

Review

Not peer-reviewed version

---

# Artificial Intelligence for Tuberculosis Screening and Detection: From Evidence to Policy and Implementation

---

[Hien Thi Thu Nguyen](#)\*, Vang Le-Quy, [Anh Tuan Dinh-Xuan](#), [Linh Nhat Nguyen](#)\*

Posted Date: 25 February 2026

doi: 10.20944/preprints202602.1545.v1

Keywords: tuberculosis; artificial intelligence; computer-aided detection (CAD); chest radiography; screening and triage



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

# Artificial Intelligence for Tuberculosis Screening and Detection: From Evidence to Policy and Implementation

Hien Thi Thu Nguyen <sup>1,2,3,\*†</sup>, Vang Le-Quy <sup>4,5,†</sup>, Anh Tuan Dinh-Xuan <sup>6</sup> and Linh Nhat Nguyen <sup>7,\*</sup>

<sup>1</sup> Department of Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark

<sup>2</sup> Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

<sup>3</sup> AVSE Global Medical Translational Research Network, Paris, France

<sup>4</sup> Novodan ApS, Aalborg, Denmark

<sup>5</sup> AVSE Global Data Science Network, Paris, France

<sup>6</sup> Department of Respiratory Medicine and Physiology, Cochin University Hospital, Paris, France

<sup>7</sup> Department for HIV, Tuberculosis, Hepatitis and Sexually Transmitted Infections, World Health Organization (WHO) Headquarters in Geneva, Switzerland

\* Correspondence: hien@rn.dk (H.T.T.N.); nguyenni@who.int (L.N.N.); Tel.: +45-6133-2955 (H.T.T.N.); +41-79-244-60-85 (L.N.N.)

† These authors contributed equally to this work and share first authorship.

## Abstract

**Background:** Artificial intelligence (AI) is increasingly used to support tuberculosis (TB) screening and diagnosis, especially computer-aided detection (CAD) applied to chest radiography (CXR). The value of these programs depends not only on diagnostic accuracy but also on threshold calibration, integration into clinical workflow, and capacity for confirmatory testing. **Methods:** We conducted a narrative state-of-the-art review of AI applications relevant to TB screening and diagnosis. We synthesize evidence from World Health Organization policy documents, independent validation initiatives, and peer-reviewed studies reporting diagnostic performance and real-world implementation outcomes. **Results:** CAD for CXR is the most mature AI application and is recommended by WHO for TB screening and triage among individuals aged  $\geq 15$  years in specific contexts. CAD-CXR can achieve sensitivity comparable to human readers, although performance varies by product, software version, population, and imaging conditions. Threshold selection is therefore a programmatic decision influencing referral volume and resource use. Independent benchmarking and local verification studies are essential to confirm performance and assess subgroup variability, including among people living with HIV and those with prior TB. Other AI approaches, including computed tomography (CT)-based imaging analysis, point-of-care ultrasound interpretation, cough or stethoscope sound analysis, clinical risk models, and genomic resistance prediction, are still at earlier stages and generally require further independent validation before routine programmatic use. **Conclusions:** AI has the potential to strengthen TB screening and diagnostic pathways, but impact should be evaluated using patient- and program-level outcomes rather than accuracy alone. Responsible scale-up requires local calibration, governance safeguards, and ongoing monitoring in real-world settings.

**Keywords:** tuberculosis; artificial intelligence; computer-aided detection (CAD); chest radiography; screening and triage

## 1. Introduction

The Tuberculosis (TB) continues to impose a substantial global health and economic burden despite the availability of effective diagnostics and treatment. Persistent gaps in case detection,

together with delays from symptom onset to treatment initiation, continue to limit the impact of national TB programs [1,2]. These challenges are further compounded by structural constraints, including shortages of trained readers, inconsistent interpretation of chest radiographs (CXR), limited capacity for confirmatory testing, and logistical barriers that disproportionately affect high-risk and underserved populations [2].

In this context, artificial intelligence (AI) has emerged as a pragmatic tool to support TB screening, and diagnosis rather than to replace clinical judgment. Current AI applications aim to standardize image interpretation, improve early sensitivity within the diagnostic cascade, and reduce delays between screening and confirmation testing [3]. The most advanced and widely implemented application to date is computer-aided detection (CAD) for digital CXR, which is used for TB screening and triage by identifying individuals who should undergo confirmatory microbiological testing [4].

When appropriately calibrated and integrated into diagnostic workflows, CAD-CXR may reduce reader workload, and support more standardized triage in high-throughput screening settings, which has contributed to its policy endorsement for adult TB screening in selected contexts [5,6]. However, improvements in diagnostic performance alone are insufficient to ensure patients benefit unless AI tools are embedded within well-functioning diagnostic and treatment pathways [7].

Beyond CXR-based screening, a range of AI-enabled approaches is under investigation across the TB diagnostic pathway, including computed tomography, point-of-care ultrasound, acoustic analysis of cough or lung sounds, clinical risk stratification models, and genomic resistance prediction [8–12]. However, these tools differ substantially in their levels of evidence, readiness for routine use, and relevance to programmatic decision-making, and most remain at the research or pilot stage [1].

