International Journal of Innovative Healthcare Research 13(4):27-36, Oct-Dec, 2025

© SEAHI PUBLICATIONS, 2025 **www.seahipublications.org** **ISSN: 2354-2950**

doi:10.5281/zenodo.17311645

AI-Powered TB Screening: Deriving an Optimal Computer-Aided Detection for Tuberculosis Score for Active Case Finding in Diverse Nigerian Communities

[1]Dr Odume Bethrand, MD, MPH.; [2]Dr Inumanye N. Ojule; [3]Emmanuel Etim Clement; [4]Daniel Okon; [5]Professor Best Ordinioha & [6]Professor Elias Aniwada

[1]Knowledge Network for Disease Control and Vigilance (KNCV) Nigeria/School of Public Health, University of Port Harcourt, Port Harcourt, Nigeria
bodume@kncvnigeria.org

[2]Department of Community Medicine, University of Port Harcourt Teaching Hospital, Rivers State, Nigeria
inumanye.ojule@uniport.edu.ng

[3]Department of Public Health, University of Port Harcourt School of Public Health, Rivers State, Nigeria
emmanuel_clement@uniport.edu.ng

[4]Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria
daniel.okon@uniport.edu.ng

[5]Department of Public Health, University of Port Harcourt School of Public Health, Rivers State, Nigeria
Best.ordinioha@uniport.edu.ng

[6]Department of Community Medicine, College of Medicine, University of Nigeria Enugu Campus, Nigeria
Elias.aniwada@unn.edu.ng

ABSTRACT
Tuberculosis remains a major public health challenge in Nigeria. Nigeria is one of the 30 high-burden TB countries, ranking 6th globally and contributing 4.6% of the world's total TB cases. One of the biggest barriers to TB control is delayed or missed diagnosis, due to limited access to early screening tools and reliance on symptom-based screening, which lacks sensitivity. To address this gap, the country's TB program is funded to deploy Portable Digital X-ray (PDX) with CAD4TB AI to improve early TB case detection. Sadly, there has not been an established cutoff point for calibration of the CAD4TB AI with which the portable digital X-ray operates in our setting. The study aimed to use the generated data from the portable digital x-ray with CAD4TB AI intervention in Nigeria across 14 states; Rivers, Cross Rivers, Akwa Ibom, Anambra, Imo, Delta, Benue, Nasarawa, Bauchi, Taraba, Plateau, Kano, Kaduna and Katsina from December 2021 to June 2022 to calibrate CAD4TB-AI based on the most appropriate thresholds for use in diverse community settings. The study was a retrospective observational study using secondary data from community-based TB screening with portable digital X-ray with AI. All Adults screened using PDX with CAD4TB AI in multiple field locations were studied. Validation involved CAD AI findings compared to molecular confirmation (GeneXpert, etc) to calculate diagnostic accuracy. Statistical Analysis included sensitivity, specificity, PPV and NPV of CAD-AI. The ROC curves for AI scoring thresholds were generated and compared to the case detection yield by age and settings. The mean CAD4 score was 43.13 with a range of scores 0 to 99. At a sensitivity of 90% and specificity of 40% (Sen 90.20% Versus Spec 39.50%), the CAD4TB threshold score of 37 was chosen for the general population screening. The findings support the recommendations by WHO for the use of AI in TB triage and set a feasible threshold score for active TB case finding using PDX with AI in Nigeria.
Keywords: Computer-Aided Detection (CAD), Tuberculosis Thresholds, Community Case Findings, Age-Based Variations

INTRODUCTION

In 2019, 10 million people were infected with TB worldwide, with an estimated three million people who did not receive a clinical diagnosis. This has created a persistent gap in TB case detection (WHO, 2021), even when the United Nations General Assembly (UNGA) in 2018 committed to diagnosing and treating 40 million people with TB in 2022. (WHO, 2021) To close the diagnostic gap and achieve this ambitious target, an urgent need exists to deploy strategies that will increase TB case finding (Zaidi et al., 2018).

