Journal of Tuberculosis Research, 2025, 13(4), 135-149
https://www.scirp.org/journal/jtr
ISSN Online: 2329-8448
ISSN Print: 2329-843X

Determining Threshold for Computer-Aided Detection (CAD) in Pre-Diagnostic Pulmonary TB Screening for Targeted Community TB Case Finding Using Portable Digital X-Ray in Nigeria

Bethrand Odume[1,2], Ojule Inumanye[1], Ordinioha Best[1], Aniwada Elias Chike[3]

[1]Department of Public Health, University of Port Harcourt, Port Harcourt, Nigeria
[2]Knowledge Network for Disease Control and Vigilance (KNCV) Nigeria, Abuja, Nigeria
[3]Department of Community Medicine, University of Nigeria, Enugu, Nigeria
Email: bethodume@gmail.com

How to cite this paper: Odume, B., Inumanye, O., Best, O. and Chike, A.E. (2025) Determining Threshold for Computer-Aided Detection (CAD) in Pre-Diagnostic Pulmonary TB Screening for Targeted Community TB Case Finding Using Portable Digital X-Ray in Nigeria. *Journal of Tuberculosis Research*, **13**, 135-149.
https://doi.org/10.4236/jtr.2025.134013

Received: October 10, 2025
Accepted: November 10, 2025
Published: November 13, 2025

Abstract

Introduction: Computer-Aided Detection technologies for TB detection were considered by the WHO as part of an update to TB screening guidelines and recommendations. However, no universal guidelines exist for selecting a decision threshold. Most CAD software products do not come with a pre-set manufacturer-recommended threshold. Several studies suggested countries and setting specific threshold. This study seeks to calibrate CAD4TB to local settings and population groups in Nigeria. **Methods:** A retrospective review of data was done for all eligible participants identified from E. presumptive Register. Data was entered into a standardized CSV spreadsheet and uploaded to the online CAD for TB detection calibration tool developed by WHO for analysis of CAD data calibration studies. CAD score threshold cut-off point was based on Area Under Curve (AUC) in relation to other test modalities. Mean CAD scores were compared using Student T test at $p < 0.005$. **Result:** The Mean CAD4 score was 43.13 (Range 0 to 99.3). Those located at health facility have highest mean 55.05 followed by those in slums 49.53 with $p < 0.001$. There were similar mean scores (between 41.0 and 49.0) with $p < 0.001$. Mean values were males 46.05 and females 44.96 with $p = 0.118$. Mean value for that previous TB treatment was 48.38 and no previous TB treatment 45.32 ($p = 0.004$). Mean value for those with TB was 67.92 versus 40.10 for those without TB ($p < 0.001$). The AUC was 0.810 [CI 0.799 to 0.822]. The chosen CAD4 Threshold cut off is 37 (Sen 90.20%, Spec 39.50%). **Conclusion:** Findings provide valuable insights into calibration of CAD4TB for the entire population. Integrating CAD4TB into routine TB detection programs can help bridge gaps in TB case detection, leading to better health outcomes globally.

Keywords

Threshold Cut-Off, Computer-Aided Detection, Pre-Diagnostic TB Screening, Nigeria

1. Introduction

Tuberculosis (TB) is the deadliest infectious bacteriological disease in the world. The World Health Organization (WHO) Global Report revealed that TB infection was responsible for the death of 1.25 million people in 2023 alone, which is despite available vaccine and cure for TB [1]. This report showed that annual TB estimates by far exceed reported cases by 3.5 million, thereby creating a diagnostic gap mostly thought to be in target populations probably not reached by the campaigns [1]. Effort has shifted focus from TB control to TB elimination not minding the drawbacks from COVID-19 pandemic. This means that, for this global dream to be achieved, there must be innovative ways, in addition to the less sensitive symptom screening (W4SS), to increase TB case detection in order to close the existing diagnostic gap and prevent further gaps. Owing to this, Chest X-Ray (CXR) with Computer-Aided Detection (CAD) has become of great relevance in this pursuit [2].

