited setting. Paediatr Int Child Health 2020;40(4):227–30.

20. Liao S, Song C, Wang X, Wang Y. A classification framework for identifying bronchitis and pneumonia in children based on a small-scale cough sounds dataset. PloS One 2022;17(10):e0275479.

21. Yellapu GD, Rudraraju G, Sripada NR, et al. Development and clinical validation of Swaasa AI platform for screening and prioritization of pulmonary TB. Sci Rep 2023;13(1):4740.

22. Sharan RV, Abeyratne UR, Swarnkar VR, Porter P. Cough sound analysis for diagnosing croup in pediatric patients using biologically inspired features. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf 2017;2017:4578–81.

23. Sharan RV, Abeyratne UR, Swarnkar VR, Porter P. Automatic Croup Diagnosis Using Cough Sound Recognition. IEEE Trans Biomed Eng 2019;66(2):485–95.

24. Alqudaihi KS, Aslam N, Khan IU, et al. Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities. IEEE Access Pract Innov Open Solut 2021;9:102327–44.

25. Chung Y, Jin J, Jo HI, et al. Diagnosis of Pneumonia by Cough Sounds Analyzed with Statistical Features and AI. Sensors 2021;21(21):7036.

26. Sharan RV, Rahimi-Ardabili H. Detecting acute respiratory diseases in the pediatric population using cough sound features and machine learning: A systematic review. Int J Med Inf 2023;176:105093.

27. Moschovis PP, Sampayo EM, Cook A, 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 2021;101:106278.

28. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health 2019;1(6):e271–97.

29. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28(1):31–8.

30. Ginsburg AS, McCollum ED. Artificial intelligence and pneumonia: a rapidly evolving frontier. Lancet Glob Health 2023;11(12):e1849–50.

31. Malhotra A, Molloy EJ, Bearer CF, Mulkey SB. Emerging role of artificial intelligence, big data analysis and precision medicine in pediatrics. Pediatr Res 2023;93(2):281–3.

32. Naydenova E, Tsanas A, Howie S, Casals-Pascual C, De Vos M. The power of data mining in diagnosis of childhood pneumonia. J R Soc Interface 2016;13(120):20160266.

33. Field MJ, Behrman RE, Children I of M (US) C on CRI. The Necessity and Challenges of Clinical Research Involving Children [Internet]. In: Ethical Conduct of Clinical Research Involving Children. National Academies Press (US); 2004 [cited 2024 Feb 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25553/

34. Moschovis PP, Sampayo E m., Porter P, et al. A Cough Analysis Smartphone Application for Diagnosis of Acute Respiratory Illnesses in Children [Internet]. In: A27. PEDIATRIC LUNG INFECTION AND CRITICAL CARE AROUND THE WORLD. Dallas, TX: American Thoracic Society; 2019 [cited 2023 Oct 22]. p. A1181–A1181.Available from: https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A1181

35. Philip A, Chawla S, Jover L, et al. Leveraging cough sounds to optimize chest x-ray usage in low-resource settings [Internet]. 2024 [cited 2024 Feb 20];Available from: http://arxiv.org/abs/2402.08789

36. 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 Lond Engl 2019;394(10200):757–79.

37. McCollum ED, Park DE, Watson NL, et al. Digital auscultation in PERCH: Associations with chest radiography and pneumonia mortality in children. Pediatr Pulmonol 2020;55(11):3197–208.

38. Park DE, Watson NL, Focht C, 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(1):e001144.

39. World Health Organization. Pocket book of hospital care for children: Guidelines for the management of common illnesses with limited resources. 2nd ed. Geneva: WHO Press; 2013.

40. info@sagebase.org SB. Synapse | Sage Bionetworks [Internet]. [cited 2024 Mar 5];Available from: https://www.synapse.org

41. AfricaNews. Kenyan scientists test AI app to diagnose tuberculosis [Internet]. Africanews. 2024 [cited 2024 Mar 1];Available from: https://www.africanews.com/2024/02/25/kenyan-scientists-test-ai-app-to-diagnose-tuberculosis/

42. Home [Internet]. ResApp Health. [cited 2024 Mar 5];Available from: https://www.resapphealth.com.au/

43. Hyfe AI - Detect & Quantify Cough [Internet]. [cited 2024 Mar 5];Available from: https://www.hyfe.ai/

44. Home - Swaasa [Internet]. 2023 [cited 2024 Mar 5];Available from: https://swaasa.ai/

45. RAIsonance [Internet]. [cited 2024 Mar 5];Available from: https://raisonance.ai/