More than 10 million people continue to fall ill with TB every year and the number has been rising since 2021. Urgent action is required to end the global TB epidemic by 2030, a goal that has been adopted by all Member States of the United Nations (UN) and the World Health Organization (WHO, 2024), Global tuberculosis report 2024. The evolving nature of the causative agent has resulted in the development of extrapulmonary TB and drug-resistant variants in addition to the respiratory burden of the disease. Currently, about 25% of the global population is either infected with active TB or has latent asymptomatic tuberculosis, with an annual incidence of 10 million new infections and 1.2 million deaths. About 15% of tuberculosis cases occur in the form of extrapulmonary infections (infection affecting other parts of the body than the lungs). Such cases have been reported to be particularly difficult to diagnose and treat and are easily spread via organ transplant (Olaleye et al., 2023)

WHO recognized Chest X-ray (CXR) interpreted by a qualified radiologist as an initial screening test given its high sensitivity (WHO, 2013). CXR can detect TB in patients who do not present with conventional TB symptoms (Onozaki et al, 2015). Conversely, lack of access to qualified radiologists and the fact that X-ray equipment is most often available only at referral hospitals, hampers wide-scale use of CXR screening at the community level in many high TB burden countries (Pande et al, 2015).

With the introduction of sensitive rapid molecular diagnostics for TB, such as Xpert MTB/RIF testing, there is a need for improved approaches for screening to reduce the number of presumptive cases referred for such tests. (Durovni et al, 2015). An increasing body of evidence from high burden countries suggests that the use of digital CXR equipment and the automated reading of CXR with Computer Aided Detection (CAD), as a pre-screening tool, in conjunction with an expensive molecular test such as Xpert, can improve case finding efforts (Van't Hoog et al, 2014). However, previous studies showed that triaging thresholds are not universal and require careful adjustment to demographic characteristics, laboratory capacities, budget, healthcare settings and study goals. (Murphy et al, 2019, Philipsen et al, 2018, Qin et al, 2019, Khan et al, 2020, Fehr et al, 2021).

Worth noting is that there is an inherent trade-off in the selection of the threshold score; a lower threshold score will maximize sensitivity of the tool to detect true TB cases among the population being screened, but will incur additional costs related to unnecessary follow-on diagnostic testing due to reduced specificity. On the other hand, a higher threshold score will reduce the volume, and thus costs, of follow-on diagnostic testing and will likely identify more severe cases, but its reduced sensitivity will result in missed cases. This buttresses the need for local calibration of CAD (World Health Organization, 2021c).

Computer-aided detection (CAD) has been recommended as a tuberculosis screening tool. However, there are limited data about its utility, specifically in a community-based setting where the targeted population and the highest burden of undetected individuals reside (Scott et al., 2025). In Nigeria, diagnostic coverage for TB case finding still stands at 40% for GeneXpert, and a greater gap still exists for AI-based diagnostic tools (NTBLCP Annual Report 2021). Chest x-ray with AI has been recommended by the WHO both for TB screening and diagnosis. Uptake of this technology has been shown to have the greatest yield in TB diagnosis; however, the use has been slow within the NTBLCP. Presently, only 12 units are in operation for TB case finding across nine states in Nigeria, despite demonstrated high yield in the KNCV pilot intervention in selected hard-to-reach locations in Nigeria.

Aim and Objectives of the Study

The main objective of this study was to use the generated data from the new portable digital X-ray with AI intervention from December 2021 to June 2022 to calibrate CAD4TB version 7 based on the most appropriate thresholds for use in diverse community settings and population groups by

comparing the performance of CAD4TB in detecting presumptive pulmonary TB against a bacteriological positive TB case in Nigeria. The specific objectives were:
1. To determine the variations in CAD4TB score with the yield of bacteriological and clinical TB diagnosis.
2. To ascertain if differences exist in CAD4TB score by age categories related to TB yield
3. To determine the CAD4TB score threshold appropriate to each targeted population group during community TB active case finding (ACF).

