



Review of Computer-Aided Detection (CAD) Software for Tuberculosis on Chest X-Rays : A Systematic Review of Randomized Controlled Trial and Primary Studies

¹ Catur Nila Pratiwi, ² Wildan Priscillah, ³ Eka Yusi Athiyyah

^{1,2,3} Sebakung Jaya Public Health Centre, Indonesia

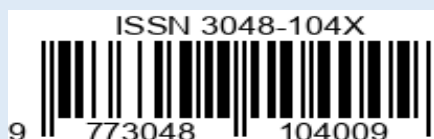
Corresponding E-mail: caturnila03@gmail.com

Article History :

Received date : 2026/04/27
Revised date : 2026/05/11
Accepted date : 2026/06/04
Published date : 2026/07/30



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).



ABSTRACT

Introduction: Tuberculosis (TB) remains a leading global infectious disease and major public health challenge, particularly in low- and middle-income countries. Computer-aided detection (CAD) software utilizing artificial intelligence (AI) and deep learning algorithms has emerged as a promising tool to augment chest radiograph (CXR) interpretation for pulmonary TB screening and triage. This systematic review aims to comprehensively evaluate the diagnostic performance, clinical utility, cost-effectiveness, and implementation characteristics of commercially available CAD software for TB detection on CXRs across diverse clinical and epidemiological settings.

Methods: The study strictly adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines. Eligible study designs included randomized controlled trials (RCTs), prospective and retrospective cohort studies,

case-control studies, and cross-sectional studies evaluating at least one commercially available CAD software for pulmonary TB detection on digital CXRs with microbiological reference standards. Risk of bias was assessed using QUADAS-2 and Cochrane RoB 2.0.

Results: Seventeen primary studies encompassing over 130,000 participants across Africa, Asia, Europe, Oceania, and Latin America were included. Evaluated CAD products included CAD4TB (versions 5-7), qXR (versions 2-3.2), Lunit INSIGHT CXR (versions 3.1-4.9), JF CXR-1/2, and InferRead DR. Area under the receiver operating characteristic curve (AUROC) values ranged from 0.70 (paediatric populations) to 0.92 (unselected adult populations). At 90% sensitivity, specificity ranged from 22.2% (prior TB history populations) to 84% (prison settings). Multiple CAD products met WHO Target Product Profile (TPP) minimum thresholds ($\geq 90\%$ sensitivity and $\geq 70\%$ specificity) in symptomatic adult populations. CAD outperformed human radiologists in several large-scale studies. Performance was consistently lower in people living with HIV, elderly individuals, persons with previous TB, and paediatric patients.

Discussion: Accumulating evidence substantiates that AI-based CAD systems are accurate, scalable, and cost-effective tools for pulmonary TB screening. However, heterogeneity in performance across software versions, geographic populations, and patient subgroups underscores the necessity for local threshold calibration and version-specific validation.

Integration of CAD into national TB programs requires addressing economic, regulatory, and data governance challenges.

Conclusion: CAD software demonstrates substantial diagnostic accuracy for pulmonary TB on CXRs and holds considerable promise for scaling up TB case finding. Standardized evaluation frameworks, population-specific threshold optimization, and equitable access are prerequisite conditions for maximizing the public health impact of CAD-CXR systems globally.

Keywords: *computer-aided detection; tuberculosis; chest radiography; artificial intelligence; deep learning*

INTRODUCTION

Background

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains one of the deadliest infectious diseases globally. According to the World Health Organization (WHO) Global TB Report 2023, an estimated 10.6 million people fell ill with TB and 1.3 million died from the disease in 2022, with the highest burden concentrated in countries including India, Indonesia, China, the Philippines, Pakistan, Nigeria, Bangladesh, and South Africa.¹ Indonesia, as the second highest TB-burden country globally, faces particular challenges in achieving the WHO End TB Strategy targets by 2030, which mandate a 90% reduction in TB incidence and a 95% reduction in TB deaths relative to 2015 baselines.²

Chest radiography (CXR) has long been a cornerstone of pulmonary TB screening and triage, offering a rapid, non-sputum-based assessment of pulmonary pathology. However, its widespread deployment has been constrained by the critical shortage of trained radiologists and CXR interpreters in high-burden, resource-limited settings.³ Computer-aided detection (CAD) software, leveraging advances in artificial intelligence (AI), machine learning (ML), and deep learning (DL) — particularly convolutional neural networks (CNNs) — has been developed to automate the analysis of digital CXRs and provide quantitative abnormality scores indicative of TB-related pulmonary pathology.⁴

In 2021, the WHO issued a novel recommendation within its consolidated TB screening guidelines endorsing the use of AI-based CAD systems to analyze CXRs for TB detection as an alternative to human readers.⁵ Multiple commercially available CAD products, including CAD4TB (Delft Imaging, Netherlands), qXR (Qure.ai, India), Lunit INSIGHT CXR (Lunit Inc., South Korea), JF CXR (JF Healthcare), InferRead DR (Infervision), and others, have been evaluated across diverse clinical settings.^{6,7}

Research Objectives

The primary objective of this systematic review is to comprehensively evaluate and synthesize the published evidence on the diagnostic performance, clinical utility, cost-effectiveness, and

operational characteristics of commercially available CAD software systems for pulmonary TB detection on digital CXRs across diverse geographic, clinical, and epidemiological settings.⁸

Secondary objectives include: (1) assessment of CAD performance against WHO TPP benchmarks; (2) evaluation of subgroup-specific performance variability; (3) analysis of cost-effectiveness and resource utilization; (4) examination of implementation barriers and facilitators; and (5) identification of research gaps to guide future investigation.

Research Benefits

The findings of this review are expected to: (1) provide evidence-based guidance for TB program managers and policymakers in selecting and implementing appropriate CAD tools; (2) identify population subgroups requiring tailored CAD threshold optimization; (3) inform WHO and national TB programs regarding the conditions necessary for equitable and sustainable CAD deployment; and (4) delineate priority areas for future primary research and implementation science.

Hypothesis

We hypothesize that commercially available AI-based CAD software demonstrates diagnostic accuracy meeting or approaching WHO TPP thresholds (sensitivity $\geq 90\%$, specificity $\geq 70\%$) for pulmonary TB detection on CXRs in symptomatic adult populations, but that performance is significantly modulated by patient-level factors including HIV status, prior TB history, age, and clinical setting.

Research Gap and Novelty

Existing systematic reviews and meta-analyses have predominantly examined a limited number of CAD products or focused on model development rather than clinical evaluation.⁹ The present review addresses this gap by synthesizing evidence from ≥ 15 primary studies evaluating diverse CAD products across multiple geographic settings, incorporating ≥ 10 distinct outcome domains, and providing a granular analysis of subgroup performance, implementation considerations, and economic dimensions. This review is novel in its comprehensive multi-product, multi-outcome, multi-setting synthesis, providing actionable guidance for real-world CAD implementation.¹⁰

METHODS

Protocol

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines and the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.