In Nigeria, diagnostic coverage for TB case finding still stands at 40% for GeneXpert and more gaps still exist for AI-based diagnostic tools [3]. Chest X-ray with AI has been recommended by WHO both for TB screening and diagnosis. In the country, the first portable digital X-ray (PDX) with CAD4TB was introduced and piloted by KNCV Nigeria through the TB LON project since December 2020 [4]. This was necessitated by the need to reach people in hard-to-reach remote communities. This pilot led to the scale up of the intervention under the USAID new tools project with additional seven PDX distributed to other states (Benue, Cross River, Delta, Katsina, Kano and Nasarawa) under the KNCV intervention [4]. 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 the Nigeria despite demonstrated high yield in the KNCV pilot intervention in selected hard-to-reach locations in Nigeria.

Chest X-ray, being a highly sensitive TB screening tool for detecting people who require diagnostic evaluation, may guide the effective allocation of diagnostic tests (GeneXpert, Truenat, TB lamp, LF Lam) to improve case detection and cost-efficiency [5]. CXR also helps in detection of asymptomatic sub-clinical TB as such leading to early detection, initiation to treatment and halting of transmission [6] [7]. Computer-Aided Detection (CAD) software products are increasingly being used as a tool to enhance the feasibility and accuracy of CXR interpretation [8] [9]. The use of Artificial Intelligence (AI) to analyze CXR for abnormalities suggestive of pulmonary TB provides an abnormality score that can be used to determine the need for follow-up on diagnostic testing for TB relative to a selected threshold. Such a technology can be used to replace or augment human expert interpretation of plain CXR when screening for pulmonary TB. Computer-aided detection (CAD) can be utilized with CXR to evade CXR's high inter- and intra-reader variability, long turnaround times and the lack of radiologists in high burden countries [10]-[13]. CAD technologies for TB detection were considered by the WHO in 2020 as part of an update to TB screening guidelines and recommendations [14]. However, 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 [15].

Currently, no official guidelines exist for selecting a decision threshold. Also, most CAD software products do not come with a pre-set manufacturer-recommended threshold that defines cut-off to be referred for diagnostic evaluation for TB. Even when some manufacturers recommend thresholds for defining images as abnormal, evaluations by the WHO Guideline development group demonstrated substantial variation in the diagnostic accuracy (sensitivity and specificity) of CAD programmes across settings, even within the same technology set to the same threshold. This calls for calibration of CAD thresholds for a given software based on setting and population in which it will be used [15]. Several studies also suggested the importance of obtaining a study-specific decision threshold, adjusted to the underlying TB prevalence and demographic parameters [2] [16]. One way to select a threshold score is by conducting on-site calibration studies [17]. This study, therefore, seeks to calibrate CAD4TB to local settings and population groups by comparing the diagnostic accuracy of CAD compared to a reference standard as this has the potential to increase active TB case finding and advance the goal of eradicating TB by 2030.

2. Methods

2.1. Study Area/Setting

This study was conducted 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 zones of Nigeria. Nigeria is a country in West Africa, and is the most populous country in Africa, with an estimated population of over 227 million according to World Bank 2023 projection. The public health service is organized into primary, secondary, and tertiary levels with responsibilities for primary health care ascribed to local governments, secondary care to states and tertiary care to the federal government. As at 2019, there are 5389 DOTS centres providing TB treatment services in Nigeria and 398 GeneXpert MTB-Rif machines were in use, supported by numerous partners and placed in all 36 states plus FCT (14). All LGAs have at least one DOTS treatment facility.

2.2. Study Design

This was a retrospective study involving review of data from electronic presumptive register generated by Portable Digital X-ray (PDX) with artificial intelligence and the web-based application for X-ray and clinical symptom evaluation platform (XMAP). The data was routinely collected across 6 states in Nigeria implementing active TB case finding.

