



A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging

Zhi-Lin Han^{1#^}, Yu-Yang Zhang^{2#^}, Jian Li^{3#}, Shan Gao^{4#}, Wei Liu⁵, Wan-Jie Yang^{6,7^}, Zhi-Heng Xing^{1,2,6,7^}

¹Department of Radiology, Haihe Hospital, Tianjin University, Tianjin, China; ²Haihe Clinical School, Tianjin Medical University, Tianjin, China;

³Artificial Intelligence Branch of the Commerce Statistics Chinese Society of China Business, Beijing, China; ⁴Department of Nanchang Zhongke Institute of Mobile Medical Imaging, Nanchang, China; ⁵Standardization Professional Branch of China Anti Tuberculosis Association, Beijing, China; ⁶Tianjin Institute of Respiratory Diseases, Tianjin, China; ⁷TCM Key Research Laboratory for Infectious Disease Prevention for State Administration of Traditional Chinese Medicine, Tianjin, China

Contributions: (I) Conception and design: ZL Han, YY Zhang; (II) Administrative support: WJ Yang, ZH Xing; (III) Provision of study materials or patients: J Li, W Liu; (IV) Collection and assembly of data: S Gao; (V) Data analysis and interpretation: ZL Han, YY Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Wan-Jie Yang, BM. Tianjin Institute of Respiratory Diseases, Tianjin, China; TCM Key Research Laboratory for Infectious Disease Prevention for State Administration of Traditional Chinese Medicine, No. 890 Jingu Road, Shuanggang Town, Jinnan District, Tianjin 300350, China. Email: yangwanjie0709@126.com; Zhi-Heng Xing, PhD. Department of Radiology, Haihe Hospital, Tianjin University, Tianjin, China; Haihe Clinical School, Tianjin Medical University, Tianjin, China; Tianjin Institute of Respiratory Diseases, Tianjin, China; TCM Key Research Laboratory for Infectious Disease Prevention for State Administration of Traditional Chinese Medicine, No. 890 Jingu Road, Shuanggang Town, Jinnan District, Tianjin 300350, China. Email: 18920696025@189.cn.

Background: Pulmonary tuberculosis (PTB) remains a global public health challenge, with 10.8 million new cases reported in 2023. Early diagnosis is crucial for controlling its spread, yet traditional sputum-based tests face limitations in turnaround time and resource availability. Chest X-ray (CXR) is a cost-effective diagnostic tool, but its use in high-tuberculosis (TB) burden regions is restricted by a shortage of radiologists. Artificial intelligence (AI)-based computer-aided detection (CAD) systems, leveraging deep learning, offer a promising solution for automated PTB detection. However, variability in diagnostic performance across AI tools and the need for scenario-specific threshold adjustments remain challenges that need to be addressed. Our meta-analysis evaluated the diagnostic accuracy of five AI-based PTB detection products, aiming to provide insights for advancing AI applications in TB screening and diagnosis.

Methods: The PubMed, Embase, Web of Science, and Cochrane Library databases were searched for literature related to CXR diagnosis of TB based on AI technology published from the establishment day of the database to December 19, 2024. The keywords were “artificial intelligence”, “tuberculosis”, “chest X-ray”, and “diagnosis”. The literature search, screening, data extraction, quality evaluation, and bias risk assessment were conducted independently by two researchers, and Stata 17.0 software (StataCorp) was used to process and analyze the data.

Results: A total of 5,651 references were retrieved, and 21 references were finally selected according to the inclusion and exclusion criteria. The meta-analysis included five software solutions for CXR analysis: JF CXR-1 (JF Healthcare, Nanchang, China), qXR (Qure.ai, Mumbai, India), Lunit INSIGHT CXR (Lunit, Seoul, South Korea), CAD4TB (Delft Imaging, ‘s-Hertogenbosch, Netherlands), and InferRead DR Chest (Infervision, Beijing, China). Their sensitivity and specificity were as follows: JF CXR-1, 86.0% and 80.0%; qXR, 90.0% and 64.0%; Lunit INSIGHT CXR, 90.0% and 63.0%; CAD4TB, 91.0% and 60.0%; InferRead

[^] ORCID: Zhi-Lin Han, 0009-0000-3363-6134; Yu-Yang Zhang, 0009-0003-6036-1027; Wan-Jie Yang, 0009-0009-5592-145X; Zhi-Heng Xing, 0000-0003-2695-9183.

DR Chest, 89.0% and 59.0%.

Conclusions: AI software has demonstrated excellent diagnostic performance in assisting the CXR diagnosis of TB and can help clinicians to make rapid and accurate decisions in screening and treating patients with TB.

Keywords: Artificial intelligence (AI); tuberculosis (TB); chest X-ray (CXR); diagnosis; meta-analysis

Submitted Mar 26, 2025. Accepted for publication May 16, 2025. Published online May 27, 2025.

doi: 10.21037/jtd-2025-604

View this article at: <https://dx.doi.org/10.21037/jtd-2025-604>

Introduction

Pulmonary tuberculosis (PTB), caused by *Mycobacterium tuberculosis* (MTB), has re-emerged as the leading cause of death from a single infectious agent globally, claiming 1.25 million lives in 2023 (1). According to the “Global

tuberculosis report 2024”, there were 10.8 million new tuberculosis (TB) cases worldwide in 2023, with an incidence rate of 134 per 100,000 population (2). For comparison, in 2014, the estimated global incidence was around 9.6 million cases. In response to the persistent burden, the World Health Organization (WHO) launched the End TB Strategy in 2014, setting ambitious targets for 2035: a 90% reduction in TB incidence, a 95% reduction in TB deaths compared to 2015 levels, and elimination of catastrophic costs due to TB. The core indicator of this strategy is to reduce the TB incidence rate to below 10 per 100,000 population (3). Therefore, early screening and diagnosis of TB remain essential for effective control of the spread of MTB and the timely treatment of TB (4).

At present, the diagnosis of most PTB relies on laboratory tests for MTB in sputum samples, such as acid-fast bacilli smear test and *Mycobacterium* culture (5). However, due to the extensive time required for laboratory examination and the lack of laboratory infrastructure in some countries and regions, the diagnostic efficiency in clinical practice is considerably restricted (6). Chest X-ray (CXR) is a widely used clinical TB detection tool with high health economic benefits (7). However, in those countries with limited medical resources and high TB burden, there is often a lack of radiologists who can accurately diagnose PTB via CXR, which may lead to the limited application of CXR (8). Interpretation of CXRs for TB diagnosis is subject to considerable variability among human readers. Inter-observer and intra-observer agreement rates are often moderate, as shown in a study by Zellweger *et al.*, which reported substantial discrepancies in the radiological assessment of TB even among trained clinicians (9).

With the rapid development of artificial intelligence (AI) technology, especially the application of deep learning in medical image analysis, AI-based computer-aided detection (CAD) system has provided a novel means to

Highlight box

Key findings

- All five artificial intelligence (AI) products in 21 clinical studies examined had excellent diagnostic performance.
- The latest versions of AI software significantly outperformed earlier versions due to advancements in algorithm development.

What is known and what is new?

- AI-assisted diagnostic tools have shown potential in supporting the detection of pulmonary tuberculosis (PTB). Previous studies have indicated that AI software can achieve performance comparable to that of experienced radiologists.
- This comprehensive meta-analysis highlights the heterogeneity in diagnostic performance across different types and versions of AI software and underscores the improvements in diagnostic accuracy of newer versions due to the integration of advanced deep learning techniques. However, limitations in meeting the World Health Organization (WHO)-recommended diagnostic standards suggest the need for further algorithm optimization and threshold adjustments.