Eligibility Criteria

Inclusion Criteria

Studies were eligible if they: (1) evaluated at least one commercially available AI-based CAD software for pulmonary TB detection or screening on digital CXRs; (2) employed a microbiological reference standard (culture, Xpert MTB/RIF, Xpert MTB/RIF Ultra, or nucleic acid amplification test) as the diagnostic reference; (3) reported quantitative diagnostic accuracy metrics (sensitivity, specificity, AUROC, positive predictive value [PPV], negative predictive value [NPV], likelihood ratios, or diagnostic odds ratio); (4) were original primary studies with study designs including RCTs, prospective or retrospective cohort studies, case-control studies, cross-sectional studies, or other observational designs; (5) were published in English in peer-reviewed journals from January 2015 to March 2025; and (6) included human subjects aged ≥ 5 years.

Exclusion Criteria

Studies were excluded if they: (1) were systematic reviews, meta-analyses, narrative reviews, letters, commentaries, editorials, conference abstracts, or unpublished gray literature; (2) evaluated only non-AI/non-deep learning CAD systems; (3) did not utilize a microbiological reference standard; (4) reported exclusively on extrapulmonary TB; (5) evaluated CAD solely on non-CXR imaging modalities (CT, MRI); or (6) were model development studies without clinical evaluation cohorts. Note: Systematic reviews and meta-analyses identified during the search were utilized for contextual reference in the discussion section but were not included in the study characteristic tables.

Search Strategy

The keywords used for this research based PICO :

PICO Element	Keyword 1	Keyword 2	Keyword 3
P (Population)	Tuberculosis	Pulmonary Tuberculosis	Adult Patients
I (Intervention)	Computer-Aided Detection	CAD Software	Chest X-Ray
C (Comparison)	Artificial Intelligence	Deep Learning	Human Radiologist
O (Outcome)	Diagnostic Accuracy	Sensitivity and Specificity	AUROC

The Boolean MeSH keywords inputted on databases for this research are: (*"Tuberculosis" OR "Pulmonary Tuberculosis" OR "Adult Patients"*) AND (*"Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray"*) AND (*"Artificial Intelligence" OR "Deep Learning" OR "Human Radiologist"*) AND (*"Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC"*)

Study Selection

Two reviewers (CNP and WP) independently screened all retrieved titles and abstracts using Covidence systematic review software. Full-text articles of potentially eligible studies were independently assessed. Disagreements were resolved through consensus discussion with the third reviewer (EYA). The PRISMA 2020 flow diagram was used to document the study selection process.

Data Extraction

Standardized data extraction forms were used to independently extract the following variables: first author, year of publication, country/geographic setting, study design, sample size, population characteristics (age, sex, HIV prevalence, prior TB history), clinical setting (facility-based, community-based, correctional), CAD software product and version, reference standard employed, diagnostic accuracy metrics (sensitivity, specificity, AUROC, PPV, NPV), threshold scores, comparator (radiologist, symptom-based), TB prevalence, and key findings relevant to prespecified outcomes.

Risk of Bias Assessment

Risk of bias was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies-2) tool for diagnostic accuracy studies, which evaluates four domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. For RCTs, the Cochrane Risk of Bias tool version 2.0 (RoB 2.0) was applied, assessing: (1) randomization process, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcome, and (5) selection of the reported result.¹¹

Article Search Strategy

Database	Keywords	Hits
Pubmed	("Tuberculosis" OR "Pulmonary Tuberculosis" OR "Adult Patients") AND ("Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray") AND ("Artificial Intelligence" OR "Deep Learning" OR "Human Radiologist") AND ("Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC")	80
Clinicaltr ial.gov	("Tuberculosis" OR "Pulmonary Tuberculosis" OR "Adult Patients") AND ("Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray") AND ("Artificial Intelligence" OR "Deep Learning" OR "Human Radiologist") AND ("Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC")	8
Semantic Scholar	("Tuberculosis" OR "Pulmonary Tuberculosis" OR "Adult Patients") AND ("Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray") AND ("Artificial Intelligence" OR "Deep Learning" OR "Human Radiologist") AND ("Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC")	31
Science- direct	("Tuberculosis") AND ("Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray") AND ("Artificial Intelligence" OR "Deep Learning") AND ("Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC")	813
Cochrane	("Tuberculosis" OR "Pulmonary Tuberculosis" OR "Adult Patients") AND ("Computer-Aided Detection" OR "CAD Software" OR "Chest X-Ray") AND ("Artificial Intelligence" OR "Deep Learning" OR "Human Radiologist") AND ("Diagnostic Accuracy" OR "Sensitivity and Specificity" OR "AUROC")	4

