DOI: 10.7860/JCDR/2025/79286.21293

Original Article

Public Health Section

Diagnostic Accuracy of Artificial Intelligence-assisted Chest X-ray Interpretation Tools for Screening of Tuberculosis: A Systematic Review and Meta-analysis

RAJU SARKAR[1], MEDHA WADHWA[2], DHAVAL PARMAR[3], APURVAKUMAR PANDYA[4], DEVANG RAVAL[5], SOMEN SAHA[6], HARSH SHAH[7], KAVITHA RAJSEKAR[8]

ABSTRACT

Introduction: Tuberculosis (TB) continues to be a major worldwide health concern, which is the leading cause of death in nations like India. Despite various efforts to combat TB, effective screening and timely diagnosis remain challenging. The use of Artificial Intelligence (AI) in computer-aided interpretation of Chest X-Rays (CXRs) has demonstrated potential in improving TB detection.

Aim: To evaluate the diagnostic accuracy of AI-assisted CXR interpretation tools for TB screening.

Materials and Methods: Authors employed Systematic Review and Meta-analysis (SR/MA) using the Preferred Reporting Items for SR/MA of Diagnostic Test Accuracy (PRISMA-DTA) guideline. A comprehensive literature search was conducted for studies published between January 2019 and December 2023, focusing on AI-assisted software's diagnostic accuracy in interpreting CXRs. Electronic databases such as Google Scholar, ScienceDirect, PubMed, and Institute of Electrical and Electronics Engineers Xplore (IEEE) were used. The raw diagnostic accuracy data, sensitivity, specificity, and Area Under the Curve (AUC) of the studies that met the inclusion criteria were examined and meta analysed to estimate pooled diagnostic accuracy measures.

Results: There were 1,825 records found in the database search. Ensuing screening and duplication removal, 170 full-text publications were assessed; 14 of them satisfied the requirements for inclusion in the SR/MA. The findings of SR highlight the important role of AI assisted diagnostic tools in faster and larger screening of patients. The meta-analysis revealed the overall sensitivity of AI assisted tools to be 92% (62.9-98.7%) while specificity was 98.2% (68.4-99.9%).

Conclusion: Although the large confidence interval questions the generalisability of the findings and consistency of the results, the present review signifies the important horizon that can be explored further for strengthening the TB elimination efforts.

Keywords: Computer aided detection, Infectious diseases, Sensitivity and specificity, Tuberculosis screening

INTRODUCTION

Infectious diseases like TB continue to be major worldwide health concerns, especially in Low- And Middle-Income Countries (LMICs). Despite remarkable developments in treatment options and medical technology, TB remains a significant burden to communities and healthcare systems around the world. Reduced TB-related morbidity and mortality rates are essential in high-prevalence countries like India, where prompt treatment and early detection are critical [1,2].

With nearly 27% of the global incidence, India has the highest TB burden of all countries. In 2022, India was predicted to have accounted for 2.8 million of the 10.6 million TB cases that occurred globally each year [1]. Effective screening and timely diagnosis remain significant barriers in the fight against this treatable but persistent disease, despite concerted efforts like Universal Access Initiatives (UAI), the use of real-time information management systems, the implementation of rapid diagnostics, and standardised treatment guidelines. With its National Tuberculosis Elimination Program (NTEP), India has set ambitious goals to eliminate TB by 2025 in response to this public health emergency [2].

Unquestionably, many efforts to improve the TB care flow have sped up early identification and treatment compliance, which is vital in lowering the morbidity and death rate associated with TB [3]. However, previous studies have also revealed a range of expenses for TB patients, including direct and indirect costs [4-8]. The goal of reducing catastrophic costs to zero aligns with the End TB strategy of the World Health Organisation (WHO). Reduction in these catastrophic costs is possible by integrating the TB care cascade with different technologies. In the context of TB screening, AI-assisted CXRs interpretation has emerged as an innovative solution. Studies have demonstrated the potential of Computer Aided Detection (CAD) equipped with AI in enhancing the sensitivity of CXR for TB screening, a crucial development for resource-constrained regions like India where the ratio of radiologists to the population is low, 1:100,000 as compared to 1:10,000 in US [9]. AI algorithms are able to recognise microscopic patterns and anomalies in CXRs, which makes early detection and rapid intervention possible. Analysing radiography images for anomalies, CAD AI software could be a solution to staffing issues. AI also offers quantitative measures of lesions associated with TB, which helps track the course of the disease and assess the effectiveness of treatment. Through improving the interpretation process, AI integration into radiology workflows frees up medical staff members to concentrate on challenging patients especially in limited resource settings. Ultimately, AI-assisted solutions have the potential to revolutionise TB detection, contributing to improved patient outcomes [10]. With improved accuracy and efficiency in the difficult field of TB diagnosis, these developments mark a considerable innovation in diagnostic procedures.

The important contribution of AI-assisted technology to TB screening and diagnosis has been shown by earlier SRs. Development research and clinical studies from 2005 to 2019 were analysed in a previous SR [10] that assessed AI-based computer algorithms for Pulmonary Tuberculosis (PTB) analysis of CXRs.

The current review explored the potential of AI assisted CXR technologies as interpretation tools for screening of TB. It also examined the efficacy of point-of-care ultrasonography devices in the detection of PTB. The diagnostic accuracy of various AI, CAD, and ML-based technologies utilised in the screening and interpretation of CXRs for PTB patients from 2019 onwards is summarised in this SR. Furthermore, the meta-analysis endeavours to measure the overall precision of the encompassed investigations with the outcomes of microbiological culture.

