narra R

Scoping Review

Machine learning-based algorithms for early detection of tuberculosis: A scoping review

Roy N. Ramadhan[1,2], Merita Arini[1*], Jihaan Farahiyah[2] and Sephia Maharani[3]

[1]Postgraduate Program in Hospital Administration, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia; [2]Faculty of Medicine, Universitas Airlangga, Surabaya, Indonesia; [3]Faculty of Medicine, Universitas Tarumanagara, Jakarta, Indonesia

*Corresponding author: merita.arini@umy.ac.id

Abstract

Tuberculosis (TB) remains a major global health challenge, particularly in low- and middle-income countries. Early and accurate detection is crucial for effective disease control, timely treatment initiation, and improved patient outcomes. Conventional diagnostic methods, such as smear microscopy and culture, are limited by suboptimal sensitivity, variable specificity, and long turnaround times. Machine learning (ML) has emerged as a promising tool to enhance diagnostic accuracy, especially through biomarker-based approaches. Techniques such as radiomics and deep learning (DL) enable early detection by identifying disease patterns in biomarker data, facilitating rapid and precise diagnosis. The aim of this study was to explore the role of ML in biomarker-based TB detection, emphasizing its potential in early diagnosis and improved screening strategies. A comprehensive review of studies on ML applications in biomarker-based early detection and diagnosis of TB was conducted using PubMed, Scopus, Web of Science, and Epistemonikos databases. Studies published up to February 3, 2025, were analyzed to evaluate the effectiveness of ML models in TB detection. ML-based diagnostic models, including those incorporating radiomics and DL, have demonstrated promising results in TB early detection and diagnosis. Studies indicate that ML techniques can enhance sensitivity and specificity in biomarker-based TB screening, enabling earlier intervention. The study highlights key challenges, such as data variability, model generalizability, and the need for standardized validation methods to ensure clinical applicability. ML offers a valuable approach for biomarker-based TB early detection and diagnosis, addressing limitations in conventional methods. While the current findings support its potential, further research is needed to optimize model performance, enhance reproducibility, and facilitate clinical integration for more effective TB control strategies.

Keywords: Tuberculosis, machine learning, biomarkers, early diagnosis, deep learning

Introduction

Tuberculosis (TB) remains a significant global health challenge, particularly in regions with high incidence rates such as Southeast Asia and Africa [1]. Despite substantial progress in TB control, the disease continues to pose a considerable burden on healthcare systems worldwide. In 2023, an estimated 10.8 million people were diagnosed with TB, with approximately 1.3 million deaths attributed to the disease [1]. Countries such as India, Indonesia, and the Philippines remain among the most affected, contributing to a significant proportion of the global TB burden. Additionally, Sub-Saharan Africa faces high morbidity and mortality rates due to TB, exacerbated by limited access to early diagnostic tools and effective treatment strategies [2].

One of the most pressing challenges in TB control is the emergence and spread of multidrug-resistant TB (MDR-TB), which is resistant to first-line anti-TB drugs such as rifampicin and

Received: July 5, 2025 - Accepted: August 27, 2025 - Published online: August 29, 2025

isoniazid. MDR-TB is particularly problematic in low- and middle-income countries (LMICs), where healthcare infrastructure often struggles with inadequate laboratory capacity, limited human resources, and delays in diagnostic turnaround times [3-4]. Late detection and misdiagnosis worsen the situation, leading to increased transmission, higher mortality, and greater financial strain on hospital management [3].

In response to these challenges, machine learning (ML) has emerged as a transformative tool for early TB detection and hospital-based infection control strategies. ML, a subset of artificial intelligence (AI), enables the analysis of large amounts of medical data, pattern recognition, and predictive modeling to improve clinical decision-making [5-6]. In TB diagnostics, ML-based algorithms have demonstrated significant potential to enhance accuracy, speed, and efficiency of early detection, particularly when integrated with hospital-based radiological imaging, laboratory tests, and electronic medical records [7].

