Proceedings of the Global Public Health Conference, Vol. 8, Issue 1, 2025, pp.135-142

ISSN 2613-8417 online
DOI: https://doi.org/10.17501/26138417.2025.8110

COMPARATIVE EVALUATION OF THE DIAGNOSTIC ACCURACY OF ARTIFICIAL INTELLIGENCE-ASSISTED TOOLS AND CONVENTIONAL TOOLS FOR PULMONARY TUBERCULOSIS SCREENING: A SYSTEMATIC REVIEW AND META-ANALYSIS

Santos CBT*, Rivera JP and Sustento VAT
Ateneo de Zamboanga University, Philippines

Abstract: Tuberculosis is a preventable and curable disease, however, it remains as the second leading cause of death worldwide. Systematic screening for TB is one of the key active approaches of the End TB Strategy. However, conventional tools for TB screening have some limitations. AI-based algorithms could be developed, which can help in improving the performance of conventional screening methods. This study intends to evaluate the diagnostic test accuracy of AI-assisted PTB screening tools by meta-analyzing existing literature, and comparison of sensitivity and specificity, as well as assessment of factors which may influence the diagnostic performance of AI-assisted TB screening tools. Literature search was done through electronic databases, reference tracking, and library search. The Preferred Reporting of Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram was used to report the selection and screening of relevant studies between 2014 to 2024. A final count of 31 studies were included in this analysis. Quality assessment was done through the use of QUADAS-C tool. Meta analysis was done through RevMan 5.4.1 and STATA 17. Sensitivity and specificity were used in this analysis. A subgroup analysis was also conducted. The AI-assisted screening tools for pulmonary tuberculosis showed a pooled sensitivity of 93.84% (95% CI: 90.88-95.88) and 83.27% (95% CI: 73.41-89.97), and a diagnostic odds ratio (DOR) of 75.829 (95% CI: 33.19-173.23). Machine Learning (ML) algorithms yielded the highest sensitivity and specificity at 95.06% (95% CI:83.57-98.64) and 91.01% (95% CI: 76.76-96.88) respectively among the AI algorithm subgroup. The results of the meta-analysis done show that AI-assisted screening tools for pulmonary tuberculosis are viable options to improve screening for pulmonary tuberculosis. More robust, multi-center clinical studies regarding the diagnostic accuracy of these AI-assisted tools must be conducted in order to ensure a more valid and generalizable study.

Keyword: diagnostic accuracy, tuberculosis, meta-analysis, screening

Introduction

Pulmonary Tuberculosis (PTB) is a highly contagious disease caused by Mycobacterium tuberculosis, leading to 7.5 million new cases globally in 2022. (World Health Organization, 2023) It disproportionately affects marginalized populations and is prevalent in 30 high-burden countries, including the Philippines, where significant cases go undetected. (Natarajan et al, 2020) Current TB screening tools have limitations, such as suboptimal sensitivity in sputum smear microscopy and the need for trained personnel for chest X-rays. (U.S. Aid for International Development, 2022) Artificial intelligence (AI) presents a promising alternative. However, challenges remain, including the need for diverse training datasets and integration into existing healthcare systems for optimal performance. (Alsdurf et al, 2021) Future research should address these gaps to improve TB detection and management. This study intends to compare AI-assisted PTB screening tools with conventional PTB screening tools by meta-analyzing existing literature.

*Corresponding Author's Email: *cboyasantos@gmail.com

Materials and Methods

This study utilizes a meta-analysis design to assess the efficacy of artificial intelligence (AI) screening methods for pulmonary tuberculosis (PTB) in low- and middle-income countries (LMICs), where the illness is disproportionately prevalent. These countries frequently have limited healthcare resources, thus reliable and cost-effective screening approaches are critical for improving public health outcomes. This study intends to address the critical need for improved diagnostic approaches in areas such as the LMICs where traditional methods may underperform.