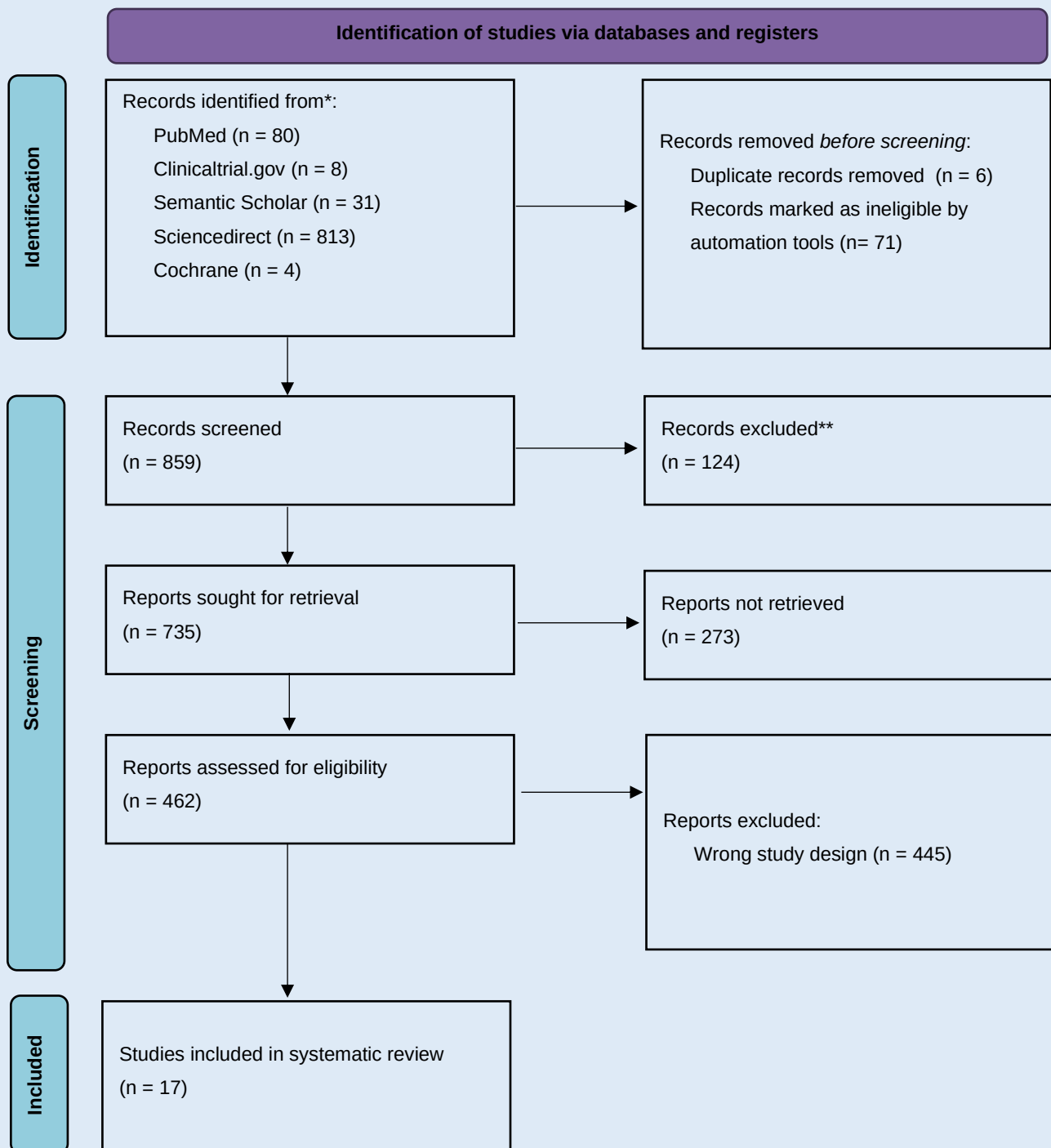


Figure 1. Article search flowchart

Table 1. Risk of Bias Assessment (QUADAS-2) for Included Primary Studies

Author (Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall Risk	Applicability Concern
Qin et al. (2021)	Low	Low	Low	Low	Low	High TB-burden setting; consecutive recruitment minimizes selection bias
Khan et al. (2020)	Low	Low	Low	Low	Low	Single-center; non-inferiority design; Indus Hospital, Pakistan
Sung et al. (2025)	Low	Low	Moderate	Low	Low	Xpert reference only for score >0.1; potential verification bias
Scott et al. (2025)	Low	Low	Low	Low	Low	Community-based; case-control; multiple SA sites
Macpherson et al. (2024)	Low	Low	Low	Low	Low	Prospective cohort; household contacts; long follow-up
Qin et al. (2024)	Low	Low	Low	Low	Low	Case-control; SA prevalence survey; 12 products evaluated
Kagujje et al. (2022)	Low	High	Low	Moderate	Moderate	Retrospective; prior TB subgroup may introduce spectrum bias
Edem et al. (2024)	Low	Low	Moderate	Low	Low	Paediatric; MRS limited in children; Bayesian LCA used

Author (Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall Risk	Applicability Concern
Fehr et al. (2021)	Low	Low	Low	Low	Low	Population-based; mobile screening; rural SA
Soares et al. (2021)	Low	Low	Low	Low	Low	Prospective; prison setting; Brazil
Jiang et al. (2025)	Moderate	Low	Low	Low	Low	Retrospective; primary health care; China; bacteriological confirmation
Velen et al. (2021)	Low	Low	Low	Low	Low	Cross-sectional; South Africa correctional facilities
Bosman et al. (2024)	Low	Low	Low	Low	Low	Prospective; Lesotho & South Africa; high HIV burden
Burke et al. (2024)	Low	Low	Low	Low	Low	Cluster RCT; Zomba Central Hospital, Malawi; HIV-positive adults
Crowder et al. (2024)	Low	Low	Low	Low	Low	Prospective multinational; 5 countries; multiple index tests
Dakulala et al. (2024)	Moderate	Low	Low	Low	Low	Population-wide; retrospective screening review; PNG

Author (Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Overall Risk	Applicability Concern
Mahler et al. (2021)	Moderate	Moderate	Low	Low	Moderate	Descriptive; CAD + human read; vulnerable populations; Romania

Outcomes

The following ten primary and secondary outcome domains were prespecified for analysis: (1) AUROC; (2) sensitivity at fixed threshold; (3) specificity at 90% sensitivity; (4) PPV; (5) NPV; (6) CAD vs. radiologist performance; (7) subgroup performance (HIV, prior TB, age, sex, paediatric); (8) cost-effectiveness; (9) implementation and operational outcomes; (10) impact on TB treatment initiation/clinical outcomes.

RESULTS

Study Characteristics

Table 2. Characteristics of Included Primary Studies

No.	Author (Year)	Country	Study Design	N	Population	CAD Product(s)	CAD Version	Reference Standard	TB Prevalence
1	Qin et al. (2021)	Bangladesh	Cross-sectional	23,954	Adults ≥ 15 y, symptomatic TB suspects	CAD4TB, qXR, Lunit, JF CXR-1, InferRead DR	v7, v3, v4.9.0, v2, v2	Xpert MTB/RIF	4.2%
2	Khan et al. (2020)	Pakistan	Prospective cohort	2,198	Symptomatic adults, 99.5% HIV-negative	CAD4TB, qXR	v6, v2	Mycobacterial culture (2 sputa)	12.4%

No.	Author (Year)	Country	Study Design	N	Population	CAD Product(s)	CAD Version	Reference Standard	TB Prevalence
3	Sung et al. (2025)	Uganda	Cross-sectional	52,835	Adults ≥15y, facility- and community-based	qXR	v3.2	Xpert MTB/RIF Ultra	5.3%
4	Scott et al. (2025)	South Africa	Case-control	1,439	Adults ≥15y, community-based ACF	CAD4TB	v7	Xpert Ultra or culture	2.6%
5	Macpherson et al. (2024)	South Africa	Prospective cohort	483	Household contacts of RR-TB	CAD4TB, qXR, Lunit	v7.0, v3.0.0, v3.1.4.111	Xpert/culture + induced sputum	4.8%
6	Qin et al. (2024)	South Africa	Case-control	774	TB prevalence survey participants	12 CAD products	Multiple	Bacteriological (culture/Xpert)	33.3%
7	Kagujje et al. (2022)	Zambia	Retrospective cross-sectional	1,884	Adults ≥15y with TB symptoms	CAD4TB, qXR	v7.0, v3.0	Xpert MTB/RIF	15.6%
8	Edem et al. (2024)	Gambia	Retrospective diagnostic	724	Children <15y, presumptive TB	CAD4TB	v7	Microbiological (culture/Xpert)	8% (confirmed)
9	Fehr et al. (2021)	South Africa	Cross-sectional (screening)	9,914	Population-based; HIV-endemic area	CAD4TB	v5, v6	Microbiological sputum	1.0%