Beyond diagnosis, ML contributes to infection control and hospital management. Predictive analytics powered by ML can help identify high-risk patients, optimize resource allocation, and improve infection surveillance systems [8]. By leveraging hospital databases and patient records, ML algorithms can detect outbreaks, predict disease progression, and support targeted interventions [8-9]. These capabilities are particularly valuable in TB-endemic regions, where limited healthcare resources require efficient case management to prevent hospital-acquired infections and community transmission.

Despite this potential, several challenges hinder ML adoption in healthcare. Issues such as data accessibility, model interpretability, cost-effectiveness, and ethical concerns remain major barriers. Healthcare professionals also require training and support to ensure effective integration of ML into clinical workflows and hospital systems. Moreover, regulatory frameworks and data privacy laws must evolve in parallel with the growing use of AI-driven technologies in healthcare.

The aim of this study was to explore the current advancements, challenges, and future prospects of ML applications in early TB detection, with a particular focus on hospital management strategies. By summarizing key findings from recent studies, this review aims to provide valuable insights for healthcare policymakers, hospital administrators, and clinicians seeking to integrate AI-driven technologies into TB diagnostic workflows and infection control programs.

Methods

Study design and selection

This study was a narrative review with a scoping approach, aimed at synthesizing evidence on ML algorithms for early TB detection. A comprehensive literature search was conducted in PubMed, Scopus, Web of Science, and Epistemonikos to identify relevant studies on the application of ML for early TB detection in hospital settings. The search was performed from January 3, 2025, until February 3, 2025. Articles were included if they were published in English and focused on ML-based approaches for TB diagnosis, infection control strategies in hospital settings, or challenges related to implementing ML-driven solutions in clinical practice. The search strategy combined Medical Subject Headings (MeSH) terms and free-text keywords, including "tuberculosis," "machine learning," "artificial intelligence," "deep learning," "diagnostic model," and "infection control".

Data extraction

The retrieved studies were independently screened by two reviewers (RNR and MA) based on titles and abstracts. Discrepancies were resolved through discussion or consultation with a third reviewer. Full-text articles were further assessed for eligibility, and studies not addressing ML applications for TB detection, hospital-based infection control, or hospital management were excluded. The study selection process and criteria are presented in **Table 1**.

Table 1. Search process and study selection criteria

Items	Specification
Date of search	January 3, 2024
Databases and other sources searched	PubMed, Scopus, Web of Science, and Epistemonikos
Search terms used	"tuberculosis," "machine learning," "artificial intelligence," "deep learning," "diagnostic model," and "infection control"
Timeframe	January 3, 2025
Inclusion and exclusion criteria	Papers in English were included and focused on ML-based approaches for TB diagnosis, infection control strategies in hospital settings, or challenges in implementing ML-driven solutions in clinical practice
Selection process	Two reviewers screened the search results independently through titles and abstracts to select the eligible ones (RNR and MA)

Conventional methods and limitations in early tuberculosis screening

Accurate and timely diagnosis of TB is essential to reduce mortality and morbidity. Some of the conventional methods are Acid-fast bacillus smear microscopy, microbial culture, while more advanced strategies include CRISPR Cas, Gene Xpert and loop-mediated isothermal amplification (LAMP). Smear microscopy has limited sensitivity and cannot detect drug resistance, whereas culture-based diagnosis, though highly specific, requires 3–4 weeks for results. Molecular biology methods are unable to distinguish live from dead *Mycobacterium tuberculosis*, and immunological assays fail to distinguish active from latent infection [10].

Traditional methods such as sputum smear microscopy and culture have played a significant role in shaping diagnostic practice. Smear microscopy, introduced in the mid-20th century, became the cornerstone of TB detection but demonstrates suboptimal sensitivity, particularly in paucibacillary and extrapulmonary cases [11]. Culture of *M. tuberculosis*, although highly specific, is hampered by its long turnaround time, often for several weeks. This delay prevents timely initiation of treatment, allowing the disease to continue to spread. In addition, culture-based methods are resource-intensive, requiring specialized facilities, skilled personnel, and controlled environments [12].