2.3. Study Population

This included all clients in all selected states who were screened using mixed screening procedures for TB active case findings irrespective of presence or absence of respiratory and constitutional symptoms for TB as documented in the electronic presumptive registers.

Inclusion Criteria

1. Individuals fifteen years and above who participated in TB screening with CXR and CAD4TB.

Exclusion criteria

1. Pregnant women.
2. Individuals who did not get screened with CXR but had sputum sample collected and tested.
3. All confirmed TB clients who were on treatment before the screening.

2.4. Sample Size Calculation

A total study was done involving all study samples retrospectively identified using data from the electronic presumptive register generated by the seven PDX machines across the six states and the web-based application for X-ray and clinical symptom evaluation platform (XMAP).

2.5. Sampling Approach

All eligible participants who have been previously confirmed as TB and non-TB cases were retrospectively identified from E. presumptive register and XMAP. All that met inclusion criteria were studied.

2.6. Data Collection

A data collection tool (Proforma) was developed and used 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 data was segregated according to local government, and states of implementation. Data on all the bacteriological positive and negative TB cases were extracted as well as tracking of the CAD4TB scores by the researcher.

2.7. Data Quality Assurance

Data entry into the E. presumptive registers were 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 were 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 it undergoes 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.

2.8. Data Management

Measurement of variables

Outcomes measures include:

1. The proportion of TB cases correctly identified as positive using CAD (sensitivity).
2. The proportion of non-TB cases correctly classified as negative using CAD (specificity).

The outcomes of interest were related to the diagnostic performance of CAD compared to a bacteriological reference standard. The predicative probabilities of TB detection by CAD at various thresholds were used to estimate cost and programmatic implications, in particular the number of TB cases missed or misdiagnosed and additional costs of follow up testing incurred or saved.

To get these results, the data collected was 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.

How to Interpret an Abnormality Score

Tuberculosis CADs are image classification algorithms because they read a CXR for abnormalities to assign a classification of TB-suggestive abnormalities being present or not. This classification is based on a score of between zero and hundred (or 0 - 100) that represents the likelihood of TB being present in a particular CXR. The output of this is a continuous numerical output that can be translated into a binary classification by selecting a cut-off point on the abnormality score (between zero and hundred), above which the "abnormalities suggestive of presence of TB" classification is assigned. A CXR with an abnormality score below the cut-off would be assigned the "abnormalities suggestive of absence of TB" classification. This score is variously referred to as the threshold score, operating point, "cut-off" or "cut point" [15] [18].

Previous studies have shown that the sensitivity and specificity of a given threshold score vary between populations and that diagnostic accuracy is also affected by factors such as age, sex, prior TB, human immunodeficiency virus (HIV) status and others. This heterogeneity presents operational challenges, as it requires users to calibrate CAD with local data to identify appropriate thresholds in their population. Likewise, the loss of explanatory information that occurs when dichotomizing a continuous variable could have negative consequences for diagnostic accuracy [9].

Receiver Operating Characteristic (ROC) Curves

The overall accuracy of a CAD software product is measured over the complete range of abnormality scores (0 - 1 for CAD4TB). ROC curves plot the CAD's true positive rate (sensitivity) against the false positive rate (1-specificity). The area under the ROC curve (AUC), between zero and one, is directly related to the overall accuracy of CAD software, with 0.5 representing the accuracy of a random guess.

2.9. Statistical Analysis

Data analysis was in line with objectives which include:

Calibration/ cut off was conducted using the online CAD for TB detection calibration tool. The analysis was based on a dichotomous classification of each participant as either a TB, or non-TB case, based on the outcomes of bacteriological testing, against a binary interpretation of CAD abnormality score (abnormal/normal) at various CAD thresholds to assess the diagnostic performance of CAD. The CAD will be scored between 0 - 100 to indicate the likelihood of TB-related lung field abnormality. However different threshold will be used as cut off for the analysis.

