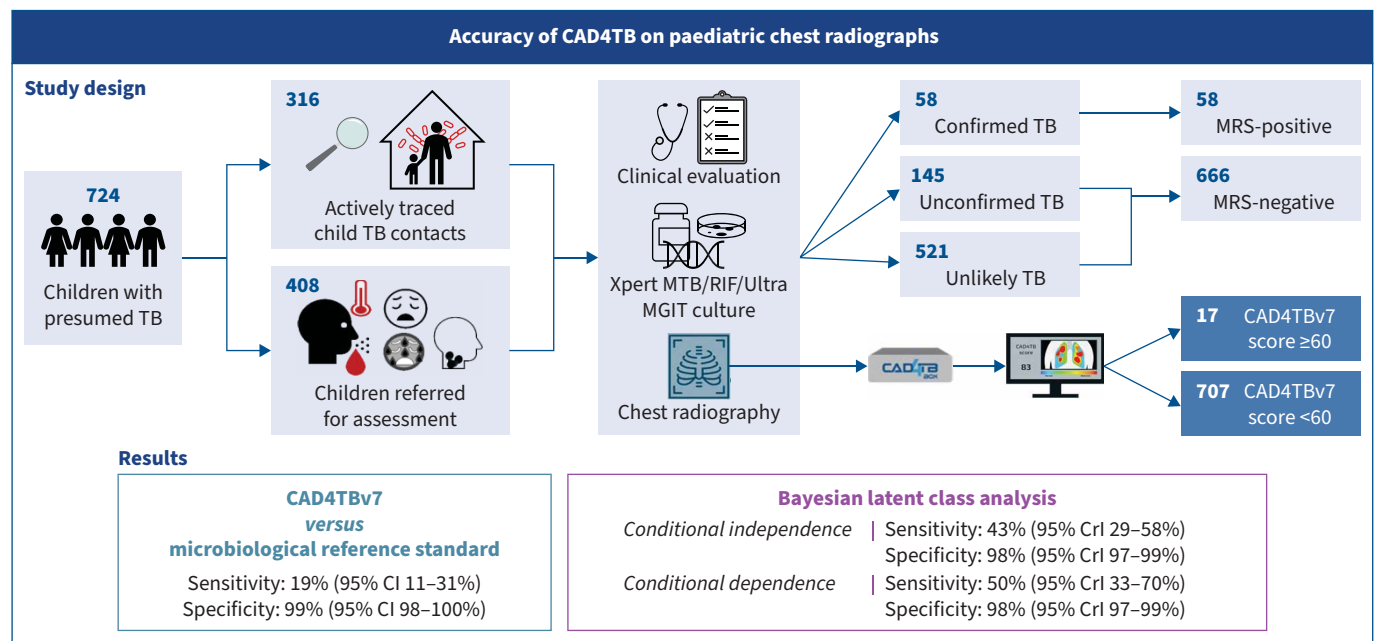




Accuracy of CAD4TB (Computer-Aided Detection for Tuberculosis) on paediatric chest radiographs

Victory Fabian Edem, Esin Nkereuwem , Schadrac C. Agbla, Sheila A. Owusu, Abdou K. Sillah, Binta Saidy, Musa B. Jallow, Audrey G. Forson, Uzochukwu Egere, Beate Kampmann and Toyin Togun 



GRAPHICAL ABSTRACT Main findings of the study. Although CAD4TBv7 (Computer-Aided Detection for Tuberculosis version 7) demonstrated high specificity, its suboptimal sensitivity underscores the crucial need for optimisation for detection of tuberculosis (TB) in children. MGIT: mycobacteria growth indicator tube; MRS: microbiological reference standard; CI: confidence interval; CrI: credible interval.



Accuracy of CAD4TB (Computer-Aided Detection for Tuberculosis) on paediatric chest radiographs

Victory Fabian Edem^{1,2,12}, Esin Nkereuwem ^{1,3,12}, Schadrac C. Agbla^{4,5}, Sheila A. Owusu^{1,3,6}, Abdou K. Sillah¹, Binta Saidy¹, Musa B. Jallow⁷, Audrey G. Forson⁸, Uzochukwu Egere⁹, Beate Kampmann^{1,3,10} and Toyin Togun ^{1,3,11}

¹Vaccines and Immunity Theme, MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Fajara, The Gambia. ²Department of Immunology, College of Medicine, University of Ibadan, Ibadan, Nigeria. ³Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. ⁴Department of Health Data Science, University of Liverpool, Liverpool, UK. ⁵Department of Infectious Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London, UK. ⁶Department of Paediatrics and Child Health, School of Medicine, University for Development Studies, Tamale, Ghana. ⁷National Leprosy and Tuberculosis Control Programme, Ministry of Health, Banjul, The Gambia. ⁸Department of Medicine, Korle Bu Teaching Hospital and University of Ghana Medical School, Accra, Ghana. ⁹Department of International Public Health, Liverpool School of Tropical Medicine, Liverpool, UK. ¹⁰Charité Centre for Global Health, Institute of International Health, Charité – Universitätsmedizin Berlin, Berlin, Germany. ¹¹TB Centre, London School of Hygiene and Tropical Medicine, London, UK. ¹²V.F. Edem and E. Nkereuwem contributed equally to this work.

Corresponding author: Toyin Togun (Toyin.Togun@lshtm.ac.uk)



Shareable abstract (@ERSpublications)

Although promising, CAD4TB (Computer-Aided Detection for Tuberculosis) has suboptimal sensitivity for detecting TB in exclusively paediatric chest radiographs, which underscores the crucial need for optimisation of CAD systems for detecting TB in children. <https://bit.ly/3yJwQ9c>

Cite this article as: 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 [DOI: 10.1183/13993003.00811-2024].

Copyright ©The authors 2024.

This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org

This article has an editorial commentary: <https://doi.org/10.1183/13993003.01709-2024>

Received: 28 April 2024
Accepted: 24 July 2024

Abstract

Background Computer-aided detection (CAD) systems hold promise for improving tuberculosis (TB) detection on digital chest radiographs. However, data on their performance in exclusively paediatric populations are scarce.