What is the implication, and what should change now?

- AI software holds great promise for rapid PTB screening and diagnosis, particularly in high-tuberculosis (TB) burden settings. However, none of the evaluated software can yet meet the WHO's diagnostic criteria (90% sensitivity and 70% specificity).
- The following initiatives should be pursued: further optimization of AI algorithms to improve specificity without compromising sensitivity; tailoring of threshold settings based on specific clinical scenarios and patient populations to enhance diagnostic applicability; completion of additional clinical studies in a diversity setting, including low-TB burden countries and pediatric populations, to ensure generalizability and robust evaluation.

diagnosing PTB. AI software can independently analyze and output its analysis results in the form of probability, thus realizing the automatic CXR-based diagnosis of PTB (10). Recently, several studies have demonstrated the potential of advanced deep learning architectures in improving diagnostic accuracy. For instance, Maruthai *et al.* proposed a multi-axis transformer-based U-net with a class-balanced ensemble model for lung disease classification using X-ray images, achieving high accuracy in detecting various lung diseases (11). Similarly, Babu *et al.* developed a lung disease classification method using optimal cross stage partial bidirectional long short-term memory, which showed promising results in classifying lung diseases from CXR images (12). Additionally, Visu *et al.* introduced an enhanced Swin transformer-based TB classification with segmentation using CXR, which further improved the accuracy of TB detection (13). The potential of AI in other diseases is also substantial (14). These studies collectively underscore the broad applicability and ongoing advancements in AI techniques for medical diagnostics.

In recent years, AI software designed to detect PTB has rapidly advanced and emerged in the global market. As of May 2024, 16 relatively complete CAD products have been applied to clinical imaging diagnosis, with more being under development (15). The WHO has updated the international TB screening policy and recommended the use of AI software as a diagnostic aid for radiologists in CXR screening and the triage of patients with PTB aged 15 years and older, with the required diagnostic performance consisting of a 90% sensitivity and a 70% specificity (16). However, a number of studies have shown that there are differences in the diagnostic results of PTB across different AI software applications; moreover, each application needs to be adjusted according to the threshold for PTB diagnosis in different clinical contexts and patient groups to obtain the optimal diagnostic accuracy (17). The threshold is the artificially set standard critical point score for determining the output result of a presence or absence of PTB. The systematic review by Zhan *et al.* indicates that many studies have used the diagnostic results of human readers as the reference standard, and these results are often higher than those obtained using microbiological criteria (such as sputum culture) (18). This discrepancy suggests that AI software has certain limitations and uncertainties in the diagnosis of PTB.

In view of this, we evaluated the performance of five AI products designed to diagnose PTB using a meta-analysis, with the aim of providing a comprehensive

summary and data support for future research in the field. We present this article in accordance with the PRISMA-DTA reporting checklist (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-2025-604/rc>), and the study was registered in the International Prospective Register of Systematic Reviews (PROSPERO) with the number CRD420250632836.

Methods

Search strategy

The PubMed, Embase, Web of Science, and Cochrane Library databases were searched for studies on the diagnosis of TB via AI software published from the establishment of the databases to December 19, 2024. There were no limitations regarding country or region. The search terms were “artificial intelligence”, “tuberculosis”, “chest X-ray”, and “diagnosis”.

Inclusion and exclusion criteria

The inclusion criteria for the literature were as follows: (I) any clinical study on AI-based software for the diagnosis of TB; (II) studies using microbiological reference standards or composite reference standards that include microbiological confirmation as a primary component; and (III) studies providing data on original diagnostic accuracy, such as sensitivity, specificity, true positive values (TPs), false positive values (FPs), false negative values (FNs), and true negative values (TNs), through which a four-fold table could be directly or indirectly established.

Meanwhile, the exclusion criteria were as follows: (I) reviews, meta-analyses, dissertations, case reports, and conference abstracts; (II) duplicate literature or data; (III) unpublished studies; and (IV) research on PTB complicated with other diseases.

Screening and data extraction

Two researchers (Z.L.H. and Y.Y.Z.) determined whether each paper met the requirements of the established inclusion and exclusion criteria by independently reviewing the title and abstract. For references that passed the initial screening, the researcher read the full text for further screening, and if there was disagreement, a third researcher (S.G.) was consulted to ensure that the reference was eligible for inclusion in the final analysis. The study data

and information were extracted independently from the included studies using standardized extraction tables. Data extracted for this study included first author's name, study country, year of publication, sample size, diagnostic criteria, AI software and version, and diagnostic accuracy. If the included studies obtained multiple sets of diagnostic results corresponding to the threshold through multiple threshold settings for the same software (the diagnostic results and diagnostic accuracy data changed accordingly to the change in threshold setting), the receiver operating characteristic (ROC) curve of each study was used to determine the extracted specificity threshold score (19) (a threshold of 90% sensitivity was obtained, which is the desired sensitivity for the TB triage test outlined by the WHO). If the sensitivity was less than 90%, the threshold score with the highest sum of sensitivity and specificity was selected (17).

Quality assessment

RevMan version 5.4 (Cochrane, London, UK) was used to evaluate the research quality of the relevant papers based on the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) independently by two researchers (Y.Y.Z. and J.L.). QUADAS-2 includes two aspects: risk of bias and applicability concerns. The risk of bias includes four parts: patient selection, index test, reference standard, and flow and timing. Applicability concerns cover the first three parts. The Cochrane tool contains 11 items. If all relevant items were evaluated as yes, the study was rated as low risk; if any item was evaluated as no, the study was rated as high risk; and if the article had insufficient information to make a definite assessment, it was rated as unclear.

Statistical methods

Stata 17.0 software (StataCorp, College Station, TX, USA) was used for data processing and analysis. TP, FP, FN, and TN were extracted from each study to establish a four-fold table for diagnostic meta-analysis. First, the heterogeneity of the included studies, which was derived from threshold or non-threshold effects, was evaluated. (I) For the threshold effect, the Spearman correlation coefficient between the logarithm of sensitivity and $(1 - \text{specificity})$ was calculated. Here, the "threshold effect" refers to the variation in diagnostic accuracy that arises when different studies use different cut-off values for the same diagnostic test. A higher cut-off value may increase specificity but decrease sensitivity, and vice versa. The Spearman correlation coefficient is used

to assess the monotonic relationship between sensitivity and $(1 - \text{specificity})$. A significant correlation (P value < 0.05) suggests that the variation in diagnostic accuracy across studies is due to the use of different cut-off values, indicating a threshold effect. In this case, the analysis results can be directly described by calculating the area under the curve (AUC), as the AUC is not affected by the choice of cut-off values. (II) For the non-threshold effect, if there was no threshold effect, heterogeneity was further assessed by the I^2 statistic and Cochran's Q test. If $I^2 > 50\%$ and $P < 0.05$, significant heterogeneity between studies was considered to be present, and a fixed-effects model was used. The sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, and area AUC of the included studies were calculated according to the selected model. To further isolate the sources of heterogeneity, subgroup analyses were performed according to different versions of the AI software. Finally, the Deek's funnel plot was used to assess publication bias.

Results

Literature screening process and retrieval

A total of 5,651 articles were identified from databases, with 1,238 identified as duplicates. After screening, 2,944 articles were excluded based on their titles and abstracts. Of the 1,349 reports we attempted to retrieve, 32 could not be obtained. After further reviewing the full texts of the 1,317 reports that were retrieved, only 21 met our criteria for inclusion in the study, as shown in *Figure 1*.