MATERIALS AND METHODS

This SR complied with the PRISMA-DTA guidelines [11]. This protocol was registered on the International Prospective Register of Systematic Reviews (PROSPERO) with registration number CRD42024508450.

Date source and search strategy: A comprehensive search plan was developed to find the appropriate studies published between January 2019 and December 2023. Electronic databases, including Science Direct, Google Scholar, PubMed, and IEEE Xplore were searched using a blend of keywords related to TB, AI, computer aided diagnosis, CXR interpretation and diagnostic accuracy [Table/Fig-1].

Database	Keywords used with Boolean operators	Studies found
PubMed	("artificial intelligence" OR "AI") AND ("tuberculosis" OR "TB") AND ("screening")	367
Google scholar	("AI" OR "artificial intelligence" OR "machine learning") AND ("chest X-ray" OR "radiograph") AND ("tuberculosis") AND ("diagnosis" OR "detection" OR "accuracy")	448
IEEE xplore	("artificial intelligence" OR "deep learning" OR "machine learning") AND ("chest X-ray" OR CXR) AND (tuberculosis OR TB) AND ("detection" OR "screening" OR "classification")	559
Science direct	("AI" OR "machine learning" OR "deep learning") AND ("chest radiograph" OR "CXR") AND (tuberculosis OR TB) AND ("diagnostic performance" OR sensitivity OR specificity)	451
Total		1825
Total after removing duplicated records		1635

[Table/Fig-1]: Search strategy and studies identified.

Study Procedure

PICOS framework

The selection of studies was guided by the PICOS criteria:

- **Population (P):** Individuals undergoing CXR for TB screening;
- **Intervention (I):** AI-assisted software tools for interpreting CXRs;
- **Comparator (C):** Human readers, microbiological confirmation, or other reference standards;
- **Outcomes (O):** Diagnostic accuracy measures, including sensitivity, specificity, and AUC;
- **Study Design (S):** Observational studies, diagnostic accuracy studies (cross-sectional, prospective, or retrospective).

Inclusion criteria: Studies were included if they met all the following conditions:

- Assessed the diagnostic accuracy of AI-assisted CXR interpretation tools for TB screening.
- Reported raw diagnostic accuracy data, including sensitivity, specificity, or AUC.
- Involved human participants screened for active PTB using chest radiographs.
- Published in peer-reviewed journals with full-text articles available in English.
- Clearly identified the source of the study population and included independently verifiable data.

Exclusion criteria: Studies were excluded if they fell into any of the following categories:

- Case reports, editorials, reviews, conference abstracts, or articles without full text.
- Studies involving mathematical modelling, economic evaluations, or segmentation/prediction-only models of TB without full diagnostic assessment.
- Studies focusing on triage tools rather than full diagnostic workflows.
- Research without published results or that lacked independent data verification.
- Studies that did not report or identify the source of the study population.
- Non-English language publications.

Study selection process: Three independent reviewers (RS, DP, and DR) screened all titles and abstracts, followed by full-text review of eligible articles. Any disagreements or conflicts during the selection process were resolved through discussion with a fourth reviewer (MW).

Data extraction: Three reviewers (RS, DP and DR) separately extracted the data using a standardised extraction form, and one reviewer (MW) confirmed the results. From each included study, the following measures were extracted: study presences (author, publication year, country), study design, sample size, patient demographics, AI-assisted software used, the reference standard for TB diagnosis, and diagnostic accuracy measures (sensitivity, specificity, AUC for microbiological references).

Quality assessment and risk of bias: The quality of the identified studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [12]. This tool evaluates the risk of bias and applicability problems in four domains: patient selection, index test, reference standard, flow and timing. Applicability issues are ranked as low, high, or uncertain after each domain is evaluated for bias. Two reviewers (RS and MW) separately assessed the quality of included studies, and disagreements were settled through discussion. The quality of data from reproducibility studies was assessed using the following criteria [Appendix 1] in an additional area that was included based on the study done in 2018 by Mokkink LB et al., [13]. Predefined guidelines were provided for addressing the QUADAS-2 questions. The reviewers came to a consensus in order to settle disagreements.

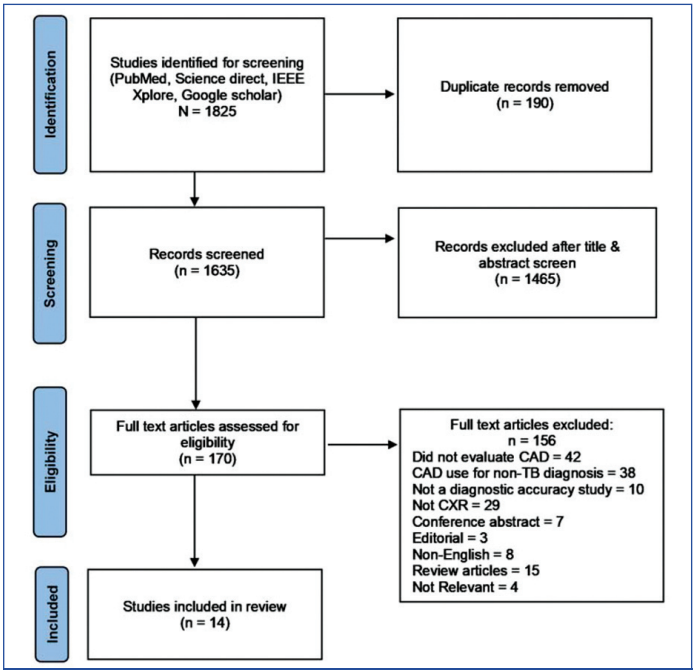
STATISTICAL ANALYSIS

Meta-analysis was conducted by comparing the number of lung anomalies (including TB) detected in comparison to the reference standard. The true positive measures were extracted from the different diagnostic measures including sensitivity, specificity, accuracy and AUC. Random-effects models were used to account for heterogeneity (i^2) between studies at 95% confidence level. Pooling of odds ratio was carried out using RevMan web software by Cochrane and diagnostic accuracies was analysed [14].