Methods We conducted a retrospective diagnostic accuracy study evaluating the performance of CAD4TBv7 (Computer-Aided Detection for Tuberculosis version 7) using digital chest radiographs from well-characterised cohorts of Gambian children aged <15 years with presumed pulmonary TB. The children were consecutively recruited between 2012 and 2022. We measured CAD4TBv7 performance against a microbiological reference standard (MRS) of confirmed TB, and also performed Bayesian latent class analysis (LCA) to address the inherent limitations of the MRS in children. Diagnostic performance was assessed using the area under the receiver operating characteristic curve (AUROC) and point estimates of sensitivity and specificity.

Results A total of 724 children were included in the analysis, with confirmed TB in 58 (8%), unconfirmed TB in 145 (20%) and unlikely TB in 521 (72%). Using the MRS, CAD4TBv7 showed an AUROC of 0.70 (95% CI 0.60–0.79), and demonstrated sensitivity and specificity of 19.0% (95% CI 11–31%) and 99.0% (95% CI 98.0–100.0%), respectively. Applying Bayesian LCA with the assumption of conditional independence between tests, sensitivity and specificity estimates for CAD4TBv7 were 42.7% (95% CrI 29.2–57.5%) and 97.9% (95% CrI 96.6–98.8%), respectively. When allowing for conditional dependence between culture and Xpert assay, CAD4TBv7 demonstrated a sensitivity of 50.3% (95% CrI 32.9–70.0%) and specificity of 98.0% (95% CrI 96.7–98.9%).

Conclusion Although CAD4TBv7 demonstrated high specificity, its suboptimal sensitivity underscores the crucial need for optimisation of CAD4TBv7 for detecting TB in children.

Introduction

Tuberculosis (TB) remains a significant global public health problem with an estimated 10.6 million incident cases and 1.3 million deaths in 2022 [1]. Although childhood TB makes up 12% and 16% of the



global TB incidence and mortality, respectively, diagnosing TB in children is challenging due to difficulties in obtaining sputum or other respiratory samples. Even when respiratory samples are successfully obtained, microbiological tests such as *Mycobacterium tuberculosis* culture or Xpert MTB/RIF Ultra are less likely to have positive results because childhood TB is frequently paucibacillary [2]. This results in the majority of childhood TB diagnoses being made presumptively, based on non-specific clinical and radiological features, and without laboratory confirmation [3].

The World Health Organization (WHO) recommends chest radiography as a useful tool to screen and triage for pulmonary TB in both children and adults [4]. However, the limited availability of specialist radiologists and the inter- and intra-reader variability associated with chest radiography limit its use, especially in TB-endemic areas [4, 5]. Computer-aided diagnosis (CAD) systems that use artificial intelligence algorithms to analyse digital chest radiography images have shown promising results for diagnosis of pulmonary TB in adults [6]. These CAD systems can compute a numerical abnormality score ranging from 0 to 100, indicating the likelihood of TB-associated abnormalities [7].

CAD4TB (Computer-Aided Detection for Tuberculosis), developed by Delft Imaging Systems ('s-Hertogenbosch, The Netherlands), is the first commercially available CAD tool. It has been shown to significantly outperform human readers for TB screening in individuals >15 years old [8]. In cohort studies that enrolled adults with symptoms suggestive of pulmonary TB, CAD4TB accurately distinguished between chest radiography images of participants with culture-positive TB and controls with an area under the receiver operating characteristic curve (AUROC) of 0.84 (95% CI 0.80–0.88) [9]. Research studies, which are almost exclusively from adult populations, have also reported promising results on the utility of CAD systems for population-level TB screening or triage in clinical settings to identify individuals requiring further confirmatory TB testing [8–11]. These findings suggest that CAD4TB may be a useful tool for detecting pulmonary TB in children where non-sputum-based approaches remain a priority in the diagnostic algorithm [12, 13].

However, to date, objective evidence of the diagnostic accuracy of all CAD systems in exclusively paediatric populations is very scarce. Children represent a unique population with distinct clinical features and radiological findings of TB, which may affect the accuracy of CAD4TB [14]. In particular, children often present with other respiratory conditions that may confound the diagnosis of TB [15, 16].

In order to address this gap in knowledge, we assessed the diagnostic accuracy of CAD4TB version 7 (CAD4TBv7) for pulmonary TB in an exclusively paediatric population using a microbiological reference standard (MRS), and we also used Bayesian latent class analysis (LCA) to address the inherent limitations of the MRS given the paucibacillary nature of TB in children.

Methods

Study participants, settings and study design

We conducted a retrospective diagnostic evaluation of CAD4TBv7 on a repository of digital chest radiography images from well-characterised cohorts of children aged <15 years with presumed pulmonary TB. The study participants were consecutively recruited *via* a comprehensive childhood TB research programme established at the Medical Research Council Unit The Gambia at the London School of Hygiene & Tropical Medicine (MRCG at LSHTM) and aimed at evaluating preventive, screening and novel diagnostic approaches in childhood TB. Between 2012 and 2022, three prospective childhood TB studies were conducted using identical protocols, including the Childhood TB Programme Grant (February 2012 to June 2017), Reach4Kids Africa (August 2017 to December 2019) and Reach4Kids Africa-2 (November 2021 to October 2022). All studies and recruitment were conducted in the Greater Banjul Area of The Gambia, a mixed urban, peri-urban and rural area. The study aims, settings, screening and recruitment procedures for these cohorts have been previously described [17–21].