Research Questions

1. What are the variations in CAD4TB score with the yield of bacteriological and clinical TB diagnosis in Nigeria?
2. What are the differences that exist in CAD4TB score by age categories related to TB yield in Nigeria?
3. What is the CAD4TB score threshold appropriate to each targeted population group during community TB active case finding (ACF) in Nigeria?

METHODS AND MATERIALS

This study was conducted using data from mixed population settings generated during active case finding in hard-to-reach remote villages, prisons, internally displaced people's camps (IDPs), urban slums and general population settings across six states in Nigeria, including Benue, Cross River, Delta, Katsina, Kano and Nasarawa states under the KNCV intervention. The selection of these states and sites for screening programs was purposive, based on the ease of the organizational operations. These states fall under three geopolitical regions of North-Central, North-West and South-South parts of Nigeria. This study adopted a retrospective case-control which was conducted under standard programmatic settings among individuals presenting for TB screening and those participating in a community-based active case finding activities study. Cases are individuals aged 15 years and above for whom a CXR and a bacteriological confirmatory diagnosis of pulmonary TB exists. Controls are individuals aged 15 years and above for whom a CXR was conducted and a TB diagnosis ruled out based on the negative bacteriological test results. It involved a review of routinely collected data on PDX implementation across six states in Nigeria, implementing active TB case finding using portable digital x-ray with CAD4TB artificial intelligence. Participants were drawn from previously conducted mixed screening activities for TB active case findings across the states, as reviewed in the electronic presumptive register. The study population included all patients in the selected states who were screened, regardless of whether they presented with respiratory or constitutional symptoms for TB, as documented in the electronic presumptive registers. Pregnant women, individuals who did not get screened with CXR but had a sputum sample collected and tested and all confirmed TB clients who were on treatment before the screening were excluded. All study samples were retrospectively identified using data from the electronic presumptive register generated by the seven PDX across the six states and the web-based application for x-ray and clinical symptom evaluation platform (XMAP). A data collection tool (Proforma) was developed for data extraction from the e-presumptive register and the XMAP portal. Trained Data clerks extracted data from the PDX e-presumptive registers from the six states. The Researcher had controlled access to the XMAP platform, which has accurate storage of all the uploaded CXR and clinical outcomes of the patients. The researcher segregated the data according to local government and states of implementation in Nigeria. Data on all the bacteriologically positive and negative TB cases were extracted, as well as tracking of the CAD4TB scores by the researcher. Demographic, clinical, and microbiological data were obtained from each study. The GeneXpert MTB/RIF Ultra (Xpert Ultra, Cepheid) nucleic acid amplification test was used in the study. Chest x-ray images were anonymized and collected in DICOM format. CAD analysis was subsequently performed using CAD4TB version 7 software (Delft Imaging, Hertogenbosch, Netherlands). Chest x-rays were additionally reviewed by expert human readers in the original studies (including PG) to detect lung abnormalities, including cavitary disease as a marker for active disease and probable infectiousness. In this study, cases were defined as participants who were tuberculosis positive, i.e, individuals diagnosed with pulmonary tuberculosis by Xpert Ultra or sputum culture positivity, or both (microbiological reference standard). The controls were defined as participants who were tuberculosis negative, i.e, individuals with both a negative Xpert Ultra and sputum culture result.

Data entry into the e-presumptive registers was done by data clerks who are already working under the KNCV Tuberculosis Local Organizations Network (TB LON) projects; as such, they are already experienced in entering the required data variables. These data are usually vetted by the State Programme Officers (SPO) or State Programme Managers (SPM) before it was submitted to the Monitoring and Evaluation (M&E) Officer in the central office, where they undergo further scrutiny. The Active Case Finding (ACF) coordinators, SPO, and SPM of the respective states were readily available to answer questions or provide clarifications to gray areas.

