INPLASY

INPLASY202590105

doi: 10.37766/inplasy2025.9.0105

Received: 25 September 2025

Published: 25 September 2025

Corresponding author:
Morro Touray

m.touray@surrey.ac.uk

Author Affiliation:
*Surrey Health Economics Centre, Faculty of Health and Medical Sciences, School of Biosciences, University of Surrey, Guildford, GU2 7XH.

Application of AI and Machine Learning for the Diagnosis of Pneumococcal Infections in Children Using Digital Chest X-ray Images – Protocol for a Scoping Review

Touray, M*[1,2]; Longo, M[2]; Gage, H[1]; Jones, B[1]; Egan, B[3]; Prada, J[4,5]; Kanteh, D[6]; Onwuchekwa, U[7].

ADMINISTRATIVE INFORMATION

Support - There was no financial support.

Review Stage at time of this submission - The review has not yet started.

Conflicts of interest - Heather Gage, Morro ML Touray, and Bernadette Egan are members of the National Institute for Health and Care Research's Applied Research Collaboration for Kent, Surrey and Sussex. BE is partially funded by the NIHR Health Determinant's Research Collaboration Surrey (HDRC Surrey). Authors have no other conflicts of interest to declare.

INPLASY registration number: INPLASY202590105

Amendments - This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 25 September 2025 and was last updated on 25 September 2025.

1. Surrey Health Economics Centre, Faculty of Health and Medical Sciences, School of Biosciences, University of Surrey, Guildford, GU2 7XH
2. School of Medicine, Cardiff University, Uhw Main Building, Heath Park, Cardiff, CF14 4XN
3. Faculty of Health and Medical Sciences, School of Psychology, University of Surrey, Guildford, GU2 7XH
4. Surrey Institute for People-Centred AI, University of Surrey, Guildford, GU2 7XH 5. School of Veterinary Medicine, University of Surrey, Guildford, GU2 7XH
6. Strategic Services, MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Atlantic Boulevard, Fajara, PO Box 273, Banjul, The Gambia
7. Centre for Vaccine Development-Mali, Ex-Institut Marchoux, Avenue Mohamed VI, Djikoroni Para, 251 Bamako, Mali.

INTRODUCTION

Review question / Objective Research question: What computer aided or AI and ML based tools are available and how effective are they for diagnosing pneumococcal infections in children using digital x-ray images?

The objectives are to:
• Identify and summarise studies utilising AI/ML or computer-aided tools for the diagnosis of pneumococcal infections in children using digital chest X-ray images.
• Evaluate the performance metrics (e.g., accuracy, sensitivity, specificity) of these computer-aided or

AI/ML models used for diagnosis of pneumococcal infections using digital chest X-ray images.
• Identify the challenges and limitations or gaps in the current research.
• Provide recommendations for future research and possible implementation of these tools.

Background Pneumococcal infections are a leading cause of pneumonia, meningitis, and sepsis all of which can be deadly. While pneumococcal vaccines have helped to reduce the burden of illness[1], pneumonia remains the biggest infectious killer of children worldwide[2-4]. This is attributable to low vaccination coverage[5], cold chain infrastructure challenges[6], and serotype replacements[7] among other things. Some countries have not introduced the vaccine at all despite studies indicating its effectiveness[8] and cost-effectiveness[9]. Every year, pneumonia is responsible for more than 725,000 deaths among children under the age of 5, including around 190,000 newborns[10], who are particularly vulnerable to infection.

Early and accurate diagnosis of pneumonia is critical for effective treatment and improved patient outcomes. As chest X-rays are the primary diagnostic tool, which relies on the expertise of radiologists, their interpretation can therefore often be subjective hence prone to errors and variability. Additionally, there is an acute shortage of skilled radiologists in less-resourced countries such as in Sub-Saharan Africa, causing delays in diagnosis. There is a pressing need for computer- aided tools that can enhance the accuracy, effectiveness, efficiency, and accessibility of this important diagnosis.