In brief, children with presumed pulmonary TB were enrolled in the studies either by active tracing of child contacts of newly diagnosed adults with infectious TB or referral of children with clinical signs and symptoms suggestive of TB for diagnostic evaluation. In these studies, presumptions of TB were defined as contact with an infectious adult index TB case and/or presence of unremitting cough >14 days and at least one of either fever, weight loss/failure to thrive, malaise/fatigue, haemoptysis, night sweats or enlarged cervical lymph nodes. Referral of children presenting at peripheral health facilities was based on the presence of the aforementioned clinical signs and symptoms suggestive of TB. At enrolment, all children with presumed pulmonary TB underwent a detailed clinical assessment for symptoms and a physical examination by the study paediatrician at the dedicated outpatient childhood TB clinic at the MRCG at LSHTM. They also had a chest radiograph taken with a CARESTREAM DRX-Ascend system,

which operated combined with a Cannon CXDI 710CW detector and had a resolution of 2800×3408 pixels. Additionally, sputum samples were collected, either spontaneously or by induction using nebulised hypertonic saline, for conventional pathogen detection tests, including the Xpert MTB/RIF or Ultra assay (Cepheid, Sunnyvale, CA, USA), liquid culture using mycobacteria growth indicator tubes (Becton Dickinson, Franklin Lakes, NJ, USA) and solid culture on Löwenstein–Jensen medium. We confirmed the presence of *M. tuberculosis* in positive cultures with acid-fast staining and MPT64 antigen detection (Abbott, Palatine, IL, USA) or GenoType MTBDRplus line probe assays (Hain Lifesciences, Nehren, Germany). The children were classified as 1) confirmed TB, 2) unconfirmed TB or 3) unlikely TB, using a combination of clinical, radiological and laboratory findings, based on the revised US National Institutes of Health classification of intrathoracic TB [22]. All children with confirmed or unconfirmed TB were referred to dedicated TB treatment centres for standard anti-TB treatment. They were also required to attend a follow-up visit 2 months later and at the end of treatment for clinical evaluation. Children with unlikely TB were treated for other respiratory diseases, with follow-up clinic visits within 4 weeks to ascertain their wellbeing.

This study is reported in accordance with the guidelines of the Standards for Reporting of Diagnostic Accuracy Studies (STARD) [23], and a complete STARD checklist is provided in the supplementary material. Ethical approvals for all studies were obtained from the Gambia Government/MRC Joint Ethics Committee. Written informed consent was obtained from each participant's parent or legal guardian, and assent was obtained from children aged ≥ 7 years at the time of enrolment. The parent or legal guardian of the participants was informed that the data collected would be stored and used for future TB diagnostic studies. Although all children enrolled in the three cohorts were eligible for this retrospective study, inclusion was based on the availability of a digital chest radiograph taken at enrolment.

Computer-aided chest radiograph analysis

Stored digital chest radiography images already available from study participants were identified using study-specific identifiers and extracted in DICOM format from the MRCG at LSHTM's Picture Archiving and Communication System (PACS). The participants' metadata embedded in the DICOM files were anonymised and transferred to a dedicated digital library for use in CAD4TB software evaluation. The digital chest radiography images were analysed using a CAD4TB offline box equipped with CAD4TBv7. DICOM files from the digital library were imported to the CAD4TB box user interface. After automated analysis, numerical abnormality scores were recorded using study-specific identifiers, with the manufacturer-recommended threshold score of ≥ 60 indicating the likelihood of TB. The study personnel who conducted the CAD4TBv7 analysis of the digital chest radiography images were blinded to the clinical data and diagnoses of the study participants. Corresponding data on demography, clinical features, laboratory investigation and TB classification were extracted from each study database and merged into a single dataset for statistical analysis.

Data analysis

Due to the secondary nature of the study, there was no formal sample size calculation. The analyses included all children with presumed pulmonary TB who were consecutively recruited into the cohort studies over the specified period and had a digital chest radiograph available. For the primary analysis based on the MRS, participants with confirmed TB were considered MRS-positive and those with unconfirmed TB and unlikely TB as MRS-negative. We calculated the AUROC and point estimates for sensitivity and specificity, with the respective 95% confidence interval (95% CI), of CAD4TBv7 at the manufacturer-recommended threshold score (≥ 60) verified against the MRS. In the subgroup analysis, sensitivity and specificity of CAD4TBv7 were similarly determined by age (children < 5 versus ≥ 5 years) and by participants' source (actively traced child TB contacts versus children referred for diagnostic evaluation).

Furthermore, using the MRS, ROC analysis was used to estimate the study-specific CAD4TBv7 analytical threshold scores when benchmarked against the WHO-endorsed Target Product Profile (TPP) for a triage test for TB at a minimum sensitivity of 90% or a minimum specificity of 70% [13].

Given the recognised limitation of the MRS for TB diagnostic accuracy studies in children [24], Bayesian LCA was performed to estimate the sensitivity, specificity and 95% credible interval (95% CrI) of the three tests in this study, namely mycobacterial culture, Xpert assay and CAD4TBv7 at the manufacturer-recommended threshold. First, we assessed the performance of the three tests assuming conditional independence. We then relaxed the assumption by considering conditional dependence between Xpert and culture as both are done using the same sputum sample and differ from the CAD4TBv7 abnormality score which relies on the digital chest radiography reading. This approach models the conditional dependence

between diagnostic tests using the covariance between tests within the diseased and disease-free populations [25]. We evaluated the goodness-of-fit of both models and reported results from the best model. Full details of the methods for the Bayesian LCA, including the prior distributions used, are provided in the supplementary material. All data were analysed using Stata/SE version 17.0 (StataCorp, College Station, TX, USA).

Results

Between 1 February 2012 and 31 October 2022, 2315 children with presumed pulmonary TB were enrolled into the three studies. Of these, 1591 children were excluded from the main analysis because they had no digital chest radiograph (n=819), were household child TB contacts considered to have latent TB based on tuberculin skin test result and symptoms screening (n=732) or they had an invalid CAD4TBv7 score (n=40); invalid CAD4TBv7 scores included scores of -1 (invalid posterior-anterior/anterior-posterior chest radiographs) or -2 (invalid/unsupported DICOM). Overall, digital chest radiography images from 724 unique children were included in the main analysis. Out of these, 316 were household TB contacts who were actively traced and 408 were referred for clinical evaluation. 58 (8%) children were classified as confirmed TB, 145 (20%) as unconfirmed TB and 521 (72%) as unlikely TB. The study flow diagram in figure 1 shows the TB case definition categories and the CAD4TBv7 classification based on the manufacturer-recommended threshold score (≥ 60) for the children included in the study.