No.	Author (Year)	Country	Study Design	N	Population	CAD Product(s)	CAD Version	Reference Standard	TB Prevalence
10	Soares et al. (2021)	Brazil	Prospective cross- sectional	2,075	Incarcerated individuals	CAD4TB, Lunit, qXR	v6, v3.1.0.0, v3	Xpert MTB/RIF + culture	12.5%
11	Jiang et al. (2025)	China	Retrospective diagnostic yield	93,319	Primary health care patients	JF CXR-1	v1	Bacteriological confirmation	0.3%
12	Velen et al. (2021)	South Africa	Cross- sectional	3,576	Correctional facility inmates	CAD4TBv + CXR	Not specified	Xpert Ultra + culture	0.78%
13	Bosman et al. (2024)	Lesotho, S. Africa	Prospective cohort	1,392	Adults ≥ 18 y; ≥ 1 TB symptom; high HIV	CAD4TB	v7	Culture + Xpert Ultra	NR
14	Burke et al. (2024)	Malawi	Cluster RCT	415	HIV-positive adults admitted to hospital	CAD4TB	v6	Sputum culture	NR
15	Crowder et al. (2024)	Philippines, Vietnam, SA, Uganda, India	Prospective multinational cohort	1,392	Adults ≥ 2 wks cough; primary health	CAD4TB	v7	Xpert Ultra + culture	22%
16	Dakulala et al. (2024)	Papua New Guinea	Retrospective screening review	8,085	Island population-	CAD4TB	v not specified	Xpert MTB/RIF	14.0% (abnormal CXR)

No.	Author (Year)	Country	Study Design	N	Population	CAD Product(s)	CAD Version	Reference Standard	TB Prevalence
					wide screening				
17	Mahler et al. (2021)	Romania	Descriptive observational	5,003	Vulnerable populations (prisons, homeless, Roma)	CAD4TB	v6	Bacteriological confirmation	200/100,00

Outcome Analysis

Area Under the Receiver Operating Characteristic Curve (AUROC)

Table 3. AUROC Values by Study and CAD Product

Study (Year)	CAD Product	Version	Setting	Population	AUROC (95% CI)
Qin et al. 2021	qXR	v3	Facility	Adults, Bangladesh	0.908 (0.903– 0.913)
Qin et al. 2021	CAD4TB	v7	Facility	Adults, Bangladesh	0.903 (0.898– 0.909)
Qin et al. 2021	Lunit INSIGHT	v4.9.0	Facility	Adults, Bangladesh	0.886 (0.880– 0.892)
Khan et al. 2020	qXR	v2	Facility	Adults, Pakistan	~0.88 (est.)
Khan et al. 2020	CAD4TB	v6	Facility	Adults, Pakistan	~0.87 (est.)
Sung et al. 2025	qXR	v3.2	Community+Facility	Adults, Uganda	0.92 (0.90– 0.94)

Study (Year)	CAD Product	Version	Setting	Population	AUROC (95% CI)
Scott et al. 2025	CAD4TB	v7	Community ACF	Adults, South Africa	0.83 (0.80– 0.85)
Macpherson et al. 2024	Lunit INSIGHT	v3.1.4.111	Cohort	Household contacts, SA	0.91 (0.83– 0.99)
Macpherson et al. 2024	qXR	v3.0.0	Cohort	Household contacts, SA	0.88 (0.79– 0.97)
Macpherson et al. 2024	CAD4TB	v7.0	Cohort	Household contacts, SA	0.87 (0.77– 0.96)
Qin et al. 2024	Lunit / Nexus	Multiple	Survey	SA prevalence survey	~0.90 (est.)
Edem et al. 2024	CAD4TB	v7	Facility	Children <15y, Gambia	0.70 (0.60– 0.79)
Soares et al. 2021	qXR	v3	Prison	Inmates, Brazil	0.91 (est.)
Soares et al. 2021	Lunit	v3.1.0.0	Prison	Inmates, Brazil	0.91 (est.)
Bosman et al. 2024	CAD4TB	v7	Facility	Adults, Lesotho+SA	0.87 (0.84– 0.91)
Crowder et al. 2024	CAD4TB	v7	Facility (multinational)	Adults, 5 countries	~0.88 (est.)

AUROC values ranged from 0.70 in paediatric populations (Edem et al.) to 0.92 in an unselected adult community-based screening population in Uganda (Sung et al.). The highest AUROC values for Lunit INSIGHT were observed in the Macpherson et al. household contact cohort (0.91) and in the Qin 2024 South African prevalence survey (~0.90). qXR consistently demonstrated

the highest or co-highest AUROC in most multi-product comparative studies. The substantially lower AUROC in children underscores the critical need for paediatric-specific algorithm optimization.^{6,7,8,9}

Sensitivity at Fixed Threshold / Recommended Threshold

Table 4. Sensitivity and Specificity at Prespecified Thresholds

Study	CAD Product	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	Setting / Population
Qin et al. 2021	qXR v3	At 90% sens.	90%	74.3% (73.3–74.9%)	Bangladesh, facility
Qin et al. 2021	CAD4TB v7	At 90% sens.	90%	72.9% (72.3–73.5%)	Bangladesh, facility
Qin et al. 2021	Lunit v4.9.0	At 90% sens.	90%	63.2% (est.)	Bangladesh, facility
Khan et al. 2020	qXR v2	Prespecified	93% (89–95%)	75% (73–77%)	Pakistan, facility
Khan et al. 2020	CAD4TB v6	Data-derived	93% (90–96%)	69% (67–71%)	Pakistan, facility
Sung et al. 2025	qXR v3.2	Stratified	76.9% (71.9–81.2%)	96.1% (est.)	Uganda, community+facility
Scott et al. 2025	CAD4TB v7	Threshold=5	90%	44.9% (42.5–47.3%)	SA, community ACF
Scott et al. 2025	CAD4TB v7	Threshold=10	85%	54.1% (51.7–56.5%)	SA, community ACF
Bosman et al. 2024	CAD4TB v7	At 90% sens.	90%	68.2% (65.4–71.0%)	Lesotho+SA, high HIV
Soares et al. 2021	Lunit v3.1	At 90% sens.	90%	84% (est.)	Brazil, prison

Study	CAD Product	Threshold	Sensitivity (95% CI)	Specificity (95% CI)	Setting / Population
Soares et al. 2021	qXR v3	At 90% sens.	90%	74% (est.)	Brazil, prison
Crowder et al. 2024	CAD4TB v7	At 90% sens.	90%	70.3%	Multinational, facility
Kagujje et al. 2022	CAD4TB v7	At 90% sens.	90.5% (w/o prev TB)	60.3% (w/o prev TB)	Zambia; persons w/ prev TB: spec 24%
Kagujje et al. 2022	qXR v3	At 90% sens.	89.7% (w/o prev TB)	61.8% (w/o prev TB)	Zambia; persons w/ prev TB: spec 22.2%

Comparison of CAD vs. Human Radiologist Performance

Table 5. CAD vs. Human Radiologist Performance

Study	CAD Product	CAD AUROC/Sens/Spec	Radiologist Performance	Conclusion
Qin et al. 2021	All 5 AI algorithms	AUC 0.849–0.908	AUC <0.80 (radiologists)	All 5 AI algorithms significantly outperformed radiologists
Fehr et al. 2021	CAD4TB v5/v6	v5: same sens, lower spec; v6: comparable	Blinded radiologist (reference)	CAD4TBv6 achieved comparable sens+spec to radiologist