The data collected were entered into a standardized CSV spreadsheet and uploaded to the online CAD for TB detection calibration tool developed by WHO to facilitate the analysis of data needed for CAD calibration studies. Further analysis was undertaken. Categorical variables were summarized using frequency and percentages and compared using the Pearson χ^2 test (χ^2 test) or Fisher's exact test. Continuous variables were summarized using Mean and standard deviation for normally distributed data or Median and Interquartile Range (IQR). Also, continuous variables like mean CAD score were compared using Student's t-test (parametric data) for two (2) groups or Analysis of Variance (ANOVA) for more than two (2) groups. The level of significance was at $p < 0.05$.

The ethical approval was sought from the National Health Research Committee (NHREC) before commencement of this study. There is no invasive procedure as the study will be done with secondary data. All necessary steps were taken to ensure the confidentiality of patients' personal information; all the results from CAD4TB and bacteriological tests in the E. presumptive register were de-identified and coded with unique identifiers and were securely protected in a password-protected file in a secure computer, which was accessible only to authorized persons.

RESULTS AND FINDINGS

The result is based on 16047 participants who had complete results on both bacteriological tests and CAD4 score out of 18529 participants that were studied. This corresponds to 86.6% of participants. They were all aged 15 years and above

Table 1: Sociodemographic characteristics and setting of participants

	Frequency (n = 16047)	Percent (100)
Location/Setting		
Health Facility	1193	7.4
Slum	4058	25.3
General population	5620	35.0
Camp	455	2.8
Hard to Reach	2644	16.5
Correctional centre	1792	11.2
Others	285	1.8
Age Cat (Years)		
15-24	2708	16.9
25-34	3205	20.0
35-44	3012	18.8
45-54	2416	15.1
55-64	2068	12.9
65+	2638	16.4
Mean (SD)	***43.45 (18.37)***	
Gender		
Male	10046	62.6
Female	6001	37.4
CAD4 Score		
Mean (SD)	***43.13 (24.10)***	
Range	***0 – 99.3***	

Table 1 shows Sociodemographic characteristics and settings of participants. A higher proportion of residents were in the general population or normal home setting, 5,620 (35.0%), followed by those in slums, 4,058 (25.3%) then hard-to-reach places 2,644 (16.5%). The least was others 285 (1.8%) followed by those who live in camp 455 (2.8%). The mean age of participants was 43.5 years with a standard deviation of 18.4 years. Those aged 25 to 34 years were the highest 3,505 (20.0%) followed by 35 to 44 years 3,012 (18.8%), while the least was 55 to 64 years 2,068 (12.9%). Males were more than Females, 10,046 (62.6%) versus 6,001 (374%). The Mean CAD4 score for participants was 43.13, with a range of scores 0 to 99.3.

Table 2: Tuberculosis characteristics and symptoms of participants

	Frequency (n = 16047)	Percent (100)
Previously Treated for TB		
Yes	802	5.0
No	15245	95.0
TB disease		
Yes	1752	10.9
No	14295	89.1
Symptoms		
Cough		
Yes	14468	90.2
No	1579	9.8
Fever		
Yes	6822	42.5
No	9225	57.5
Weight Loss		
Yes	4630	28.9
No	11417	71.1
Night Sweat		
Yes	5294	33.0
No	10753	67.0

Table 2 shows TB characteristics and symptoms of participants. About 802 (5.0%) were previously treated for TB. Among all the participants, 1,752 (10.9%) have TB disease. Based on symptoms: 14,468 (90.2%) have cough, 6,822 (42.5%) have fever, 4,630 (28.9%) have weight loss and 5,294 (33.9%) have night sweats.