The median (interquartile range (IQR)) age of all children included in the analysis was 5.0 (2.8–8.4) years, with 321 (44%) children aged <5 years and 342 (47%) females. Overall, 35 (5%) of the enrolled children were living with HIV (CLHIV). The most common symptom was cough for >2 weeks that was reported in 465 (64%) of the enrolled children, followed by fever in 440 (61%). Detailed demographic and clinical characteristics of all children and by TB classification are presented in table 1.

The median (IQR) CAD4TBv7 score for all the children in the study was 34 (19–43). The median CAD4TBv7 score showed a statistically significant trend across the TB case definitions, with the median (IQR) score highest in children with confirmed TB (46 (28–53)) compared to children with unconfirmed TB (41 (34–46)) and unlikely TB (30 (16–40)) ($p < 0.001$). 19% (11/58) of children with confirmed TB had CAD4TBv7 scores ≥ 60 , relative to 2% (3/145) and 1% (3/521) in the unconfirmed and unlikely TB groups, respectively (figure 2a).

Given that there was only one HIV-positive child among the children with confirmed TB, we could not analyse the distribution of CAD4TBv7 by HIV status. However, among children with unconfirmed TB, the

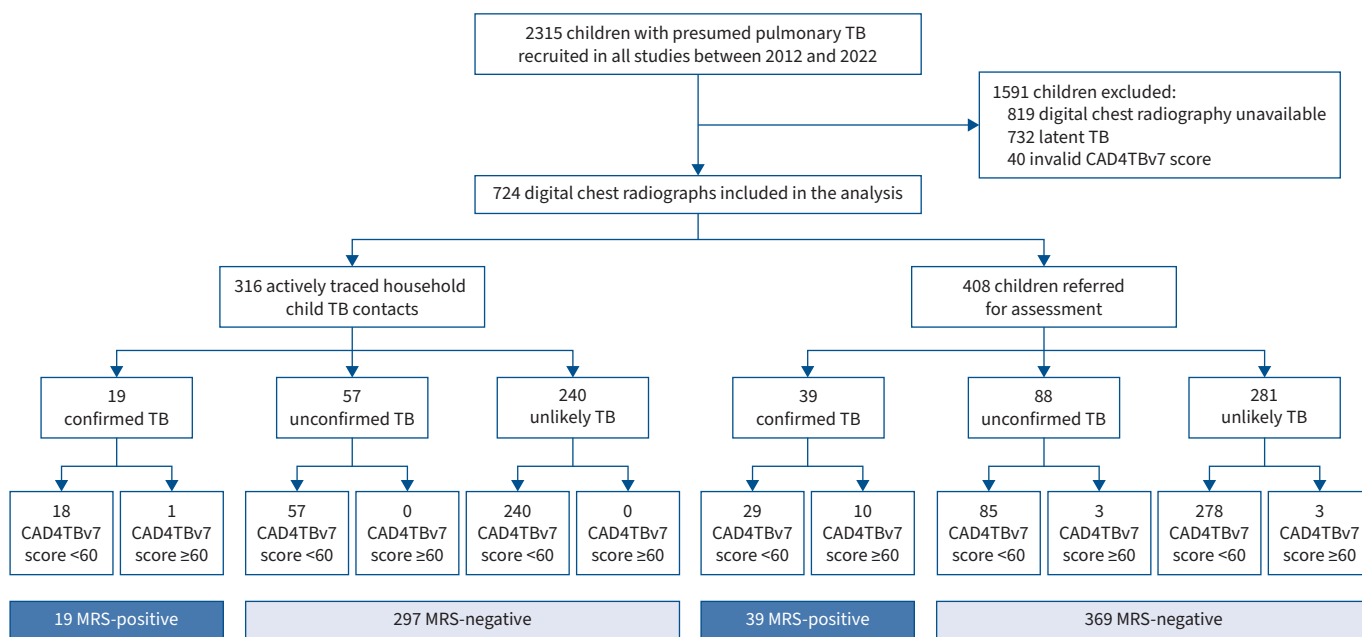


FIGURE 1 STARD diagram reporting the flow of participants in the study. TB: tuberculosis; MRS: microbiological reference standard.

TABLE 1 Demographic and clinical characteristics of participants included in the analysis

	All (n=724)	Confirmed TB (n=58)	Unconfirmed TB (n=145)	Unlikely TB (n=521)
Age, years	5.0 (2.8–8.4)	6.1 (3.2–11.7)	4.6 (2.4–7.0)	5.6 (2.8–8.6)
<5 years	321 (44.3)	21 (36.2)	79 (54.5)	221 (42.4)
≥5 years	403 (55.7)	37 (63.8)	66 (45.5)	300 (57.6)
Sex				
Female	342 (47.2)	25 (43.1)	66 (45.5)	251 (48.2)
Male	382 (52.8)	33 (56.9)	79 (54.5)	270 (51.8)
CLHIV	35 (4.8)	1 (1.7)	25 (17.2)	9 (1.7)
Symptoms				
Cough	465 (64.2)	41 (70.7)	83 (57.2)	341 (65.5)
Fever	440 (60.8)	38 (65.5)	102 (70.3)	300 (57.6)
Lethargy	184 (25.4)	28 (48.3)	45 (31.0)	111 (21.3)
Weight loss	437 (60.4)	44 (75.9)	95 (65.5)	298 (57.2)
Underweight	228 (31.5)	26 (44.8)	63 (43.5)	139 (26.7)
Previous TB	4 (0.6)	0	2 (1.4)	2 (0.4)
Actively traced	316 (43.7)	19 (32.8)	57 (39.3)	240 (46.1)
CAD4TBv7 score	34.1 (19.4–42.8)	45.9 (27.9–53.3)	41.3 (34.0–45.6)	30.4 (15.9–39.6)

Data are presented as median (interquartile range) or n (%). TB: tuberculosis; CLHIV: children living with HIV.

median (IQR) CAD4TBv7 score in CLHIV was significantly higher than the score in HIV-negative children (45 (42–48) versus 40 (31–45); $p < 0.001$). HIV status did not significantly impact CAD4TBv7 scores in children with unlikely TB (figure 2b).