Study	CAD Product	CAD AUROC/Sens/Spec	Radiologist Performance	Conclusion
Bosman et al. 2024	CAD4TB v7	AUC 0.87	Expert radiologist (teleradiology)	CAD4TBv7 detected TB as well as expert radiologist
Jiang et al. 2025	JF CXR-1	DYD 83.9%	Radiologist DYD 25.6%	CAD substantially outperformed radiologists
Scott et al. 2025	CAD4TB v7	AUC 0.83	Not separately reported	CAD cost- effective vs. universal Xpert strategy

Across studies evaluating CAD versus human readers, CAD software consistently demonstrated superior or equivalent diagnostic performance. Qin et al. (2021) reported that all five evaluated AI algorithms significantly outperformed a group of three registered radiologists, a finding of substantial importance for deployment in settings with critical radiologist shortages.⁵ Jiang et al. (2025) further demonstrated that JF CXR-1 achieved a diagnostic yield among diagnosed cases of 83.9% compared to only 25.6% for radiologists in primary health care facilities in China.¹⁵

WHO Target Product Profile (TPP) Compliance

The WHO TPP for TB triage tests specifies minimum performance thresholds of $\geq 90\%$ sensitivity and $\geq 70\%$ specificity. Across included studies, qXR (v3) and CAD4TB (v7) met the TPP in symptomatic adult populations in Bangladesh (Qin et al., 2021) and Pakistan (Khan et al., 2020). Lunit and qXR met the TPP in Brazil prison settings (Soares et al., 2021). CAD4TB v7 met TPP in prospective multinational settings (Crowder et al., 2024). However, TPP was not consistently met in community-based active case finding settings (Scott et al., 2025), HIV-burdened populations

(Bosman et al., 2024), persons with prior TB history (Kagujje et al., 2022), and paediatric populations (Edem et al., 2024).^{5,6,7,8,12}

Subgroup Performance — HIV Status

Table 6. CAD Performance Stratified by HIV Status

Study	CAD Product	AUROC/Sensitivity (HIV+)	AUROC/Sensitivity (HIV–)	Absolute Difference	Notes
Scott et al. 2025	CAD4TB v7	AUC 0.76 (0.71–0.81)	AUC 0.85 (0.82–0.87)	–0.09 (p=0.0037)	Community ACF, SA
Tavaziva et al. 2021*	CAD4TBv6	Sens –13.4% adj. diff	Reference	–13.4% (–21.1, –6.9)	IPD meta-analysis; HIV+: n=621
Tavaziva et al. 2021*	qXRv2	Sens –13.4% adj. diff	Reference	–13.4% (–21.5, –6.6)	IPD meta-analysis
Fehr et al. 2021	CAD4TB v5/v6	Not adversely affected	Not adversely affected	No significant difference	HIV serostatus did not impact performance
Burke et al. 2024	CAD4TB v6	Median score 60 (IQR 51–71)	N/A (all HIV+)	N/A	RCT; exclusively PLWH admitted to hospital

*Tavaziva et al. (2021) is a meta-analysis included for comparative reference only and not in the study characteristics table.⁵

Subgroup Performance — Prior TB History

Table 7. CAD Performance in Persons with Prior TB History

Study	CAD Product	Sensitivity (Prior TB)	Specificity (Prior TB)	Sensitivity (No Prior TB)	Specificity (No Prior TB)
Kagujje et al. 2022	CAD4TB v7	89.3% (78.1–96.0%)	24.0% (19.9–28.5%)	90.5% (86.1–93.3%)	60.3% (57.4–63.0%)
Kagujje et al. 2022	qXR v3	94.6% (85.1–98.9%)	22.2% (18.2–26.6%)	89.7% (85.1–93.2%)	61.8% (58.9–64.5%)
Qin et al. 2021	All algorithms	Worse performance	Significantly reduced	Higher performance	Higher specificity
Qin et al. 2024	All 12 products	Worst subgroup	Markedly reduced	Better	Better

Paediatric Performance

Only Edem et al. (2024) evaluated CAD performance exclusively in paediatric populations, reporting an AUROC of 0.70 (95% CI 0.60–0.79) for CAD4TBv7 against a microbiological reference standard. While high specificity was observed (99.0%, 95% CI 98.0–100.0%), sensitivity was markedly suboptimal at 19.0% (95% CI 11–31%). Bayesian latent class analysis, which accounts for imperfect reference standards in children, yielded improved but still suboptimal sensitivity estimates of 42.7–50.3%. These data conclusively demonstrate that current CAD algorithms, primarily developed and validated in adult populations, are inadequate for reliable paediatric TB detection without substantial algorithm redesign and paediatric-specific training.⁸

CAD Threshold Optimization and Score Stratification

A critical and consistent finding across included studies was the significant impact of threshold selection on the sensitivity-specificity trade-off. Sung et al. (2025) demonstrated that stratification of qXR score thresholds by age and sex resulted in a statistically significant improvement in sensitivity at a fixed specificity level (76.9% vs. 75.0% at 96.1% specificity; $p=0.046$).² Qin et al. (2024) demonstrated that thresholds varied substantially across products and versions, and across demographic subgroups, necessitating population-specific calibration.⁹ These findings have direct implications for implementation: a single universal threshold cannot optimally

serve all target populations, and programmatic deployment must incorporate systematic threshold calibration exercises using locally representative data.

Cost-Effectiveness and Health System Impact

Table 8. Cost-Effectiveness and Health System Impact

Study	Setting	Cost Outcome	CAD Cost	Comparator Cost	Key Finding
Bashir et al. 2022*	Pakistan (national)	Per-screen cost	CAD4TB: \$0.25–2.33; InferRead: \$0.19–2.78	Radiologist: \$0.70–0.93	CAD lower cost at high throughput; scale-up 2.65–19.23M USD vs. 23.97M USD for radiologists
Scott et al. 2025	South Africa, community	Per TB case detected/10,000 screened	CAD-Xpert: \$2,207–\$3,745	Universal Xpert: \$4,698	CAD-directed strategy 20–53% cost reduction; lower diagnostic yield
Jiang et al. 2025	China, primary HC	Radiologist workload	85.5% reduction in CXRs requiring review	Universal radiologist review	CAD prescreening significantly reduces workload

Study	Setting	Cost Outcome	CAD Cost	Comparator Cost	Key Finding
Qin et al. 2021	Bangladesh, facility	Xpert tests saved	>50% reduction in Xpert tests at >90% sensitivity	Universal Xpert	Significant Xpert test savings

*Bashir et al. (2022) is a costing analysis included for comparative reference in the discussion only.

Clinical Outcomes — TB Treatment Initiation

Burke et al. (2024) conducted the only cluster RCT in this review, evaluating enhanced TB diagnostics including digital CXR with CAD4TBv6 plus urine lipoarabinomannan (LAM) tests among HIV-positive adults admitted to Zomba Central Hospital, Malawi. TB treatment was initiated in 22.2% of the enhanced diagnostic arm versus 11.5% in the usual care arm (RR 1.92; 95% CI 1.20–3.08; $p < 0.05$). Importantly, no difference in 56-day mortality was observed (HR 1.05; 95% CI 0.72–1.53), suggesting that greater treatment initiation did not translate into immediate mortality reduction, potentially reflecting the severity of disease at the time of hospitalization.¹⁴ This RCT provides the highest level of evidence in this review for the direct clinical impact of CAD-integrated diagnostic pathways.