Table 3: Comparison of Mean CAD4 scores of participants based on Location/setting and Sociodemographic characteristics

	Mean	Std. Dev	ANOVA	p value
Location/Setting				
Health Facility	55.05	17.96		
Slum	49.53	21.83		
General population	37.17	22.45	67.412	<0.001`
Camp	42.91	24.13		
Hard to Reach	46.82	24.02		
Correctional centre	47.65	21.34		
Others				
Age Cat (Years)				
15-24	41.25	20.71		
25-34	46.30	24.44		
35-44	46.52	23.33	9.485	<0.001
45-54	45.34	23.10		
55-64	46.13	22.74		
65+	48.50	21.58		
			T test	
Male	46.05	22.93	1.561	0.118
Female	44.96	22.55		

Table 3 shows a comparison of Mean CAD4 scores of participants based on location/setting and Sociodemographic characteristics. Those located at health facilities have the highest mean value of 55,05 followed by those in slums mean value of 49.53 then those at correctional centres mean value of 47.65, with the least among general population (normal home) mean value of 37.17. There was a statistically significant difference in mean CAD4 score among them p < 0.001. Based on age there were similar mean scores (between 41.0 and 49.0). Those aged 65 years and above have highest mean value of 48.50 followed by those aged 35 to 44 years mean value of 46.52 with the least was those aged 15 to 24 years mean value of 41.25. There was equally a statistically significant difference in mean CAD4 score among them p < 0.001. For gender the mean values were 46.05 and 44.96 for males and females respectively. There was no statistically significant difference in mean CAD4 score between them p = 0.118.

Table 4: Comparison of Mean CAD4 scores based on TB characteristics of participants

	Mean	Std. Dev	T test	p value
Previously Treated for TB				
Yes	48.38	23.54	2.857	0.004
No	45.32	22.69		
TB disease				
Yes	67.92	22.41	48.88	<0.001
No	40.10	23.09		
Symptoms				
Cough				
Yes	45.44	22.93	2.433	0.015
No	48.64	20.79		
Fever				
Yes	48.45	22.06	9.438	< 0.001
No	42.14	23.24		
Weight Loss				
Yes	45.66	22.80	1.616	0.106
No	40.12	18.35		
Night Sweat				
Yes	44.77	23.68	2.312	0.132
No	45.99	22.46		

Table 4 shows comparison of Mean CAD4 scores on TB characteristics of participants. The mean value was higher for those previously treated for TB compared to those not previously treated (48.38 versus 45.32). There was a statistically significant difference in mean CAD4 score between them p = 0.004. Also, the mean value was higher for those with TB compared to those without TB (67.92 versus 40.10). There was an equally significant difference in the mean CAD4 score between them p < 0.001. Based on symptoms, there were statistically significant differences in mean CAD4 score for those with cough mean value 45.44 versus no cough mean score 48.64 (p = 0.015) and those with fever mean score 48.45 versus no fever 42.14 (p < 0.001) but not statistically significant difference for those with weight loss mean score 45.66 versus no weight loss mean score 40.12 (p = 0.106) and those with night sweat mean score 44.77 versus no night sweat mean score 45.99 (p = 0.132).

Table 5: Distribution of AUC, sensitivity, specificity and CAD4 score of overall participants

	Sensitivity %	Specificity %	CAD4 Score	AUC (95% CI)
Overall				
Sen (approx. 90%) ǁ Spec	***90.20***	39.50	37	0.810 (0.799 – 0.822)
Sen ǁ Spec (approx. 70%)	80.70	***69.70***	49	

**Sen – Sensitivity* **Spec - Specificity*

Table 5 shows the Overall AUC of participants. The AUC was 0.810 with a confidence Interval of 0.799 to 0.822. The CAD4 Threshold cut off is 37 (Sen 90.20%, Spec 39.50%), considering the missed cases of 13 versus and True Positive (TP) of 116 against missed cases of 25 and TP of 104 (Sen 80.70%, Spec 69.70%).