Summary of the basic characteristics and results of the included articles

A total of 21 clinical articles were included. Five CAD products were analyzed, including JF CXR-1 (JF Healthcare, Nanchang, China), qXR (Qure.ai, Mumbai, India), Lunit INSIGHT CXR (Lunit, Seoul, South Korea), CAD4TB (Delft Imaging, 's-Hertogenbosch, Netherlands), and InferRead DR Chest (Infervision, Beijing, China). The studies were from 15 countries, including Bangladesh, Brazil, Cameroon, China, India, Malawi, Nepal, Pakistan, the Philippines, South Africa, Tanzania, the United Kingdom, the United States, Vietnam, and Zambia. There were a few differences in the specific diagnostic criteria used across the studies. Ten studies used two or more composite diagnostic criteria. Nearly three-quarters of the studies

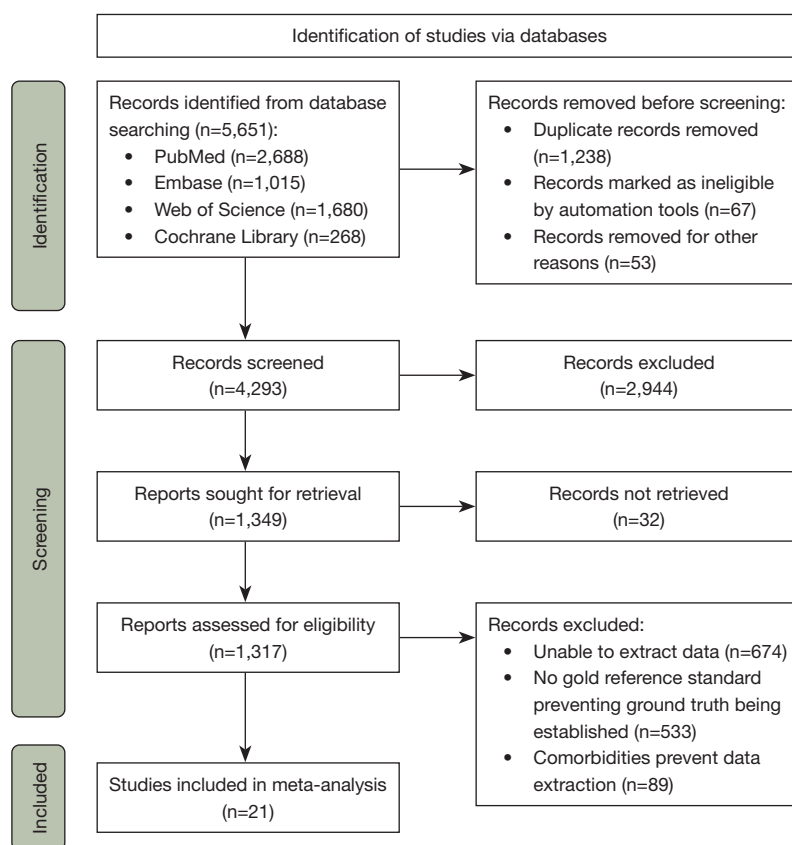


Figure 1 Flowchart of literature selection.

were conducted in 2020 or later, and this growth in research recent years reflects the increase in public interest in the field of AI. *Table 1* shows the basic characteristics of the included articles.

Assessment of study quality

For patient selection, all studies were assigned low bias and high applicability scores due to their prospective design and continuous enrollment. For the index test, the diagnostic results of the AI output were independent of prior knowledge of the microbiological test results, and the critical thresholds were prespecified. Consequently, 12 clinical studies were assigned a low risk of bias and high applicability score. However, six studies were assigned a high risk of bias and low applicability score, primarily due to unclear patient selection, lack of blinding, or insufficient reporting. In some of these studies, diagnostic thresholds were determined post hoc based on study data, which may influence the interpretation of sensitivity and specificity

estimates. The risk of bias was uncertain in three studies. For the reference standard, all studies were determined to have low bias and high applicability scores because the results of the reference standard were obtained via the analysis and testing of sputum samples when the diagnostic results of the AI software were unknown. In terms of flow and timing, due to the inconsistency of diagnostic criteria, 12 clinical studies were determined to have low bias and high applicability scores, and nine studies had a high or uncertain risk of bias. The QUADAS-2 risk-of-bias results are shown in *Table 2* and *Figure 2*.

Test of heterogeneity

- (I) Threshold effects: the Spearman correlation coefficients of JF CXR-1, qXR, Lunit INSIGHT CXR, CAD4TB, and InferRead DR Chest were 0.40 ($P=0.60$), 0.40 ($P=0.19$), 0.43 ($P=0.34$), -0.009 ($P=0.97$), and -0.500 ($P=0.667$), respectively, indicating that there was no significant threshold effect.

Table 1 The basic characteristics of the included articles

Case	Year	Author	Country	AI software & version	Diagnostic criteria	Sample size	Diagnostic metrics
1	2014	Breuninger <i>et al.</i> (20)	Tanzania	CAD4TB (v3)	MTB culture, AFB smear	861	Sn, Sp, PPV, NPV
2	2016	Melendez <i>et al.</i> (21)	South Africa	CAD4TB (v3)	Xpert MTB/RIF	392	AUC, Sn, Sp, PPV, NPV
3	2017	Rahman <i>et al.</i> (22)	Bangladesh	CAD4TB (v3)	Xpert MTB/RIF	17,066	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
4	2018	Zaidi <i>et al.</i> (23)	Pakistan	CAD4TB (v3)	Xpert MTB/RIF	6,090	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
5	2018	Melendez <i>et al.</i> (24)	United Kingdom	CAD4TB (v5)	MTB culture, clinical diagnosis	38,961	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
6	2019	Qin <i>et al.</i> (25)	Nepal, Cameroon	Lunit INSIGHT CXR (v4.7.2), qXR (v2), CAD4TB(v6)	Xpert MTB/RIF	1,196	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
7	2019	Philipsen <i>et al.</i> (26)	Philippines	CAD4TB (v5)	Xpert MTB/RIF	10,755	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
8	2020	Khan <i>et al.</i> (27)	Pakistan	qXR (v2), CAD4TB (v6)	MTB culture	2,198	ACC, Sn, Sp, PPV, NPV
9	2020	Nash <i>et al.</i> (28)	India	qXR (v2)	Xpert MTB/RIF, MTB culture, AFB smear	929	AUC, Sn, Sp
10	2020	Murphy <i>et al.</i> (29)	Pakistan	CAD4TB (v3, v4, v5, v6)	Xpert MTB/RIF	5,565	TP, FP, TN, FN, AUC, Sn, Sp
11	2021	Qin <i>et al.</i> (30)	Bangladesh	Lunit INSIGHT CXR (v4.9.0), qXR (v3), CAD4TB (v7), InferRead DR (v2), JF CXR-1 (v2)	Xpert MTB/RIF	23,954	AUC, Sn, Sp
12	2021	Codlin <i>et al.</i> (31)	Vietnam	Lunit INSIGHT CXR (v3.1.0), qXR (v3), CAD4TB (v7), InferRead DR Chest (v1), JF CXR-1 (v3.0)	Xpert MTB/RIF	1,032	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
13	2021	Twabi <i>et al.</i> (32)	Malawi	qXR (v2)	Xpert MTB/RIF, MTB culture	774	TP, FP, TN, FN, AUC, Sn, Sp
14	2022	Tavaziva <i>et al.</i> (33)	Pakistan	Lunit INSIGHT CXR (v3.1.0)	Xpert MTB/RIF, MTB culture	2,190	TP, FP, TN, FN, AUC, ACC, Sn, Sp
15	2022	Qin <i>et al.</i> (17)	Bangladesh	qXR (v2, v3), CAD4TB (v6, v7)	Xpert MTB/RIF	12,890	AUC, Sn, Sp
16	2022	Liao <i>et al.</i> (34)	China	JF CXR-1 (v2)	Xpert MTB/RIF, MTB culture, AFB smear, clinical diagnosis	2,543	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
17	2023	Gelaw <i>et al.</i> (35)	United States	Lunit INSIGHT CXR (v4.9.0), qXR (v2), CAD4TB (v6)	Xpert MTB/RIF, MTB culture	1,769	AUC, Sn, Sp
18	2023	Soares <i>et al.</i> (36)	Brazil	Lunit INSIGHT CXR (v3.1.0), qXR (v3), CAD4TB (v6)	Xpert MTB/RIF, MTB culture	7,081	AUC, Sn, Sp, PPV, NPV
19	2023	Kagujje <i>et al.</i> (37)	Zambia	qXR (v3), CAD4TB (v7)	Xpert MTB/RIF	1,432	TP, FP, TN, FN, AUC, Sn, Sp
20	2023	Yang <i>et al.</i> (38)	China	JF CXR-1 (v2)	Bacteriological confirmation, clinical diagnosis	1,121	TP, FP, TN, FN, AUC, ACC, Sn, Sp, PPV, NPV
21	2024	Qin <i>et al.</i> (39)	South Africa	Lunit INSIGHT CXR (v4.9), qXR (v3), CAD4TB (v7), InferRead DR Chest (v1), JF CXR-1 (v2)	Xpert MTB/RIF, MTB culture	774	AUC, Sn, Sp