The study consists of randomized controlled trials and observational studies that focus on high-risk groups or suspected tuberculosis cases, with interventions provided by AI-based systems such as machine learning and deep learning. Conventional screening methods like symptom assessment, chest radiography, and WHO-recommended diagnostic tests are used as comparators. The primary outcomes to be evaluated are diagnostic performance indicators such as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Studies published in English or with official English translations throughout the last decade (July 2014 to July 2024) are included, however case reports, editorials, and studies lacking original data or relevant performance metrics are excluded.

A thorough literature search was conducted across numerous electronic databases, including PubMed/MEDLINE and ScienceDirect, in search of relevant research published between 2014 and 2024. The search method used a combination of keywords and medical subject headings in accordance with PRISMA guidelines. Additional papers were discovered through reference searches. To collect information on study characteristics, intervention details, and outcome measures, data was extracted using a standardized form. The data were analyzed using RevMan 5.4.1 software, and the risk of bias was assessed using the QUADAS-C tool. To guarantee consistency and generalizability, the study will include pooled estimates of diagnostic performance parameters as well as subgroup analyses.

A Certificate of Exemption from Review was granted due to the study's use of publicly available data. This study intends to benefit marginalized groups by offering insights that can help enhance healthcare treatments in these vulnerable communities.

Results and Discussion

The systematic review process, based on the PRISMA guidelines, began with a search of 80,648 papers from databases such as PubMed and ScienceDirect. After removing 124 duplicates, 44,193 studies were deemed ineligible due to criteria like language and publication date. An additional 34,377 studies were irrelevant, leaving 1,954 for title and abstract screening. Of these, 1,839 were excluded as irrelevant. The reviewers retrieved full-text copies of 132 studies, but 17 were inaccessible. Ultimately, 115 articles were screened, with 2 lacking a conventional TB comparator, and 78 studies from non-low and middle-income countries were excluded, focusing the review on relevant research. The review includes 31 suitable articles chosen based on predetermined inclusion and exclusion criteria. The studies' designs vary, with 19 cross-sectional studies, 6 cohort studies, 1 case-control study, 2 retrospective multicenter trials, 1 experimental study, 1 comparative effectiveness research study, and one computational

analysis. This variety of study types provides an extensive overview of the efficacy of AI-assisted systems for pulmonary tuberculosis (PTB) screening.

All included studies were conducted in low- and middle-income countries (LMICs), reflecting the global burden of tuberculosis in regions with limited healthcare resources. The majority of the studies originated from India (10 studies), followed by Pakistan (6 studies) and Tanzania (3 studies), with additional studies conducted in Nepal, Zambia, the Philippines, Bangladesh, Cambodia, Gambia, Sri Lanka, and Vietnam. The participant demographics varied, with sample sizes ranging from 104 to 23,954, primarily focusing on individuals presumed to have TB through symptom-based screening or healthcare referrals. Some studies also targeted high-risk populations, including people living with HIV, patients with diabetes mellitus, prisoners, and migrants. Benchmark datasets, notably Montgomery County (MC) and Shenzhen (SZ), were commonly used in six studies.

The majority of studies (87.10%) used chest X-rays (CXR) as the primary diagnostic method for PTB. Other diagnostic procedures included culture analysis, sputum analysis, and WHO-recommended tests such as the Xpert MTB/Rif. The AI models utilized in these studies were divided into two categories: Deep Learning (DL) and Machine Learning (ML), with 24 investigations using DL methods. Notable AI tools included the CAD4TB software, which was evaluated in 12 investigations, and Convolutional Neural Networks (CNN), which were investigated in five studies. Additional AI tools used included Normalization Free Network (NFNet), Egret Swarm Optimization Algorithm, and Adapted Monarch Butterfly Optimization Integrated Deep Belief Network. Most studies compared AI-assisted screening tools against human readers for CXR (25 studies), while others used Xpert MTB/Rif, culture, or sputum analysis as comparators.