Table 6: Distribution of AUC, sensitivity, specificity and CAD4 score of participants based on Age category

Age Categories (Years)	Sensitivity %	Specificity %	CAD4 Score	AUC (95% CI)
15-24				
Sen (approx. 90%) ǁ Spec	90.00	28.80	***29***	0.801 (0.766 – 0.836)
Sen ǁ Spec (approx. 70%)	78.40	69.40	46	
25-34				
Sen (approx. 90%) ǁ Spec	90.20	49.10	***40***	0.837(0.814 - 0.859)
Sen ǁ Spec (approx. 70%)	83.10	70.30	47	
35-44				
Sen (approx. 90%) ǁ Spec	90.40	41.50	***38***	0.810(0.786 – 0.835)
Sen ǁ Spec (approx. 70%)	80.80	69.60	49	
45-54				
Sen (approx. 90%) ǁ Spec	90.10	41.90	***38***	0.830(0.800 – 0.859)
Sen ǁ Spec (approx. 70%)	83.30	69.40	50	
55-64				
Sen (approx. 90%) ǁ Spec	90.10	32.10	***36***	0.758(0.722 – 0.793)
Sen ǁ Spec (approx. 70%)	71.70	69.50	53	
65+				
Sen (approx. 90%) ǁ Spec	90.20	43.10	***41***	0.795(0.768 – 0.823)
Sen ǁ Spec (approx. 70%)	80.70	69.60	52	

**Sen – Sensitivity* **Spec - Specificity*

Table 6 shows the distribution of AUC of participants based on age of participants. The AUC was high for all ages with the highest value of 0.830 among those aged 45 to 54 years and the lowest value 0.758 among those aged 55 to 64 years. However, AUC was > 0.800 for all ages < 55 years but < 0.800 for those 55 years and above. The CAD4 Threshold cut Off scores were 29 for those aged 15-24 years, 40 for 25-34 years, 38 for 35-44 years and 45-54 years, 36 for 55-64 years and 41 for 65+ years, corresponding respectively to a sensitivity of approximately 90%.

DISCUSSION OF FINDINGS

The Current study reported that the Mean CAD4 score for participants was 43.13, with a range of scores 0 to 99. Overall, at a sensitivity of about 90% and corresponding specificity of 40% (sen 90.20% Versus spec 39.50%) the CAD4TB threshold score of 37 is chosen. This is informed by the fact that when trying to maximize case detection, high sensitivity is more important than high specificity. When trying to maximize the efficiency of the algorithm and preserve costly confirmatory test cartridges, specificity may be prioritized even during screening or triage. At this cut off missed TB cases was 13 (9.80%), True Positive (TP) 116 and attracted about 1,529,856 dollars for confirmatory test. Likewise, at Specificity of about 70% CAD4TB and corresponding sensitivity of about 80% (sen 80.30% versus soec 71.20%) CAD4TB score was 49. At this cut off missed TB cases was 25 (19.30%), TP 104 and attracted about 860,273 dollars for confirmatory test. This is not in line with WHO's minimum target accuracy for TB triage tests, which is ≥90% sensitivity and ≥70% specificity. This score demonstrated high accuracy and meets the desired standard value for use in screening and triaging for TB.

This study found CAD to have good accuracy (AUC>0·80), which is in line with previous findings of other studies. 23–27 Nevertheless, CAD had a diffident specificity which might be ascribed to the uncertainty of chest x-ray features of active tuberculosis, especially in settings with asymptomatic tuberculosis, and a history of previous tuberculosis. In 2024, a study that appraised CAD accuracy during ACF for tuberculosis suggested that not including individuals with normal appearing chest x-rays or who were asymptomatic severely underestimates specificity. Even though CAD accuracy might be better characterized by assuming participants who were not microbiologically tested were tuberculosis negative, there is a high risk of missing individuals with asymptomatic or subclinical tuberculosis. This buttresses the necessity for Threshold determination when implementing CAD. A high threshold might miss people with tuberculosis, whereas a low threshold will detect most individuals with tuberculosis but with greater costs due to increased confirmatory microbiological testing.