This article is a narrative, state-of-the-art review of AI applications for TB screening and diagnosis. We synthesize evidence from international policy guidance, independent validation initiatives, and peer-reviewed studies reporting diagnostic performance, implementation experience, and patient- and program-level outcomes [4,13]. Particular emphasis is placed on CAD-enabled chest radiography, given its current policy endorsement and programmatic uptake [1], while emerging AI modalities are interpreted cautiously in light of existing evidence gaps and implementation challenges.

## 2. Materials and Methods

This article presents a narrative, state-of-the-art review of the application of AI for TB screening and diagnosis. The review was not conducted as a systematic review and does not follow a formal meta-analytic protocol. Rather, it aims to synthesize and interpret the current evidence base in a way that is relevant to policy, clinical practice, and programmatic implementation.

Evidence was identified and synthesized from multiple complementary sources to minimize selection bias and ensure coverage of both policy-level guidance and peer-reviewed evidence. These sources included World Health Organization (WHO) policy documents and technical guidance related to TB screening and diagnosis, as well as WHO materials addressing the use of AI-enabled software as a medical device. We also reviewed outputs from independent validation initiatives, including international benchmarking platforms used to evaluate the diagnostic performance of computer-aided detection (CAD) software for chest radiography.

Peer-reviewed studies were included when they reported diagnostic accuracy, implementation experience, or program-level outcomes of AI applications relevant to TB care. To enhance transparency and reproducibility, sources were identified using a structured search and screening approach across WHO repositories, independent CAD benchmarking platforms (including FIND resources), and bibliographic databases. Bibliographic searches were conducted in PubMed/MEDLINE and Google Scholar for studies published from 2005 to January 2026 using combinations of terms including “tuberculosis”, “computer-aided detection”, “chest radiography”, “artificial intelligence”, “deep learning”, “screening”, “triage”, “cost-effectiveness”, and

“implementation”. Materials were selected based on relevance to programmatic decision-making, maturity of evidence, and alignment with current international guidance.

Although WHO guidance informs several sections of this review, the scope is not limited to AI applications for which WHO recommendations currently exist. Emerging tools at earlier stages of evidence development are included where they illustrate future potential, with explicit attention to current limitations. Across all sections, priority is given to issues relevant to clinical safety, governance, equity, and patient- and program-level outcomes rather than algorithmic performance metrics alone.

### 3. CAD-CXR for Tuberculosis Screening and Triage

#### 3.1. Policy Landscape and Scope of WHO Recommendations

The WHO recommends the use of CAD software to interpret digital chest radiographs for TB screening and triage among individuals aged 15 years and older in populations where systematic screening is recommended [3,4]. CAD is not intended to function as a stand-alone diagnostic test. Instead, it is meant to be embedded within diagnostic algorithms that include confirmatory microbiological testing and appropriate clinical oversight [3].

WHO recommendations for CAD-CXR are conditional and based on low to moderate certainty of evidence. Guidance emphasizes that the intended use, target population, and position of CAD within national screening pathways should be explicitly defined [3,4]. Current recommendations do not extend to children, reflecting insufficient evidence regarding diagnostic accuracy and clinical impact in pediatric populations [3].

Priority deployment settings include community-based active case finding, high-throughput outpatient departments, prisons, and other congregate or high-risk environments where screening coverage is constrained by human reader capacity [2,3].

#### 3.2. Evidence Underpinning Policy Recommendations

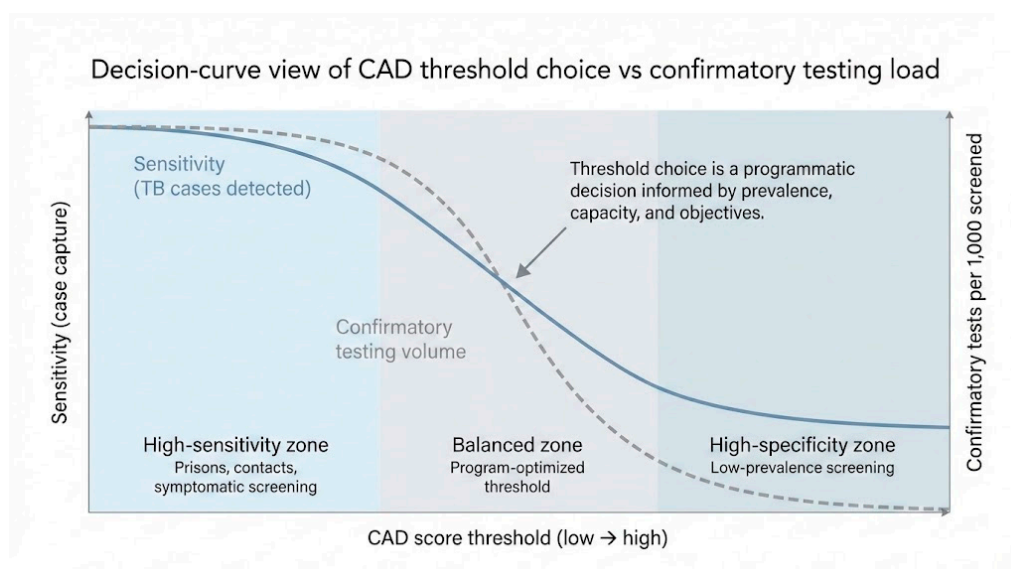
##### 3.2.1. Diagnostic Performance in Screening Contexts

In TB screening and triage, diagnostic performance is best interpreted in terms of sensitivity and its downstream implications for case detection and time to treatment, rather than relying primarily on summary accuracy metrics such as the area under the receiver operating characteristic curve (AUC). Screening tools are designed to identify individuals who should proceed to confirmatory testing, and their practical value depends on how effectively they capture people with disease while managing the workload placed on health systems [3,4].