Challenges extend beyond sensitivity and turnaround time. In resource-constrained settings, the infrastructure necessary for sophisticated laboratory techniques is often lacking, hampering widespread adoption of this diagnostic approach [10]. In addition, reliance on empirical treatment in the absence of rapid and accurate diagnostic tools can lead to the emergence of drug-resistant strains of tuberculosis, increasing the complexity of tuberculosis management.

Conventional TB screening has several limitations, including low sensitivity, long turnaround times, and the inability to differentiate between strains. These limitations can lead to delayed diagnosis and treatment, which can contribute to the spread of the disease [13]. As the limitations of traditional methods become increasingly apparent, there is a critical need to change the diagnostic paradigm. This challenge paves the way for newer diagnostic tests such as the Xpert MTB/RIF Ultra system.

Machine learning (ML) application in early tuberculosis screening

The introduction of machine learning

ML is driving major advances in the diagnosis and detection of health conditions, ranging from population-level emergencies to individual immune responses. Current applications of ML in healthcare primarily support physicians and analysts by enhancing decision-making, identifying trends, and developing predictive disease models [14]. In large medical organizations, ML-based approaches have also been applied to improve efficiency in organizing electronic health records, interpreting medical imaging, monitoring patients, detecting irregularities in blood samples and organ function, and even in robot-assisted surgery. Artificial intelligence (AI) and ML applications are increasingly vital for the progression of healthcare, offering faster diagnosis, greater accuracy, and streamlined workflows [15].

Feature variables for machine learning models in early TB screening

In the health sector, ML has been applied to address challenges in public health, medical image analysis, and diagnostic support systems [16]. Previous study developed ML approaches for TB diagnosis, using thoracic imaging as the main data source. Advances in this field have enabled better detection of thoracic diseases, including TB, asthma, cancer, and pneumonia. Researchers have widely used specialized ML models to enhance TB diagnosis by utilizing available data, such as clinical records and molecular biology data [17].

More broadly, ML has been widely applied to assist medical diagnosis through radiographic images analysis. TB research has increasingly incorporated deep learning (DL) architectures trained on large datasets to provide image-based diagnostic scenarios. For example, the image-CLEF dataset was created to facilitate TB type identification and treatment resistance using computed tomography images; radiographic images have also been used to support clinical decision-making by healthcare professionals [16].

The present study proposes ML techniques as a “tool in the loop” for TB diagnosis, where health professionals retain primary decision-making authority but are supported by ML-based insights derived from limited available data. Such approach is particularly in resource-limited settings, where ML can complement existing protocols and help overcome infrastructure constraints.

ML models for early TB screening

In the case of TB, diagnosis is based on respiratory symptoms, followed by testing of suspected patients using serial sputum smears. Although this test is simple, its clinical utility requires careful consideration. From an ML perspective, applications often involved biomedical data characterized by high uncertainty and incompleteness [18], and strategies beyond direct ML are sometimes necessary. In this context, “ML in the loop” (MLL) was investigated; its effectiveness depends on how the ML tool is integrated into clinical workflows. Researchers have analyzed the workflows required to improve outcomes, while others have studied the doctor-in-the-loop approaches that account for the critical role of healthcare professionals in system performance [18]. Today, model performance alone is no longer sufficient; generalizability and functionality during human interactions are equally important. Assessing these broader aspects of performance provides valuable insight into decision-making and operational requirements that must be considered in system design [16].

The ML process for TB diagnostic support examined in this study is presented in **Figure 1**. Initially, a patient presenting with respiratory symptoms visits a medical center for consultation or emergency care. After an initial assessment, the patient is referred to the internal medicine department for further evaluation. If symptoms persist, the medical team requests three primary diagnostic tests for pulmonary TB: sputum smear microscopy, sputum culture, and molecular testing (GenXpert®). When test results suggest infection, the patient is started on antituberculosis therapy. In the case described, although the test results were definitive, no conclusive diagnosis was made. Nevertheless, anti-tuberculosis treatment was initiated. At this stage, ML was applied to support the diagnostic process and assist the medical staff [16].