Using the MRS in the overall study population, CAD4TBv7 differentiated children with confirmed TB from children with unconfirmed and unlikely TB with an AUROC of 0.70 (95% CI 0.60–0.79). In the subgroup analysis using the MRS, the AUROC of CAD4TBv7 in children aged <5 and 5–14 years was 0.75 (95% CI 0.62–0.89) and 0.68 (95% CI 0.56–0.80), respectively. Furthermore, the AUROC in children referred for evaluation and the actively traced child TB contacts was 0.73 (95% CI 0.62–0.85) and 0.62 (95% CI 0.46–0.77), respectively (figure 3). It is pertinent to note that the 95% confidence intervals of the point estimates for AUROC in the subgroup analysis were wide and overlapping.

The overall point estimates for sensitivity and specificity of CAD4TBv7 at the manufacturer-recommended threshold, verified against the MRS, were 19% (95% CI 11–31%) and 99% (95% CI 98–100%), respectively. In the subgroup analysis also using the MRS, sensitivity of CAD4TBv7 in children <5 and 5–14 years was 5% (95% CI 1–23%) and 27% (95% CI 15–43%), respectively. CAD4TBv7 demonstrated a sensitivity of 26% (95% CI 15–41%) among referred children and 5% (95% CI 1–24%) among the

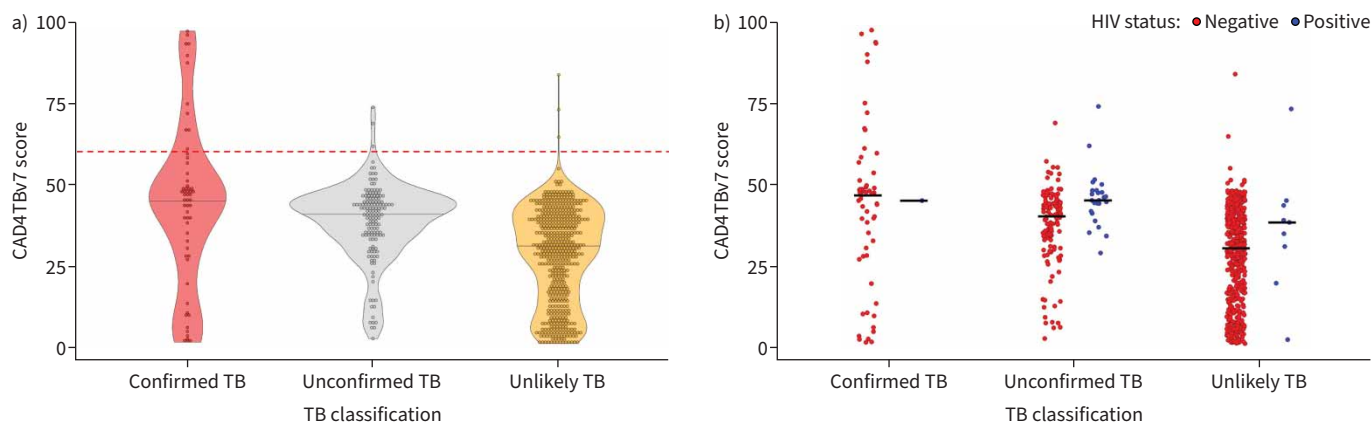


FIGURE 2 a) Violin plots of CAD4TBv7 scores stratified by diagnostic categories. The manufacturer-recommended threshold score (≥ 60) is indicated by the dashed red line. b) Scatter plots of CAD4TBv7 scores by diagnostic categories, stratified by HIV status. TB: tuberculosis.

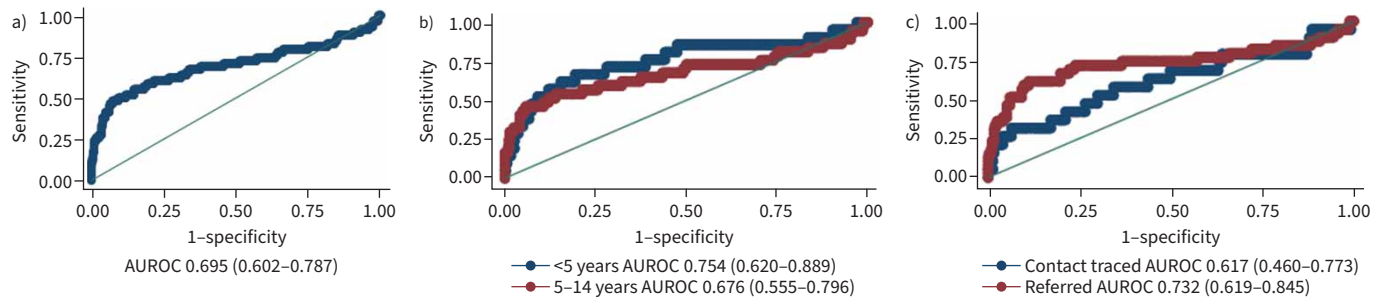


FIGURE 3 Area under the receiver operating characteristic curve (AUROC) with 95% confidence interval of CAD4TBv7 using the manufacturer-recommended threshold and verified against the microbiological reference standard: a) overall, b) by age category and c) by source (contact traced *versus* referred).

actively traced child TB contacts. The 95% confidence intervals of the point estimates for sensitivity in the subgroup analysis were also observed to be wide and overlapping. The point estimates for specificity were, however, comparable between the groups (figure 4).

