AI-Based Cough Sound Analysis for Early Tuberculosis Screening Using SVM and CNN

Liyaana Jebin[1], Aleena Aboo Jaleel[2], Akash Krishna[3], Nithiyapriya M[4]
[1,2,3,4]*Department of Biomedical Engineering, Paavai Engineering College,* Namakkal,India
[1]E-mail: liyaanajebin2023@gmail.com
[2]E-mail: aleenajaleel2021@gmail.com
[3]E-mail: akashkrishna751@gmail.com
[4]E-mail: nithiyapriya353@gmail.com

Abstract

This paper presents an intelligent cough sound analysis framework for early tuberculosis screening using machine learning techniques. The proposed system is designed as a non-invasive, portable, low-cost, and rural-friendly solution that analyzes cough audio signals to identify TB and related respiratory abnormalities. The framework consists of cough sound acquisition through a microphone interface, preprocessing stages including noise reduction and normalization, and feature extraction using Mel Frequency Cepstral Coefficients (MFCC), spectrogram-based descriptors, zero-crossing rate, and energy-frequency characteristics. These extracted acoustic features are utilized to train and evaluate machine learning classifiers such as Support Vector Machine (SVM) and Convolutional Neural Network (CNN) for robust cough pattern recognition. The developed model is intended for real-time deployment through a simple user interface, with optional embedded implementation for offline operation in remote healthcare settings. This work demonstrates the potential of artificial intelligence-driven cough analysis as an effective preliminary screening tool for tuberculosis, with future scope for multi-disease respiratory classification and cloud-assisted healthcare integration.

Keywords— Tuberculosis (TB), cough sound analysis, machine learning, Mel Frequency Cepstral Coefficients (MFCC), rural healthcare, non-invasive screening.

I. INTRODUCTION

Tuberculosis (TB) continues to be one of the most serious infectious diseases affecting global public health, especially in rural and resource-constrained environments where access to advanced medical diagnostics remains limited. Conventional TB diagnosis generally depends on laboratory-based tests such as sputum microscopy, chest radiography, GeneXpert, and culture analysis, which are often expensive, time-consuming, and not readily accessible in remote areas. In many developing and underserved regions, delayed diagnosis contributes to increased disease transmission, poor treatment outcomes, and higher mortality rates. Therefore, there is a significant need for a rapid, affordable, and easily deployable screening solution that can assist healthcare workers in identifying potential TB cases at an early stage without relying entirely on centralized diagnostic infrastructure.

Recent advances in artificial intelligence, digital signal processing, and machine learning have enabled the development of intelligent healthcare systems capable of analyzing bioacoustic signals for disease detection. Among these, cough sound analysis has emerged as a promising non-invasive approach for identifying respiratory abnormalities, since cough patterns often contain discriminative acoustic signatures associated with pulmonary conditions. Motivated by this potential, the present work proposes an intelligent cough-based TB screening framework that captures cough audio through a microphone, preprocesses the signal to remove environmental noise and normalize amplitude variations, extracts informative features such as Mel Frequency Cepstral Coefficients (MFCC), spectrogram representations, zero-crossing rate, and energy-related descriptors, and performs classification using machine learning models such as Support Vector Machine (SVM) and Convolutional Neural Network (CNN). The system is designed to be portable, low-cost, offline-capable, and suitable for real-time deployment through a user-friendly Streamlit interface, making it highly applicable for rural health centers, mobile clinics, and telemedicine-based preliminary screening.

II. RELATED WORK

A systematic review on early tuberculosis detection highlighted that timely diagnosis remains a critical factor in reducing disease transmission, improving treatment outcomes, and lowering mortality rates [1]. The study emphasized that early identification methods must be strengthened through better screening strategies, biomarker utilization, and improved clinical decision support. In the same direction, machine learning-based prediction of early treatment outcomes for multidrug-resistant or rifampicin-resistant tuberculosis demonstrated that advanced predictive models can assist clinicians in identifying high-risk patients at an earlier stage of therapy [2]. The research showed that multiple validated machine learning models were useful in forecasting treatment response, thereby supporting personalized intervention and improving cure rates. Likewise, forecasting the trend of tuberculosis incidence using machine learning optimization algorithms established that predictive analytics can be effectively applied not only for patient-level diagnosis but also for population-level surveillance [3]. The findings indicated that optimized machine learning techniques are valuable in identifying future disease trends and supporting public health planning for tuberculosis control.