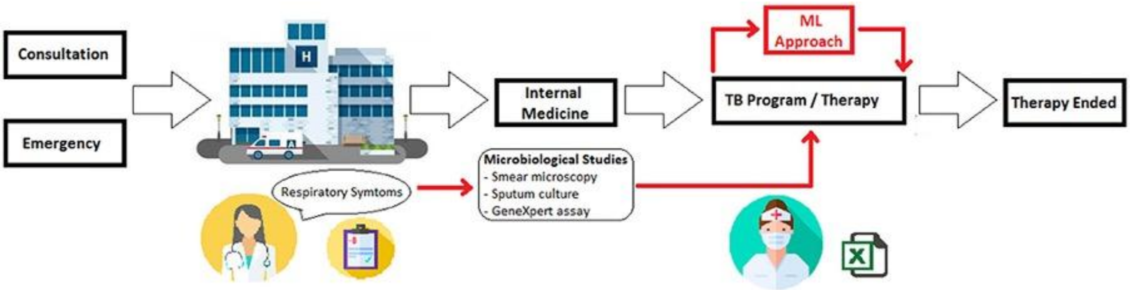

Figure 1. ML process for TB diagnostic support, illustrating the point at which ML was applied to assist medical staff in the diagnostic workflow [16].

Convolutional neural network (CNN) represents a powerful integration of math, biology, and computer science. It has been a significant invention in image processing and machine vision.

The conception of CNN depicts the probability of the data (images) into certain classes on the same basis as the human brain works [19]. For example, when a human looks at an image, they classify it into different types based on its characteristic features. Similarly, the computer algorithm can organize the image by identifying the low-level features, such as curves and edges, at initial levels and gradually generating more intense concepts with high-level features through a series of hidden layers (convolutional layer, pooling layer, fully connected layer) [20]. Similarly, a CNN model learns distinguishing features from the TB dataset and tries to classify a new input image as a TB or non-TB case (**Figure 2**).

