TB or not TB: does AI have an answer for children?

Jacob Creswell[1], Rachel L. Byrne [2] and Tushar Garg [1]

[1]Stop TB Partnership, Geneva, Switzerland. [2]Liverpool School of Tropical Medicine, Liverpool, UK.

Corresponding author: Jacob Creswell (jacobc@stoptb.org)

Shareable abstract (@ERSpublications)
Artificial intelligence is being developed and evaluated for reading paediatric chest radiography. The TB community needs more high-quality studies to move the evidence base for global policy forward. https://bit.ly/47rRvFm

Cite this article as: Creswell J, Byrne RL, Garg T. TB or not TB: does AI have an answer for children? *Eur Respir J* 2024; 64: 2401709 [DOI: 10.1183/13993003.01709-2024].

Received: 28 Aug 2024
Accepted: 8 Sept 2024

Introduction

In 2021, the World Health Organization (WHO) issued recommendations for the use of artificial intelligence (AI), affirming the technology could be used instead of human readers for the interpretation of chest radiograph (CXR) images for the screening and triage of tuberculosis (TB), marking a historic precedent for a guideline on AI in health [1]. The forward-looking guidance highlights the utility of CXR in TB, despite past reluctance from WHO to use it in TB programmes [2]. Employing CXR in a diagnostic algorithm can identify people with subclinical TB who are asymptomatic but have lung pathology suggestive of TB, and refer them for testing and diagnosis [3]. CXR can also greatly reduce the number of confirmatory tests required, which can save costs for TB programmes and prevent unnecessary inconvenience for patients [4]. The limitations to the use of CXR for TB have generally been due to a lack of equipment or a trained reader [5, 6]. Radiologists, or even well-trained physicians, are not always easy to find, especially in high TB burden areas outside of major cities [7].

The use of AI for interpreting CXR images has come a long way in a short period of time. Early studies of the first software showed promise when first appearing just a decade ago [8], but the technology fell short when compared to expert radiologists and the comparatively high cost was an issue [9]. Since 2017, several studies have documented good performance of multiple AI solutions for the interpretation of CXRs in the screening or triage of TB [10–13], with newer versions continuing to improve [14, 15]. Nonetheless, the technology is still relatively new, and many questions about it remain unanswered, especially regarding its performance and programmatic implementation.

What is the correct comparison for AI?

CXR produces an image that needs to be interpreted by a radiologist or physician. AI provides CXR interpretation in lieu of such a read; thus, evaluations of the technology should include a human reader comparison. Such a human comparison is critical in decision-making when evaluating AI products, as we know that human readers have high levels of both intra- and inter-reader variability, meaning one reader often interprets images differently to others, and the same reader may also interpret the same image differently upon reading again [16]. AI products can provide consistency to the reads, but different AI products may provide different interpretations for the same images.

When evaluations of AI benchmark performance of CXR to Xpert or culture, the sensitivity and specificity of CXR is measured against microbiological evidence of TB, not the interpretation of images. While AI has been shown to have variable performance in different populations, in such comparisons, human readers often do not perform well either and so, a head-to-head evaluation is critical [10, 17–19]. Ultimately, the programmatic decision is whether AI should be used in the absence of a human to interpret CXR, not as a replacement for microbiological testing.

What are the use cases?

The initial WHO recommendations focus on TB screening and triage use cases among adults aged 15 years and above [1]. However, there are many variations and setups through which implementers can

deploy AI for CXR [20]. Currently, most of the field implementation is in active case-finding efforts where large volumes of CXRs must be interpreted quickly, and often in remote regions without typical healthcare infrastructure [21]. Some programmes conduct screening without any human reader on site [22], while in other settings, it is a legal requirement to have a radiologist interpret it with the assistance of AI [23]. AI can be used in TB prevalence surveys or active case-finding campaigns where CXR plays a critical role in identifying people with TB who are not symptomatic [24]. AI can also be used together with human readers to improve performance, as a quality control measure, and to reduce workload for experienced readers in high volume situations [25]. It is important to note that most people with chest conditions will not have TB, and providing an interpretation of the CXR beyond TB could be ultimately far more impactful for both patients and health systems with limited resources.