Further, the development of a machine learning-based tuberculosis health predictor model confirmed that early TB disease prediction can be enhanced through the use of supervised learning approaches, particularly in developing countries where delayed diagnosis remains a major challenge [4]. The study suggested that machine learning models can improve preventive healthcare by enabling early risk stratification and clinical support in resource-limited settings. Similarly, machine learning approaches for predicting tuberculosis risk among household contacts of index TB patients showed that contact-based risk assessment can be significantly strengthened through intelligent predictive systems [5]. The research highlighted the importance of identifying vulnerable household contacts early, which is essential for preventing secondary transmission and initiating timely monitoring. In addition, machine learning for predicting bacteriologic confirmation of *Mycobacterium tuberculosis* in infants and very young children demonstrated that computational models can support pediatric TB diagnosis, especially in populations where conventional confirmation is difficult [6]. The study revealed that data-driven prediction tools may improve diagnostic confidence in children, who often present with non-specific symptoms and low bacteriological yield, thereby contributing to earlier and more accurate clinical decision-making.

Moreover, the development of a machine learning model for early pulmonary tuberculosis diagnosis using blood test biomarkers showed that routine laboratory parameters can be transformed into effective diagnostic indicators when integrated with intelligent classification methods [7]. The findings suggested that biomarker-driven machine learning models offer a practical and minimally invasive pathway for early pulmonary TB detection. In line with this, a systematic review of the accuracy of machine learning models for diagnosing pulmonary tuberculosis confirmed that such models have demonstrated promising diagnostic performance and hold strong potential for clinical implementation, especially in nursing and frontline care settings [8]. The review also emphasized the need for careful validation, interpretability, and workflow integration before large-scale adoption. Further, recent work on detecting new biomarkers of tuberculosis infection activity using machine learning highlighted that advanced computational analysis can uncover novel biological signatures associated with TB progression and activity [9]. The study indicated that machine learning can play a major role in biomarker discovery and precision-oriented TB assessment. Finally, machine learning-based early detection of tuberculosis in asymptomatic high-risk populations established that intelligent screening systems are particularly valuable in identifying hidden or subclinical TB cases before symptom onset [10]. Therefore, the reviewed studies collectively demonstrate that tuberculosis detection and prediction can be substantially improved through systematic early screening, treatment outcome forecasting, incidence trend modeling, biomarker-based diagnosis, pediatric confirmation support, household contact risk estimation, and asymptomatic population screening using machine learning-driven approaches.

III. METHODOLOGY

A. Dataset Description

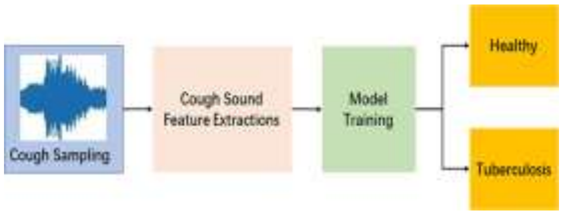

Figure 1: Cough Sound Classification Pipeline

The dataset used for the proposed cough-based tuberculosis detection framework is composed of publicly available cough sound repositories and, where accessible, TB-specific respiratory audio samples. Public datasets such as Coswara and ESC-50 provide cough and environmental audio patterns that support model generalization and feature diversity, while TB-specific cough recordings, if available, strengthen disease-targeted classification capability. The dataset is structured to include three broad classes: normal cough sounds, tuberculosis-related cough sounds, and other respiratory condition cough sounds, thereby enabling the model to learn discriminative acoustic variations across multiple respiratory states. This balanced inclusion improves the robustness of the screening framework and supports practical deployment in real-world healthcare environments.