Diagnostic performance was estimated using sensitivity (Se), specificity (Sp), Positive Predictive Value (PPV) and Negative Predictive Values (NPV) using the number of false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN) of CAD performance at various thresholds compared to positive bacteriological test with either GeneXpert, Truenat or TB LAM. Further, TP, TN, FP, and FN were used to construct a Receiver Operator Characteristic (ROC) curve to illustrate how sensitivity and specificity varied across the range of possible CAD thresholds.

Additionally, the mean CAD scores were compared using Student T test at significance level of $p < 0.005$.

2.10. Ethical Considerations

The ethical approval was sought from University of Port Harcourt Ethic Review Committee (UPH/CEREMAD/REC/MM95/057) before commencement of this study. There was no invasive procedure as the study was done with secondary data. All necessary steps were taken to ensure confidentiality of patient's personal information, all the results from CAD4TB and bacteriological test in the E. presumptive register.

3. Results

Table 1. Sociodemographic characteristics and setting of participants.

	Frequency (n = 16,047)	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	10,046	62.6
Female	6001	37.4
Previously Treated for TB		
Yes	802	5.0
No	15,245	95.0
TB disease		
Yes	1752	10.9
No	14,295	89.1
CAD4 Score		
Mean (SD)	43.13 (24.10)	
Range	0 - 99.3	

Table 1 shows the sociodemographic characteristics and settings of participants. Higher proportion of residents were in general population or normal home setting

5620 (35.0%) followed by those in slums 4058 (25.3%) then hard to reach areas 2644 (16.5%). The mean age of participants was 43.45 years. those aged 25 to 34 years were slightly higher in proportion 3505 (20.0%). Males were more than Females 62.6% versus 37.4%. About 802 (5.0%) were previously treated for TB. In all 1752 (10.9%) have TB disease. The Mean CAD4 score for participants was 43.13 with a range of scores O to 99.3.

Table 2. Comparison of mean CAD4 scores based on characteristics of participants.

	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		
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	
Gender				
Male	46.05	22.93	1.561	0.118
Female	44.96	22.55		
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		

Table 2 shows Comparison of Mean CAD4 scores based on characteristics of participants. Those located at health facility have highest mean 55.05 followed by those in slums 49.53 then those at correctional centres 47.65 with the least among the general population. 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). 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 among them p = 0.118. The mean value was higher for those previously treated for TB compared to those not previously treated (48.38 versus 45.32). There was equally a statistically significant difference in mean CAD4 score among 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 equally significant difference in the mean CAD4 score among them p < 0.001.

Table 3. Distribution of AUC, sensitivity, specificity and CAD4 score of overall participants.

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

*Sen—Sensitivity *Spec—Specificity.

Table 3 shows Overall AUC of participants. The AUC was 0.810 with 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%).