To estimate study-specific CAD4TBv7 analytical thresholds, fixing the sensitivity at the WHO-endorsed minimum sensitivity for a triage test of 90.1% (95% CI 81.0–97.1%) resulted in a corresponding specificity of 6.9% (95% CI 5.1–9.1%) at an analytical threshold of 4.6. Furthermore, benchmarking the specificity at the minimum specificity for a triage test of 70.1% (95% CI 66.5–73.6%) yielded a CAD4TBv7 sensitivity of 62.1% (95% CI 48.4–75.5%) at an analytical threshold of 40.2 (table 2).

When we applied the Bayesian LCA with the assumption of conditional independence between tests, the point estimates for sensitivity of CAD4TBv7, Xpert and culture overall were 42.7% (95% CrI 29.2–57.5%), 80.1% (95% CrI 60.9–93.6%) and 72.8% (95% CrI 54.2–88.5%), respectively. Similar results were observed when allowing for conditional dependence between Xpert and culture, with CAD4TBv7 demonstrating a sensitivity of 50.3% (95% CrI 32.9–70.0%), which was relatively lower than the sensitivity of Xpert (74.6%, 95% CrI 50.6–91.9%) and culture (62.3%, 95% CrI 39.7–83.4%) (table 3). The sensitivity of CAD4TBv7 was also relatively lower than that of Xpert and culture among referred children and actively traced child contacts, irrespective of conditional assumptions between the tests. However, the sensitivity of Xpert was consistently higher than that of culture overall and among referred children irrespective of conditional assumptions between the tests, but not among actively traced child contacts (table 3). With the Bayesian LCA, all three tests in this study demonstrated comparably high specificities (>94%) overall and among the referred children and actively traced child contacts, irrespective of the assumed conditional relationship between tests (table 3). Full details of the results of the Bayesian LCA are provided in the supplementary material.

CAD4TBv7	True positive	False positive	False negative	True negative	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
HIV-negative	11	3	46	620	0.19 (0.11–0.31)	1.00 (0.99–1.00)	■—	■
HIV-positive	0	3	1	31	0.00 (0.00–0.79)	0.91 (0.77–0.97)	■—	■—
Age <5 years	1	1	20	298	0.05 (0.01–0.23)	1.00 (0.99–1.00)	■—	■
Age 5–14 years	10	5	27	361	0.27 (0.15–0.43)	0.99 (0.97–1.00)	■—	■
Contact traced	1	0	18	297	0.05 (0.01–0.24)	1.00 (0.99–1.00)	■—	■
Referred	10	6	29	363	0.26 (0.15–0.41)	0.98 (0.96–0.99)	■—	■
Overall	11	6	47	660	0.19 (0.11–0.31)	0.99 (0.98–1.00)	■—	■

FIGURE 4 Sensitivity and specificity of CAD4TBv7 using the manufacturer-recommended threshold and verified against the microbiological reference standard.

TABLE 2 Accuracy estimates (%) and CAD4TBv7 threshold score at World Health Organization (WHO) Target Product Profile (TPP) targets for a triage test

WHO TPP minimum requirement	Sensitivity (95% CI)	Specificity (95% CI)	CAD4TBv7 threshold score
Sensitivity set at 90%	90.1 (81.0–97.1)	6.9 (5.1–9.1)	4.6
Specificity set at 70%	62.1 (48.4–74.5)	70.1 (66.5–73.6)	40.2

Discussion

Our study evaluated the diagnostic accuracy of CAD4TBv7 for detecting pulmonary TB in Gambian children aged <15 years, verified against the MRS of positive Xpert or mycobacterial culture results. We addressed the limitation of the MRS in children by also using Bayesian LCA. Against the MRS, CAD4TBv7 classified children with confirmed TB from those with unconfirmed and unlikely TB with an AUROC of 0.70 and demonstrated a suboptimal sensitivity of 19.0% but a high specificity of 99.0%. Notably, Bayesian LCA confirmed the consistently lower sensitivity of CAD4TBv7 compared to Xpert and culture, irrespective of the assumed conditional relationships between the tests. However, all three tests demonstrated a comparably high specificity that was >94%.

The findings from our study provide objective evidence to support the WHO's recommendation that the currently available CAD solutions are not suitable for children <15 years old [26]. The Stop TB Partnership recently provided a list of CAD products that developers have intended for use in children as young as 2 years [27]. However, data on the systematic assessments of their unbiased performance in exclusive populations of children aged <15 years are lacking.

Our finding that CAD4TBv7 classified children with confirmed TB with an AUROC of 0.70 is similar to the report from another paediatric study that aimed to measure and optimise the performance of CAD4TBv7 [28]. The authors in that study used a radiological reference standard set by human readers and also did not report corresponding sensitivity and specificity estimates, thus limiting our ability to compare findings in detail. However, the suboptimal sensitivity of CAD4TB as reported in our study, using a MRS and the more robust statistical approach of Bayesian LCA, contrasts with the growing evidence supporting CAD4TB use for TB screening and triage in adults [8–11]. This difference might be due to the distinct radiological features of TB in children compared to adults.

TABLE 3 Estimated sensitivity and specificity of CAD4TBv7, Xpert and culture as well as the prevalence of tuberculosis in the study from Bayesian latent class analysis