ACC, accuracy rate; AFB, acid-fast bacilli; AI, artificial intelligence; AUC, area under the curve; FN, false negative value; FP, false positive value; MTB, *Mycobacterium tuberculosis*; NPV, negative predictive value; PPV, positive predictive value; RIF, rifampicin; Sn, sensitivity; Sp, specificity; TN, true negative value; TP, true positive value.

Table 2 QUADAS-2 risk assessment results of the included studies

First author	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Breuninger <i>et al.</i> (20)	Low	High	Low	High	Low	Low	Low
Melendez <i>et al.</i> (21)	Low	Low	Low	Unclear	Low	Low	Low
Rahman <i>et al.</i> (22)	Low	High	Low	Low	Low	High	Low
Zaidi <i>et al.</i> (23)	Low	High	Low	Unclear	Low	High	Low
Melendez <i>et al.</i> (24)	Low	Unclear	Low	Low	Low	Low	Low
Qin <i>et al.</i> (25)	Low	High	Low	Low	Low	High	Low
Philipsen <i>et al.</i> (26)	Low	Low	Low	High	Low	Low	Low
Khan <i>et al.</i> (27)	Low	Low	Low	Low	Low	Low	Low
Nash <i>et al.</i> (28)	Low	Unclear	Low	Low	Low	Low	Low
Murphy <i>et al.</i> (29)	Low	Low	Low	Unclear	Low	Low	Low
Qin <i>et al.</i> (30)	Low	High	Low	Low	Low	Low	Low
Codlin <i>et al.</i> (31)	Low	Unclear	Low	Low	Low	Low	Low
Twabi <i>et al.</i> (32)	Low	Low	Low	Low	Low	Low	Low
Tavaziva <i>et al.</i> (33)	Low	Low	Low	Low	Low	Low	Low
Qin <i>et al.</i> (17)	Low	Low	Low	Unclear	Low	Low	Low
Liao <i>et al.</i> (34)	Low	High	Low	Low	Low	High	Low
Gelaw <i>et al.</i> (35)	Low	Low	Low	Unclear	Low	Low	Low
Soares <i>et al.</i> (36)	Low	Low	Low	High	Low	Low	Low
Kaguije <i>et al.</i> (37)	Low	Low	Low	Low	Low	Low	Low
Yang <i>et al.</i> (38)	Low	Low	Low	Unclear	Low	Low	Low
Qin <i>et al.</i> (39)	Low	Low	Low	Low	Low	Low	Low

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

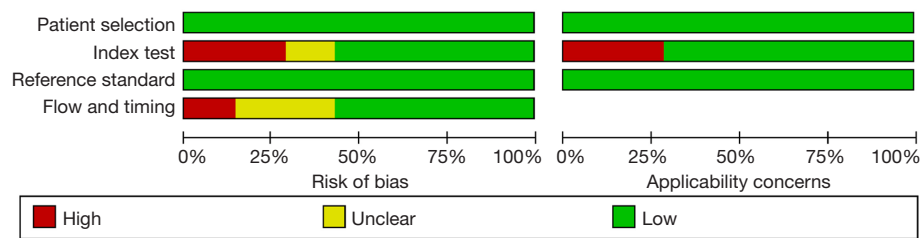


Figure 2 Bar graph of risk assessment of the QUADAS-2 scale for the included articles. QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2.

(II) Non-threshold effects: according to the I^2 statistics and the results of Cochran's Q test of sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and

diagnostic odds ratio of all five types of AI software, the fixed-effects model was only used for the combination of Lunit INSIGHT CXR as this produced a lower

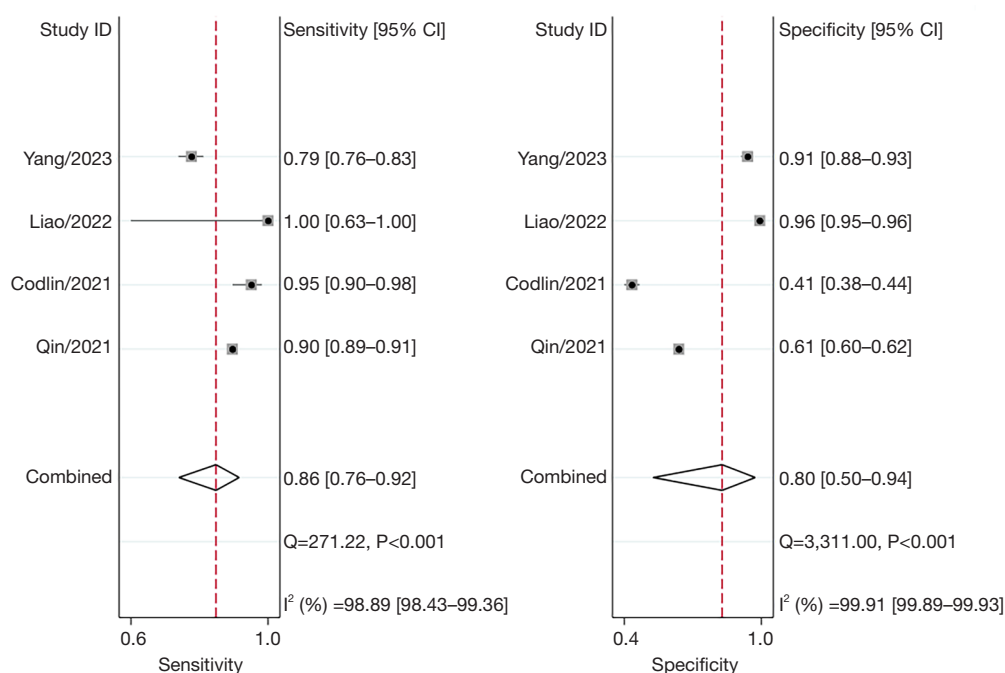


Figure 3 The combined sensitivity and specificity of JF CXR-1 in the diagnosis of PTB. CI, confidence interval; PTB, pulmonary tuberculosis.

heterogeneity in sensitivity, while the random-effects model was used for the other combinations.