RESULTS

A total of 1,825 studies were found in the database search. [Table/Fig-2] presents studies selected as per PRISMA guidelines [11]. Total 170 full-text articles were evaluated for eligibility following the removal of duplicates and the screening of titles and abstracts. Ultimately, the SR/MA contained fourteen articles that satisfied the inclusion criteria [15-28]. Fourteen studies with a combined sensitivity of 92.0% and specificity of 98.2% were included in the meta-analysis. Significant

differences in approaches, populations, AI models, and reference standards were evident in the research' considerable heterogeneity (i^2=99%). This implies that results should be evaluated cautiously and restricts the findings' generalisability. For future research to be more consistent, standardised procedures are required.



[Table/Fig-2]: Flow diagram adapted from Page MJ et al., 2021 [11].

Features of the Study

The included studies were carried out in India, Pakistan, the Republic of Korea, Brazil, Bangladesh, Indonesia, the United States of America, and China, and were published between 2019 and 2023 [Table/Fig-3] [15-28]. The sample sizes ranged from 272 to 23,954 participants. The AI-assisted software and/or computer assisted technologies used in the studies included qXR, CAD4TB, Lunit INSIGHT, RF-HOGADM, DNN, UNet model, Xception model, CNN, and DCNN model. The reference standards for TB diagnosis included the results of microbiological culture [Table/Fig-3].

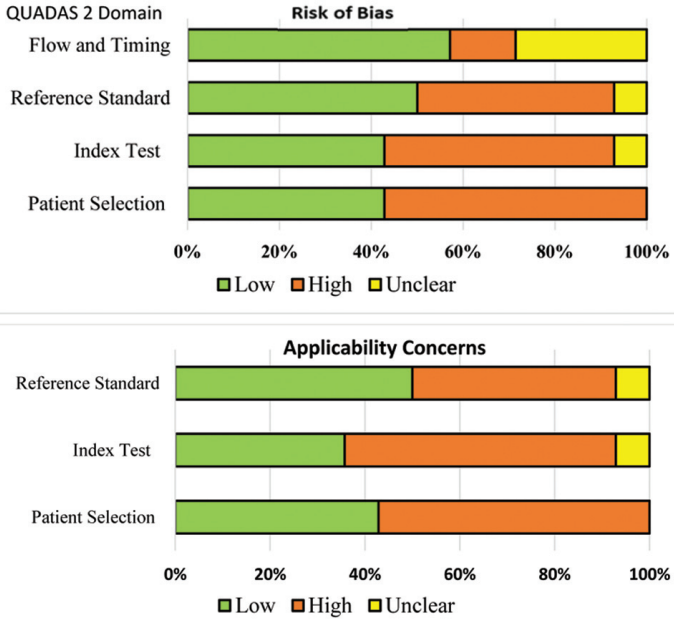
Quality Assessment

The quality assessment using the QUADAS-2 tool revealed a low to moderate risk of bias across the included studies [Table/Fig-4]. In the risk of bias, regarding patient selection, six studies [15,17,21,24,26,27] had a low-risk of bias, and eight studies [16,18-20,22,23,25,28] had a high-risk of bias. While for the index test, the risk of bias was found to be low in six studies, high in seven studies and unclear in one study. The reference standard was also found to be low in seven studies, high in six studies and unclear in one study. Regarding the applicability concern using the QUADAS-2 tool, it was seen that there were seven studies with low concerns in patient selection, eight with high concerns. Further, five studies have low concern and eight high concerns in index test and with respect to reference standard, seven studies had low applicability concerns, six had high concerns and one had unclear concerns.