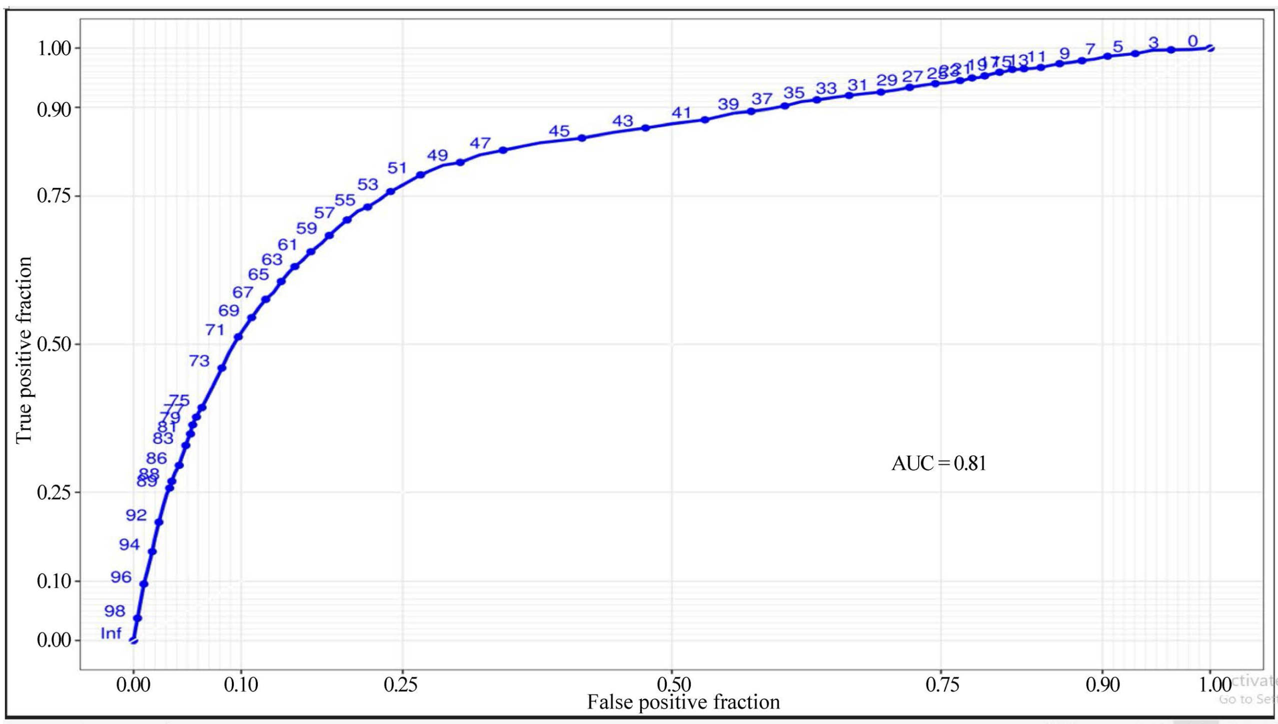

Figure 1. Plot showing overall AUC for all participants.

Figure 1 shows that AUC was 0.810.

Table 4. CAD4 cut off for participants.

cut-off	Population	Sens	Spec	CADpos	TP	FN	FP	TN	Missed cases	% of missed	Total cost for confirmatory test
31	100,000	92.60%	30.60%	71.90%	119	10	71,825	39.00%	10	7.40%	$1,726,678
32	100,000	92.40%	32%	70.70%	119	10	70,579	41	10	7.60%	$1,696,766
33	100,000	92%	33.50%	69.30%	119	10	69,134	43	10	8%	$1.662.068
34	100,000	91.70%	34.80%	68.10%	118	11	68,007	45	11	8.30%	$1,634,997
35	100,000	91.30%	36.50%	66.50%	118	11	66,387	47	11	8.70%	$1,596,111
36	100,000	91%	38%	65.20%	117	12	65,066	49	12	9%	$1,564,405
37	100,000	90.20%	39.50%	63.70%	116	13	63,628	51	13	9.80%	$1.529.856
38	100,000	89.70%	41%	62.40%	116	13	62,245	53	13	10.30%	$1.496.654
39	100,000	89.30%	42.60%	60.90%	115	14	60,743	55	14	10.70%	$1.460.609
46	100,000	84%	62.30%	42.80%	108	21	42,685	80	21	16%	$1.027.033
47	100,000	82.80%	65.70%	39.60%	107	22	39,502	85	22	17.20%	$950.61
48	100,000	82%	67.90%	37.60%	106	23	37,478	88	23	18%	$902.00
49	100,000	80.70%	69.70%	35.80%	104	25	35,741	90	25	19.30%	$860,273
50	100,000	80.30%	71.20%	34.40%	104	25	34,283	92	25	19.70%	$825,276
51	100,000	78.60%	73.30%	32.30%	101	28	32,229	95	28	21.40%	$775.92
52	100,000	77.10%	74.80%	30.90%	99	30	30,785	96	30	22.90%	$741,223
53	100,000	75.80%	76.10%	29.50%	98	31	29,447	98	31	24.20%	$709.07
54	100,000	74.40%	77.10%	28.50%	96	33	28,389	100	33	25.60%	$683.64

Table 4 shows that at a sensitivity of 90.20 % the corresponding specificity was 39.50% and CAD4TB threshold score of 37. At this score 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 69.70% the corresponding sensitivity was 80.70% and CAD4TB score was 49. At this score missed TB cases was 25 (19.30%), TP 104 and attracted about 860,273 dollars for confirmatory test.