Owing that this is a screening test with the aim to maximize case detection, high sensitivity is more important than high specificity. The disparity in missed cases is enormous to ignore, bearing in mind that a person with active pulmonary tuberculosis (TB) can potentially infect a significant number of people, with estimates ranging from 10 to 15 individuals per year if left untreated. This implies 120 to 180 persons are at risk based on the difference alone. (WHO, NTBLCP). The chosen threshold was guided by the main goals of the program, balancing reduction in confirmatory testing with risk of missed cases, and considerations of cost-effectiveness. 7,8,24 medRxiv

Receiver Operating Characteristic (ROC) curves measure the overall performance of the CAD software; however, in choosing an acceptable cut off for CAD4 TB, there is an inherent trade-off in the selection of the threshold score; a lower threshold score will maximize sensitivity of the tool to detect true TB cases among the population being screened but will incur additional costs related to unnecessary follow-on diagnostic testing due to reduced specificity. On the other hand, a higher threshold score will reduce the volume, and thus costs, of follow-on diagnostic testing and will likely identify more severe cases, but its reduced sensitivity will result in missed cases. This buttresses the need for local calibration of CAD. (WHO, 2021)

These study findings are good and underscore the brilliant performance of CAD4TB. It demonstrates its potential to enhance TB triage, particularly in resource-limited settings in identification of TB cases, support timely treatment and reduce TB transmission. This finding has a far-reaching implication. It can help support programmatic decision-making in the implementation of CAD Algorithms. The thresholds identified may be useful as a starting point, although updated versions of CAD algorithms may require re-assessment. Moreover, the chosen threshold should also be guided by

the main goals of the program, balancing reduction in confirmatory testing with risk of missed cases, and considerations of cost-effectiveness.(Bashir et al., 2022)..

This study establishes that optimal thresholds for Computer-Aided Detection (CAD) of tuberculosis are not universal but are highly specific to the population and context, providing crucial data for Nigeria's high-tuberculosis-burden setting. It found that the ideal CAD threshold varies significantly by demographic factors, with different cut-offs for age groups, gender (40 for males vs. 32 for females), and individuals with a history of TB (48) compared to new cases (36). The research confirms that CAD performance is lower in key subgroups like older individuals and those previously treated for TB, and it aligns with global studies showing that fixed, manufacturer-set thresholds are suboptimal. Using an incorrect threshold risk missing true cases or overburdening the health system with false positives. Therefore, the study concludes that for effective and efficient TB screening, local pilot testing and recalibration of CAD thresholds are essential for each new software version and specific population, rather than relying on standardized cut-offs.

CONCLUSION

Findings suggest that CAD systems could be a useful tool for TB screening programs in high throughput programmes such as active case-finding interventions where a quick decision on further testing is critical, especially if trained readers are scarce and/or costs are high. It provides valuable insights into calibration of CAD4TB for the entire population and subpopulations or settings, offering a robust framework for its deployment in diverse settings. CAD algorithms can achieve or exceed the minimum target accuracy for a TB triage test, with improvement when using setting- or population-specific thresholds. Continuous refinement of CAD4TB, based on real-world data, is essential to improving its accuracy and reliability. Integrating CAD4TB into routine TB detection programs, frontline health workers can help bridge gaps in TB case detection, leading to better health outcomes globally.

RECOMMENDATIONS

Based on the discussion of findings and conclusions of this study, the following recommendations were made:

1. The National Tuberculosis Programme should formally adopt this threshold as the standard for current CAD4TB deployments. This will ensure the interruption of TB transmission is prioritized by minimizing missed cases, which is crucial in a high-burden setting.
2. Given that CAD performance varies by demographic and clinical subgroups and with each software update, a formal protocol for ongoing recalibration should be established.
3. To make an informed decision on the trade-off between sensitivity and specificity, a comprehensive cost-effectiveness analysis is urgently needed. This analysis should weigh the financial cost of confirmatory testing for false positives against the long-term public health and economic impact of missed TB cases, including onward transmission and additional treatment costs.

REFERENCES

Bashir, S., Kik, S. V., Ruhwald, M., Khan, A., Tariq, M., Hussain, H., & Denkinger, C. M. (2022). Economic analysis of different throughput scenarios and implementation strategies of computer-aided detection software as a screening and triage test for pulmonary TB. *PLoS ONE, 17*(12), 1–16.