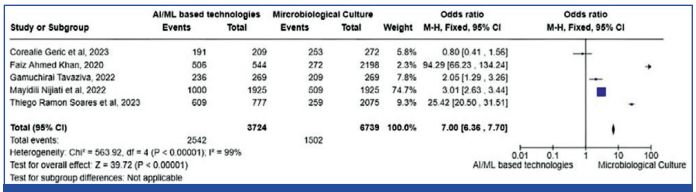
Author, publication year, [Reference]	Country where CXR completed	Study design	Imaging modality	Computer software/ model	Sample size	Reference standard	Accuracy measures	DL/ ML	Key findings
Khan FA et al., 2020 [16]	Pakistan	Prospective study on diagnostic accuracy	CXR	qXR V2, CAD4TB V6	N=2,198	Microbiological culture	Sensitivity, specificity	No/ No	The 93% sensitivity of qXRv2 and CAD4TBv6 to WHO guidelines supports its use in HIV-negative patients for TB triage.
Herman B et al., 2021 [17]	Indonesia	Cross-sectional	CXR	CUHAS ROBUST ANN Model	N=644	DST	Sensitivity, specificity, accuracy	No/ Yes	With 88% accuracy and 84% sensitivity, the CUHAS-ROBUST ANN model identified RR-TB, assisting in screening in situations when GeneXpert is not available.
Qin ZZ et al., 2021 [18]	Bangladesh	Prospective, multicentre diagnostic	CXR	CAD4TB, InferRead DR, JF CXR-1, Lunit INSIGHT CXR, qXR	N=23,954	Xpert MTB/RIF and culture	Specificity	No/ No	Outperforming radiologists, qXR and CAD4TB achieved WHO triage targets with AUCs of 90.81% and 90.34%.
Kotei E et al., 2022 [19]	USA, China	Experimental using ensemble and deep learning techniques	CXR (U-Net Segmentation, CNN Classification)	CNN	N=1,500	National Library of Medicine (NLM) dataset and Shenzhen dataset	Accuracy, AUC, sensitivity, specificity	Yes/ No	For TB detection, a stacked ensemble using CNNs and U-Net obtained 98.38% accuracy.
Mahbub MK et al., 2022 [20]	Not reported	Cross-sectional	CXR	DNN	N=15,675	Kaggle data, Known CXR data	Accuracy, AUC, Sensitivity, Specificity	Yes/ No	DNN detected TB, pneumonia, and COVID-19 from CXRs with up to 100% accuracy.
Nijiati M et al., 2022 [15]	China	Retrospective, population-based study	CXR	DCNN model	N=9,628	Microbiological culture	Accuracy, AUC, sensitivity, specificity	Yes/ No	Outperforming VGG and AlexNet, the ResNet-based AI model was able to detect PTB from CXR with an accuracy of 96.73%.
Tavaziva G et al., 2022 [22]	Pakistan	Prospective study	CXR	Lunit insight 3.1.0.0,	N=2,190	Microbiological culture	Sensitivity, specificity	No/ No	LUNIT AI showed 88.1% sensitivity and 69.9% specificity for detecting culture-confirmed TB.
Acharya V et al., 2022 [23]	India, China	Cross-sectional study using a deep learning model for TB detection from Chest X-Rays (CXR)	CXR	CNN model	N=11,200	TBX11k and Kaggle data	Accuracy, AUC, sensitivity, specificity	Yes/ No	Using NFNets and Score-CAM for AI-assisted TB identification from CXRs, the study showed up to 96.91% accuracy, proving its efficacy as a clinical assistance tool.

Bhandari M et al., 2022 [24]	Not reported	Cross-sectional	CXR	CNN	N=7,132	CXR dataset	Specificity sensitivity, Precision, F1 score, and Recall	Yes/ No	DL model with XAI assistance classified COVID-19, pneumonia, TB, and normal CXRs with 94.31% accuracy.
Ramon Soares T et al., 2023 [21]	Brazil	Cross-sectional	CXR	CAD4TB version 6, Lunit version 3.1.0.0 and qXR version 3	N=2,075	Xpert MTB/RIF and culture	AUC, Sensitivity, specificity	No/ No	All models performed poorly in inmates with a history of TB, however Lunit and qXR satisfied WHO triage criteria for TB screening in prisons with AUCs up to 0.91.
Geethamani R and Ranichitra A 2023 [25]	Not reported	Cross-sectional	CXR	RF-HOGADM	N=1,840	X-ray data	Accuracy, AUC, Sensitivity, specificity	No/ No	RF-HOG model accurately detects TB from Chest X-Rays (CXR) using HOG features and Random Forest.
Geric C et al., 2023 [26]	Pakistan	Cross-sectional study using CAD for CXR	CXR	qXR V3	N=272	Microbiological culture	Sensitivity	No/ No	Compared to CXR, CDTS-based AI CAD increased TB and pneumonia sensitivity by 5.4% and 8.7%, respectively, with higher accuracy.
Kim K et al., 2023 [27]	Republic of Korea	Comparative study of AI CAD systems	CXR, CDTS	CXR AI CAD, CDTS AI CAD	N=948	NA	Sensitivity, accuracy	Yes/ No	Compared to CXR, CDTS-based AI CAD increased the sensitivity of TB and pneumonia identification by 5.4% and 8.7%, respectively, while improving accuracy.
Sharma V et al., 2023 [28]	US, China	Cross-sectional study	CXR	UNet model, Xception model	N=1400	Kaggle - Chest X-Ray (CXR) masks and labels dataset	Accuracy	Yes/ No	Xception model detected TB in CXRs with an accuracy of 99.29% and an AUC of 0.99.

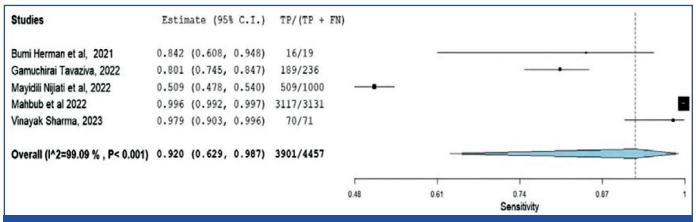
[Table/Fig-3]: The descriptive analysis involves the study approaches used [15-28].
ANN: Artificial neural network; CNN: Convolutional neural network; DCNN: Deep convolutional neural network; MTB: Mycobacterium tuberculosis; and RIF: Resistance to Rifampin; RF-HOGADM: Random forest-histogram of oriented gradients abnormality detection model; DL/ML: Deep learning/Machine learning



[Table/Fig-4]: Risk of bias summary plot using the QUADAS-2 tool.



[Table/Fig-5]: Forest plot representing development study accuracy (odds) measures.