Evidence synthesized for WHO policy development indicates that contemporary CAD-CXR products can be configured to achieve sensitivity levels comparable to those of human readers when used for adult TB screening [3,5]. However, performance varies across products, software versions, populations, and imaging conditions, and no single operating point appears optimal across all settings [3].

##### 3.2.2. Threshold Calibration as a Programmatic Decision

A defining feature of CAD-CXR systems is the ability to select an operating threshold that determines which images are classified as abnormal and referred for further testing. Threshold selection is therefore not simply a technical choice but a programmatic decision that reflects local priorities, TB prevalence, and available diagnostic capacity [4]. Thresholds that prioritize high sensitivity increase case capture but also raise confirmatory testing volume, whereas more specific thresholds reduce workload at the risk of missed cases [4]. The programmatic implications of this trade-offs are illustrated schematically in Figure 1.



**Figure 1. Decision-curve view of CAD threshold choice versus confirmatory testing load.** Computer-aided detection (CAD) systems for chest radiography allow programs to select an operating threshold that determines which individuals are referred for confirmatory tuberculosis testing. Lower thresholds prioritize sensitivity and maximize case capture but increase confirmatory testing volume, whereas higher thresholds reduce workload at the cost of missed cases. Optimal threshold selection is therefore a programmatic decision informed by local TB prevalence, screening objectives, and diagnostic capacity.

WHO guidance recommends that programs define a sensitivity target appropriate to the screening context and calibrate CAD threshold locally using real-world data, rather than treating CAD output as a fixed or universally applicable score [4,13].

### 3.2.3. Subgroup Performance and Sources of Variability

Diagnostic performance of CAD-CXR is not uniform across populations. Individuals with previous TB, or chronic lung disease may generate higher false-positive scores. Conversely, people living with HIV and older adults may present with atypical radiographic features that affect sensitivity [3,14]. Device type, image acquisition protocols, and image quality further contribute to variability.

Disaggregated reporting and local verification are therefore essential to identify potential biases and mitigate inequities in access to diagnosis [2,3].

### 3.2.4. Linking Diagnostic Performance to Patient- and Population-Level Outcomes

While much of the literature on AI for TB focuses on diagnostic accuracy, a smaller but growing body of evidence reports outcomes more directly relevant to patients and TB programs. Selected studies reporting patient- and population-level outcomes of AI-enabled TB interventions are summarized in Table 1, including evidence on screening throughput, confirmatory testing yield, cost per case detected [6,15,16].

**Table 1.** Studies reporting patient- and program-level outcomes of AI-enabled tuberculosis interventions.

Study	Setting and population	AI application	Study design	Key patient / program-level outcomes	Main findings
Moodley et al., 2022	Primary health clinics,	CAD-enabled digital CXR	Prospective implementation study	Screening throughput;	CAD-supported CXR improved screening efficiency and throughput in routine clinic

	South Africa	for TB screening		confirmatory testing yield	settings, with acceptable referral volumes for confirmatory testing.
Velen et al., 2022	Correctional facilities, South Africa	CAD-enabled digital CXR	Prospective screening evaluation	TB yield; referral volume; operational feasibility	Use of CAD in prisons identified additional TB cases compared with symptom-based screening and supported high-volume screening in a congregate setting.
Garg et al., 2025	Community-based active case finding, Nigeria	Ultraportable CXR with CAD	Economic evaluation alongside implementation	Cost per TB case detected; program costs	CAD-enabled screening was associated with lower cost per TB case detected than symptom-based screening in settings with substantial asymptomatic TB.
Qin et al., 2019	Facility- and community-based screening, multiple countries	CAD-CXR	Comparative diagnostic study with operational implications	Inter-reader variability; workflow implications	CAD reduced inter-reader variability and achieved diagnostic performance comparable to human readers, supporting its potential use as a standardized triage aid in screening workflows.
Signorell et al., 2025 (protocol)	Community screening, Lesotho and South Africa	CAD alone vs. CAD + point-of-care CRP	Paired screening positive pragmatic trial (protocol)	Time to treatment initiation; cost-effectiveness	Designed to evaluate downstream patient- and program-level outcomes beyond diagnostic accuracy; results pending.
Bartl et al., 2025	HIV care cohorts, sub-Saharan Africa	ML-based clinical TB risk model	Retrospective cohort analysis	Incident TB risk stratification	Risk model identified PLHIV at higher risk of developing TB than conventional screening approaches, suggesting potential to improve prioritization for testing or preventive therapy.
Kagujje et al., 2023	Adults with prior TB, Zambia	CAD-CXR for triage	Diagnostic accuracy study with subgroup analysis	False-positive referrals; subgroup performance	CAD performance differed in individuals with prior TB due to residual lung changes, highlighting implications for referral volume and threshold calibration.

Evidence linking AI deployment to programmatic response and patient-important outcomes (PIOs), including relapse-free cure, survival, health-related quality of life, and long-term functional recovery, remains limited. Pragmatic implementation studies and ongoing trials are increasingly prioritizing endpoints beyond diagnostic accuracy, including time to treatment initiation, treatment adherence, and downstream treatment outcomes [17].