4. Discussion

The WHO recommends that CAD can replace a human reader of Xray in TB investigation. This expectedly will be in two broad contexts; for screening and for triage with both aims to establish whether an individual should receive confirmatory diagnostic tests. CAD4TB produces an abnormality score that can be used to signal probable TB and trigger further TB diagnostic evaluation relative to a selected threshold [7]. However, effective integration of CAD4TB into routine programming requires the calibration of CAD products. This calibration is based on the local context, intended use as well as decision making around the goals for screening and acceptable costs [15]. Based on this, the WHO sets the target accuracy for TB triage tests at ≥90% sensitivity and ≥70% specificity.

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 was 37 which was chosen as the acceptable cut off for general population. 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. Conversely, at specificity of about 70% CAD4TB and corresponding sensitivity of about 80% (sen. 80.70% versus spec. 69.70%) CAD4TB score was 49. At this cut off missed TB cases was 25 (19.30%), TP of 104 and attracted about 860,273 dollars for confirmatory test. This chosen cut-off is not in line with WHO's minimum target accuracy for TB triage tests which is ≥90% sensitivity and ≥70% specificity. Not minding, this score demonstrated high accuracy and meets the desired standard value for use in screening and triaging for TB. Owing that this is a screening test with 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 people are at risk based on the difference alone. 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 [18]-[20].

This 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 [18] [19]. A study that compared versions of CAD4TB software reported that no threshold from any version was able to match the performance target by the WHO of ≥90% sensitivity and ≥70% specificity. The study therefore identified one threshold for each CAD4TB version that most closely matched the radiologist sensitivity at 80.8% lower specificity than the radiologist. The selected threshold ranged from 20 to 47 [15] [21].

At this cut-off from the current study, the Area under Curve (AUC) from Receiver Operating Characteristic plot was 0.810 which is good. 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 [15].

A study that determined CAD4 thresholds through secondary data analysis of Lesotho national tuberculosis prevalence survey a CAD4TBv7 threshold chose cut-off 13 at a sensitivity of 89.7% (95% CI 84.6 - 94.8) and a specificity of 74.2% (73.6 - 74.9), close to World Health Organization (WHO) target product profile criteria [21]. The difference in cut off from current study may be due to study setting and design. While the previous study was secondary data which is prone to missing data and assumptions, the current data was a validated primary data that addressed the shortcomings of study in Lesotho. The previous study also used multiple imputations to estimate tuberculosis status for those eligible and in that study, those not tested as well as those assumed not eligible for testing were categorized as negative which is not the case with current study.

Another study on accuracy of computer-aided chest X-ray screening in the Kenya National Tuberculosis prevalence survey indicated that accuracy from their test was non-inferior to WHO-recommended minimum values; CAD4TBv6 was reported to have sensitivity of 0.93 [0.90 - 0.96] $p < 0.0001$, specificity 0.69 [0.67 - 0.71], $p = 0.0003$) [22]. This study used a fixed cut-off probably from manufacturers and comparison was against gold standard while current study determined a tentative cut-off for the country. Equally, a previous study on iterative evaluation of mobile computer assisted digital chest X-ray screening for TB in Nigeria reported use of CAD scores ≥80 for persons initially spot -sputum negative and found 11 additional TB cases (6.3%) after 121 person-years of follow-up. It further documented that high CAD scores can identify subclinical TB and those at risk of progression to bacteriologically confirmed TB disease in the near term [23]. However, these identified cases may be due to latent infection, re-infection or newly infected bearing in mind the mode of transmission. Also, the study was not designed to calibrate and identify CAD threshold for cut-off. This most likely accounts for difference with the current study which showed decrease in True positives and increased in missed cases.