Sample	Independence assumption	Diagnostic test	Sensitivity (95% CrI)	Specificity (95% CrI)	Prevalence (95% CrI)
All children	Independence between all the three tests	CAD4TBv7	42.7 (29.2–57.5)	97.9 (96.6–98.8)	5.7 (3.9–8.0)
		Xpert	80.1 (60.9–93.6)	97.9 (96.4–98.7)	
		Culture	72.8 (54.2–88.5)	97.6 (96.1–98.7)	
	Conditional dependence between Xpert and culture only	CAD4TBv7	50.3 (32.9–70.0)	98.0 (96.7–98.9)	4.7 (2.8–7.4)
		Xpert	74.6 (50.6–91.9)	96.5 (94.5–98.3)	
		Culture	62.3 (39.7–83.4)	96.1 (94.1–97.8)	
Referred children	Independence between all the three tests	CAD4TBv7	47.1 (32.4–62.4)	96.3 (94.1–97.9)	8.1 (5.4–11.4)
		Xpert	80.3 (61.0–93.8)	96.7 (93.4–98.5)	
		Culture	70.4 (50.2–87.7)	97.6 (95.7–99.0)	
	Conditional dependence between Xpert and culture only	CAD4TBv7	53.2 (35.5–72.1)	96.4 (94.1–98.0)	6.9 (4.1–10.7)
		Xpert	73.0 (48.5–91.2)	94.7 (91.7–97.3)	
		Culture	60.2 (37.0–82.7)	95.7 (93.0–98.0)	
Actively traced children	Independence between all the three tests	CAD4TBv7	48.1 (30.0–67.1)	97.3 (95.2–98.7)	5.2 (3.0–8.4)
		Xpert	68.4 (40.0–90.7)	97.9 (95.8–99.2)	
		Culture	73.3 (46.0–92.5)	96.3 (93.6–98.2)	
	Conditional dependence between Xpert and culture only	CAD4TBv7	50.6 (31.0–70.9)	97.3 (95.1–98.7)	4.8 (2.5–8.2)
		Xpert	63.4 (33.7–88.8)	97.0 (94.4–98.8)	
		Culture	66.7 (37.5–90.1)	95.0 (91.9–97.4)	

Data are presented as %. CrI: credible interval.

We also report that CAD4TBv7 demonstrated a sensitivity of only 5.0% in children aged <5 years compared to 27.0% among children aged 5–14 years. This strongly suggests the need for age-specific considerations in developing and implementing CAD for TB detection in children. While current CAD algorithms perform well in identifying abnormalities within the lung parenchyma, their ability to detect abnormalities in the lymph nodes and large airways, which are frequently observed in young children with TB, is unclear [28]. This limited capability may hinder the overall performance of CAD algorithms in diagnosing childhood TB.

While CAD4TBv7 shows high specificity in our study, its observed low sensitivity significantly limits its clinical utility for risk stratification as a triage tool among children with presumed TB, as many children with TB could be missed.

The WHO recommends selecting CAD thresholds for likelihood of TB on digital chest radiography images based on various factors, including demographics, laboratory capacity and programmatic goals [26]. In our study, benchmarking the performance of CAD4TBv7 to either the WHO-endorsed TPP minimum sensitivity of 90% or specificity of 70% for a triage test resulted in significant trade-offs. Benchmarking the sensitivity at $\geq 90\%$ resulted in a specificity of only 6.9% at a CAD4TBv7 analytical threshold of ≥ 4.6 and the specificity set at $\geq 70\%$ yielded a sensitivity of 62.1% at an analytical threshold of ≥ 40.2 . These findings therefore support the WHO recommendation of context-specific analytical thresholds.

It is also notable that in our study, using Bayesian LCA, we found that the sensitivity of the Xpert assay frequently exceeds, or at the minimum is comparable to, the sensitivity of mycobacterial culture within different conditional assumptions between the tests. These results could reflect the nature of our study population being ambulant children with presumptive TB who were enrolled in an outpatient childhood TB clinic. Such children could have milder TB disease with lower bacterial load relative to children who are hospitalised with TB. Our findings, therefore, support the assertions made in some other studies that Xpert assay could outperform mycobacterial culture, particularly in the context of lower prevalence of TB and possible paucibacillary TB disease [29, 30].

Our study has some limitations. We did not evaluate the performance of CAD4TBv7 in comparison to a human reader given the secondary nature of our study in which we reported findings from a retrospective study; we do not have expert human reader reports for the library of digital chest radiographs used for the analysis. A number of digital chest radiographs taken before the deployment of PACS at the study site were unavailable and could not be included in the analysis. However, given that the proportion of children in the various TB classes in our study is similar to that observed in the individual studies from which data were obtained, as well as in studies of childhood TB in TB-endemic settings, this is unlikely to have introduced selection bias. Our study included a relatively small number of children with confirmed TB, which is reflected in the wide 95% confidence intervals when using the MRS in the overall or subgroup analyses; however, this further highlights the challenge of bacteriological confirmation of TB and its implication in using the MRS in diagnostic accuracy studies of TB in children, and justifies our decision to use the Bayesian LCA to address this inherent limitation of MRS. Co-infection with HIV could influence the diagnostic accuracy of CAD4TB. In our study, we also had relatively small numbers of HIV-infected children, with only one child co-infected with HIV among the children with confirmed TB. This reflects the epidemiology of HIV in The Gambia, which has a low/concentrated HIV prevalence; as such, our findings might not be generalisable to settings with high or generalised HIV prevalence.

In conclusion, we have shown that the sensitivity of CAD4TBv7 for detecting childhood pulmonary TB is suboptimal. This study provides objective evidence in support of the need to refine the existing CAD algorithms and to develop paediatric-specific CAD systems that could aid TB detection in children. Creating extensive and well-characterised libraries of paediatric digital chest radiography images from settings with diverse HIV and TB epidemiology is essential for optimising the current CAD systems and for developing the much needed paediatric-specific CAD systems in the future.

Acknowledgement: We acknowledge the support and contributions of the data management team at the MRCG at LSHTM. We thank the children whose digital chest radiography images were used in this study, and their parents/guardians and families. We also thank the field, clinic and laboratory teams who facilitated the recruitment of the children.

Data availability: Data from this study could be made available after publication and upon reasonable request.

Ethics statement: Ethical approvals were obtained from the Gambia Government/MRC Joint Ethics Committee.

Author contributions: V.F. Edem and T. Togun conceptualised the study. V.F. Edem, E. Nkereuwem, S.A. Owusu and T. Togun contributed to the study design, implementation and data acquisition. E. Nkereuwem, A.K. Sillah, B. Saidu, U. Egere, B. Kampmann and T. Togun contributed to the data generation. V.F. Edem, E. Nkereuwem, S.C. Agbla and T. Togun analysed and interpreted the data. V.F. Edem and E. Nkereuwem wrote the first draft of the manuscript with input from A.G. Forson, U. Egere, B. Kampmann and T. Togun. All authors contributed to the revision and corrections on multiple iterations of the manuscript.