Emerging evidence suggests that CAD-supported screening can improve program efficiency, but improvements in diagnostic performance alone are unlikely to translate into better patient

outcomes unless AI tools are embedded within well-functioning diagnostic and referral pathways [3,7].

### 3.3. Validation, Benchmarking, and Local Verification

Independent validation is a critical safeguard for the safe and effective deployment of CAD-CXR. Performance estimates derived from enriched datasets or vendor-curated data may overstate accuracy relative to real-world screening population, where disease prevalence is lower and presentations are more heterogeneous [4,18].

Independent benchmarking platforms provide standardized mechanism for evaluating CAD products prior to procurement and scale-up. The FIND Validation Platform assesses CAD software using de-identified chest radiographs linked to clinical, microbiological, and expert reference standards derived predominantly from low- and middle-income countries [13]. Key features of independent and reproducible validation are summarized in Box 1.

Independent benchmarking does not replace the need for local verification. WHO guidance recommends short, prospective local validation studies using routine screening populations to confirm that selected thresholds achieve predefined sensitivity targets and to quantify resulting specificity and confirmatory testing volume [4,13]. Given that AI software evolves over time, explicit version control and re-verification following substantive updates are essential to maintain performance and patient safety [4].

**Box 1.** Validation Platform and why “independent and reproducible” matters. (finddx.org).

#### **Box 1. Independent validation platforms and why reproducibility matters**

- Independent evaluation avoids bias from vendor-developed datasets
- Uses real-world chest X-rays from multiple low- and middle-income countries
- Enables fair comparison of multiple CAD products under identical conditions
- Supports disaggregated performance analysis by key subgroups
- Provides reproducible evidence for policy, procurement, and version control

### 3.4. Economics and Procurement

For TB screening interventions, economic value is determined not only by diagnostic performance, but by also increasing detection and reducing delays to diagnosis at acceptable cost [2]. CAD-enabled chest radiography influences both referral patterns and resource utilization, making confirmatory testing volume a key cost driver [4].

Economic evaluations suggest that CAD-enabled screening may be cost-effective or cost-saving in certain contexts, particularly where asymptomatic TB is prevalent and screening throughput is high [15]. However, cost-effectiveness is highly context-dependent and sensitive to TB prevalence, testing costs, and operational design [2].

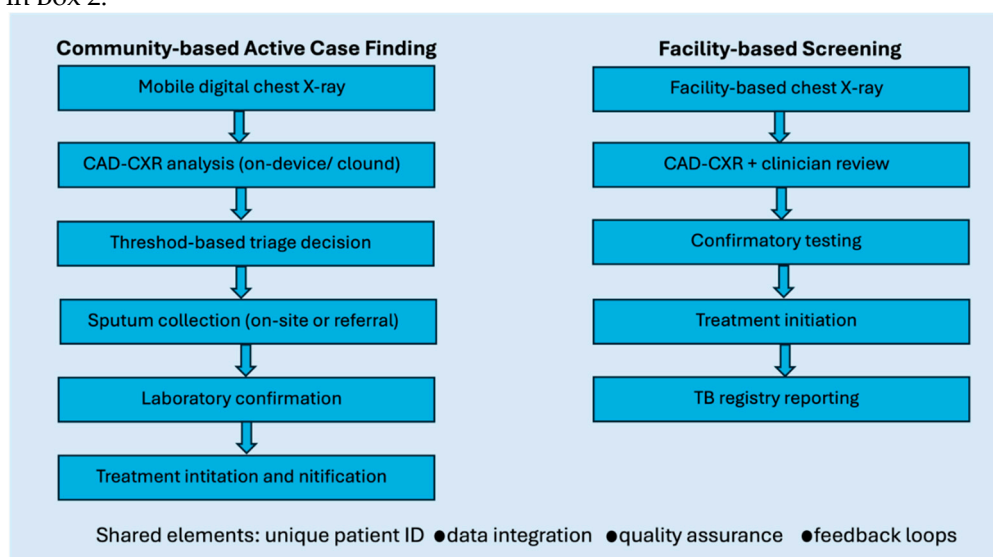
Procurement models for CAD software vary, including per-read licensing, per-device licensing, and bundled agreements. Programs should consider total cost of ownership, including hardware maintenance, software updates, training, and data governance, to ensure sustainability and value for money [2].

### 3.5. Implementation, Governance and Equity

In TB care, the real-world impact of AI depends on both diagnostic performance and how well tools are integrated into routine health system workflows. CAD-CXR must be embedded within clearly defined diagnostic pathways linking screening to confirmatory testing, and treatment initiation [2,7]. The translation of AI outputs into clinical action across screening and facility-based workflows is illustrated in Figure 2.

WHO guidance on AI in health emphasizes principles of safety, transparency, accountability, equity, and sustainability [18]. Operationalizing these principles requires human oversight at key decision points, audit trails documenting AI outputs and downstream actions, and transparent documentation of software versions and performance.

Equity considerations are central. Programs should monitor CAD performance by subgroup and adjust thresholds or workflows to mitigate unintended disparities [14,18]. Robust data governance, clear contractual arrangements, and continuous performance monitoring are essential to maintain trust and effectiveness over time. A practical governance checklist adapted for TB screening is provided in Box 2.