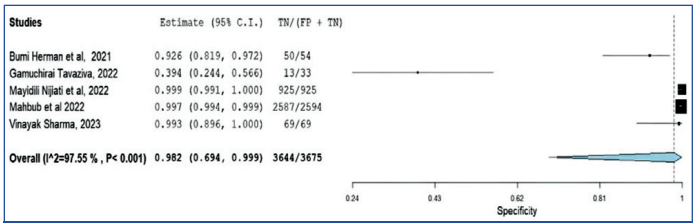
[Table/Fig-6]: Forest plot of sensitivity measures of AI based CXR technologies.

Diagnostic Accuracy

The meta-analyses for accuracy, sensitivity, and specificity comprised eight studies [15,16,19-21,24,26,28]. [Table/Fig-5] shows the study characteristics including authors, publication years, and pertinent data such as case numbers, total participants, weights, and odds ratios with confidence intervals, the analysis revealed a pooled odds ratio of 7.00 (CI 6.32, 7.70; p<0.00001) indicating that higher odds of diagnosis through the microbiological culture as compared to the AI-assisted technology. Despite the notable heterogeneity among studies, as evidenced by the high i^2 value (99%), the overall effect remains statistically significant (Z=563.92, p=<0.00001). The variability of the pooled odds ratio varies from 6.36 to 7.70.

The overall sensitivity of the included studies is shown in [Table/Fig-6]. The overall effect was found to be significant as the p-value is less than 0.001. The high i^2 value indicates high variability in included studies which could be attributed to factors such as the specific lung disease being diagnosed, co-morbidities that may have affected the pulmonary region, the type of AI algorithm employed, the quality and size of the datasets used to train the models, and variations in the reference standard methods used in different studies. The overall sensitivity of these tools was 92.0% and varies from 62.9% to 98.7%.

Similarly, in [Table/Fig-7], the overall specificity of the included studies was 98.2% varying from 69.4% to 99.9% indicating the promising diagnostic accuracy of the AI assisted diagnostic methods. The higher range of confidence interval signifies high variability and emphasises the need for further research to improve the consistency and generalisability of AI-assisted diagnostic tools with set standards.



[Table/Fig-7]: Forest plot representing specificity measures of AI based CXR technologies.

DISCUSSION

This SR is an attempt to review the different developmental studies that have aimed to assess the diagnostic accuracy of CAD aided (AI-assisted) diagnostic tools in comparison to the gold standard. The technology-based diagnostic tools aim to provide faster interpretation and screening of TB. Different studies have quoted different software and methodologies to assess the diagnostic accuracy of these technologies. A study from the Republic of Korea developed a CAD system utilising CXR and Contact Digital Tomosynthesis (CDTS) with high sensitivity and accuracy [27]. Another study from Pakistan, employed qXR V3 without deep learning methods [26]. Geethamani R and Ranichitra A utilised RF-HOGADM without deep learning techniques [25], while Khan FA et al., in Pakistan used qXR V2 and CAD4TB V6 without machine learning [16]. Sharma V et al., in the US and China used UNet and Xception models with good accuracy [28]. Other studies from various countries utilised different models and datasets with varying levels of success, indicating the potential of DL and ML methods in refining TB detection from CXRs.

In the realm of TB detection through CXRs, the effectiveness of CAD systems has been scrutinised across multiple studies [Table/Fig-8] [15-25,27,29-31]. Kim K et al., pioneered a CAD aided CXR in South Korea, showcasing a sensitivity of 72.8% and an AUC of 0.87 [27]. Geethamani R and Ranichitra A reported robust sensitivity (97.4%) and specificity (97.2%) for their CAD model [25]. While in Pakistan explored qXR and CAD4TB, yielding sensitivities of 93% for both, their specificities trailed at 75% and 69%, respectively, highlighting the variability in performance across different CAD systems [16].

Author, publication year, reference	Sensitivity (%)	Specificity (%)	AUC (%)
Khan FA et al., 2020 [16]	qXR: 93, CAD4TB: 93	qXR: 75, CAD4TB: 69	Not reported
Herman B et al., 2021 [17]	84.0	90.0	87.0
Qin ZZ et al., 2021 [18]	CAD4TB: 90 Lunit INSIGHT CXR: 90.1 qXR: 90.2	CAD4TB: 91.5 Lunit INSIGHT CXR: 88.8 qXR: 92.6	CAD4TB: 90.34 Lunit INSIGHT CXR: qXR: 90.81
Kotei E et al., 2022 [19]	AlexNet 93.20, VGG 94.2, ResNet 95.5	AlexNet 97.08, VGG 95.78, ResNet 98.05	ACC AlexNet 95.06, VGG 94.96, ResNet 96.73
Mahbub MK et al., 2022 [20]	99.6	99.9	99.76
Nijiati M et al., 2022 [15]	95.50	98.05	96.73
Tavaziva G et al., 2022 [22]	At threshold scores with Sn 15: 88.1, 30: 87.7; 45: 86.6	At threshold scores with Sn 15: 57.9; 30: 64.3; 45: 69.9	Not reported
Acharya V et al., 2022 [23]	91.81	98.42	96.91 accuracy, 99.38 AUC,
Bhandari M et al., 2022 [24]	99.53	90.37	94.95
Ramon Soares T et al., 2023 [21]	CAD4TB: 80.7 Lunit: 79.9 qXR: 74.5	CAD4TB: 82.7 Lunit: 89.8 qXR: 89.4	0.88 to 0.91.
Geethamani R and Ranichitra A 2023 [25]	97.4	97.2	97.3
Kim K et al., 2023 [27]	72.8, CXR-based AI CAD	93.4	87.4
Harris M et al., 2019 [29]	-	-	Deep-learning 0.91 Machine-learning 0.82
Bigio J et al., 2021 [30]	98.89	98.7	98.8
Anis S et al., 2020 [31]	97.3	100	99.0