The strength of this study is that it included participants with cough and those who did not have cough and met other screening criteria for TB testing. This provides more generalizable results and powered the study to assess threshold identification by both countries and subgroups. The thresholds identified from this study may be useful as a starting point, although updated versions of CAD algorithms may require re-assessment which is a limitation.

5. 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 their accuracy and reliability. Integrating CAD4TB into routine TB detection programs can help bridge gaps in TB case detection, leading to better health outcomes globally.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] WHO (2024) Global Tuberculosis Report 2024. https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2024

[2] Zaidi, S.M.A., Habib, S.S., Van Ginneken, B., Ferrand, R.A., Creswell, J., Khowaja, S., *et al.* (2018) Evaluation of the Diagnostic Accuracy of Computer-Aided Detection of Tuberculosis on Chest Radiography among Private Sector Patients in Pakistan. *Scientific Reports*, **8**, Article No. 12339. https://doi.org/10.1038/s41598-018-30810-1

[3] Federal Ministry of Health (2022) Draft Annual Report National Tuberculosis, Leprosy and Buruli Ulcer Control. 1-34.

[4] Odume, B., Chukwu, E., Fawole, T., Nwokoye, N., Ogbudebe, C., Chukwuogo, O., *et al.* (2022) Portable Digital X-Ray for TB Pre-Diagnosis Screening in Rural Communities in Nigeria. *Public Health Action*, **12**, 85-89. https://doi.org/10.5588/pha.21.0079

[5] WHO (2016) Chest Radiography in Tuberculosis. World Health Organization, 1-44.

[6] Horne, D., Kohli, M., Zifodya, J., Schiller, I., Dendukuri, N., Tollefson, D., Schu-macher, S., Ochodo, E., Pai, M. and Steingart, K. (2019) Xpert MTB/RIF and Xpert MTB/RIF Ultra for Pulmonary Tuberculosis and Rifampicin Resistance in Adults. *Cochrane Database of Systematic Reviews*, **5**, 1-34.

[7] Frascella, B., Richards, A.S., Sossen, B., Emery, J.C., Odone, A., Law, I., *et al.* (2020) Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases*, **73**, e830-e841. https://doi.org/10.1093/cid/ciaa1402

[8] Tavaziva, G., Harris, M., Abidi, S.K., Geric, C., Breuninger, M., Dheda, K., *et al.* (2021) 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. *Clinical Infectious Diseases*, **74**, 1390-1400. https://doi.org/10.1093/cid/ciab639

[9] Geric, C., Tavaziva, G., Breuninger, M., Dheda, K., Esmail, A., Scott, A., *et al.* (2024) Breaking the Threshold: Developing Multivariable Models Using Computer-Aided Chest X-Ray Analysis for Tuberculosis Triage. *International Journal of Infectious Diseases*, **147**, Article ID: 107221. https://doi.org/10.1016/j.ijid.2024.107221

[10] Menzies, N.A., Wolf, E., Connors, D., Bellerose, M., Sbarra, A.N., Cohen, T., *et al.* (2018) Progression from Latent Infection to Active Disease in Dynamic Tuberculosis Transmission Models: A Systematic Review of the Validity of Modelling Assumptions. *The Lancet Infectious Diseases*, **18**, e228-e238.

https://doi.org/10.1016/s1473-3099(18)30134-8

[11] Emery, J.C., Richards, A.S., Dale, K.D., McQuaid, C.F., White, R.G., Denholm, J.T., *et al.* (2021) Self-Clearance of *Mycobacterium tuberculosis* Infection: Implications for Lifetime Risk and Population At-Risk of Tuberculosis Disease. *Proceedings of the Royal Society B: Biological Sciences*, **288**, Article ID: 20201635. https://doi.org/10.1098/rspb.2020.1635