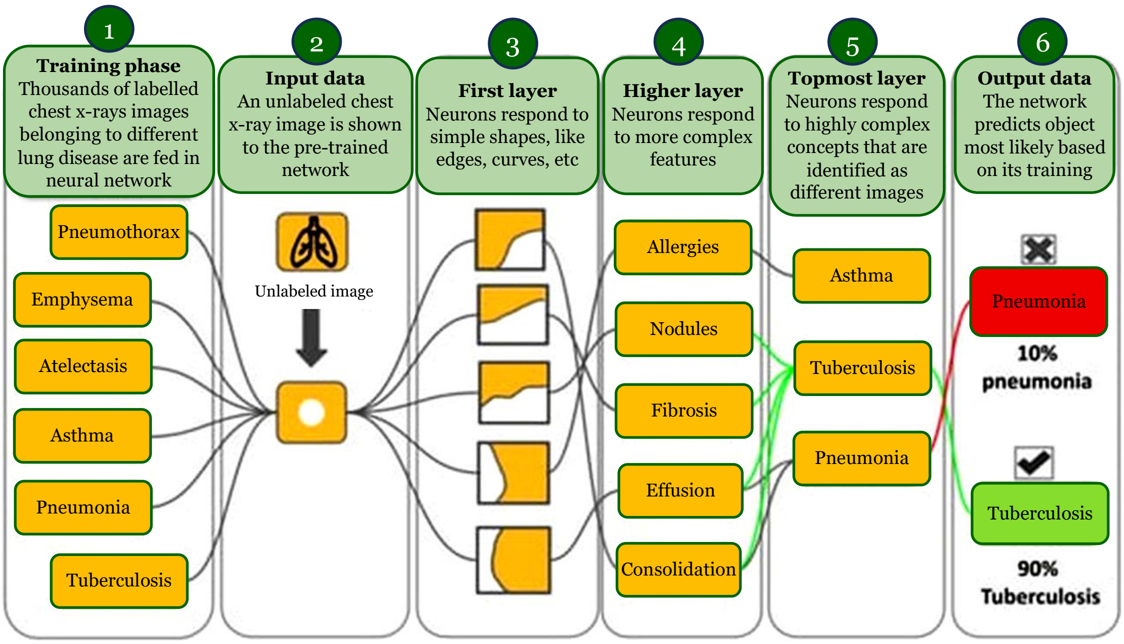

Figure 2. Workflow of a convolutional neural network (CNN) for tuberculosis (TB) detection. Input chest X-ray images are processed through successive layers to extract low- and high-level features, which are then mapped to disease categories. The output layer predicts the probability of TB versus non-TB cases, providing final classification.

Challenges and prospects of ML in early TB screening

ML algorithms encompass a wide range of techniques, many of which have already surpassed human expertise in specific tasks. Their application to TB diagnosis is therefore highly promising. AI-based approaches in TB diagnosis are mainly integrated for early disease diagnosis, focusing on TB symptoms and drug resistance. Despite significant successes in the past, the implementation of AI in TB diagnosis is still challenging due to quality acquisition; thus, adequate data and problem solving remain a challenge. Data compilation is easy if we look at it in the field of computer vision, as the selected data is highly reliable and forms a large dataset. Unfortunately, it is not the same in disease diagnosis due to many reasons that will hinder the production of good quality and sufficient data to form a dataset. On one hand, data collected from different experimental sources is highly dependent on the variance of the applied experimental conditions, which may give different or contradictory results. This is due to the complex human biological system, which exhibits different symptoms and drug resistance for the same disease and drug. On the other hand, a wide variety of data is available, depending on the nature of the requirements and the need to build a high-quality dataset [19].

One solution to this perplexing problem is to develop algorithms that can handle different or insufficient datasets. Another important challenge is to utilize and adhere to the principle of removing unwanted data to achieve further refinement and a set of task-oriented standards. In 2017, DL models were pre-trained on millions of images using the AlexNet and GoogLeNet datasets, which contain many images, including chest radiographs. These models require significant computer storage space to work efficiently, even for a small task of TB detection using chest radiograph images [21]. At the same time, a dataset was proposed that requires less storage memory and cheaper processors because it only requires storage in mega-FLOPs, unlike the AlexNet and GoogLeNet datasets that require storage in Giga–FLOPs [22].

Deep learning methods integrated with neuro-fuzzy algorithms, genetics, and other artificial immune systems are the most promising subdomains for TB diagnosis. These systems improve sensitivity, specificity, and accuracy by incorporating biological information. However, limitations such as overfitting (modeling error where a function fits too well to a limited set of data points) and lack of interpretability of CNN remain significant challenges [23]. This overfitting is mainly due to the limited data set for accurate TB diagnosis and despite this, many studies related to DL and integrated systems have been conducted, which can help overcome these limitations. Furthermore, technological advancements have led to the development of visualization methods.

Conclusion

ML-based algorithms have shown promising potential for enhancing the early detection of TB, offering improved diagnostic accuracy, efficiency, and accessibility, especially in resource-limited settings. Despite these advancements, challenges such as data quality, model generalizability, and integration into clinical workflows remain significant barriers. Future research should focus on optimizing algorithms with diverse and high-quality datasets, improving interpretability, and ensuring seamless integration into healthcare systems. With continuous advancements, machine learning has the potential to revolutionize TB diagnostics and contribute to global TB elimination efforts.

Acknowledgments

The authors have nothing to declare.

Competing interests

The authors declare no competing interests.

Funding

This study received no external funding.

Underlying data

All underlying data have been presented in this article.

Declaration of artificial intelligence use

We hereby confirm that no artificial intelligence (AI) tools or methodologies were utilized at any stage of this study, including during data collection, analysis, visualization, or manuscript preparation. All work presented in this study was conducted manually by the authors without the assistance of AI-based tools or systems.