[Table/Fig-8]: Measures of accuracy derived from development studies [15-25,27, 29-31].
Sn: Sensitivity; Sp: Specificity; AlexNet: Alexander network; AUC: Area under the curve; CAD: Computer-aided detection; CAD4TB: Computer-aided detection for tuberculosis; ResNet: Residual network; VGG: Visual geometry group

On the other end of the spectrum, Mahbub MK et al., achieved remarkable sensitivity (99.61%) and specificity (99.91%) with a DNN-based approach, and an accuracy of 99.76%, setting a benchmark for future comparisons [20]. Tavaziva G et al., observed the fluctuating sensitivity (ranging from 88.1% to 86.6%) and specificity (ranging from 57.9% to 69.9%) at various threshold scores for their CAD system, emphasising the importance of threshold optimisation for performance enhancement [22]. Meanwhile, various studies identified the similar performance metrics across different CAD models, suggesting comparable efficacy in detecting TB from CXRs [17,24].

The summary of findings of different studies included in the SR emphasises the variability in the performance of different CAD and AL/ML based models. This ensures that with increasing technological advancements, the future of these technologies in the screening and interpretation of TB looks promising. However, the studies included for SR encompassed different tools (AI-assisted) and compared the accuracy with microbiological, GeneXpert test or CXR, leading to high variability in the review. A meta-analysis was also carried out using odds ratios and pooled sensitivity which allowed for a comprehensive evaluation of the aggregated data from multiple studies. The meta-analysis highlights the pooled odds ratio of different studies to more than 1 (7.00), indicating that there are seven times more odds of detection of lung anomalies through microbiological culture than AI assisted CXR technologies.

Different studies have shown varying odds ratios, it was observed that an odds ratio of less than one highlighting that artificial neural networks can be used to replace screening at the primary care level where the GeneXpert is not available [16,26]. A study by Mahbub MK et al., concluded equal chance of detection of lung anomalies using AI/ML assisted technologies and gold standard as the odds ratio was 1 while a higher odds ratio was observed in other studies [15,19-21,24,28].

The pooled sensitivity provided a more robust estimate of the CAD aided AI systems overall performance in identifying TB cases as compared to the microbiological culture. The pooled sensitivity was lower than the pooled specificity (98.2% in comparison to 92%) indicating promising diagnostic methods through the use of AI-assisted tools. It also reported in a similar SR/MA that the pooled sensitivity was 94% (89-96%) and the specificity was 95% (91-97%) [15].

These results suggest that AI-assisted CXR technology could be a trustworthy medical imaging diagnosis tool for TB screening; especially within limited resource settings. However, in order to make it easier to incorporate this innovative technology into normal clinical practice, it was highlighted the critical need for uniform reporting requirements, country or geography specific standardised AI algorithm models to compare particularly for AI-specific trials and multicentre diagnostic clinical trials [32].

The findings of this SR/MA provide valuable new insights into the diagnostic accuracy of AI-assisted CXR interpretation tools for TB screening. The high odds ratio, sensitivity, and specificity show that these instruments have the potential to better patient outcomes and TB detection rates through earlier diagnosis and treatment. To maximise the application of AI in clinical practice and uncover the parameters influencing accuracy, more research is necessary, as evidenced by the variances in diagnostic performance between

studies. Additionally, the successful application of AI-assisted CXR interpretation tools in TB control initiatives depends on resolving issues with algorithm validation, tracking permission, and integration into current healthcare systems.

Implications for Future Studies

Future research through large scale, multi-centric studies should verify the diagnostic accuracy of AI-assisted CXR interpretation tools against gold standards across a range of demographics and contexts. In addition, longitudinal research is required to evaluate these technologies' long-term efficiency and financial efficacy in actual clinical settings. The Standards for Reporting Diagnostic Accuracy (STARD) and Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) are two examples of existing reporting guidelines for diagnostic accuracy studies (like any other clinical studies) that should be adapted to include AI-specific amendments, promoting standardised and transparent reporting practices for particular pathology or anomalies [14,27]. It is important to carefully evaluate and look into how imaging parameters affect imaging results and the ultimate diagnosis. Model details and code should be publicly available as part of development studies to promote research repeatability and transparency. Further studies are needed to explore the potential superiority of AI-assisted imaging tools in the early diagnosis of extra PTB or paediatric TB using Computed Tomography (CT) imaging or sonography.

Strengths and Limitations and Future Directions

The present SR/MA attempts to organise the findings of different studies that have used different CAD aided AI-assisted tools for a comprehensive diagnostic measure. The included studies, which were conducted in several regions of the world, provide insight into the many development models that are available globally following the COVID-19 pandemic. The many comparators that the included development studies used are also summarised in the report. The overall usefulness of these tools in comparison to the gold standard is indicated by the pooled diagnostic accuracy. This SR/MA have certain restrictions, despite the encouraging results.

Heterogeneity in diagnostic accuracy estimations may have resulted from differences in the included studies' sample sizes, study designs, and use of AI-assisted tools. Moreover, the results' applicability was limited because the majority of the research was conducted in different geographic locations. It is also important to remember that some model training and validation procedures used CXRs from particular datasets or places, which may have resulted in an overestimation of the diagnostic efficiency.