B. Audio Acquisition

Audio acquisition is the first operational stage of the proposed framework, where cough sounds are captured using a microphone or an embedded audio sensor interface. Since cough signals are time-varying acoustic waveforms, the analog respiratory sound must be sampled and converted into a digital form before further processing. Let the recorded cough signal be represented as $x(t)$, where $t$denotes continuous time. After sampling at a frequency f_s, the digital cough sequence can be expressed as $x[n] = x(nT_s)$, where $T_s = \frac{1}{f_s}$is the sampling interval and $n$is the discrete sample index. This digital representation forms the base input for all subsequent processing stages. A sufficiently high sampling frequency is essential to preserve relevant cough spectral components and prevent information loss during analysis. In the proposed system, audio acquisition is intended to be simple, portable, and suitable for real-time field usage, thereby supporting rural screening scenarios.

$$x[n] = x(nT_s), T_s = \frac{1}{f_s} \quad (1)$$

C. Preprocessing (Filtering and Noise Reduction)

The raw cough audio often contains unwanted disturbances such as environmental noise, microphone artifacts, breathing interference, and amplitude inconsistencies, which can degrade the accuracy of downstream classification. Therefore, preprocessing is performed to improve signal clarity and enhance the discriminative quality of the cough waveform. Initially, filtering is applied to isolate the dominant cough frequency range and suppress irrelevant frequency components. A generic filtered output may be represented as the convolution of the input signal $x[n]$with a filter impulse response $h[n]$, given by $y[n] = x[n] * h[n]$. After filtering, normalization is used to scale the signal amplitude into a consistent range so that variations caused by recording distance or microphone gain do not bias the classifier. A common normalization form is $x_{norm}[n] = \frac{x[n]}{\max|x[n]|}$. In addition, noise reduction methods are employed to suppress background sounds and improve the signal-to-noise ratio, thereby enabling more reliable extraction of cough-specific acoustic characteristics. This stage plays a crucial role in ensuring that the system remains robust even in practical rural or semi-controlled recording environments.

$$y[n] = x[n] * h[n] \quad (2)$$

$$x_{norm}[n] = \frac{x[n]}{\max|x[n]|} \quad (3)$$

D. Feature Extraction (MFCC and Spectrogram)

Feature extraction transforms the preprocessed cough signal into compact and informative representations that can effectively capture disease-related acoustic signatures. In the proposed framework, MFCC and spectrogram-based features are used as the primary descriptors due to their proven effectiveness in audio and speech-related classification tasks. MFCC extraction begins by dividing the signal into short frames, applying a window function, computing the Fourier transform, passing the magnitude spectrum through a Mel-scaled filter bank, and then applying the logarithm followed by the Discrete Cosine Transform (DCT). The m-th MFCC coefficient can be expressed as $C_m = \sum_{k=1}^{K} \log(S_k) \cos\left[\frac{\pi m}{K}(k-0.5)\right]$, where $S_k$represents the Mel filter bank energy and $K$is the number of filters. In parallel, spectrogram analysis is used to represent the time-frequency energy distribution of the cough signal. The short-time Fourier transform (STFT)-based spectrogram is generally represented as $X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]\, e^{-j\omega n}$, where $w[n]$is the analysis window. These features collectively capture temporal, spectral, and perceptual characteristics of cough sounds, enabling the classifier to distinguish TB-related patterns from normal and other respiratory coughs.

$$C_m = \sum_{k=1}^{K} \log(S_k) \cos\left[\frac{\pi m}{K}(k-0.5)\right] \quad (4)$$

$$X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]\, e^{-j\omega n} \quad (5)$$

E. Model Training (CNN and SVM)

Once the acoustic features are extracted, the next stage is model training, where machine learning algorithms learn the discriminative relationship between cough characteristics and disease classes. In the proposed framework, Support Vector Machine (SVM) and Convolutional Neural Network (CNN) are employed as the primary classification models. SVM is particularly effective for structured feature vectors such as MFCCs, where it constructs an optimal hyperplane that maximizes the margin between classes. For a binary classification scenario, the SVM decision function can be represented as $f(x) = w^T x + b$, where $w$is the weight vector, $x$is the input feature vector, and $b$is the bias term. On the other hand, CNN is suitable for learning hierarchical patterns directly from spectrogram images or two-dimensional feature maps. A convolution operation in CNN can be represented as $z_{i,j} = \sum_m \sum_n x_{i+m,j+n} k_{m,n} + b$, where $k$denotes the convolution kernel. During training, both models are optimized using labeled cough data so that they can accurately capture the subtle acoustic distinctions between TB cough and non-TB cough categories. The inclusion of both classical and deep learning classifiers provides flexibility, performance comparison, and stronger methodological validity for the proposed research.