**Figure 2. Pixels-to-patients workflow for CAD-enabled tuberculosis screening in active case finding and facility-based settings.** The figure illustrates how computer-aided detection (CAD) for chest radiography can be integrated into end-to-end TB care pathways. In community-based active case finding and facility-based screening, CAD output informs triage decisions that trigger confirmatory testing, linkage to treatment, and reporting to national TB registries. Programmatic impact depends on effective integration of imaging, laboratory services, clinical oversight, and data systems rather than algorithm performance alone.

**Box 2.** Governance checklist from WHO principles adapted to TB screening. (World Health Organization).

**Box 2. Governance checklist for AI-enabled tuberculosis screening**

- Define intended use, target populations, and clinical role of AI
- Ensure human oversight and accountability for AI-informed decisions
- Use independently validated tools and conduct local verification
- Monitor performance across subgroups to promote equity
- Protect data privacy and define data ownership
- Plan for sustainability, version updates, and post-market surveillance

## 4. Beyond CXR: CT, Ultrasound, Cough Sound, and Digital Stethoscopes

### 4.1. AI-Assisted Computed Tomography

AI-assisted analysis of chest computed tomography (CT) images has been explored mainly for disease detection, severity assessment, and differential diagnosis in pulmonary tuberculosis. Several deep learning models have demonstrated the ability to detect active TB, quantify disease burden, and distinguish TB from other lung conditions such as malignancy or inactive TB disease. Reported correlations between AI-derived severity scores and radiologist assessments suggest that these tools may be useful in selected clinical situations, particularly where disease extent or treatment response requires more detailed characterization [11,19].

At the same time, the role of CT-based AI in routine TB screening or broader programmatic use remains limited. CT imaging is resource-intensive, involves higher radiation exposure than chest radiography, and is generally unavailable in peripheral or community-based settings where screening needs are greatest. Current evidence therefore supports CT-based AI primarily as a research-stage or referral-level diagnostic tool, with potential value in specific clinical contexts rather than as a scalable screening solution.

### 4.2. Point-of-Care Ultrasound

Point-of-care ultrasound (POCUS) has attracted interest as a TB diagnostic aid because it is portable, relatively low-cost, and does not involve ionizing radiation. AI-assisted interpretation of lung ultrasound images has been proposed as a way to reduce operator dependence and variability, potentially enabling wider use by non-specialist providers.

However, the available evidence for POCUS in pulmonary TB remains limited. A systematic review of lung ultrasound for TB diagnosis reported highly heterogeneous sensitivity estimates, very limited specificity data, and substantial risk of bias across studies, reflecting variation in reference standards, operator expertise, and image acquisition protocols [9]. At present, there is insufficient evidence to support routine programmatic use of POCUS, with or without AI, for TB screening or diagnosis. Further progress will likely require standardized acquisition protocols, clearer interpretation criteria, and prospective evaluation within defined diagnostic pathways.

### 4.3. Cough Sound Analysis

AI-based analysis of cough sounds has emerged as a novel, non-invasive approach for TB triage and, potentially, disease monitoring. Recent work has explored deep learning approaches using curated cough audio datasets linked to clinical and microbiological reference testing, supported by

wider use of mobile recording tools and improved data collection pipelines. Compared with earlier proof-of-concept studies, these efforts reflect a gradual shift toward more structured dataset development and more rigorous evaluation of model performance [12].

Despite this progress, reported diagnostic performance of cough-based AI models varies widely across studies, devices, and recording environments. Many evaluations rely on internal validation conducted by technology developers, which may increase the risk of optimistic performance estimates and limit confidence in reproducibility. In addition, a substantial proportion of studies use case-control designs, where the selection of both cases and controls may not reflect the spectrum of disease and non-disease presentations seen in routine screening populations, thereby limiting generalizability. While cough analysis may hold promise as a low-cost triage or monitoring tool, particularly in community-based contexts, current evidence supports its classification as a pilot-stage technology requiring further independent validation and prospective assessment within routine screening pathways.

#### *4.4. Digital Stethoscope and Lung Sound Analysis*

AI-assisted analysis of lung sounds captured using digital stethoscopes has also been investigated for TB detection. Experimental studies suggest that deep learning models can classify lung sounds with high accuracy under controlled conditions, potentially reducing the inter-observer variability inherent in traditional auscultation [20].

Translation of these findings into clinically meaningful decision support, however, remains challenging. Lung sound interpretation by AI systems is not yet sufficiently robust or interpretable for routine use, particularly in primary care settings where background noise, device variability, and mixed pathology are common. At present, digital stethoscope-based AI for TB should be regarded as early research-stage, with no established role in screening or diagnosis.

#### *4.5. AI-Enabled Data Analytic Tools for Tuberculosis Risk Stratification*

AI-enabled analytic models are increasingly being explored to estimate an individual's risk of developing active tuberculosis using routinely collected clinical, demographic, and laboratory data. Unlike imaging-based tools, these approaches operate at the level of patient-level data and may support decisions regarding screening intensity, confirmatory testing, or preventive therapy across different populations.

People living with HIV (PLHIV) represent one of the most extensively studied use cases for such models, given their substantially elevated TB risk even in settings with widespread antiretroviral therapy [21]. Conventional approaches to TB risk assessment in PLHIV, including symptom screening, tuberculin skin testing, and interferon-gamma release assays, have limited ability to predict incident disease and may be challenging to implement consistently at scale. In this context, cohort-based studies suggest that machine-learning models using routinely collected data may provide improved discrimination compared with standard screening tools, potentially enabling more targeted testing or preventive therapy [8].