Combined analysis results

The combined sensitivity and specificity of the AI software were 86.0–91.0% and 59.0–80.0%, respectively. The sensitivity and specificity for the five software applications, respectively, were as follows: JF CXR-1, 86.0% [95% confidence interval (CI): 76.0–92.0%] and 80.0% (95% CI: 50.0–94.0%); qXR, 90.0% (95% CI: 87.0–92.0%) and 64.0% (95% CI: 55.0–73.0%); Lunit INSIGHT CXR, 90.0% (95% CI: 89.0–91.0%) and 63.0% (95% CI: 51.0–74.0%); CAD4TB, 91.0% (95% CI: 90.0–92.0%) and 60.0% (95% CI: 53.0–66.0%); and InferRead DR Chest, 89.0% (95% CI: 86.0–91.0%) and 59.0% (95% CI: 55.0–62.0%) (Figures 3–7).

Subgroup analyses

Subgroup analyses were performed for different versions of each AI software application. At least two studies were included in the meta-analysis, and some studies were excluded from relevant subgroup analyses due to missing

information. Therefore, only three AI software applications including qXR (v2 and v3), Lunit INSIGHT CXR (v3.1.0 and v4.9.0), and CAD4TB (v3, v5, v6, and v7), were included. The diagnostic outcomes for all five AI software applications included in the study, as well as the subgroup analysis results for the aforementioned three software applications, are summarized in Table 3.

Publication bias

Publication bias was assessed using Deek's funnel plot. The scatter points of each study were distributed symmetrically on the left and right sides of the midline. There was no significant publication bias in the included studies on JF CXR-1 ($P=0.25$), qXR, ($P=0.32$), Lunit INSIGHT CXR ($P=0.90$), CAD4TB ($P=0.84$), and InferRead DR Chest ($P=0.24$).

Discussion

The early screening and diagnosis of PTB are critical to the prevention and control of TB (40). We conducted a meta-analysis evaluating the efficacy of five AI software applications (JF CXR-1, qXR, Lunit INSIGHT CXR,

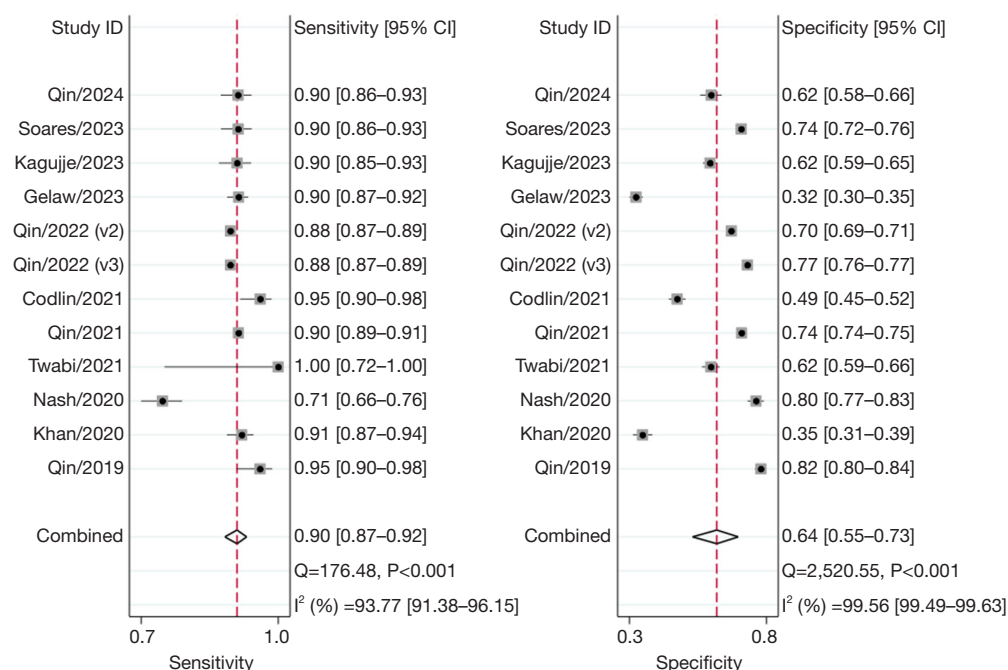


Figure 4 The combined sensitivity and specificity of qXR in the diagnosis of PTB. CI, confidence interval; PTB, pulmonary tuberculosis.

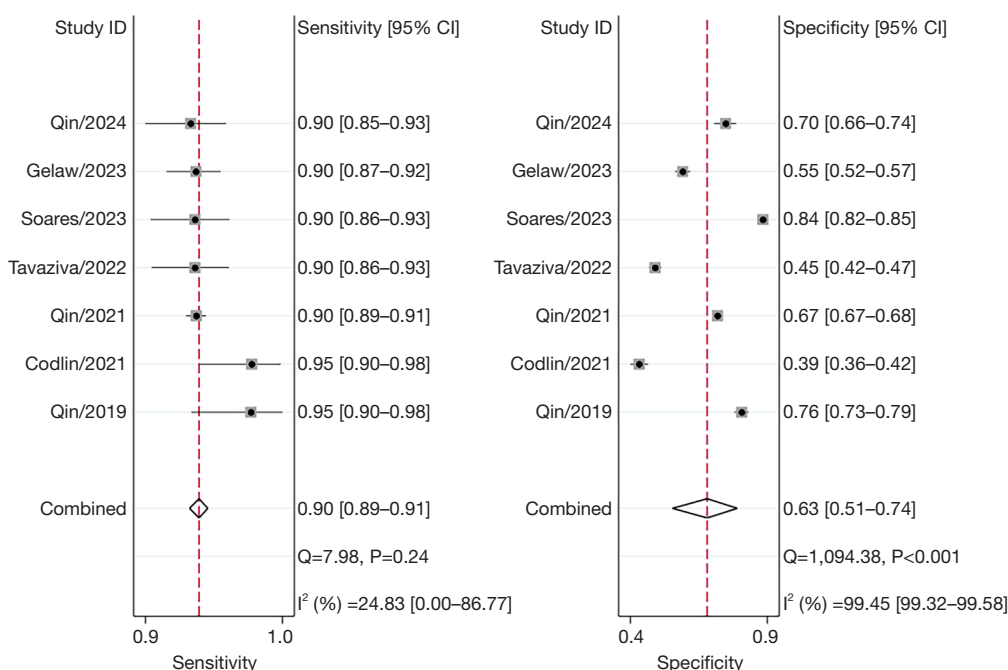


Figure 5 The combined sensitivity and specificity of Lunit INSIGHT CXR in the diagnosis of PTB. CI, confidence interval; PTB, pulmonary tuberculosis.