How to choose the best threshold score

Human readers generally provide a dichotomous conclusion: whether findings are consistent with TB or not. In the application of AI for CXR interpretation, the reading produces an abnormality score as a continuous variable on a 0–1 or 1–100 scale. These outputs represent the likelihood of having TB, with positive predictive values of microbiological test positivity increasing as abnormality scores increase [26, 27]. This also offers programmes flexibility in deployment, depending on the objectives and limitations of the intervention. For instance, a TB prevalence survey may want more relaxed criteria (a lower threshold) to ensure as many people as possible with TB can be identified, while an active case-finding initiative with a single molecular testing platform may need to limit the number of people sent for testing due to logistical and throughput constraints, and so may choose a higher threshold. Therefore, the optimal threshold is context-specific [28], and dependent upon programmatic variables such as budget and time [20]. Flexibility to choose an "optimal" threshold is an additional advantage of AI systems, as human readers will likely struggle to adapt consistent reading approaches to different settings and instructions. Although AI products may provide a default threshold score, local threshold calibration is strongly encouraged [26].

What is the utility in childhood TB diagnosis?

The current WHO guidelines on AI to interpret CXR do not include children aged below 15 years [1]. Children with TB are missed in far greater proportions than adults, with current estimates indicating that around half of incident TB cases are being missed, which directly contributes to high mortality among children [29, 30]. The difficulties of linking more children to TB care are well-documented, and include poorly sensitive tests, a reliance on sputum or other respiratory specimen testing, limited contact-tracing and a paucity of physicians confident in making a TB diagnosis [31]. Having a CXR suggestive of TB is a major part of the consensus diagnostic algorithm for children with TB [32], yet many children and their caregivers have to travel long distances to find a physician to read paediatric CXR images and assist in diagnosis. In many countries, childhood TB diagnosis is concentrated in a few select hospitals [33, 34]. AI algorithms that could support TB diagnosis in children in decentralised sites could potentially be a huge step forward in improving access to life-saving treatment.

Before their deployment in paediatric cohorts, AI products require training data. Current versions of AI for TB have used millions of adult CXR images for development, which are insufficient for child-focused software, given that the presentation of TB in adults is often much different from TB in children (figure 1) [35]. The lack of data for AI development is likely due to difficulty in obtaining labelled paediatric CXRs with a TB diagnosis or ground truth, highlighting the need for establishing paediatric CXR libraries to train AI software. We hope access to annotated paediatric CXR data through ongoing research, including CAPTURE (Catalyzing AI for Pediatric Tuberculosis Research) [36] and Start4All (Start Taking Action For TB Diagnosis) [37], can accelerate AI product development. Additionally, current AI tools are only able to interpret a postero-anterior or antero-posterior film, and not lateral films, which can provide additional support to clinical decision-making [35].

There have been only a handful of studies looking at the performance of AI for TB in children [38, 39], and in this issue of *European Respiratory Journal*, Edem *et al.* [40] present the findings of a study from the Gambia, in which they collected CXR images from children with presumptive TB over a 10-year period. The study evaluated the performance of the latest version of CAD4TB (Computer-Aided Detection for Tuberculosis, Version 7; Delf Imaging, Hertogenbosch, the Netherlands) against a microbiological reference standard. The authors also used Bayesian latent class analysis (BLCA) to address the shortcomings in data completeness. They found that the software did not have high sensitivity at pre-defined abnormality thresholds.

The study marks a step in the right direction for the potential use of AI to help with childhood TB diagnosis and highlights several important points to include in future studies. First, future studies should