However, PLHIV represent only one potential application. Similar risk stratification approaches could be extended to other high-risk groups, including individuals with prior TB, close contacts of infectious cases, or populations in congregate settings. Across use cases, most models have been developed and evaluated within specific datasets, and their transportability across settings, health systems, and epidemiological contexts remains uncertain.

Before routine implementation, AI-based risk models will likely require external validation, local calibration, and clearly defined clinical pathways that specify how risk scores inform action. Without these safeguards, analytic complexity may increase without corresponding gains in patient-important outcomes.

#### *4.6. AI-Assisted Interpretation of Genomic Data on TB Drug-Resistance*

Advances in whole-genome sequencing (WGS) have transformed the surveillance and management of drug-resistant TB, but interpretation of sequencing data remains analytically demanding and resource-intensive. AI and machine-learning approaches have been applied to support the identification of resistance-associated mutations, improve prediction of phenotypic resistance, and accelerate analysis pipelines for both clinical care and public health surveillance.

WHO's 2023 catalogue of mutations in *Mycobacterium tuberculosis* provides a standardized reference linking genetic variants to resistance phenotypes and underpins many current analytic tools. Building on this foundation, machine-learning models trained on large genomic datasets have shown improved accuracy in predicting resistance to first- and second-line TB drugs and in identifying rare or previously under-characterized resistance-associated mutations [10,22,23].

Despite these advances, AI-supported genomic interpretation should be viewed as decision support rather than a substitute for established laboratory and clinical expertise. Model performance depends on the quality and representativeness of training data, the completeness of mutation catalogues, and the transparency of analytic pipelines. Ongoing version control, auditability, and periodic re-evaluation remain essential, particularly as resistance patterns and reference catalogues evolve.

#### *4.7. Summary of Readiness and Programmatic Implications Across Modalities*

In contrast to CAD-enabled chest radiography, which is policy-endorsed and already in use at scale, AI applications involving CT, ultrasound, cough analysis, and digital stethoscopes remain at earlier and more heterogeneous stages of development. Although each modality addresses important gaps in TB care, the current evidence base does not support routine programmatic deployment. Clear differentiation between policy-ready tools and research-stage innovations is therefore essential to avoid misapplication and to guide responsible investment, evaluation, and future research.

Compared with imaging-based AI tools, risk models for PLHIV and AI-assisted genomic resistance interpretation address more focused, but clinically critical, decision points. Risk stratification tools may help prioritize preventive therapy and diagnostic testing among PLHIV, while genomic AI tools can enhance the speed and scale of drug-resistance surveillance and support regimen selection in specialized settings.

At present, both AI-based risk stratification models and AI-assisted genomic resistance interpretation tools should be regarded as emerging but increasingly mature applications. Although evidence supporting their technical performance and potential clinical utility is growing, broader programmatic adoption will depend on prospective evaluations demonstrating impact on patient-important outcomes, such as TB incidence, treatment success, and mortality, as well as successful integration into existing clinical and laboratory workflows.

## **5. Future Directions and Research Agenda**

Despite substantial progress in the application of AI to TB screening and diagnosis, important evidence gaps remain. Addressing these gaps will be essential if AI-enabled tools are to contribute meaningfully to TB control while avoiding unintended harms or inefficient use of limited resources.

### *5.1. Expanding Evidence to Priority Populations*

Current policy recommendations for CAD-enabled chest radiography are limited to individuals aged 15 years and older, reflecting persistent evidence gaps in children. Dedicated studies in pediatric populations are therefore a priority, with attention to age-specific radiographic patterns, disease presentation, and the suitability of available reference standards. Stronger evidence is also needed for adult subgroups in whom diagnostic performance and clinical impact may differ from general screening populations, including people living with HIV, individuals with previous TB, and older adults. In addition, future research should examine how CAD-enabled screening workflows manage

non-TB abnormalities and whether AI tools can support appropriate referral pathways for differential diagnosis of other clinically important conditions detected on chest radiography.

Future studies would benefit from routine disaggregation of results by these subgroups and from evaluating whether tailored thresholds or complementary screening strategies improve both equity and effectiveness in specific contexts.

### *5.2. Moving Beyond Accuracy to Patient- and Population-Level Important Outcomes*

Most published evaluations of AI-enabled TB tools continue to emphasize diagnostic accuracy, yet accuracy alone provides an incomplete picture of impact. Greater priority should be given to studies that assess patient- and population-level outcomes, such as time from screening to treatment initiation, treatment uptake coverage, impact on TB incidence and mortality, and cost effectiveness of interventions.

Pragmatic study designs, including cluster-randomized trials, stepped-wedge evaluations, and paired screen-positive studies, appear well suited to assessing these outcomes in real-world settings. Such approaches can capture how AI tools interact with health system capacity, diagnostic workflows, and patient behavior, generating evidence that is more directly relevant to policy and programmatic decision-making.

### *5.3. Strengthening Evidence for Emerging AI Modalities*

AI applications beyond chest radiography, including point-of-care ultrasound, cough sound analysis, digital stethoscope data, and multimodal approaches, require more rigorous and standardized evaluation. Priority research needs for these modalities include the development of standardized acquisition and interpretation protocols, independent external validation using representative datasets, and prospective studies embedded within routine screening or care pathways.