The quality of the included studies was assessed using the QUADAS-C tool, which evaluates risk of bias across four domains: patient selection, index test, reference standard, and flow and timing. This assessment identified a high or uncertain risk of bias in over half of the studies, highlighting inconsistencies in reporting and the need for improved quality assessment tools tailored for AI-based diagnostic accuracy studies. A comprehensive review of existing literature indicated significant risks of bias across multiple domains. A high or unknown risk of bias was identified in 57.5% of the patient selection domain and 26% of the index test domain. The challenges that health systems encounter in achieving diagnostic needs were highlighted, with an emphasis on AI systems as complementary diagnostic tools and an acknowledgement of the lack of comprehensive quality assessment standards for AI investigations. Furthermore, the need for prospective research to prove AI diagnostic effectiveness while eliminating bias reporting was emphasized. The included studies exhibited various biases, with 28 classified as having an uncertain risk of bias. Only two studies showed low risk, while one showed significant risk due to insufficient patient screening criteria and inconsistent reporting on participant flow. A total of 20 studies were flagged for selection bias, with three assessed as high risk. Concerns were raised about the clarity of patient recruitment and selection criteria in some research, underlining the need of well-defined techniques for improving study validity. Measurement bias was observed in studies that lacked explicit validation protocols, potentially compromising the reproducibility of results. Variability in diagnostic performance was observed when multiple CAD software were used, resulting in discrepancies. Verification bias was also found in research that lacked comprehensive discussion on the limitations of reference standards utilized by AI systems, which could

affect diagnostic accuracy. Timing bias was prevalent in 25 studies that lacked clear reporting on participant flow and timing of tests. Several studies were classified as high risk due to insufficient information on when tests were administered and whether all individuals were evaluated using the same reference standard. A lack of detailed explanations for patient exclusions raised concerns about the findings' reliability. A meta-analysis was conducted to compare the diagnostic accuracy of AI-assisted screening tools for pulmonary tuberculosis (PTB) to traditional screening approaches. The AI tools evaluated include those that use Deep Learning (DL) and Machine Learning (ML) algorithms, whereas conventional standards specified by the WHO include symptom screening, chest radiography, and other diagnostic procedures. Using Review Manager 5.4.1 and STATA version 17, the meta-analysis demonstrated a pooled sensitivity of 93.84% (95% CI: 90.88-95.88) and specificity of 83.27% (95% CI: 73.41-89.97) for AI-assisted tools. Notably, the results using STATA, which uses a bivariate model to account for study heterogeneity, differed slightly from those obtained with RevMan due to the latter's simpler pooling procedures. Recalculating the pooled sensitivity and specificity in Microsoft Excel yielded estimates of 91.57% (95% CI: 91.12–91.93) and 74.17% (95% CI: 73.80–74.17), respectively, demonstrating how the bivariate model accounts for variation between studies. The bivariate analysis yielded important parameter estimates. The estimated log odds ratio for sensitivity was 2.72 (95% CI: 2.30-3.15), indicating accurate detection of true positives, while the log odds ratio for specificity was 1.60 (95% CI: 1.02-2.19), indicating the test's ability to detect true negatives. Variance in sensitivity and specificity suggested moderate variability across studies. The correlation between sensitivity and specificity was determined to be 0.34 (95% confidence interval: -0.04-0.63), illustrating a moderately positive association. The diagnostic odds ratio (DOR) was calculated to be 75.829 (95% CI: 33.19-173.23), indicating the tests' significant discriminatory capacity. However, DOR values must be interpreted cautiously, as they do not account for false negatives and false positives. Subgroup analyses were performed based on AI algorithms and diagnostic modalities.

Pooled sensitivity in studies utilizing ML algorithms was 95.06% (95% CI: 83.57-98.64), while specificity was 91.01% (95% CI: 76.76-96.88). In contrast, studies using DL algorithms demonstrated a pooled sensitivity of 93.99% (95% CI: 90.49-96.26) and specificity of 80.27% (95% CI: 67.12-89.03). For studies that used a combination of algorithms, pooled sensitivity and specificity were 90.48% (95% CI: 88.26-92.70) and 89.06% (95% CI: 86.74-91.38), respectively.