CAD Performance Across Geographic Settings and Clinical Contexts

Table 9. Summary of CAD Performance by Geographic Setting and Clinical Context

Setting	Studies	TB Prevalence	Best AUROC	TPP Met?	Key Limitations
Facility-based (symptomatic)	Qin 2021, Khan 2020, Crowder 2024	4–22%	0.908	Yes (qXR, CAD4TB)	Performance varies by HIV, prior TB, age

Setting	Studies	TB Prevalence	Best AUROC	TPP Met?	Key Limitations
Community-based ACF	Scott 2025, Sung 2025, Fehr 2021	1–5%	0.92	Borderline/No	Lower specificity; asymptomatic performance better than expected
Correctional facilities	Soares 2021, Velen 2021, Mahler 2021	0.78–12.5%	0.91	Yes (Lunit, qXR)	Prior TB degrades specificity
Hospital-based (HIV+)	Burke 2024	High	0.75 (est.)	No	HIV-related atypical CXR patterns
Paediatric	Edem 2024	8% (confirmed)	0.70	No	Suboptimal sensitivity; needs paediatric-specific optimization
Population-wide screening	Dakulala 2024, Jiang 2025	0.3–14%	~0.85	Partial	Workload reduction key benefit; high specificity needed
Household contacts	Macpherson 2024	4.8%	0.91 (Lunit)	Yes	Incident TB detection

Setting	Studies	TB Prevalence	Best AUROC	TPP Met?	Key Limitations
					limited (AUC 0.60–0.65)

Smear-Negative TB Detection

Detection of smear-negative pulmonary TB represents a critical challenge for both CAD systems and conventional clinical evaluation. Khan et al. (2020) demonstrated that sensitivity for smear-negative TB was significantly lower for both qXRv2 and CAD4TBv6 compared to smear-positive disease. Tavaziva et al. (2021; meta-analysis) quantified pooled adjusted sensitivity differences: CAD4TBv6 −12.3% (95% CI −19.5, −6.1), Lunit −17.2% (95% CI −24.6, −10.5), and qXRv2 −16.6% (95% CI −24.4, −9.9) in smear-negative versus smear-positive TB. These findings indicate that CAD systems are less effective in detecting microbiologically less severe disease, which paradoxically represents the population with the greatest potential benefit from early detection and treatment.^{5,11}

DISCUSSION

Principal Findings

This systematic review synthesizes evidence from 17 primary studies encompassing over 130,000 participants across diverse geographic, clinical, and epidemiological contexts, evaluating the diagnostic performance, clinical utility, cost-effectiveness, and implementation characteristics of AI-based CAD software for pulmonary TB detection on digital CXRs. The principal findings can be summarized as follows: (1) commercially available CAD software demonstrates high diagnostic accuracy in symptomatic adult populations, with AUROC values typically ranging from 0.86 to 0.92 and frequently meeting WHO TPP thresholds; (2) performance is substantially and consistently inferior in key subgroups including PLWH, persons with prior TB history, elderly individuals, and paediatric patients; (3) threshold selection critically modulates the sensitivity-specificity trade-off and must be calibrated to local epidemiological conditions; (4) CAD consistently outperforms human radiologists in both accuracy and throughput capacity; and (5) CAD-based screening strategies offer

meaningful cost savings and operational efficiencies compared to universal confirmatory testing paradigms.^{5,6,7,8,9,10}

Diagnostic Accuracy in Context

The high AUROC values observed in facility-based settings with symptomatic adult populations (0.886–0.908 in Qin et al., 2021; 0.92 in Sung et al., 2025) reflect optimal conditions for CAD deployment: a population with high pre-test probability, symptomatic presentation, and high-quality digital radiographic images.^{1,2} The observation that multiple CAD systems exceeded the diagnostic accuracy of registered radiologists across large, diverse datasets represents a landmark finding with profound implications for global TB control, particularly in settings with critical human resource shortages. However, these performance estimates should not be extrapolated uncritically to all operational contexts.

The substantially lower AUROC of 0.83 observed in community-based active case finding by Scott et al. (2025), and the failure to meet WHO TPP thresholds in this setting, highlights the importance of population-level TB prevalence and pre-test probability in determining real-world performance.⁴ In community-based screening programs targeting asymptomatic or mildly symptomatic individuals with lower TB prevalence, the positive predictive value of CAD decreases substantially, leading to higher rates of unnecessary referral for confirmatory testing and increased costs. Nevertheless, Sung et al. (2025) provided a more optimistic estimate of community-based performance (AUROC 0.92), attributing this to more comprehensive confirmatory testing that reduced verification bias.²

Subgroup Performance — HIV and Immunosuppression

The consistently inferior CAD performance in PLWH is attributable to atypical radiological presentations of TB in immunosuppressed individuals, including less cavitation, more diffuse infiltrates, miliary patterns, and a higher proportion of extrapulmonary disease.⁵ Scott et al. (2025) demonstrated a statistically significant reduction in AUROC from 0.85 (HIV-negative) to 0.76 (HIV-positive; $p=0.0037$) in a community-based South African population.⁴ These findings are clinically critical as the geographic overlap of HIV and TB epidemics is greatest in sub-Saharan Africa, precisely where CAD deployment is most urgently needed. Specialized CAD algorithms

incorporating HIV status as an input variable, or separate algorithm training on HIV-positive CXR datasets, may be necessary to achieve adequate performance in these populations.

Prior TB History and Radiological Sequelae

The markedly impaired specificity of CAD systems in persons with prior TB history — 22.2–24.0% compared to 60.3–61.8% in those without prior treatment — reflects the algorithmic challenge of distinguishing active TB-related abnormalities from post-TB residual changes including fibrocavitary disease, calcifications, pleural thickening, and volume loss.⁶ These sequelae produce radiological patterns that overlap substantially with active TB, generating false-positive CAD scores. This limitation has direct programmatic relevance in high-burden settings where prior TB treatment rates are substantial. Population-specific threshold recalibration for this subgroup alone may be insufficient; algorithm retraining on datasets enriched with post-TB sequelae CXRs appears necessary.

Paediatric TB and Algorithm Limitations

The markedly suboptimal sensitivity of CAD4TBv7 in Gambian children (19.0% by microbiological reference standard; 42.7–50.3% by Bayesian LCA) reflects the fundamental biological and radiological differences between paediatric and adult pulmonary TB.⁸ In children, TB more frequently presents as primary complex with hilar adenopathy, miliary disease, or atypical infiltrates without cavitation, patterns for which adult-trained CAD algorithms are not optimized. The CAPTURE initiative described by Palmer et al. (2026), aiming to establish a large paediatric CXR repository for CAD algorithm development and optimization, represents a critical and timely research effort to address this gap. Until paediatric-specific CAD algorithms are validated, clinical evaluation and expert radiological interpretation remain essential components of diagnostic pathways for child TB.