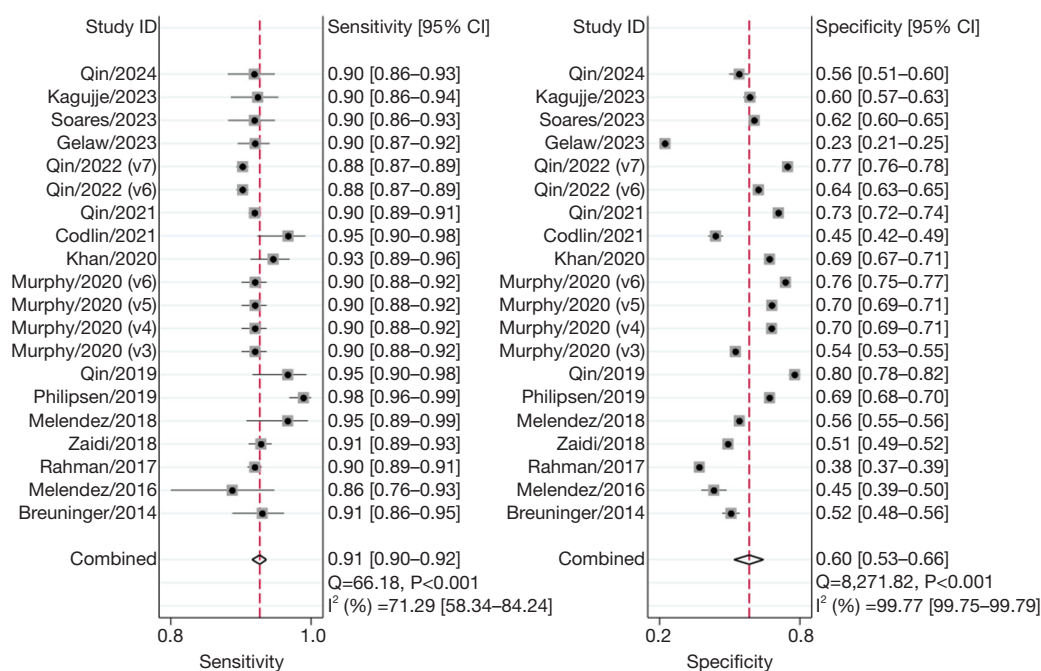


Figure 6 The combined sensitivity and specificity of CAD4TB in the diagnosis of PTB. CI, confidence interval; PTB, pulmonary tuberculosis.

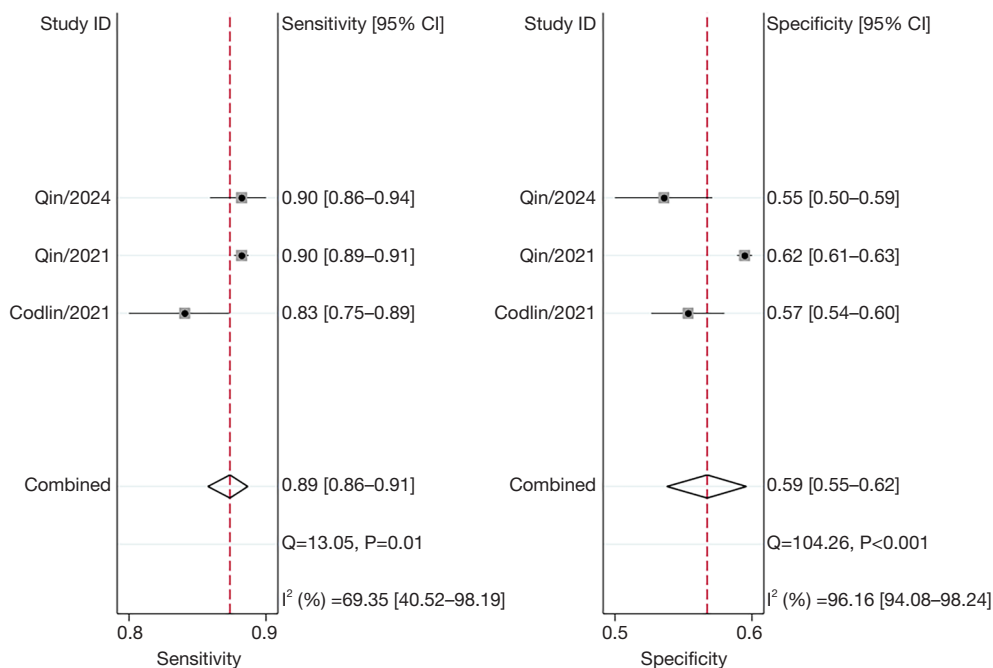


Figure 7 The combined sensitivity and specificity of InferRead DR Chest in the diagnosis of PTB. CI, confidence interval; PTB, pulmonary tuberculosis.

Table 3 Diagnostic accuracy test evaluation index of the included studies

Software	Sensitivity (95% CI), %	Specificity (95% CI), %	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Diagnostic odds ratio [95% CI]	AUC (95% CI)
JF CXR-1 (all)	86.0 (76.0–92.0)	80.0 (50.0–94.0)	4.2 (1.6–11.4)	0.18 (0.13–0.24)	24 [12–49]	0.90 (0.87–0.92)
qXR (all)	90.0 (87.0–92.0)	64.0 (55.0–73.0)	2.5 (2.0–3.2)	0.16 (0.12–0.21)	16 [10–24]	0.89 (0.86–0.91)
v2	89.0 (82.0–94.0)	62.0 (44.0–77.0)	2.3 (1.5–3.6)	0.18 (0.10–0.30)	13 [6–30]	0.87 (0.84–0.90)
v3	91.0 (89.0–93.0)	67.0 (59.0–74.0)	2.8 (2.2–3.4)	0.14 (0.12–0.16)	20 [15–26]	0.91 (0.88–0.93)
Lunit INSIGHT CXR (all)	90.0 (89.0–91.0)	63.0 (51.0–74.0)	2.5 (1.8–3.4)	0.15 (0.13–0.18)	16 [10–27]	0.91 (0.88–0.93)
v3.1.0	91.0 (88.0–93.0)	63.0 (32.0–80.0)	2.2 (1.2–4.0)	0.15 (0.10–0.23)	14 [6–37]	0.91 (0.89–0.94)
v4.9.0	90.0 (89.0–91.0)	64.0 (56.0–71.0)	2.5 (2.0–3.1)	0.15 (0.13–0.18)	16 [12–23]	0.90 (0.88–0.93)
CAD4TB (all)	91.0 (90.0–93.0)	60.0 (53.0–66.0)	2.3 (1.9–2.7)	0.14 (0.12–0.18)	16 [11–22]	0.90 (0.87–0.93)
v3	90.0 (89.0–91.0)	48.0 (43.0–53.0)	1.7 (1.6–1.9)	0.20 (0.17–0.24)	9 [7–11]	0.90 (0.87–0.92)
v5	96.0 (93.0–98.0)	62.0 (55.0–68.0)	2.5 (2.1–3.0)	0.07 (0.03–0.13)	37 [16–86]	0.86 (0.83–0.89)
v6	90.0 (88.0–92.0)	63.0 (47.0–77.0)	2.4 (1.6–3.8)	0.15 (0.11–0.22)	16 [8–34]	0.91 (0.88–0.93)
v7	91.0 (89.0–93.0)	63.0 (52.0–73.0)	2.5 (1.9–3.2)	0.14 (0.12–0.16)	18 [13–24]	0.91 (0.88–0.93)
InferRead DR Chest (all)	89.0 (86.0–91.0)	59.0 (55.0–62.0)	2.1 (1.9–2.4)	0.19 (0.15–0.25)	11 [8–16]	0.82 (0.79–0.85)

AUC, area under the curve; CI, confidence interval.