Further subgroup analysis was carried out using diagnostic samples. The sensitivity of chest X-rays ranged from 64% to 100%, with the pooled sensitivity and specificity estimated at 93.75% (95% CI: 90.41-95.98) and 84.24% (95% CI: 73.53-91.14). Other diagnostic techniques, such as Xpert MTB/Rif and culture, showed a pooled sensitivity of 93.67% (95% CI: 85.88-97.30) and specificity of 75.42% (95% CI: 50.75-90.13). The majority of studies (28) were classified as having unclear risk of bias, with pooled sensitivity of 94.30% (95% CI: 91.14-96.37) and pooled specificity of 84.28% (95% CI: 73.68-91.13). In studies with low risk of bias, pooled sensitivity was 89.92% (95% CI: 89.30-90.54) and specificity was 72.33% (95% CI: 71.50-73.16). Sensitivity analysis revealed minimal changes in pooled sensitivity (93.92%) and specificity (83.48%) when excluding high-risk bias studies, indicating stable results. The analysis also highlighted a trade-off between sensitivity and specificity, particularly when excluding specific AI algorithms from the analysis. This study conducted a meta-analysis to evaluate the diagnostic performance of AI-assisted screening tools for pulmonary tuberculosis (PTB) in low and

middle-income countries, analyzing 31 studies with a total of 74,430 participants. Quality assessments were performed using the QUADAS-C tool, revealing that AI-assisted tools achieved a pooled sensitivity of 93.84% and specificity of 83.27%, meeting WHO guidelines for TB screening. In comparison, traditional symptom screening yielded lower sensitivity (42% to 71%) and higher specificity (64% to 94%), while chest X-rays (CXR) showed a sensitivity of 85% and specificity of 96%. The molecular test Xpert MTB/Rif had a sensitivity of 69% and specificity of 99%.35 Among diagnostic modalities, CXR and other samples showed similar sensitivities, but CXR had significantly higher specificity, with Machine Learning algorithms demonstrating the highest sensitivity (95.06%) and specificity (91.01%) within the AI subgroup. A comprehensive evaluation of published studies comparing the diagnostic accuracy of AI-assisted tools for screening pulmonary tuberculosis (PTB) was conducted through a meta-analysis of 31 studies published between 2014 and 2024. Out of 80,648 records identified from four online databases (EBSCOHost, PubMed, ScienceDirect, and ELibrary USA), these studies were selected to assess performance metrics such as sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). AI-assisted screening tools for pulmonary tuberculosis (PTB) have the potential to enhance existing conventional screening methods, particularly in resource-poor areas with limited healthcare access. These tools can provide point-of-care access to disadvantaged populations and help healthcare workers manage more patients efficiently, thereby maximizing the effectiveness of TB screening programs. However, the results indicate that various AI-assisted tools exhibit different sensitivity and specificity relationships, highlighting the necessity for careful assessment of methodologies when selecting and implementing these tools to ensure optimal diagnostic accuracy for specific populations. Despite these promising developments, there are limitations to the current research. More comprehensive studies are essential to minimize bias in evaluating the diagnostic accuracy of evolving AI-assisted tools. Additionally, investigations into turnaround times and cost-effectiveness are crucial for health economists and policymakers in designing effective TB screening programs.37,38 The lack of detailed demographic data in many studies included in the meta-analysis also underscores the need for further research to explore differences in effectiveness across diverse populations, as the characteristics of study participants were often not explicitly described.

Conclusion

The meta-analysis indicates that AI-assisted screening tools for pulmonary tuberculosis (PTB) are promising options for enhancing screening processes, particularly in high-risk populations within resource-poor areas, such as the Philippines. To improve the validity and generalizability of these findings, more robust multi-center clinical studies are necessary to minimize potential biases. A thorough review and quality assessment of the studies are essential, focusing on key domains such as patient selection and the flow and timing of diagnostic tests, as suggested by the QUADAS-C framework. The insights gained from this study could inform the implementation of TB screening programs in vulnerable populations, ultimately aiming to reduce morbidity and mortality associated with tuberculosis. To effectively integrate AI-assisted screening tools into clinical practice, it is crucial to provide proper training and technical support to healthcare workers, especially those in grassroots settings. Future research should focus on the diagnostic accuracy of these tools across diverse demographics, geographical areas, and clinical contexts, as well as their long-term impact on patient

outcomes and cost-effectiveness. Collaborative efforts among health researchers, professionals, and AI developers could refine AI algorithms for tuberculosis screening. Additionally, adequate resource allocation from governments and NGOs is vital for the development and implementation of these tools. Establishing evidence-based guidelines for the use of AI-assisted tools and conducting long-term evaluations will ensure their safety and effectiveness, ultimately contributing to the WHO's goal of a TB-free world.