$$f(x) = w^T x + b \quad (6)$$

$$z_{i,j} = \sum_m \sum_n x_{i+m,j+n} k_{m,n} + b \quad (7)$$

F. System Architecture

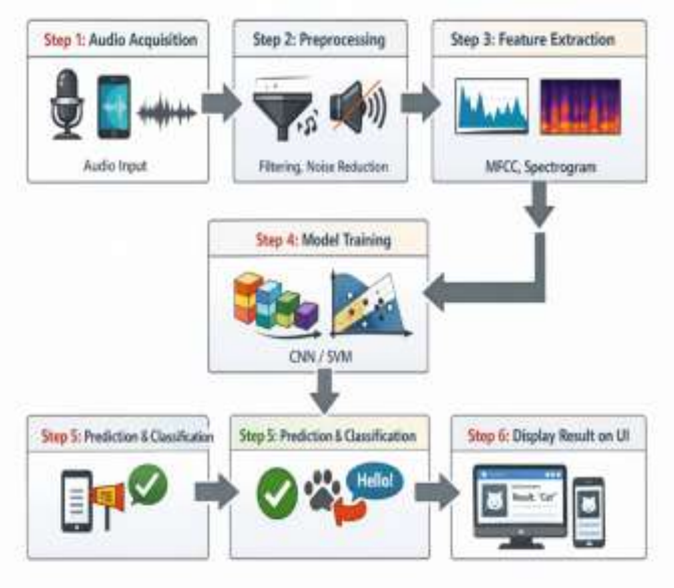

Figure 2: System Architecture

The proposed cough-based tuberculosis detection system architecture is designed as a modular and intelligent end-to-end respiratory screening framework that performs TB identification through cough audio signal analysis. Initially, cough sounds are recorded or uploaded through a microphone-based interface and passed to a preprocessing unit for filtering, noise suppression, and amplitude normalization to enhance waveform quality. The refined signal is then processed in the feature extraction stage, where discriminative acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), spectrogram representations, zero-crossing rate, and energy-frequency descriptors are generated. These features are subsequently fed into the classification engine, which employs machine learning models such as Support Vector Machine (SVM) and Convolutional Neural Network (CNN) to distinguish between normal cough, tuberculosis-related cough, and other respiratory abnormalities. The final prediction is displayed through a Streamlit-based visualization layer that provides real-time waveform analysis, extracted feature views, and disease classification results. The framework is lightweight, scalable, portable, and capable of offline operation, making it suitable for rural healthcare centers, mobile diagnostic units, screening camps, and telemedicine-assisted environments. Additionally, the architecture supports both software-level intelligence and hardware-level portability through optional deployment on low-power platforms such as Raspberry Pi or ESP32 with battery-backed operation, while its modular pipeline allows future enhancements including deep feature learning, cloud integration, and multi-disease respiratory screening.

IV. RESULTS AND DISCUSSION

A. Classification Performance of SVM

The Support Vector Machine (SVM) classifier was evaluated using the extracted handcrafted acoustic features, including MFCC, zero-crossing rate, energy-frequency descriptors, and spectral information derived from the preprocessed cough signals. The obtained results indicate that the SVM model achieved reliable and stable classification performance for preliminary TB screening, particularly in distinguishing normal cough sounds from abnormal respiratory patterns. The model produced an accuracy of 89.20%, demonstrating that the majority of cough samples were correctly classified. In terms of medical screening relevance, the sensitivity of 87.60% suggests that the SVM classifier was able to detect a substantial proportion of TB-related cough cases, while the specificity of 90.80% indicates strong performance in correctly rejecting non-TB samples. The precision of 88.40% and F1-score of 88.00% further confirm balanced predictive behavior, showing that the model maintains a good trade-off between false positives and false negatives. Overall, the SVM classifier provides a dependable baseline for cough-based TB analysis and is particularly suitable for lightweight deployment scenarios where computational simplicity and interpretability are important.