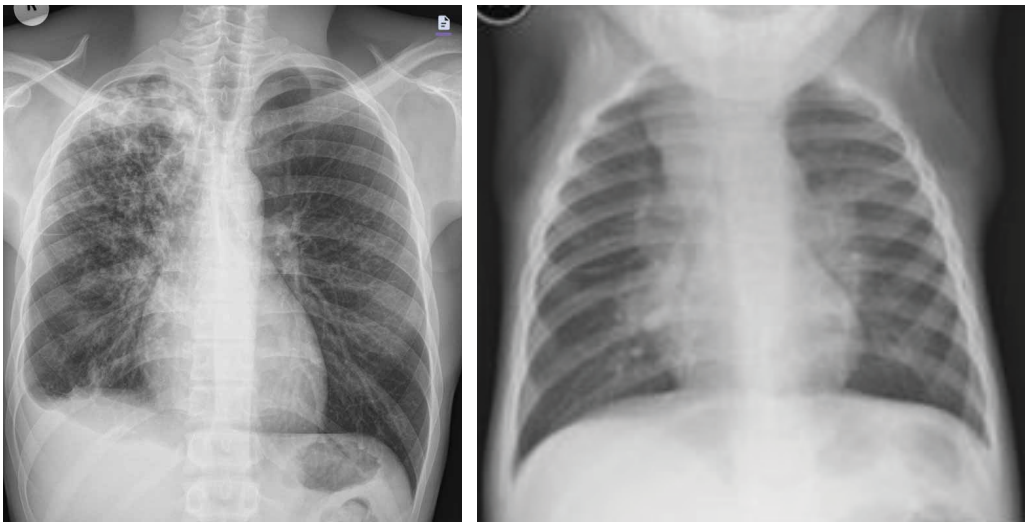

FIGURE 1 Chest radiographs of an adult (left) and child (right) with tuberculosis (TB). The adult radiograph shows typical post-primary TB with opacities in the right upper zone and multiple cavities with a pleural effusion on the right. In contrast, the child radiograph shows bilateral opacities in the perihilar regions, suggestive of enlarged perihilar lymph nodes: the hallmark radiological feature of paediatric TB. Images courtesy of M. Palmer at the Desmond Tutu TB Centre, Stellenbosch University, Cape Town, South Africa.

include head-to-head comparisons with human readers for programmatic guidance. These should also explore variability of human readers in CXR interpretation as this may be even more striking when interpreting CXR in children. Recent work from Zambia has shown high levels of variability among expert radiologists when reading childhood CXR films [41]. Evaluations should include performance in differentiating severity of TB among children, as non-severe TB can be treated with a shorter regimen.

Second, it is important to control for age in the evaluation of paediatric computer-aided detection products. Just as the CXR images of adults and children look different, the radiographs of children aged under 15 years will vary widely as well, particularly in the youngest age group. While some AI companies claim their products can be used in children as young as 2 years old, others specify a limit of 5 or 7 years and above [42, 43]. If age is included in the CXR metadata for children younger than 5 years old, CAD4TBv7 will generate an error, while it will give a warning about interpretation if age is removed. Both the image of the thorax and the presentation of TB will differ greatly depending on the age and development of the child. Surprisingly, Edem *et al.* [40] documented a marginally worse area under the curve in older children than in those aged under 5 years, although with overlapping confidence intervals. This warrants further exploration, because disease presentation in adolescents is likely to resemble that found in an adult. Additionally, it may be beneficial to define populations as infants (aged less than 1 year old), children (1–4 and 5–9 years) and young adolescents (10–14 years) to better understand this variability.

Third, in this study, as with many other published results, the majority of children with TB are not microbiologically confirmed. With the power that deep learning can bring to the field to improve performance, we need to better understand what the software currently misses, and also what expert human readers miss. The authors evaluated CAD4TB against a microbiological reference standard, but not a composite standard. The AI software (and human readers) will identify people with TB that laboratory tests will miss [44]. Follow-up studies that include children diagnosed on clinical grounds can provide additional information to programmes and policymakers.

Although Edem *et al.* [40] found AI to have a low sensitivity (19%) while focusing on a pre-determined threshold score of 60, the software produced a sensitivity of 62.1% and a specificity of 70% using a threshold of 40. Further, despite BLCA showing worse performance for the AI compared to Xpert and culture, consideration of the use case is important. It is unlikely that CXR would be used in the same way as it is used in active case-finding for adults, that is, to screen children *out* of a testing algorithm. Most programmes will aim to provide children with several different tests to get the best clinical picture possible, as per the National Institutes of Health consensus definition [32]. In this sense, CXR is being used as an additional diagnostic tool, not a screening or triage test. Therefore, performance among children with

microbiologically confirmed TB is less useful from a decision-to-treat standpoint. CXR performance for children (either with human or AI interpretation) should be evaluated as part of an algorithm for treatment decisions. Programmes need to know how well AI software performs when CXR results become a key part of a treatment decision algorithm, *i.e.* can it be used to decide if the image is consistent with TB when faced with negative microbiological results?

Conclusion

With sufficient data to consider guideline recommendations for children, the issue for implementers and policymakers will be how CXR (with AI) fits into the diagnostic pathway, and what the value of the findings are, ideally when compared to what humans would do. Edem *et al.* [40] have started to build an evidence base, and future research must take the subject of AI for CXR image interpretation in children further to answer the basic question: can AI support or replace a human reader to help inform a decision on treatment? If the answer is "yes", it could greatly help provide more access to better diagnostic algorithms for children with presumptive TB.

Conflict of interest: J. Creswell, R.L. Byrne and T. Garg report grants from UNITAID.