Acknowledgements

The author would like to extend his heartfelt gratitude to his family and close friends for their unwavering support throughout his academic journey. The author extends special thanks to his adviser and co-authors, Dr. Jejunee P. Rivera and Dr. Vladimir Alexis Sustento, and the MPH panelists for their guidance and insights, as well as to his professors for their encouragement and assistance. He appreciates his research assistants, Dr. Aisha A. Abdulajid, Dr. Maryam S. Munabirul, and Dr. Mudzralyn T. Pajar for technical support and his peers for their camaraderie during this challenging process. Acknowledgment also goes to the university faculty and staff for their resources, the campus ministry for their support, and his spiritual mentors, the Jesuits, for their inspiration. Finally, the author expresses gratitude to the triune God for His grace and blessings that made this research possible, along with everyone who contributed to its success.

Declaration of Interest Statement

The authors declare that they have no conflict of interests.

References

Acharya, V., Dhiman, G., Prakasha, K., Bahadur, P., Choraria, A., M. S., et al. (2022). AI-assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model. Computational Intelligence and Neuroscience, 2022, 1–19.

Alsdurf, H., Empringham, B., Miller, C., & Zwerling, A. (2021). Tuberculosis screening costs and cost-effectiveness in high-risk groups: A systematic review. BMC Infectious Diseases, 21(1), 935.

Ayaz, M., Shaukat, F., & Raja, G. (2021). Ensemble learning-based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. Physiological Engineering and Science in Medicine, 44(1), 183–194.

Bhandari, M., Shahi, T. B., Siku, B., & Neupane, A. (2022). Explanatory classification of CXR images into COVID-19, pneumonia, and tuberculosis using deep learning and XAI. Computational Biology and Medicine, 150, 106156.Chauhan, A., Chauhan, D., & Rout, C. (2014). Role of GIST and PHOG features in computer-aided diagnosis of tuberculosis without segmentation. PLOS ONE, 9(11), e112980.

Breuninger, M., Van Ginneken, B., Philipsen, R. H. H. M., Mhimbira, F., Hella, J. J., Lwilla, F., et al. (2014). Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: A validation study from sub-Saharan Africa. PLOS ONE, 9(9), e106381.

Chauhan, A., Chauhan, D., & Rout, C. (2014). Role of GIST and PHOG features in computer-aided diagnosis of tuberculosis without segmentation. PLOS ONE, 9(11), e112980.

Codlin, A. J., Dao, T. P., Vo, L. N. Q., Forse, R. J., Van Truong, V., Dang, H. M., et al. (2021). Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. Scientific Reports, 11(1), 23895.

Das, D., Santosh, K. C., & Pal, U. (2021). Inception-based deep learning architecture for tuberculosis screening using chest X-rays. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3612–3619). IEEE. https://doi.org/10.1109/ICPR48806.2021.9412748

Dasanayaka, C., & Dissanayake, M. B. (2021). Deep learning methods for screening pulmonary tuberculosis using chest X-rays. Computational Methods in Biomechanics and Biomedical Engineering Imaging and Visualization, 9(1), 39–49.

Govindarajan, S., & Swaminathan, R. (2021). Extreme learning machine-based differentiation of pulmonary tuberculosis in chest radiographs using integrated local feature descriptors. Computational Methods and Programs in Biomedicine, 204, 106058.

Habib, S. S., Rafiq, S., Zaidi, S. M. A., Ferrand, R. A., Creswell, J., Van Ginneken, B., et al. (2020). Evaluation of computer-aided detection of tuberculosis on chest radiography among people with diabetes in Karachi, Pakistan. Scientific Reports, 10(1), 6276.