[12] Singh, M., Pujar, G.V., Kumar, S.A., Bhagyalalitha, M., Akshatha, H.S., Abuhaija, B., *et al.* (2022) Evolution of Machine Learning in Tuberculosis Diagnosis: A Review of Deep Learning-Based Medical Applications. *Electronics*, **11**, Article 2634. https://doi.org/10.3390/electronics11172634

[13] Behr, M.A., Edelstein, P.H. and Ramakrishnan, L. (2024) Rethinking the Burden of Latent Tuberculosis to Reprioritize Research. *Nature Microbiology*, **9**, 1157-1158. https://doi.org/10.1038/s41564-024-01683-0

[14] World Health Organization (2020) Tuberculosis and COVID-19: Considerations for Tuberculosis Care. World Health Organisation. https://www.who.int/docs/default-source/documents/tuberculosis/infonote-tb-covid-19.pdf

[15] World Health Organisation (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://iris.who.int/server/api/core/bitstreams/cfe76bb2-73ee-4067-ad4f-fccc23d5f47a/content

[16] Rahman, M.T., Codlin, A.J., Rahman, M.M., Nahar, A., Reja, M., Islam, T., *et al.* (2017) An Evaluation of Automated Chest Radiography Reading Software for Tuberculosis Screening among Public- and Private-Sector Patients. *European Respiratory Journal*, **49**, Article ID: 1602159. https://doi.org/10.1183/13993003.02159-2016

[17] Fehr, J., Gunda, R., Siedner, M.J., Hanekom, W., Ndung´u, T., Grant, A., *et al.* (2023) CAD4TB Software Updates: Different Triaging Thresholds Require Caution by Users and Regulation by Authorities. *The International Journal of Tuberculosis and Lung Disease*, **27**, 157-160. https://doi.org/10.5588/ijtld.22.0437

[18] Qin, Z.Z., Ahmed, S., Sarker, M.S., Paul, K., Adel, A.S.S., Naheyan, T., *et al.* (2021) Tuberculosis Detection from Chest X-Rays for Triaging in a High Tuberculosis-Burden Setting: An Evaluation of Five Artificial Intelligence Algorithms. *The Lancet Digital Health*, **3**, e543-e554. https://doi.org/10.1016/s2589-7500(21)00116-3

[19] Bashir, S., Kik, S.V., Ruhwald, M., Khan, A., Tariq, M., Hussain, H., *et al.* (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**, e0277393. https://doi.org/10.1371/journal.pone.0277393

[20] Jha, Y., Nations, U. and Ruhwald, M. (2021) Digital Chest Radiography and Computer Aided Detection (CAD) Solutions for Tuberculosis Diagnostics. Technology Lanscape Analysis. FIND, Geneva.

[21] Vanobberghen, F., Keter, A.K., Jacobs, B.K.M., Glass, T.R., Lynen, L., Law, I., *et al.* (2023) Computer-Aided Detection Thresholds for Digital Chest Radiography Interpretation in Tuberculosis Diagnostic Algorithms. *ERJ Open Research*, **10**, 00508-2023. https://doi.org/10.1183/23120541.00508-2023

[22] Khan, F.A., Majidulla, A., Tavaziva, G., Nazish, A., Abidi, S.K., Benedetti, A., *et al.* (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**, e573-e581.

https://doi.org/10.1016/s2589-7500(20)30221-1

[23] Eneogu, R.A., Mitchell, E.M.H., Ogbudebe, C., Aboki, D., Anyebe, V., Dimkpa, C.B., *et al.* (2024) Iterative Evaluation of Mobile Computer-Assisted Digital Chest X-Ray Screening for TB Improves Efficiency, Yield, and Outcomes in Nigeria. *PLOS Global Public Health*, **4**, e0002018. https://doi.org/10.1371/journal.pgph.0002018