References

1 World Health Organization. WHO Consolidated Guidelines on Tuberculosis: Module 2: Screening: Systematic Screening for Tuberculosis Disease. Geneva, World Health Organization, 2021. Available from: https://apps.who.int/iris/handle/10665/340255
2 World Health Organization Communicable Diseases Cluster. What is DOTS? A Guide to Understanding the WHO-recommended TB Control Strategy Known as DOTS. Geneva, World Health Organization, 1999. Available from: https://iris.who.int/handle/10665/65979
3 Coussens AK, Zaidi S, Allwood B, *et al.* Classification of early tuberculosis states to guide research for improved care and prevention: an international Delphi consensus exercise. *Lancet Respir Med* 2024; 12: 484–498.
4 Philipsen R, Sánchez C, Maduskar P, *et al.* Automated chest-radiography as a triage for Xpert testing in resource constrained settings: a prospective study of diagnostic accuracy and costs. *Sci Rep* 2015; 5: 12215.
5 Kiguli-Malwadde E, Byanyima R, Kawooya MG, *et al.* An audit of registered radiology equipment resources in Uganda. *Pan Afr Med J* 2020; 37: 295.
6 Gathuru LM, Elias GDO, Pitcher RD. Analysis of registered radiological equipment in Kenya. *Pan Afr Med J* 2021; 40: 205.
7 Hricak H, Abdel-Wahab M, Atun R, *et al.* Medical imaging and nuclear medicine: a Lancet Oncology Commission. *Lancet Oncol* 2021; 22: e136–e172.
8 Muyoyeta M, Maduskar P, Moyo M, *et al.* The sensitivity and specificity of using a computer aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka Zambia. *PLoS One* 2014; 9: e93757.
9 Rahman MT, Codlin AJ, Rahman MM, *et al.* An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. *Eur Respir J* 2017; 49: 1602159.
10 Qin ZZ, Sander MS, Rai B, *et al.* Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Sci Rep* 2019; 9: 15000.
11 Qin ZZ, Ahmed S, Sarker MS, *et al.* Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. *Lancet Digit Health* 2021; 3: e543–e554.
12 Codlin AJ, Dao TP, Vo LNQ, *et al.* Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. *Sci Rep* 2021; 11: 23895.
13 Qin ZZ, Van der Walt M, Moyo S, *et al.* Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software. *Lancet Digit Health* 2024; 6: e605–e613.
14 Murphy K, Habib SS, Zaidi SMA, *et al.* Computer aided detection of tuberculosis on chest radiographs: an evaluation of the CAD4TB v6 system. *Sci Rep* 2020; 10: 5492.
15 Qin ZZ, Barrett R, Ahmed S, *et al.* Comparing different versions of computer-aided detection products when reading chest X-rays for tuberculosis. *PLoS Digit Health* 2022; 1: e0000067.
16 Garland LH. Studies on the accuracy of diagnostic procedures. *Am J Roentgenol Radium Ther Nucl Med* 1959; 82: 25–38.
17 Steiner A, Mangu C, van den Hombergh J, *et al.* Screening for pulmonary tuberculosis in a Tanzanian prison and computer-aided interpretation of chest X-rays. *Public Health Action* 2015; 5: 249–254.
18 Fehr J, Konigorski S, Olivier S, *et al.* Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. *NPJ Digit Med* 2021; 4: 106.