Conflict of interest: The authors have no potential conflicts of interest to disclose.

Support statement: This study was supported by a European and Developing Countries Clinical Trials Partnership (EDCTP) grant to the West African Networks of Excellence for TB, AIDS and Malaria (WANETAM) (EDCTP-RegNet2015-1049). The childhood TB studies that generated the stored digital chest radiographs used in this study were funded by a UK MRC programme grant (MR/K011944/1 to B. Kampmann) and a UKRI-GCRF grant (MR/P024270/1 to B. Kampmann). The funder of the study had no role in the study design, data acquisition, data analysis, data interpretation or writing of the manuscript. Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 World Health Organization. Global tuberculosis report 2023. www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023 Date last accessed: 25 January 2024.
- 2 Basu Roy R, Whittaker E, Seddon JA, *et al.* Tuberculosis susceptibility and protection in children. *Lancet Infect Dis* 2019; 19: e96–e108.
- 3 Dodd PJ, Gardiner E, Coghlan R, *et al.* Burden of childhood tuberculosis in 22 high-burden countries: a mathematical modelling study. *Lancet Glob Health* 2014; 2: e453–e459.
- 4 World Health Organization. Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches. 2016. <https://iris.who.int/handle/10665/252424> Date last accessed: 25 January 2024.
- 5 Tripti P, Madhukar P, Faiz Ahmad K, *et al.* Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J* 2015; 46: 1816–1819.
- 6 Kulkarni S, Jha S. Artificial intelligence, radiology, and tuberculosis: a review. *Acad Radiol* 2020; 27: 71–75.
- 7 Stop TB Partnership. Screening and triage for TB using computer-aided detection (CAD) technology and ultra-portable X-ray systems: a practical guide. 2023. www.stoptb.org/resources-implementing-cad-and-xray/cad-and-x-ray-practical-implementation-guide Date last accessed: 2 February 2024.
- 8 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.
- 9 Breuninger M, van Ginneken B, Philipsen RHHM, *et al.* Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. *PLoS One* 2014; 9: e106381.
- 10 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.
- 11 Tavaziva G, Harris M, Abidi SK, *et al.* Chest X-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: an individual patient data meta-analysis of diagnostic accuracy. *Clin Infect Dis* 2022; 74: 1390–1400.
- 12 Nkereuwem E, Kampmann B, Togun T. The need to prioritise childhood tuberculosis case detection. *Lancet* 2021; 397: 1248–1249.
- 13 Denkinger CM, Kik SV, Cirillo DM, *et al.* Defining the needs for next generation assays for tuberculosis. *J Infect Dis* 2015; 211: Suppl. 2, S29–S38.
- 14 García-Basteiro AL, López-Varela E, Augusto OJ, *et al.* Radiological findings in young children investigated for tuberculosis in Mozambique. *PLoS One* 2015; 10: e0127323.
- 15 Graham SM, Ahmed T, Amanullah F, *et al.* Evaluation of tuberculosis diagnostics in children: 1. Proposed clinical case definitions for classification of intrathoracic tuberculosis disease. Consensus from an expert panel. *J Infect Dis* 2012; 205: Suppl. 2, S199–S208.
- 16 Marais BJ, Graham SM, Cotton MF, *et al.* Diagnostic and management challenges for childhood tuberculosis in the era of HIV. *J Infect Dis* 2007; 196: Suppl. 1, S76–S85.
- 17 Egere U, Sillah A, Togun T, *et al.* Isoniazid preventive treatment among child contacts of adults with smear-positive tuberculosis in The Gambia. *Public Health Action* 2016; 6: 226–231.
- 18 Nkereuwem E, Togun T, Gomez MP, *et al.* Comparing accuracy of lipoarabinomannan urine tests for diagnosis of pulmonary tuberculosis in children from four African countries: a cross-sectional study. *Lancet Infect Dis* 2021; 21: 376–384.

- 19 Sabi I, Olomi W, Nkereuwem E, *et al.* Diagnosis of paediatric TB using Xpert MTB/RIF Ultra on fresh respiratory samples. *Int J Tuberc Lung Dis* 2022; 26: 862–868.
- 20 Togun T, Hoggart CJ, Agbla SC, *et al.* A three-marker protein biosignature distinguishes tuberculosis from other respiratory diseases in Gambian children. *EBioMedicine* 2020; 58: 102909.
- 21 Togun TO, Egere U, Sillah AK, *et al.* Contribution of Xpert MTB/RIF to the diagnosis of pulmonary tuberculosis among TB-exposed children in The Gambia. *Int J Tuberc Lung Dis* 2015; 19: 1091–1097.
- 22 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.
- 23 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351: h5527.
- 24 DiNardo AR, Detjen A, Ustero P, *et al.* Culture is an imperfect and heterogeneous reference standard in pediatric tuberculosis. *Tuberculosis* 2016; 101S: S105–S108.
- 25 Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; 57: 158–167.
- 26 World Health Organization. WHO consolidated guidelines on tuberculosis: module 2: systematic screening for tuberculosis disease. 2001. www.who.int/publications/i/item/9789240022676 Date last accessed: 25 February 2024.
- 27 Stop TB Partnership. Introduction to CAD products for TB. 2024. <https://ai4hlth.wixsite.com/website-1> Date last accessed: 10 August 2024.
- 28 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.
- 29 Nguyen HV, de Haas P, Nguyen HB, *et al.* Discordant results of Xpert MTB/Rif assay and BACTEC MGIT 960 liquid culture to detect *Mycobacterium tuberculosis* in community screening in Vietnam. *BMC Infect Dis* 2022; 22: 506.
- 30 Shi J, Dong W, Ma Y, *et al.* GeneXpert MTB/RIF outperforms mycobacterial culture in detecting *Mycobacterium tuberculosis* from salivary sputum. *Biomed Res Int* 2018; 2018: 1514381.