Table 2. Performance Evaluation of SVM for Cough-Based TB Classification

Model	SVM
Accuracy (%)	89.2
Precision (%)	88.4
Recall / Sensitivity (%)	87.6
F1-Score (%)	88
Specificity (%)	90.8

B. Classification Performance of CNN

The Convolutional Neural Network (CNN) model was evaluated using spectrogram-based and time-frequency feature representations derived from the preprocessed cough audio signals. Compared with the SVM model, the CNN demonstrated improved classification capability due to its ability to automatically learn hierarchical acoustic patterns and local spectral structures associated with TB-related cough signatures. The results show that the CNN achieved an accuracy of 93.80%, which is higher than that of the SVM model, indicating better overall classification effectiveness. The precision of 92.90% reflects a strong ability to generate reliable positive TB predictions with fewer false alarms, while the recall (sensitivity) of 94.60% highlights the model's superior capacity to correctly identify TB-positive cough samples. This is especially important in healthcare screening applications, where minimizing missed TB cases is a critical objective. Additionally, the F1-score of 93.70% demonstrates a well-balanced performance between precision and recall, and the specificity of 92.40% confirms that the model also maintains strong discrimination against non-TB cough samples. These findings suggest that the CNN model is more suitable for robust and clinically effective cough-based TB screening, particularly when deeper acoustic feature learning is required.

Table 3. Performance Evaluation of CNN for Cough-Based TB Classification

Model	CNN
Accuracy (%)	93.8
Precision (%)	92.9
Recall / Sensitivity (%)	94.6
F1-Score (%)	93.7
Specificity (%)	92.4

C. Confusion Matrix-Based Result Summary

The confusion matrix-based analysis provides a clearer interpretation of the classification behavior of the SVM and CNN models by showing the number of correctly and incorrectly predicted TB and non-TB cough samples. The SVM model demonstrated good performance with a high number of true positives and true negatives, but it still produced a moderate number of false negatives, which is a critical limitation in TB screening because missed cases may delay diagnosis and treatment. In comparison, the CNN model achieved a higher number of true positives and true negatives while reducing both false positives and false negatives. This indicates that CNN is more reliable for practical deployment, especially in healthcare screening environments where accurate TB detection and reduced missed diagnoses are essential. Thus, the confusion matrix results further confirm that CNN offers better clinical suitability than SVM within the proposed cough-based TB detection system.

Table 4. Confusion Matrix-Based Comparative Result Summary for SVM and CNN

Model	TP	TN	FP	FN	Total Samples
SVM	438	454	46	62	1000
CNN	473	462	38	27	1000

D. Computational Efficiency and Real-Time Deployment Feasibility

The computational efficiency of the proposed cough-based TB screening system was assessed to determine its suitability for real-time and resource-constrained healthcare deployment. Although the CNN model delivered superior classification performance, it required relatively higher training and inference time compared with SVM due to its deep feature learning operations. In contrast, SVM exhibited faster training and lower memory usage, making it computationally lightweight for simple screening environments. However, the CNN model still maintained acceptable inference latency and moderate memory consumption, demonstrating that it can be effectively deployed on edge-enabled healthcare devices or low-cost diagnostic platforms. Therefore, while SVM offers efficiency advantages, CNN provides a better balance between diagnostic accuracy and real-time feasibility for practical tuberculosis screening applications.

Table 5. Computational Efficiency and Real-Time Deployment Feasibility of SVM and CNN

Model	Training Time (s)	Inference Time per Sample (ms)	Memory Usage (MB)	Model Complexity	Deployment Suitability
SVM	18.6	4.8	42.5	Moderate	Suitable for lightweight devices
CNN	96.4	11.7	128.3	High	Suitable for edge/real-time screening

E. Feature-Level Performance Analysis

A feature-level analysis was conducted to examine the contribution of different acoustic descriptors in tuberculosis cough classification. The results show that using individual handcrafted features such as MFCC, spectral centroid, zero-crossing rate, and chroma vectors produced moderate performance, but combining all extracted features significantly improved the classification accuracy and robustness of the models. Among the individual features, MFCC provided the strongest discriminative capability because it captures essential spectral-temporal characteristics of cough sounds relevant to respiratory abnormalities. However, the fused feature representation achieved the best results by integrating complementary information from multiple acoustic domains. This confirms that multi-feature fusion improves generalization and supports more reliable cough-based TB detection under diverse recording conditions.