19 Kazemzadeh S, Yu J, Jamshy S, *et al.* Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists. *Radiology* 2023; 306: 124–137.
20 Creswell J, Vo LNQ, Qin ZZ, *et al.* Early user perspectives on using computer-aided detection software for interpreting chest X-ray images to enhance access and quality of care for persons with tuberculosis. *BMC Global Public Health* 2023; 1: 30.
21 Zaidi SMA, Habib SS, Van Ginneken B, *et al.* Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on chest radiography among private sector patients in Pakistan. *Sci Rep* 2018; 8: 12339.
22 John S, Abdulkarim S, Usman S, *et al.* Comparing tuberculosis symptom screening to chest X-ray with artificial intelligence in an active case finding campaign in Northeast Nigeria. *BMC Global Public Health* 2023; 1: 17.
23 Innes AL, Martinez A, Gao X, *et al.* Computer-aided detection for chest radiography to improve the quality of tuberculosis diagnosis in Vietnam's district health facilities: an implementation study. *Trop Med Infect Dis* 2023; 8: 488.
24 Law I, Floyd K, *et al.* National tuberculosis prevalence surveys in Africa, 2008–2016: an overview of results and lessons learned. *Tropical Med Int Health* 2020; 25: 1308–1327.
25 Dvijotham K, Winkens J, Barsbey M, *et al.* Enhancing the reliability and accuracy of AI-enabled diagnosis *via* complementarity-driven deferral to clinicians. *Nat Med* 2023; 29: 1814–1820.
26 World Health Organization, UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. Determining the Local Calibration of Computer-Assisted Detection (CAD) Thresholds and Other Parameters: a Toolkit to Support the Effective Use of CAD for TB Screening. Geneva, World Health Organization, 2021. Available from: https://iris.who.int/handle/10665/345925
27 Stop TB Partnership. Screening and Triage for TB using Computer-Aided Detection (CAD) Technology and Ultra-portable X-ray Systems: A Practical Guide. www.stoptb.org/resources-implementing-cad-and-xray/cad-and-x-ray-practical-implementation-guide
28 Vanobberghen F, Keter AK, Jacobs BKM, *et al.* Computer-aided detection thresholds for digital chest radiography interpretation in tuberculosis diagnostic algorithms. *ERJ Open Res* 2024; 10: 00508-2023.
29 World Health Organization. Global Tuberculosis Report 2023. Geneva, World Health Organization, 2023. Available from: https://iris.who.int/handle/10665/373828
30 Jenkins HE, Yuen CM, Rodriguez CA, *et al.* Mortality in children diagnosed with tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2017; 17: 285–295.
31 Reuter A, Hughes J, Furin J. Challenges and controversies in childhood tuberculosis. *Lancet* 2019; 394: 967–978.
32 Graham SM, Cuevas LE, Jean-Philippe P, *et al.* Clinical case definitions for classification of intrathoracic tuberculosis in children: an update. *Clin Infect Dis* 2015; 61: Suppl. 3, S179–S187.
33 Graham SM, Sekadde MP. Case detection and diagnosis of tuberculosis in primary-care settings. *Paediatr Int Child Health* 2019; 39: 84–87.
34 Rahman MT, Malik AA, Amanullah F, *et al.* Improving tuberculosis case detection in children: summary of innovations and findings from 18 countries. *J Pediatric Infect Dis Soc* 2022; 11: Suppl. 3, S117–S124.
35 Palmer M, Seddon JA, Goussard P, *et al.* Diagnostic CXR Atlas for Tuberculosis in Children: A Guide to Chest X-ray Interpretation. Paris, International Union Against Tuberculosis and Lung Disease, 2022.
36 Palmer M. The CAPTURE Consortium: developing a large image repository to support CAD evaluation and development for childhood TB. *Int J Tuberc Lung Dis* 2023; 27: Suppl. 1, S27.
37 Byrne RL, Wingfield T, Adams ER, *et al.* Finding the missed millions: innovations to bring tuberculosis diagnosis closer to key populations. *BMC Global Public Health* 2024; 2: 33.
38 Palmer M, Seddon JA, Van Der Zalm MM, *et al.* Optimising computer aided detection to identify intra-thoracic tuberculosis on chest x-ray in South African children. *PLoS Glob Public Health* 2023; 3: e0001799.
39 Naidoo J, Shelmerdine SC, Ugas-Charcape CF, *et al.* Artificial intelligence in paediatric tuberculosis. *Pediatr Radiol* 2023; 53: 1733–1745.
40 Edem VF, Nkereuwem E, Agbla SC, *et al.* Accuracy of CAD4TB (Computer-Aided Detection for Tuberculosis) on paediatric chest radiographs. *Eur Respir J* 2024; 64: 2400811.
41 Muyoyeta M. Performance of CAD systems in children aged 5–14 years in Lusaka, Zambia. *Int J Tuberc Lung Dis* 2023; 27: Suppl. 1, S27.
42 Foundation for Innovative New Diagnostics (FIND). AI-assisted Diagnostics for TB, COVID-19, and Other Respiratory Diseases in Low- and Middle-income Countries. Geneva, Foundation for Innovative New Diagnostics, 2024. Available from: www.finddx.org/wp-content/uploads/2024/05/20240530_rep_ai_assisted_dx_FV_EN.pdf
43 Stop TB Partnership, AI4HLTH, Foundation for Innovative New Diagnostics (FIND). AI Products for Tuberculosis Healthcare. Date last accessed: 27 August 2024. www.ai4hlth.org
44 Macpherson L, Kik SV, Quartagno M, *et al.* Diagnostic accuracy of chest X-ray Computer Aided Detection software and blood biomarkers for detection of prevalent and incident tuberculosis in household contacts followed up for 5 years. *medRxiv* 2024; preprint [https://doi.org/10.1101/2024.06.30.24309731].