Equally important is the transparent reporting of negative or neutral findings. Without this, there is a risk of premature adoption based on incomplete evidence, as well as inefficient allocation of research and development resources.

### *5.4. Governance, Safety, and Lifecycle Evaluation*

As AI-enabled tools evolve through software updates and retraining, future research will need to address governance questions across the full product lifecycle. This includes methods for post-market surveillance, detection of performance drift, that is, unintended changes in diagnostic accuracy over time due to shifts in population characteristics, imaging practices, or software updates, assessment of subgroup bias over time, and evaluation of how humans and AI systems interact in clinical workflows.

Comparative studies of deployment models, such as cloud-based versus edge-based processing, or centralized versus decentralized governance, may also yield practical insights into trade-offs related to privacy, reliability, and sustainability across different implementation settings.

### *5.5. Integrating AI into Comprehensive TB Care*

Looking ahead, the greatest potential value of AI may lie in its integration across multiple stages of the TB care cascade rather than in isolated applications. Multimodal models that combine imaging, clinical, laboratory, and programmatic data could support earlier detection, treatment initiation and monitoring, and post-TB care. At the same time, such approaches introduce additional challenges related to data interoperability, governance, and clinical accountability. Addressing these challenges alongside technical development is essential before consideration of broader adoption.

## **6. Conclusions**

Artificial intelligence has progressed from experimental use to practical application in tuberculosis screening and diagnosis, with CAD-enabled chest radiography now supported by international policy and real-world deployments. When used appropriately, CAD-CXR can standardize interpretation, increase screening throughput, and support earlier identification of individuals who require confirmatory testing.

However, diagnostic accuracy alone is insufficient to determine real-world impact. Programmatic value depends on independent validation, local calibration, integration with diagnostic pathways, and robust governance and monitoring. Beyond chest radiography, several emerging AI applications show promising but require further independent evidence before routine use. If deployed responsibly and evaluated using patient- and population-level outcomes rather than accuracy alone, AI has the potential to contribute meaningfully to more timely, efficient, and equitable TB diagnosis.

**Author Contributions:** Conceptualization, L.N.N. and H.T.T.N.; literature review, formal analysis, visualization, figure development, and writing original draft preparation, H.T.T.N. and V.Q.L.; critical revision of the manuscript, L.N.N. and A.T.D.-X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The authors thank Dennis Falzon and Cecily Miller for carefully reviewing the manuscript and providing helpful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Storla, D.G.; Yimer, S.; Bjune, G.A. A Systematic Review of Delay in the Diagnosis and Treatment of Tuberculosis. *BMC Public Health* **2008**, *8*, 1–9, doi:10.1186/1471-2458-8-15.
2. Team, W.; Global Programme on Tuberculosis and Lung Health (GTB) *Global Tuberculosis Report 2025*; 2025;
3. Team, W.; Global Programme on Tuberculosis and Lung Health (GTB), G.R.C. *WHO Consolidated Guidelines on Tuberculosis: Module 2: Screening: Systematic Screening for Tuberculosis Disease*; 2021;
4. Team, W.; (GTB), G.P. on T. and L.H. *Use of Computer-Aided Detection Software for Tuberculosis Screening: WHO Policy Statement*; 2025;
5. 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.; et al. Using Artificial Intelligence to Read Chest Radiographs for Tuberculosis Detection: A Multi-Site Evaluation of the Diagnostic Accuracy of Three Deep Learning Systems. *Sci. Rep.* **2019**, *9*, 1–10, doi:10.1038/s41598-019-51503-3.
6. Moodley, N.; Velen, K.; Saimen, A.; Zakhura, N.; Churchyard, G.; Charalambous, S. Digital Chest Radiography Enhances Screening Efficiency for Pulmonary Tuberculosis in Primary Health Clinics in South Africa. *Clin. Infect. Dis.* **2022**, *74*, 1650–1658, doi:10.1093/cid/ciab644.
7. Subbaraman, R.; Nathavitharana, R.R.; Mayer, K.H.; Satyanarayana, S.; Chadha, V.K.; Arinaminpathy, N.; Pai, M. Constructing Care Cascades for Active Tuberculosis: A Strategy for Program Monitoring and Identifying Gaps in Quality of Care. *PLoS Med.* **2019**, *16*, 1–18, doi:10.1371/journal.pmed.1002754.
8. Bartl, L.; Zeeb, M.; Kälin, M.; Loosli, T.; Notter, J.; Furrer, H.; Hoffmann, M.; Hirsch, H.H.; Zangerle, R.; Grabmeier-Pfistershammer, K.; et al. Machine Learning-Based Prediction of Active Tuberculosis in People with HIV Using Clinical Data. *Clin. Infect. Dis.* **2025**, *81*, 521–530, doi:10.1093/cid/ciaf149.
9. Bigio, J.; Kohli, M.; Klinton, J.S.; MacLean, E.; Gore, G.; Small, P.M.; Ruhwald, M.; Weber, S.F.; Jha, S.; Pai, M. Diagnostic Accuracy of Point-of-Care Ultrasound for Pulmonary Tuberculosis: A Systematic Review. *PLoS One* **2021**, *16*, 1–14, doi:10.1371/journal.pone.0251236.
10. Bradley, P.; Gordon, N.C.; Walker, T.M.; Dunn, L.; Heys, S.; Huang, B.; Earle, S.; Pankhurst, L.J.; Anson, L.; De Cesare, M.; et al. Rapid Antibiotic-Resistance Predictions from Genome Sequence Data for *Staphylococcus Aureus* and *Mycobacterium Tuberculosis*. *Nat. Commun.* **2015**, *6*, doi:10.1038/ncomms10063.