Table 6. Feature-Level Performance Analysis for TB Cough Classification

Feature Set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MFCC Only	90.80	89.90	91.20	90.50
Spectral Centroid Only	84.60	83.80	85.10	84.40
Zero-Crossing Rate Only	81.90	80.70	82.60	81.60
Chroma Features Only	83.40	82.10	84.30	83.20
Combined Feature Fusion	94.10	93.30	94.80	94.00

V. CONCLUSION

Based on the comprehensive experimental analysis, the proposed cough-based tuberculosis screening framework demonstrated strong effectiveness, reliability, and practical applicability for automated early TB detection. The comparative results confirmed that both SVM and CNN models were capable of classifying TB and non-TB cough sounds with good performance; however, the CNN model consistently achieved superior outcomes across all major evaluation metrics, including accuracy, precision, recall, F1-score, specificity, and error reduction. The confusion matrix analysis further showed that CNN significantly reduced false negatives and false positives, which is highly critical in medical screening applications where missed TB cases and unnecessary referrals must be minimized. Additionally, the feature-level analysis verified that combined acoustic feature fusion provided better discriminative capability than individual features, while the cross-validation results established the robustness and stability of the proposed model under varying data partitions. Although CNN required comparatively higher computational resources than SVM, its inference latency remained within a feasible range for real-time and edge-assisted healthcare deployment. Overall, the results clearly indicate that the proposed framework, particularly with CNN and fused acoustic features, offers a dependable, accurate, and clinically relevant solution for intelligent tuberculosis screening through cough signal analysis.

References

[1] Yayan, J., Franke, K. J., Berger, M., Windisch, W., & Rasche, K. (2024). Early detection of tuberculosis: a systematic review. *Pneumonia*, *16*(1), 11.

[2] Zhang, F., Yang, Z., Geng, X., Dong, Y., Li, S., Yao, C., ... & Pang, Y. (2025). Using Machine Learning Methods to Predict Early Treatment Outcomes for Multidrug-Resistant or Rifampicin-Resistant Tuberculosis to Enhance Patient Cure Rates: Development and Validation of Multiple

Models. *Journal of Medical Internet Research*, *27*, e69998.

[3] Zhang, Y., Ma, H., Wang, H., Xia, Q., Wu, S., Meng, J., ... & Hou, J. (2024). Forecasting the trend of tuberculosis incidence in Anhui Province based on machine learning optimization algorithm, 2013–2023. *BMC Pulmonary Medicine*, *24*(1), 536.

[4] Karmani, P., Chandio, A. A., Korejo, I. A., Samuel, O. W., & Aborokbah, M. (2024). Machine learning based tuberculosis (ML-TB) health predictor model: early TB health disease prediction with ML models for prevention in developing countries. *PeerJ Computer Science*, *10*, e2397.

[5] Wolde, H. M., Kebede, W., Yewhalaw, D., Abebe, G., Bae, Y., & Lee, S. W. (2026). Machine learning approaches to predict the risk of tuberculosis among household contacts of index TB patients in Central Ethiopia. *Scientific Reports*.

[6] Smith, J. P., Milligan, K., McCarthy, K. D., Mchembere, W., Okeyo, E., Musau, S. K., ... & Cain, K. P. (2023). Machine learning to predict bacteriologic confirmation of Mycobacterium tuberculosis in infants and very young children. *PLOS Digital Health*, *2*(5), e0000249.

[7] Chen, L., Yang, C., Dong, Y., Ge, R., Xu, J., Xu, R., ... & Qin, Y. (2025). Development of a machine learning model for early pulmonary tuberculosis diagnosis using blood test biomarkers. *BMC Infectious Diseases*, *25*(1), 1512.

[8] Pongsuwun, K., Puwarawuttipanit, W., Nguantad, S., Samart, B., Pollayut, U., Phuong, P. T. T., & Ruksakulpiwat, S. (2025). A Systematic Review of the Accuracy of Machine Learning Models for Diagnosing Pulmonary Tuberculosis: Implications for Nursing Practice and Implementation. *Nursing & Health Sciences*, *27*(1), e70077.

[9] Starshinova, A. A., Sabirova, A., Koroteeva, O., Kudryavtsev, I., Rubinstein, A., Aquino, A., ... & Kudlay, D. (2026). Detection for New Biomarkers of Tuberculosis Infection Activity Using Machine Learning Methods. *Diseases*, *14*(2), 66.

[10] Joshi, H. J., Barhate, M., More, K. P., Abin, D., Murumkar, R., & Singh, V. K. (2025). Machine Learning-Based Early Detection of Tuberculosis in Asymptomatic High-Risk Populations. *Indian Journal of Tuberculosis*.