Hooda, R., Sofat, S., Kaur, S., Mittal, A., & Meriaudeau, F. (2017). Deep learning: A potential method for tuberculosis detection using chest radiography. In 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 497–502). IEEE. https://doi.org/10.1109/ICSIPA.2017.8120663

Khan, F. A., Majidulla, A., Tavaziva, G., Nazish, A., Abidi, S. K., Benedetti, A., et al. (2020). Chest X-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: A prospective study of diagnostic accuracy for culture-confirmed disease. Lancet Digital Health, 2(11), e573–e581.

Melendez, J., Van Ginneken, B., Maduskar, P., Philipsen, R. H. H. M., Reither, K., Breuninger, M., et al. (2015). A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest X-rays. IEEE Transactions on Medical Imaging, 34(1), 179–192.

Murphy, K., Habib, S. S., Zaidi, S. M. A., Khowaja, S., Khan, A., Melendez, J., et al. (2020). Computer-aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. Scientific Reports, 10(1), 5492.

Muyoyeta, M., Maduskar, P., Moyo, M., Kasese, N., Milimo, D., Spooner, R., et al. (2014). The sensitivity and specificity of using a computer-aided diagnosis program for automatically scoring chest X-rays of presumptive TB patients compared with Xpert MTB/RIF in Lusaka, Zambia. PLOS ONE, 9(4), e93757.

Nash, M., Kadavigere, R., Andrade, J., Sukumar, C. A., Chawla, K., Shenoy, V. P., et al. (2020). Deep learning, computer-aided radiography reading for tuberculosis: A diagnostic accuracy study from a tertiary hospital in India. Scientific Reports, 10(1), 210.

Natarajan, A., Beena, P. M., Devnikar, A. V., & Mali, S. (2020). A systemic review on tuberculosis. Indian Journal of Tuberculosis, 67(3), 295–311.

Philipsen, R. H. H. M., Sánchez, C. I., Melendez, J., Lew, W. J., & Van Ginneken, B. (2019). Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: A cohort study. International Journal of Tuberculosis and Lung Disease, 23(7), 805–810.

Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S. S., Naheyan, T., et al. (2021). Tuberculosis detection from chest X-rays for triaging in a high tuberculosis-burden setting: An evaluation of five artificial intelligence algorithms. Lancet Digital Health, 3(9), e543–e554.

Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mazhar, R., et al. (2020). Reliable tuberculosis detection using chest X-ray with deep learning, segmentation, and visualization. IEEE Access, 8, 191586–191601.

Sharma, A., Sharma, A., Malhotra, R., Singh, P., Chakrabortty, R. K., Mahajan, S., et al. (2021). An accurate artificial intelligence system for the detection of pulmonary and extrapulmonary tuberculosis. Tuberculosis, 131, 102143.

Steiner, A., Mangu, C., Van Den Hombergh, J., Van Deutekom, H., Van Ginneken, B., Clowes, P., et al. (2015). Screening for pulmonary tuberculosis in a Tanzanian prison and computer-aided interpretation of chest X-rays. Public Health Action, 5(4), 249–254.

Ullah, R., Khan, S., Chaudhary, I. I., Shahzad, S., Ali, H., & Bilal, M. (2020). Cost-effective and efficient screening of tuberculosis disease with Raman spectroscopy and machine learning algorithms. Photodiagnosis and Photodynamic Therapy, 32, 101963.

U.S. Agency for International Development. (2022). Philippines tuberculosis roadmap overview, fiscal year 2023. https://www.usaid.gov/global-health/health-areas/tuberculosis/resources/news-and-updates/global-accelerator-end-tb/tb-roadmaps/philippines-tuberculosis-roadmap-overview-fiscal-year-2022

Zaidi, S. M. A., Habib, S. S., Van Ginneken, B., Ferrand, R. A., Creswell, J., Khowaja, S., et al. (2018). Evaluation of the diagnostic accuracy of computer-aided detection of tuberculosis on chest radiography among private sector patients in Pakistan. Scientific Reports, 8(1), 12339.