CONCLUSION(S)

AI-assisted software demonstrated exceptionally high pooled sensitivity and specificity, indicating that it could be useful for large-scale screening initiatives and for the diagnosis of TB in medical radiographical imaging. Nonetheless, a great deal of heterogeneity was noted among the studies, with notable differences in methodology, reporting, and design. In order to evaluate the consistency and diversity of these trials across various populations and circumstances, standardised reporting guidelines are desperately needed, especially for AI-specific trials and multicenter clinical trials. Finally, we can improve the detection of TB cases and support worldwide efforts to eradicate the disease by utilising AI technology. AI-based screening technologies have the potential to sustain community screening programs. These affordable and practical alternatives are an important part of the integrated TB care delivery process, especially for those with limited resources.

Data availability: We have not gathered primary data. All data generated or analysed during the review are synthesised in this published article and additional information related to the study is provided in supplementary files.

REFERENCES

[1] Global Tuberculosis Report 2023, Geneva: World Health Organization; 2023 [cited 2025 Apr 5]. Available from: https://iris.who.int/bitstream/handle/10665/373828/9789240083851-eng.pdf?sequence=1.

[2] National TB Elimination Programme India TB Report 2024, New Delhi: Central TB Division, Ministry of Health and Family Welfare; 2024. Available from: http://www.tbcindia.gov.in.

[3] Muniyandi M, Rajeswari R, Rani B. Costs to patients with tuberculosis treated under dots programme. Indian J of Tuberc. 2005;52:188-96. Available from: https://dhr.gov.in/sites/default/files/COSTS%20TO%20PATIENTS%20WITH%20TUBERCULOSIS%20TREATED.pdf.

[4] Ananthakrishnan R, Muniyandi M, Jeyaraj A, Palani G, Sathiyasekaran BWC. Expenditure pattern for TB treatment among patients registered in an urban government DOTS program in Chennai City, South India. Tuberc Res Treat. 2012;2012:747924. Doi: 10.1155/2012/747924.

[5] Tanimura T, Jaramillo E, Weil D, Raviglione M, Lönnroth K. Financial burden for tuberculosis patients in low- and middle-income countries: A systematic review. Eur Respir J. 2014;43(6):1763-75. Doi: 10.1183/09031936.00193413.

[6] Muniyandi M, Lavanya J, Karikalan N, Saravanan B, Senthil S, Selvaraju S, et al. Estimating TB diagnostic costs incurred under the National Tuberculosis Elimination Programme: A costing study from Tamil Nadu, South India. Int Health. 2021;13(6):536-44.

[7] Rupani MP, Cattamanchi A, Shete PB, Vollmer WM, Basu S, Dave JD. Costs incurred by patients with drug-susceptible pulmonary tuberculosis in semi-urban and rural settings of Western India. Infect Dis Poverty. 2020;9(1):144.

[8] Sarin R, Vohra V, Singla N, Thomas B, Krishnan R, Muniyandi M. Identifying costs contributing to catastrophic expenditure among TB patients registered under RNTCP in Delhi metro city in India. Indian J Tuberc. 2019;66(1):150-57.

[9] Kalyanpur A. Commentary - radiology in India: The next decade. Indian J Radiol Imaging. 2008;18(3):191-92. Doi: 10.4103/0971-3026.41869.

[10] Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, Macaskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): Explanation, elaboration, and checklist. BMJ. 2020;370:m2632. Doi: 10.1136/bmj.m2632.

[11] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. Doi: 10.1136/bmj.n71.

[12] Whiting PF. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36. Doi: 10.7326/0003-4819-155-8-201110180-00009.

[13] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171-79.

[14] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. Doi: 10.1136/bmj.h5527.

[15] Nijiati M, Ma J, Hu C, Tuersun A, Abulizi A, Kelimu A, et al. Artificial intelligence assisting the early detection of active pulmonary tuberculosis from chest X-rays: A population-based study. Front Mol Biosci. 2022;9:874475. Doi: 10.3389/fmolb.2022.874475.

[16] Khan FA, Majidulla A, Tavaziva G, Nazish A, Abidi SK, Benedetti A, 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.

[17] Herman B, Sirichokchatchawan W, Pongpanich S, Nantasenamat C. Development and performance of CUHASROBUST application for pulmonary rifampicin-resistance tuberculosis screening in Indonesia. PLoS One. 2021;16(3):e0249243. Doi.org/10.1371/journal.pone.0249243.

[18] Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, 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.

[19] Kotei E, Thirunavukarasu R. Ensemble technique coupled with deep transfer learning framework for automatic detection of tuberculosis from chest X-ray radiographs. Healthcare (Switzerland). 2022;10(11):2335. Doi: 10.3390/healthcare10112335.

[20] Mahbub MK, Biswas M, Gaur L, Alenezi F, Santosh KC. Deep features to detect pulmonary abnormalities in chest X-rays due to infectious diseaseX: Covid-19, pneumonia, and tuberculosis. Inf Sci (N Y). 2022;592:389-401. Doi: 10.1016/j.ins.2022.01.062.

[21] Ramon Soares T, Dias de Oliveira R, Liu YE, da Silva Santos A, Cesar Pereira dos Santos P, Ravena Soares Monte L, 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. 2023;(17):100388. Doi: 10.1016/j.lana.2022.100388.

[22] Tavaziva G, Majidulla A, Nazish A, Saeed S, Benedetti A, Khan AJ, et al. Diagnostic accuracy of a commercially available, deep learning-based chest X-ray interpretation software for detecting culture-confirmed pulmonary tuberculosis. Int J Infect Dis. 2022;122:15-20. Doi: 10.1016/j.ijid.2022.05.037.