How to cite

Ramadhan RN, Arini M, Farahiyah J, Maharani S. Machine learning-based algorithms for early detection of tuberculosis: A scoping review. Narra R 2025; 1 (2): e7 - http://doi.org/10.52225/narrar.v1i2.7.

References

1. World Health Organization. Global tuberculosis report 2024. Geneva: World Health Organization; 2024.
2. Ardiansyah DN, Suryawati C, Adi MS. Provision of resources in the implementation of tuberculosis-multi drugs resistance treatment service in "X" hospital. CoMPHI 2020;1(2):51-57.
3. Guo R, Passi K, Jain CK. Tuberculosis diagnostics and localization in chest X-rays via deep learning models. Front Artif Intell 2020;3:583427.

4. Wijiseno B, Arini M, Listiowati E. Healthcare workers' acceptance of the integrated tuberculosis-COVID-19 screening in central Java Private Hospitals, Indonesia. J Taibah Univ Med Sci 2023;18(6):1311-1320.
5. 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(1):5492.
6. Tang N, Yuan M, Chen Z, *et al.* Machine learning prediction model of tuberculosis incidence based on meteorological factors and air pollutants. Int J Environ Res Public Health 2023;20(5):3910.
7. Syahbani MR, Shamshirband S, Golzari HS, *et al.* Diagnosing tuberculosis with a novel support vector machine-based artificial immune recognition system. Iran Red Crescent Med J 2015;17(4):e24557.
8. Salcedo J, Rosales M, Kim JS, *et al.* Cost-effectiveness of artificial intelligence monitoring for active tuberculosis treatment: A modeling study. PLoS One 2021;16(7):e0254950.
9. Rodrigues MMS, Barreto-Duarte B, Vinhaes CL, *et al.* Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment. BMC Public Health 2024;24(1):1385.
10. Dong B, He Z, Li Y, *et al.* Improved conventional and new approaches in the diagnosis of tuberculosis. Front Microbiol 2022;13:924410.
11. Desikan P. Sputum smear microscopy in tuberculosis: Is it still relevant? Indian J Med Res 2013;137(3):442-444.
12. Ghodbane R, Raoult D, Drancourt M. Dramatic reduction of culture time of Mycobacterium tuberculosis. Sci Rep 2014;4:4236.
13. Yadav S, Rawal G, Jeyaraman M, Jeyaraman N. Advancements in tuberculosis diagnostics: a comprehensive review of the critical role and future prospects of Xpert MTB/RIF ultra technology. Cureus 2024;16(3):e57311.
14. Gaonkar B, Davatzikos C. Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. NeuroImage 2013;78:270-283.
15. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics 2021;22(4):291-300.
16. Chen M, Decary M. Artificial intelligence in healthcare: An essential guide for health leaders. Healthc Manage Forum 2020;33(1):10-18.
17. Awaysheh A, Wilcke J, Elvinger F, *et al.* Review of medical decision support and machine-learning methods. Vet Pathol 2019;56(4):512-525.
18. Xin D, Ma L, Liu J, *et al.* Accelerating human-in-the-loop machine learning: Challenges and Opportunities. In : Proceedings of the Second Workshop on Data Management for End-To-End Machine Learning. New York: Association for Computing Machinery (DEEM'18); 2018.
19. Shin HC, Roth HR, Gao M, *et al.* Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 2016;35(5):1285-1298.
20. Zhang Q, Zhang M, Chen T, *et al.* Recent advances in convolutional neural network acceleration. Neurocomputing 2019;323:37-51.
21. Lakhani P, Sundaram B. Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017;284(2):574-582.
22. Pasa F, Golkov V, Pfeiffer F, *et al.* Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. Sci Rep 2019;9(1):6268.
23. Singh M, Pujar GV, Kumar SA, *et al.* Evolution of machine learning in tuberculosis diagnosis: A review of deep learning-based medical applications. Electronics 2022;11(17):2634.