CAD4TB, and InferRead DR Chest) for diagnosing PTB based on CXR images, which included 21 clinical studies. The results showed that all five AI products exhibited high sensitivity and modest specificity. These findings indicate that AI software applications are capable of accurately identifying PTB from CXR images, potentially reducing the time required for diagnosis and improving the efficiency of TB screening programs, particularly in settings where access to radiologists or advanced diagnostic equipment is limited. However, the modest specificity observed in some AI applications suggests that additional microbiological tests may be necessary to minimize false positives. Considering these findings, AI software can serve as a valuable tool integrated into clinical practice for the rapid diagnosis of PTB. The findings of this study are similar to those of the systematic review and meta-analysis conducted by Hua *et al.* (10). Qin *et al.* also reported that the accuracy of CAD in detecting PTB was significantly better than that of some experienced radiologists (30). In addition, in a nationwide systematic TB screening conducted in Papua New Guinea, CAD4TB was able to complete 98.6% of the CXR diagnosis work, essentially replacing traditional manual screening (41). Therefore, faster and more accurate judgments in the screening and diagnosis of PTB can be facilitated by these AI products, thereby significantly improving overall

diagnostic efficiency and clinical outcomes.

However, in our study, significant heterogeneity was found in the sensitivity and specificity across these five types of AI software, which may be due to the different models and frameworks employed, as well as the various versions of the same software. The results of subgroup analysis showed that the more recent versions of qXR and CAD4TB perform significantly better than their earlier versions. For example, early versions of CAD4TB (e.g., v3) had a sensitivity of 90.0%, which is commendable for identifying individuals with PTB. However, the specificity was relatively low at 48.0%, indicating a higher rate of false positives. In contrast, the more recent version of CAD4TB (v7) achieved a sensitivity of 91.0% and a specificity of 63.0%, demonstrating a notable improvement in diagnostic accuracy. This demonstrates that as AI software evolves through version updates, it achieves enhanced sensitivity and specificity, leading to improved diagnostic performance. In addition, Murphy *et al.* reported that compared with versions 3, 4, and 5, CAD4TB v6 had significantly improved accuracy and higher sensitivity, along with a lower use price (29). Qin *et al.* comprehensively evaluated the software performance of the latest versions of two AI software applications (qXR and CAD4TB), reporting a sensitivity of 90.0%; moreover, they found that the newer versions

of qXR and CAD4TB had a significantly improved ability to detect PTB as compared with the earlier versions (30). These performances may be related to the use of different types of classifiers and algorithm upgrades in AI software. For example, CAD4TB v3, an earlier version, uses the k-nearest neighbors algorithm, which performs well in some cases but still has certain limitations when encountering high-dimensional medical images (10). With the development of deep learning technology, the newer version of CAD4TB uses a convolutional neural network to detect complex and small image abnormalities in CXR through automatic feature extraction technology, thus performing well in computer vision tasks (42). The latest version of Lunit INSIGHT CXR also uses advanced deep learning technology to further improve the accuracy of PTB detection (21). Therefore, although the diagnostic accuracy of different AI software applications and different versions of the same software vary, the newer versions possess marked advantages over the old ones in diagnostic performance.

After meta-analysis correction, although the more recent AI software demonstrated high accuracy in detecting PTB, none of them met the diagnostic criteria recommended by the WHO. At a sensitivity of 90.0%, the specificity of qXR, Lunit INSIGHT CXR, and CAD4TB was 64.0%, 63.0%, and 60.0%, respectively. This finding underscores the need for continued optimization of AI algorithms and threshold settings to ensure stable and reliable diagnostic performance in diverse settings. Appropriate threshold setting can significantly improve the detection rate of PTB, especially in settings with high TB burden or a lack of laboratory infrastructure, which is important for rapid diagnosis and control of PTB transmission (39). However, the threshold setting may be influenced by a variety of factors. It should be adjusted according to specific clinical needs and the prevalence of the disease. For instance, during the process of determining the best threshold CAD4TB v7, a community-based study in South Africa set the threshold at 20 in order to approximate the output of AI software to the radiologist's 81.0% sensitivity and 57.0% specificity (43). In contrast, in a study of nearly 24,000 participants in Bangladesh, in order to reach the standard recommended by the WHO, researchers adjusted the threshold to 50 (30). Of note, the study participants were mostly patients with symptomatic TB who voluntarily attended or were referred to a TB screening center. These studies highlight the importance of tailoring thresholds to specific clinical needs and disease prevalence in different application scenarios and patient

populations. In clinical application, medical institutions need to dynamically adjust and optimize the threshold according to the characteristics of different populations and clinical needs so as to maximize the ability of AI software in actual clinical work and further improve the diagnostic accuracy.

This study was subject to several limitations. First, the volume of literature in this field is limited, and some software versions were not included in the analysis, which might have reduced the stability and universality of the results and not fully reflect the application of the software. Second, the research samples mostly focused on adult patients, and there was a lack of evaluation of children; thus, the diagnostic accuracy of AI software in patients of different age groups remains relatively unclear. Third, while the included studies employed different reference standards—such as microbiological, clinical, or composite criteria—these diagnostic standards are all widely recognized and accepted in clinical practice for identifying active TB. Although this variation may have introduced some heterogeneity in sensitivity and specificity estimates, we believe that the validity and diagnostic accuracy of these reference standards are sufficient to justify their inclusion in the pooled analysis. We also performed subgroup analysis where feasible and acknowledged the variation in *Table 1*. Thus, we consider that the impact on the overall findings is limited, though future studies may benefit from standardized reference definitions to enhance comparability. Finally, the bulk of the studies were conducted in countries with a high burden of TB, which may limit the generalizability of the findings in countries with a low burden. Given these limitations, future studies should focus on addressing these issues. More comprehensive literature reviews and inclusion of additional software versions are needed to enhance the stability and universality of the results.

Conclusions

This meta-analysis highlights the considerable potential of AI software in diagnosing PTB, particularly in large-scale screening, as demonstrated by its excellent diagnostic performance. These findings underscore the value of AI as a diagnostic aid in resource-limited and high-TB burden settings.

However, the current evidence base remains limited due to the small number of studies and variability in application scenarios and patient populations. More clinical studies are needed to evaluate AI performance across various contexts

and to address specific challenges, such as optimizing algorithms and adjusting diagnostic thresholds for different populations.

Future research should focus on enhancing AI algorithms and refining threshold settings to ensure reliable and adaptable diagnostic performance. These efforts will be critical for maximizing the utility of AI software in TB management and supporting strategic decisions to combat TB on a global scale.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the PRISMA-DTA reporting checklist. Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-2025-604/rc>

Peer Review File: Available at <https://jtd.amegroups.com/article/view/10.21037/jtd-2025-604/prf>