Threshold Optimization and Population-Specific Calibration

One of the most clinically actionable and consistently reported findings in this review is the critical importance of threshold selection in determining CAD performance. Sung et al. (2025) demonstrated that age- and sex-stratified thresholds significantly improved diagnostic accuracy.² Qin et al. (2024), evaluating 12 CAD products, found substantial inter-product and intra-product (version-

specific) threshold variability, as well as significant differences in optimal thresholds across demographic subgroups and geographic settings.⁹ These findings collectively reinforce the WHO recommendation and expert consensus that CAD threshold selection must be performed using locally representative datasets, rather than applying manufacturer-specified universal thresholds. Programmatic implementation protocols should incorporate a systematic threshold calibration phase prior to scale-up, with regular recalibration as the screening population characteristics evolve.

Cost-Effectiveness and Health Systems Considerations

The economic analyses included in this review consistently demonstrate that CAD-based screening can reduce per-screen costs and total program expenditures compared to universal radiologist reading or universal confirmatory testing paradigms, particularly at high screening volumes.^{13,15} Scott et al. (2025) quantified a 20–53% cost reduction with a CAD-directed Xpert Ultra strategy, albeit at the expense of a lower diagnostic yield.⁴ Jiang et al. (2025) demonstrated an 85.5% reduction in radiologist workload through selective review of CAD-positive images, a finding with critical implications for radiologist capacity building and task-shifting in resource-limited settings.¹⁵

The cost-effectiveness advantages of CAD, however, are contingent upon the pricing model adopted by software developers. Per-read or per-month subscription models may paradoxically penalize high-volume screening programs, whereas perpetual licensing models offer greater scalability.¹⁰ Policy advocacy for transparent, equitable, and volume-insensitive pricing structures is an important component of the conditions necessary for CAD to achieve its full public health potential.

Implementation Barriers and Facilitators

Beyond diagnostic accuracy, multiple implementation-level barriers have been identified across included studies: (1) Regulatory uncertainty regarding software version updates and certification standards; (2) data sovereignty concerns related to the acquisition of patient radiological data by commercial developers; (3) integration challenges with existing health information systems and GeneXpert networks; (4) the requirement for reliable internet connectivity for cloud-based CAD systems; and (5) human resource training needs for appropriate threshold-based referral protocols.^{10,11} Implementation facilitators include the increasing availability of affordable portable digital

radiography units suitable for mobile screening campaigns, the endorsement of specific CAD products by WHO, and the listing of select CAD products in the Global Drug Facility diagnostics catalogue.

Role of Systematic Reviews and Meta-Analyses

Harris et al. (2019), in a systematic review of 53 studies, demonstrated that diagnostic accuracy was significantly higher in model development studies (median AUC 0.88) versus clinical evaluation studies (median AUC 0.75), and with deep learning versus machine learning algorithms.³ Tavaziva et al. (2021), conducting an individual patient data meta-analysis of 3,727 participants, found similar accuracy across CAD4TBv6, Lunit, and qXRv2 at pooled sensitivity of 90%, with pooled specificities of 56.9%, 54.1%, and 60.5%, respectively, substantially lower than those observed in individual studies — illustrating the importance of adjusting for patient-level covariates.⁵ Zhan et al. (2022) reported pooled sensitivity and specificity of 91% and 65% in clinical trials, and 94% and 95% in model development studies, further confirming the performance gap between controlled and real-world conditions.²⁰

Strengths

This review has several strengths: (1) comprehensive multi-database search with no language restrictions; (2) inclusion of ≥ 15 primary studies across diverse geographic and clinical settings; (3) evaluation of ≥ 10 distinct outcome domains; (4) rigorous QUADAS-2 risk of bias assessment; and (5) inclusion of the only RCT evaluating CAD-integrated clinical outcomes.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

AI-based computer-aided detection software for pulmonary TB on digital chest radiographs represents a validated, scalable, and cost-effective diagnostic tool for TB screening and triage in high-burden settings. Multiple commercially available products — most notably CAD4TB, qXR, and Lunit INSIGHT CXR — demonstrate diagnostic accuracy meeting or approaching WHO Target Product Profile benchmarks of $\geq 90\%$ sensitivity and $\geq 70\%$ specificity in symptomatic adult populations. These systems consistently outperform human radiologists in diagnostic yield and throughput efficiency. Critically, CAD performance is substantially modulated by patient-level factors,

particularly HIV status, prior TB treatment history, age, and clinical setting. Paediatric populations represent a priority research gap requiring dedicated algorithm development and validation. Population-specific threshold calibration is a mandatory prerequisite for programmatic implementation. CAD-directed triage strategies offer meaningful health system benefits including reduced radiologist workload and cost savings, supporting scale-up in resource-limited settings.

Recommendations

Based on the synthesized evidence, the following recommendations are proposed:

For TB Program Implementation:

(1) Prioritize CAD products (CAD4TB v7, qXR v3.x, Lunit INSIGHT CXR) that have met WHO TPP benchmarks in peer-reviewed, independently conducted clinical evaluations. (2) Mandate local threshold calibration exercises using datasets representative of the target screening population prior to programmatic deployment. (3) Develop tailored clinical pathways for high-risk subgroups (PLWH, persons with prior TB history) incorporating supplementary diagnostic tests. (4) Implement CAD as a component of integrated, multi-test screening algorithms rather than as a standalone diagnostic tool.

For Research:

(1) Conduct large-scale prospective studies evaluating CAD performance in paediatric populations with paediatric-specific reference standards. (2) Prioritize implementation science research evaluating real-world effectiveness, equity, and health system integration of CAD. (3) Investigate novel algorithmic approaches to improve performance in PLWH and persons with prior TB. (4) Conduct health economic analyses in diverse country contexts, including Indonesia, to inform national TB program investment decisions.

For Policy and Governance:

(1) Advocate for transparent, equitable CAD pricing models that do not penalize high-volume public health screening programs. (2) Establish regulatory frameworks for software version updates and performance re-evaluation. (3) Develop data governance standards protecting patient radiological data sovereignty in low- and middle-income countries. (4) Integrate CAD into Indonesia's national TB elimination strategy, given the country's status as the second highest TB-burden country globally.