11. Ma, L.; Wang, Y.; Guo, L.; Zhang, Y.; Wang, P.; Pei, X.; Qian, L.; Jaeger, S.; Ke, X.; Yin, X.; et al. Developing and Verifying Automatic Detection of Active Pulmonary Tuberculosis from Multi-Slice Spiral CT Images Based on Deep Learning. *J. Xray. Sci. Technol.* **2020**, *28*, 939–951, doi:10.3233/XST-200662.
12. Rajasekar, S.J.S.; Balaraman, A.R.; Balaraman, D.V.; Mohamed Ali, S.; Narasimhan, K.; Krishnasamy, N.; Perumal, V. Detection of Tuberculosis Using Cough Audio Analysis: A Deep Learning Approach with Capsule Networks. *Discov. Artif. Intell.* **2024**, *4*, doi:10.1007/s44163-024-00179-4.
13. FIND Validation Platform for AI-Based Diagnostic Evaluation. **2023**.
14. Kagujje, M.; Kerkhoff, A.D.; Nteeni, M.; Dunn, I.; Mateyo, K.; Muyoyeta, M. 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.* **2023**, *76*, E894–E901, doi:10.1093/cid/ciac679.
15. Garg, T.; John, S.; Abdulkarim, S.; Ahmed, A.D.; Kirubi, B.; Rahman, M.T.; Ubochioma, E.; Creswell, J. Implementation Costs and Cost-Effectiveness of Ultraportable Chest X-Ray with Artificial Intelligence in Active Case Finding for Tuberculosis in Nigeria. *PLOS Digit. Heal.* **2025**, *4*, 1–13, doi:10.1371/journal.pdig.0000894.
16. Velen, K.; Sathar, F.; Hoffmann, C.J.; Hausler, H.; Fononda, A.; Govender, S.; Lerefolo, M.; Govender, A.; Charalambous, S. Digital Chest X-Ray with Computer-Aided Detection for Tuberculosis Screening within Correctional Facilities. *Ann. Am. Thorac. Soc.* **2022**, *19*, 1313–1319, doi:10.1513/AnnalsATS.202103-380OC.
17. Signorell, A.; Van Heerden, A.; Ayakaka, I.; Jacobs, B.K.; Antillon, M.; Tediosi, F.; Verjans, A.; Brugger, C.; Harkare, H.V.; Labhardt, N.D.; et al. Effectiveness and Cost-Effectiveness of Community-Based TB Screening Algorithms Using Computer-Aided Detection (CAD) Technology Alone Compared with CAD Combined with Point-of-Care C Reactive Protein Testing in Lesotho and South Africa: Protocol for a Paired Screen-Positive Trial. *BMJ Open* **2025**, *15*, 1–14, doi:10.1136/bmjopen-2024-093989.
18. Team, W.; Chief Scientist and Science Division (SCI), H.E.& G. (HEG) *Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models*; 2025;
19. Yan, C.; Wang, L.; Lin, J.; Xu, J.; Zhang, T.; Qi, J.; Li, X.; Ni, W.; Wu, G.; Huang, J.; et al. A Fully Automatic Artificial Intelligence-Based CT Image Analysis System for Accurate Detection, Diagnosis, and Quantitative Severity Evaluation of Pulmonary Tuberculosis. *Eur. Radiol.* **2022**, *32*, 2188–2199, doi:10.1007/s00330-021-08365-z.
20. Mangione, S.; Nieman, L.Z. Pulmonary Auscultatory Skills during Training in Internal Medicine and Family Practice. *Am. J. Respir. Crit. Care Med.* **1999**, *159*, 1119–1124, doi:10.1164/ajrccm.159.4.9806083.
21. Lawn, D.S.D.; Wood, R.; Cock, K.M. De; Kranzer, K.; Lewis, J.J.; Gavin J Churchyard Antiretrovirals and Isoniazid Preventive Therapy in the Prevention of HIV-Associated Tuberculosis in Settings with Limited Health-Care Resources. *Lancet Infect Dis* **2010**, *10*, 489–498.
22. Jamal, S.; Khubaib, M.; Gangwar, R.; Grover, S.; Grover, A.; Hasnain, S.E. Artificial Intelligence and Machine Learning Based Prediction of Resistant and Susceptible Mutations in Mycobacterium Tuberculosis. *Sci. Rep.* **2020**, *10*, 1–16, doi:10.1038/s41598-020-62368-2.
23. Pruthi, S.S.; Billows, N.; Thorpe, J.; Campino, S.; Phelan, J.E.; Mohareb, F.; Clark, T.G. Leveraging Large-Scale Mycobacterium Tuberculosis Whole Genome Sequence Data to Characterise Drug-Resistant Mutations Using Machine Learning and Statistical Approaches. *Sci. Rep.* **2024**, *14*, 1–10, doi:10.1038/s41598-024-77947-w.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.