Artificial Intelligence (AI) and Machine Learning (ML), computer-aided tools, offer the potential to automate and enhance the analysis of chest x-rays. They have now emerged as technologies that are transforming medical imaging. For example, deep learning, such as Convolutional Neural Networks (CNNs), have been highly effective in image classification, object detection, video analysis and classifying various diseases from medical images[11]. However, the routine application of AI/ML specifically for pneumococcal infection diagnosis is still not fully explored. There is therefore the need for a comprehensive review to map existing research, evaluate the performance of AI/ML systems, and identify challenges and opportunities to inform future work. By synthesising the current evidence, this scoping review aims to inform the adoption or development of accurate and clinically applicable AI/ML tools for pneumococcal infection diagnosis in children and their subsequent evaluation in a large Randomised Controlled Trial.

Rationale I. The Problem - A Critical Diagnostic Gap in Child Health:
Pneumococcal pneumonia remains a leading cause of childhood mortality globally, with the highest burden in low-resource settings and among low-socioeconomic populations. Pneumococcal diagnosis relies heavily on availability of radiologists to examine and interpret chest X-rays, but a critical shortage of radiological expertise leads to inconsistent interpretation and delayed treatment. This diagnostic bottleneck represents a significant and preventable barrier to accessing timely appropriate care and improving child health outcomes.

II. The Potential Solution - AI-Driven Diagnostic Support:
Artificial Intelligence (AI) and Machine Learning (ML) have demonstrated remarkable potential to automate the analysis of medical images, offering a path to rapid, accurate, and scalable interpretation of chest X-rays. Such tools could function as a force multiplier for healthcare professionals, providing expert-level diagnostic support in settings where specialists are unavailable particularly in primary care settings.

III. The Knowledge Gap - A Lack of Synthesised Evidence:
Despite these potentials, the applications of AI and ML for diagnosing pneumococcal infections in children is not fully explored. It is unclear which models and systems exist, how they perform, or what challenges impede their real-world implementation. A systematic synthesis of this evidence is therefore needed to guide future research and responsible computer aided diagnostic tool development.

IV. The Proposed Response - A Scoping Review to Map the Landscape:
This scoping review aims to address these gaps by systematically identifying and evaluating existing AI and ML tools for diagnosing paediatric pneumococcal infection using digital chest X-ray images. By mapping the current evidence, assessing model performance, and identifying key research and implementation barriers, this review will provide a foundational roadmap. The findings will directly inform the development of robust, equitable AI technologies for diagnosing pneumococcal infections in children and guide future clinical trials, ultimately contributing to faster, more accurate diagnosis and improved

survival for children in the United Kingdom and worldwide.

METHODS

Strategy of data synthesis Search Strategy

We will search the following electronic databases:

• PubMed
• IEEE Xplore
• Scopus
• Web of Science

The search terms will include the following keywords and Boolean operators:

("pneumococcal infections" OR "pneumococcal pneumonia" OR "Streptococcus pneumoniae" OR “pneumonia” OR “bacterial pneumonia” OR “pneumococcus”) AND (“Computer-based” OR “computer-aided” OR "Artificial Intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "computer-aided diagnosis") AND (“X-ray” OR "chest X-ray" OR "radiograph" OR "medical imaging") AND ("diagnosis" OR "detection" OR "classification")

There will not be any restricted timeframe, and we will include studies and articles published in English language only. The study types will be pilot and feasibility studies, and RCTs, and articles including peer-reviewed articles, conference papers, and preprints.

Eligibility criteria
Inclusion and exclusion criteria

Inclusion Criteria:

• Studies focusing on the use of AI/ML and computer-aided tools for diagnosing pneumococcal infections in children.
• Primary end point being the use of chest X-ray images for diagnosis as the primary data source.
• Studies reporting performance metrics such as accuracy, sensitivity, and specificity
• Articles and conference papers that have been peer reviewed including pilot and feasibility studies, RCTs, preprints and evidence reviews.
• Studies published in English.