[23] Acharya V, Dhiman G, Prakasha K, Bahadur P, Choraria A, Sushobhitha M, et al. AI-Assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model. Comput Intell Neurosci. 2022;2022(1):2399428. Doi: 10.1155/2022/2399428.

[24] Bhandari M, Shahi TB, Siku B, Neupane A. Explanatory classification of CXR images into COVID-19, pneumonia and tuberculosis using deep learning and XAI. Comput Biol Med. 2022;150:106156. Doi: 10.1016/j.compbiomed.2022.106156.

[25] Geethamani R, Ranichitra A. Enhancing tuberculosis detection: Leveraging RF-HOG model for automated diagnosis from chest X-ray images. Procedia Computer Science. 2023;230:21-32. Doi: 10.1016/j.procs.2023.12.057.
[26] Geric C, Majidulla A, Tavaziva G, Nazish A, Saeed S, Benedetti A, et al. Artificial intelligence-reported chest X-ray findings of culture-confirmed pulmonary tuberculosis in people with and without diabetes. J Clin Tuberc Other Mycobact Dis. 2023;31:100365. Doi: 10.1016/j.jctube.2023.100365. PMID: 37095759.
[27] Kim K, Lee JH, Je Oh S, Chung MJ. AI-based computer-aided diagnostic system of chest digital tomography synthesis: Demonstrating comparative advantage with X-ray-based AI systems. Comput Methods Programs Biomed. 2023;240:107643. Doi: 10.1016/j.cmpb.2023.107643.
[28] Sharma V, Nillmani, Gupta SK, Shukla KK. Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. Intelligent Medicine. 2023;4(2):104-13. Doi.org/10.1016/j.imed.2023.06.001.
[29] Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, 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(9):e0221339. Doi: 10.1371/journal.pone.0221339. PMID: 31479448.
[30] Bigio J, Kohli M, Klinton JS, MacLean E, Gore G, Small PM, et al. Diagnostic accuracy of point-of-care ultrasound for pulmonary tuberculosis: A systematic review. PLoS One. 2021;16(5):e0251236. Doi: 10.1371/journal.pone.0251236. PMID: 3396163.
[31] Anis S, Lai KW, Chuah JH, Ali SM, Mohafez H, Hadizadeh M, et al. An overview of deep learning approaches in chest radiograph. IEEE Access. 2020;8:182347-54. Doi.org/10.1109/ACCESS.2020.3028390.
[32] Zhan Y, Wang Y, Zhang W, Ying B, Wang C. Diagnostic accuracy of the artificial intelligence methods in medical imaging for pulmonary tuberculosis: A systematic review and meta-analysis. J Clin Med. 2022;12(1):303. Doi: 10.3390/jcm12010303. PMID: 36615102.

PARTICULARS OF CONTRIBUTORS:
1. Scientist II, Department of Health Technology Assessment (HTA)-Resource Hub, Indian Institute of Public Health Gandhinagar, Gujarat, India.
2. Assistant Professor, Department of Health Policy, Management and Behavioural Science, Indian Institute of Public Health Gandhinagar, Gujarat, India.
3. Senior Health Economist, Department of Health Technology Assessment (HTA)-Resource Hub, Indian Institute of Public Health Gandhinagar, Gujarat, India.
4. Assistant Professor, Department of Health Policy, Management and Behavioural Science, Indian Institute of Public Health Gandhinagar, Gujarat, India.
5. Scientist II, Department of Health Technology Assessment (HTA)-Resource Hub, Indian Institute of Public Health Gandhinagar, Gujarat, India.
6. Professor, Department of Public Health Programme and HTA Resource Hub, Indian Institute of Public Health Gandhinagar, Gujarat, India.
7. Assistant Professor, Department of Public Health Science, Indian Institute of Public Health Gandhinagar, Gujarat, India.
8. Scientist-F, ICMR, Department of Health Research Ministry of Health and Family Welfare, Government of India, New Delhi, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Medha Wadhwa,
Indian Institute of Public Health Gandhinagar, Gujarat, India.
E-mail: mwadhwa@iiphg.org

PLAGIARISM CHECKING METHODS: [Jain H et al.]
- Plagiarism X-checker: Mar 27, 2025
- Manual Googling: May 26, 2025
- iThenticate Software: May 28, 2025 (9%)

ETYMOLOGY: Author Origin

EMENDATIONS: 6

AUTHOR DECLARATION:
- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? Yes
- Was informed consent obtained from the subjects involved in the study? NA
- For any images presented appropriate consent has been obtained from the subjects. NA

Date of Submission: **Mar 12, 2025**
Date of Peer Review: **Apr 03, 2025**
Date of Acceptance: **May 30, 2025**
Date of Publishing: **Aug 01, 2025**

APPENDIX 1

Additional reproducibility domain for quality assessment

1. Was the time interval between the repeated tests appropriate? (Yes/No/Unclear)
2. Were the test conditions similar for the repeated tests (type of administration, environment, instructions)? (Yes/No/Unclear)
3. Was a Kappa score calculated? (Yes/No/Unclear).

Like the risk of bias questions in QUADAS-2, the answer to the question “Could the reproducibility data be biased?” (Low-risk/High-risk/Unclear-risk) was based on the answers to the three reproducibility criteria questions.

Based on the work of Mokkink LB et al., 2018 [13].