Funding: This work was supported by the Tianjin Major Science and Technology Projects and Engineering Projects (to W.J.Y.) (No. 24ZXKJGX00070), the Tianjin Science and Technology Plan Project (to W.J.Y.) (No. 23JCZDJC00970), the Tianjin Health Science and Technology Project (to W.J.Y.) (No. TJWJ2024ZD009), and the Tianjin Second Batch of Health Industry High-Level Talent Selection and Training Project (to Z.H.X.) (No. TJSJMYXYC-D2-012).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://jtd.amegroups.com/article/view/10.21037/jtd-2025-604/coif>). W.J.Y. reports funding support from the Tianjin Major Science and Technology Projects and Engineering Projects (No. 24ZXKJGX00070), the Tianjin Science and Technology Plan Project (No. 23JCZDJC00970), and the Tianjin Health Science and Technology Project (No. TJWJ2024ZD009). Z.H.X. reports funding support from the Tianjin Second Batch of Health Industry High-Level Talent Selection and Training Project (No. TJSJMYXYC-D2-012). The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related

to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Goletti D, Meintjes G, Andrade BB, et al. Insights from the 2024 WHO Global Tuberculosis Report - More Comprehensive Action, Innovation, and Investments required for achieving WHO End TB goals. *Int J Infect Dis* 2025;150:107325.
- World Health Organization. Global tuberculosis report 2024. Geneva: World Health Organization; 2024.
- World Health Organization. The End TB Strategy. Geneva: World Health Organization; 2015.
- Ding C, Ji Z, Zheng L, et al. Population-based active screening strategy contributes to the prevention and control of tuberculosis. *Zhejiang Da Xue Xue Bao Yi Xue Ban* 2022;51:669-78.
- Lewinsohn DM, Leonard MK, LoBue PA, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. *Clin Infect Dis* 2017;64:e1-e33.
- Chen X, Hu TY. Strategies for advanced personalized tuberculosis diagnosis: Current technologies and clinical approaches. *Precis Clin Med* 2021;4:35-44.
- Hwang EJ, Park S, Jin KN, et al. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clin Infect Dis* 2019;69:739-47.
- Pande T, Pai M, Khan FA, et al. Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J* 2015;46:1816-9.
- Zellweger JP, Heinzer R, Touray M, et al. Intra-observer and overall agreement in the radiological assessment of tuberculosis. *Int J Tuberc Lung Dis* 2006;10:1123-6.
- Hua D, Nguyen K, Petrino N, et al. Benchmarking the diagnostic test accuracy of certified AI products

- for screening pulmonary tuberculosis in digital chest radiographs: Preliminary evidence from a rapid review and meta-analysis. *Int J Med Inform* 2023;177:105159.
11. Maruthai S, Thanarajan T, Ramesh T, et al. Multi-axis transformer based U-Net with class balanced ensemble model for lung disease classification using X-ray images. *J Xray Sci Technol* 2025;33:540-52.
 12. Babu T, Sam Kumar GV, Kartheesan L, et al. Lung disease classification in chest X-ray images using optimal cross stage partial bidirectional long short term memory. *J Xray Sci Technol* 2025;33:501-15.
 13. Visu P, Sathiya V, Ajitha P, et al. Enhanced swin transformer based tuberculosis classification with segmentation using chest X-ray. *J Xray Sci Technol* 2025;33:167-86.
 14. Balasubramani J, Surendran R. Advanced Deep Learning Framework for Diagnosing Autism Spectrum Disorder Through Facial Expression Analysis. In: 2025 Fifth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE; 2025:1-7.
 15. Stop TB Partnership and FIND. AI products for tuberculosis healthcare. Accessed May 9, 2024. Available online: <https://www.ai4hlth.org/>
 16. World Health Organization. WHO consolidated guidelines on tuberculosis: Module 2: Screening: Systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021.
 17. Qin ZZ, Barrett R, Ahmed S, et al. Comparing different versions of computer-aided detection products when reading chest X-rays for tuberculosis. *PLOS Digit Health* 2022;1:e0000067.
 18. Zhan Y, Wang Y, Zhang 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;12:303.
 19. Tavaziva G, Harris M, Abidi SK, 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* 2022;74:1390-400.
 20. Breuninger M, van Ginneken B, Philipsen RH, et al. Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. *PLoS One* 2014;9:e106381.
 21. Melendez J, Sánchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep* 2016;6:25265.
 22. Rahman MT, Codlin AJ, Rahman MM, et al. An evaluation of automated chest radiography reading software for tuberculosis screening among public- and private-sector patients. *Eur Respir J* 2017;49:1602159.
 23. Zaidi SMA, Habib SS, Van Ginneken B, et al. Evaluation of the diagnostic accuracy of Computer-Aided Detection of tuberculosis on Chest radiography among private sector patients in Pakistan. *Sci Rep* 2018;8:12339.
 24. Melendez J, Hogeweg L, Sánchez CI, et al. Accuracy of an automated system for tuberculosis detection on chest radiographs in high-risk screening. *Int J Tuberc Lung Dis* 2018;22:567-71.
 25. Qin ZZ, Sander MS, Rai B, 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;9:15000.
 26. Philipsen RHHM, Sánchez CI, Melendez J, et al. Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: a cohort study. *Int J Tuberc Lung Dis* 2019;23:805-10.
 27. 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:e573-81.
 28. Nash M, Kadavigere R, Andrade J, et al. Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India. *Sci Rep* 2020;10:210.
 29. Murphy K, Habib SS, Zaidi SMA, et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Sci Rep* 2020;10:5492.
 30. Qin ZZ, Ahmed S, Sarker MS, 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:e543-54.
 31. Codlin AJ, Dao TP, Vo LNQ, et al. Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. *Sci Rep* 2021;11:23895.
 32. Twabi HH, Semphere R, Mukoka M, et al. Pattern of abnormalities amongst chest X-rays of adults undergoing computer-assisted digital chest X-ray screening for tuberculosis in Peri-Urban Blantyre, Malawi: A cross-sectional study. *Trop Med Int Health* 2021;26:1427-37.
 33. Tavaziva G, Majidulla A, Nazish A, 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.
34. Liao Q, Feng H, Li Y, et al. Evaluation of an artificial intelligence (AI) system to detect tuberculosis on chest X-ray at a pilot active screening project in Guangdong, China in 2019. *J Xray Sci Technol* 2022;30:221-30.
 35. Gelaw SM, Kik SV, Ruhwald M, et al. Diagnostic accuracy of three computer-aided detection systems for detecting pulmonary tuberculosis on chest radiography when used for screening: Analysis of an international, multicenter migrants screening study. *PLOS Glob Public Health* 2023;3:e0000402.
 36. Soares TR, Oliveira RD, Liu YE, 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.
 37. Kagujje M, Kerkhoff AD, Nteeni M, 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* 2023;76:e894-901.
 38. Yang Y, Xia L, Liu P, et al. A prospective multicenter clinical research study validating the effectiveness and safety of a chest X-ray-based pulmonary tuberculosis screening software JF CXR-1 built on a convolutional neural network algorithm. *Front Med (Lausanne)* 2023;10:1195451.
 39. Qin ZZ, Van der Walt M, Moyo S, 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;6:e605-13.
 40. Yayan J, Franke KJ, Berger M, et al. Early detection of tuberculosis: a systematic review. *Pneumonia (Nathan)* 2024;16:11.
 41. Dakulala P, Kal M, Honjepari A, et al. Evaluation of a population-wide, systematic screening initiative for tuberculosis on Daru island, Western Province, Papua New Guinea. *BMC Public Health* 2024;24:959.
 42. Cao XF, Li Y, Xin HN, et al. Application of artificial intelligence in digital chest radiography reading for pulmonary tuberculosis screening. *Chronic Dis Transl Med* 2021;7:35-40.
 43. Fehr J, Gunda R, Siedner MJ, et al. CAD4TB software updates: different triaging thresholds require caution by users and regulation by authorities. *Int J Tuberc Lung Dis* 2023;27:157-60.
- (English Language Editor: J. Gray)

Cite this article as: Han ZL, Zhang YY, Li J, Gao S, Liu W, Yang WJ, Xing ZH. A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. *J Thorac Dis* 2025;17(5):3223-3237. doi: 10.21037/jtd-2025-604