Durovni, B., Saraceni, V., Cordeiro-Santos, M., Cavalcante, S., Silva, D. R., & Golub, J. E. (2015). Correction: Impact of replacing smear microscopy with Xpert MTB/RIF for diagnosing tuberculosis in Brazil: A stepped-wedge cluster-randomized trial. *PLoS Medicine, 12*(12), e1001928.

Fehr, J., Gunda, R., Siedner, M. J., Hanekom, W., Ndung'u, T., Grant, A., Lippert, C., & Wong, E. B. (2023). CAD4TB software updates: Different triaging thresholds require caution by users and regulation by authorities. *The International Journal of Tuberculosis and Lung Disease, 27*(2), 157–160. https://doi.org/10.5588/ijtld.22.0437

Khan, F. A., Majidulla, A., Tavaziva, G., Nazish, A., Abidi, S. K., Benedetti, A., Menzies, D., Johnston, J. C., Khan, A. J., & Saeed, S. (2020). Chest x-ray analysis with deep learning-

based software as a triage test for pulmonary tuberculosis: A prospective study of diagnostic accuracy for culture-confirmed disease. *The Lancet Digital Health, 2*(11), e573–e581. https://doi.org/10.1016/S2589-7500(20)30221-1

Olaleye, S. A., Balogun, O. S., & Adusei-Mensah, F. (2023). Bibliometric structured review of tuberculosis in Nigeria. *African Health Sciences, 23*(2), 139–160.

Onozaki, I., Law, I., Sismanidis, C., Zignol, M., Glaziou, P., & Floyd, K. (2015). National tuberculosis prevalence surveys in Asia, 1990–2012: An overview of results and lessons learned. *Tropical Medicine & International Health, 20*(9), 1128–1145.

Pande, T., Vasquez, N. A., Cazabon, D., Creswell, J., Brouwer, M., Ramis, O., Stevens, R. H., Ananthakrishnan, R., Qayyum, S., Alphonsus, C., Oga-Omenka, C., Nafade, V., Sen, P., & Pai, M. (2020). Finding the missing millions: Lessons from 10 active case finding interventions in high tuberculosis burden countries. *BMJ Global Health, 5*(12), 1–5. https://doi.org/10.1136/bmjgh-2020-003835

Philipsen, R., Van Ginneken, B., & Melendez, J. (2018). Computer-aided detection of tuberculosis: CAD4TB. *Unpublished manuscript*.

Qin, Z. Z., Sander, M. S., Rai, B., Titahong, C. N., Sudrungrot, S., Laah, S. N., Adhikari, L. M., Carter, E. J., Puri, L., Codlin, A. J., & Creswell, J. (2019). Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific Reports, 9*(1), 1–10. https://doi.org/10.1038/s41598-019-51503-3

Scott, A. J., Perumal, T., Pooran, A., Oelofse, S., Jaumdally, S., Swanepoel, J., Gina, P., & Mthiyane, T. (2025). Clinical evaluation of computer-aided digital x-ray detection of pulmonary tuberculosis during community-based screening or active case-finding: A case–control study. *The Lancet Global Health, 13*(3), e517–e527. https://doi.org/10.1016/S2214-109X(24)00516-3

van't Hoog, A. H., Onozaki, I., & Lönnroth, K. (2014). Choosing algorithms for TB screening: A modelling study to compare yield, predictive value and diagnostic burden. *BMC Infectious Diseases, 14*(1), 1–12. https://doi.org/10.1186/1471-2334-14-532

World Health Organization. (2021). *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*. https://apps.who.int/iris/handle/10665/345122

Zaidi, S. M. A., Habib, S. S., Van Ginneken, B., Ferrand, R. A., Creswell, J., Khowaja, S., & Khan, A. (2018). Evaluation of the diagnostic accuracy of computer-aided detection of tuberculosis on chest radiography among private sector patients in Pakistan. *Scientific Reports, 8*(1), 1–9. https://doi.org/10.1038/s41598-018-30810-1.