Exclusion Criteria:

• Studies not focused on pneumococcal infections or viral pneumonia infections.
• Non-imaging-based or non-digital x-ray image based diagnostic methods (e.g., laboratory tests, clinical algorithms).
• Editorials, opinion pieces, and non-English publications.
• Studies that do not have performance metrics or sufficient methodological detail.

Source of evidence screening and selection Data Screening and Extraction:

Following the search, all identified citations will be collated and uploaded into the Systematic Review Data Repository (SRDR) platform of the US Agency for Healthcare Research and Quality, Rockville, MD 20857 where duplicates will be removed. This will be followed by assessment of the citations where titles and abstracts will be screened by Touray and a Research Assistant against the described inclusion and exclusion criteria. Where Touray and the Research Assistant cannot agree on the inclusion or exclusion of a publication, Longo or another member of the research team will screen the citation to support decisions as to whether to include it or not.

The potentially relevant articles will then be retrieved in full, and their citation details imported into the SRDR platform. The full text of selected citations will be assessed in detail against the inclusion criteria by a Research Assistant and Touray. Reasons for exclusion of sources of evidence at full text considerations that do not meet the inclusion criteria will be recorded and reported in the scoping review. Any disagreements that arise between the reviewers at each stage of the screening process will be resolved through discussion, or with another member of the research team.

A structured data extraction form will be developed in the SRDR platform following PRISMA-ScR guidelines[12]. The data that will be collected through the data extraction form include:

• Study details: author, year, country, study design,
• Target population studied,
• Computer-aided or AI/ML techniques used: model architecture, training methods,
• Dataset: Source, size, labelling, and availability,
• Performance metrics: Accuracy, sensitivity, and specificity, etc,
• Challenges and limitations,
• Outcomes and results.

Data management The Surrey Health Economics Centre will oversee all aspects of data management, ensuring full compliance with the

University of Surrey's established data governance policies and protocols.

Reporting results / Analysis of the evidence Data analysis and presentation
The results of the search and the study inclusion process will be reported in full in the final scoping review and presented in a PRISMA flow diagram[13]. The extracted data will be synthesised thematically to address the review objectives. The key themes that will be included include - types of AI/ML tools used, performance metrics and their variability, dataset characteristics and limitations, challenges in clinical implementation if provided, and future research directions. The results will also be tabulated guided by a data extraction form. A synthesised narrative will accompany the tabulated results, and this will analyse, describe and discuss how the results relate to the scoping review objectives.

Language restriction English.

Country(ies) involved United Kingdom, The Gambia and Mali.

Keywords Scoping review, pneumococcal infections, diagnosis, chest x-ray, artificial intelligence, machine learning, computer aided, children.

Dissemination plans Dissemination Our findings will be published in an appropriate peer-reviewed journal. It will also be presented in conferences and shared with relevant stakeholders such as researchers and policy makers.

Contributions of each author
Author 1 - Morro Touray - Conceptualised the study and drafted the protocol and contributed to its review.
Email: m.touray@surrey.ac.uk
Author 2 - Mirella Longo - Drafted the protocol and contributed to its review.
Email: longom1@cardiff.ac.uk
Author 3 - Heather Gage - Reviewed the protocol.
Email: h.gage@surrey.ac.uk
Author 4 - Bridget Jones - Reviewed the protocol.
Email: b.e.jones@surrey.ac.uk
Author 5 - Bernadette Egan - Reviewed the protocol.
Email: m.egan@surrey.ac.uk
Author 6 - Joaquin Prada - Contributed to the review of the protocol.
Email: j.prada@surrey.ac.uk
Author 7 - Dembo Kanteh - Contributed to the review of the protocol.
Email: dembo.kanteh@lshtm.ac.uk
Author 8 - Uma Onwuchekwa - Contributed to the review of the protocol.
Email: uonwuche@som.umaryland.edu