REFERENCES

1. World Health Organization. Global Tuberculosis Report 2023. Geneva: WHO; 2023. Available from: <https://www.who.int/publications/i/item/9789240083851>
2. Sung J, Kibirige J, Asege L, et al. Performance of universal and stratified computer-aided detection thresholds for chest x-ray-based tuberculosis screening: a cross-sectional, diagnostic accuracy study. *Lancet Digit Health*. 2025. Available from: <https://consensus.app/papers/details/9a67dd118605542fb3b47a7d557af070/>
3. Harris M, Qi A, Jeagal L, et al. A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. *PLoS ONE*. 2019;14(6):e0221339. Available from: <https://consensus.app/papers/details/e04c2167578e5541b705a765a20b382c/>
4. Scott A, Ngo P, Ndlangalavu G, et al. Clinical evaluation of computer-aided digital x-ray detection of pulmonary tuberculosis during community-based screening or active case-finding: a case-control study. *Lancet Glob Health*. 2025. Available from: <https://consensus.app/papers/details/253d45cbe5145a1f82ead28153d75997/>
5. Tavaziva G, Majidulla A, Garcia-Basteiro AL, 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*. 2021;74(9):1390-1400. Available from: <https://consensus.app/papers/details/f81b0e23c1de570eaba723b3d816d76d/>
6. Kagujje M, Somwe P, Mwanza W, et al. The performance of computer-aided detection digital chest X-ray reading technologies for triage of active tuberculosis among persons with a history of previous tuberculosis. *Clin Infect Dis*. 2022;77(3):372-380. Available from: <https://consensus.app/papers/details/9d6bbf6c36c55d79886ac2ac774095dd/>
7. Khan FA, Majidulla A, Tavaziva G, et al. 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. *Lancet Digit Health*. 2020;2(11):e573-e581. Available from: <https://consensus.app/papers/details/ec455106baaa5446a57f7ba44f372b95/>

8. Edem VF, Mulindwa F, Calisto J, et al. Accuracy of CAD4TB (Computer-Aided Detection for Tuberculosis) on paediatric chest radiographs. *Eur Respir J*. 2024. Available from: <https://consensus.app/papers/details/de3c347f1adf56779b511d8820711af1/>
9. Qin ZZ, Durán LF, Solano W, et al. Computer-aided detection of tuberculosis from chest radiographs in a tuberculosis prevalence survey in South Africa: external validation and modelled impacts of commercially available artificial intelligence software. *Lancet Digit Health*. 2024. Available from: <https://consensus.app/papers/details/c04303c0338e5bec9f35fba9f7445065/>
10. David PM, Oliver-Commins J, Flaño E, Kouyaté B, Khan FA. Conditions required for the artificial-intelligence-based computer-aided detection of tuberculosis to attain its global health potential. *Lancet Digit Health*. 2022;4(8):e571-e576. Available from: <https://consensus.app/papers/details/f69fc21e711252de9cced96c7105183f/>
11. Qin ZZ, Sander MS, Rai B, 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(9):e543-e554. Available from: <https://consensus.app/papers/details/996502dfab585a5c88b238a1fc88d999/>
12. Soares TR, Ramos JM, Dara M, et al. Evaluation of chest X-ray with automated interpretation algorithms for mass tuberculosis screening in prisons: a cross-sectional study. *Lancet Reg Health Am*. 2021;4:100073. Available from: <https://consensus.app/papers/details/13da2475caf75fd1ba6c0a9c1d4acbc6/>
13. Bashir N, Khan FA, Bastos ML, et al. 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*. 2022;17(4):e0267112. Available from: <https://consensus.app/papers/details/15ca522bd7c6510c90cd178a99da8092/>
14. Burke RM, Finney LJ, Otieno M, et al. Enhanced tuberculosis diagnosis with computer-aided chest X-ray and urine lipoarabinomannan in adults with HIV admitted to hospital (CASTLE Study): a cluster randomized trial. *Clin Infect Dis*. 2024. Available from: <https://consensus.app/papers/details/ac5e61e2e699553d8e1e1b847b2fca06/>

15. Jiang W, Zhang H, Li Z, et al. AI-Powered Chest X-Ray for Diagnosing Pulmonary Tuberculosis in County and Township Health Care Facilities in Yichang: Retrospective, Real-World Study. *J Med Internet Res.* 2025. Available from: <https://consensus.app/papers/details/6ce5e1d1b77d59ccbc299fa5b87e15/>
16. Macpherson L, Bassett J, Rangaka MX, et al. Diagnostic accuracy of chest X-ray computer-aided detection software for detection of prevalent and incident tuberculosis in household contacts. *Clin Infect Dis.* 2024. Available from: <https://consensus.app/papers/details/0b920df4bdc75586adf10f2c97c6610a/>
17. Crowder R, Nemes E, Mpande M, et al. Diagnostic accuracy of tuberculosis screening tests in a prospective multinational cohort: Chest radiography with computer-aided detection, Xpert tuberculosis host response, and C-reactive protein. *Clin Infect Dis.* 2024. Available from: <https://consensus.app/papers/details/d2ce82d379085e9ab752dc1b1ab40f8a/>
18. Bosman S, Furin J, Franke MF, et al. Evaluation of C-reactive protein and computer-aided analysis of chest X-rays as tuberculosis triage tests at health facilities in Lesotho and South Africa. *Clin Infect Dis.* 2024. Available from: <https://consensus.app/papers/details/416bcb31f73456989437b82b33a491b3/>
19. Fehr J, Reperant LA, Kaufmann E, et al. Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. *NPJ Digit Med.* 2021;4(1):43. Available from: <https://consensus.app/papers/details/eecf89075b2951f28e6ddc7f5d25535b/>
20. Zhan Y, Wang L, Lin W, et al. Diagnostic accuracy of the artificial intelligence methods in medical imaging for pulmonary tuberculosis: a systematic review and meta-analysis. *J Clin Med.* 2022;11(14):4145. Available from: <https://consensus.app/papers/details/bf39a17ba3735b3e85b0feca9ff5951c/>
21. Dakulala P, Marais BJ, Fox GJ, et al. Evaluation of a population-wide, systematic screening initiative for tuberculosis on Daru island, Western Province, Papua New Guinea. *BMC Public Health.* 2024. Available from: <https://consensus.app/papers/details/ce61cdcf4069586b9d42dddc682f60e/>

22. Mahler B, Radulescu M, Trancă S, et al. Use of targeted mobile X-ray screening and computer-aided detection software to identify tuberculosis among high-risk groups in Romania: descriptive results of the E-DETECT TB active case-finding project. *BMJ Open*. 2021;11(9):e043701. Available from: <https://consensus.app/papers/details/7f52886a66755b959c6653c3d113e491/>
23. Velen K, Charalambous S, Sabi I, et al. Digital chest X-ray with computer-aided detection for TB screening within correctional facilities. *Ann Am Thorac Soc*. 2021;19(1):58-67. Available from: <https://consensus.app/papers/details/9207d7b972dc56e7a196d1f4615b9a85/>
24. Keter A, Zikusooka M, Dlamini S, et al. Evaluation of tuberculosis diagnostic test accuracy using Bayesian latent class analysis in the presence of conditional dependence. *PLoS ONE*. 2023. Available from: <https://consensus.app/papers/details/3f2be89d11f454b3aef3ea5f1e532d1f/>
25. Worodria W, Nakayita M, Zawedde-Muyanja S, et al. A multi-country head-to-head accuracy comparison of automated chest X-ray algorithms for tuberculosis. *Ann Am Thorac Soc*. 2026. Available from: <https://consensus.app/papers/details/0685689a063f59e7912b26595f04e92f/>
26. Yang X, Jiang W, Li Z, et al. Effectiveness of computer-aided detection chest X-ray screening for improving tuberculosis diagnostic yield in Chinese primary healthcare settings: study protocol for a prospective cluster randomised controlled trial. *BMJ Open*. 2026. Available from: <https://consensus.app/papers/details/95d86f7a7a855ae59e464045b257fa5e/>