Mrs.S.Subha, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Post-TB Lung Risk Forecaster: A Real-Time Clinical Decision Support System.

Mrs.S.Subha [1], Abinaya M [2], Dharshini S [3], Hari priya J [4]
Department of Computer Science and Engineering
Kongunadu College of Engineering and Technology
Tamilnadu,India

Abstract- Post-tuberculosis (post-TB) lung disease has emerged as a significant cause of long-term respiratory impairment, especially in regions with high TB prevalence. Detecting individuals at risk of developing chronic pulmonary complications—such as fibrosis, persistent airflow obstruction, and recurrent infections—remains a critical gap in clinical practice. This paper presents a Post-TB Lung Risk Forecaster, a real-time clinical decision support system aimed at predicting the likelihood of adverse lung outcomes in patients previously treated for tuberculosis. The system leverages machine learning techniques to analyze integrated data, including patient demographics, medical history, imaging results, lung function tests, and treatment responses, to produce personalized risk assessments. Its intuitive interface allows healthcare providers to obtain actionable insights during routine clinical workflows, supporting early intervention and tailored patient management. The model is developed and validated using longitudinal clinical datasets, showing strong predictive performance and practical relevance. By promoting early detection and preventive care, this system can help reduce long-term complications, enhance patient outcomes, and support efficient use of healthcare resources, contributing to improved post-TB care strategies.

Keywords: Post-tuberculosis lung disease, decision support system, machine learning, risk assessment, respiratory complications, real-time prediction, personalized healthcare.

I. INTRODUCTION

Tuberculosis (TB) continues to be a major global health concern, with millions of new infections occurring each year. While advancements in diagnostic methods and treatment regimens have improved cure rates, the long-term health effects experienced by many TB survivors are often underrecognized. Increasing evidence shows that a significant proportion of individuals who complete TB treatment develop lasting respiratory complications, collectively known as post-tuberculosis (post-TB) lung disease. These conditions may include pulmonary fibrosis, bronchiectasis, chronic airflow limitation, and a heightened risk of recurrent respiratory infections, all of which can severely impact quality of life and increase the demand for ongoing medical care.

The burden of post-TB lung disease is especially pronounced in regions with high TB prevalence, where healthcare systems may face challenges such as delayed diagnosis, insufficient follow-up, and limited awareness of long-term complications. Conventional clinical practice tends to emphasize microbiological cure, often overlooking the need for continued respiratory assessment after treatment completion. Consequently, many patients are only diagnosed with chronic lung damage at advanced stages, when interventions are less effective. Identifying at-risk individuals early is therefore crucial to enable timely management and reduce disease progression.

With the rapid evolution of digital healthcare technologies, new possibilities have emerged to enhance clinical decision-making processes. In particular, machine learning approaches have shown considerable promise in processing large and complex medical datasets to uncover patterns and predict future health outcomes. By combining multiple data inputs—such as patient demographics, clinical records, imaging findings, lung function tests, and treatment history—predictive systems can offer more accurate and individualized risk evaluations.

In this context, the present study proposes the Post-TB Lung Risk Forecaster, a real-time clinical decision support system developed to estimate the risk of adverse pulmonary outcomes in individuals with a history of TB. Utilizing advanced machine learning models, the system generates personalized risk scores that can assist clinicians in making evidence-based decisions during routine care. Its intuitive and accessible interface allows for easy integration into existing healthcare workflows, promoting efficient and timely clinical interventions.

By encouraging a shift from reactive treatment to proactive risk evaluation, this approach seeks to close the gap between TB recovery and long-term respiratory health management. Ultimately, the proposed system aims to enhance patient outcomes, minimize chronic complications, and contribute to more effective and resource-efficient healthcare delivery.

II. RELATED WORK

Predicting the Risk of Pulmonary Embolism in Patients with Tuberculosis Using Machine Learning Algorithms

Authors: Haobo Kong, Yong Li, Ya Shen, Jingjing Pan, Min Liang, Zhi Geng, Yanbei Zhang, et al.develops machine learning models to predict pulmonary embolism risk among tuberculosis patients. Using data from multiple clinical centers, the researchers applied algorithms such as Random Forest, Support Vector Machine, and XGBoost to identify high-risk individuals. Key predictors included D-dimer levels, smoking status, comorbidities, and clinical symptoms. The Random Forest model demonstrated superior performance with high accuracy and generalizability. The study emphasizes the importance of early risk stratification to prevent life-threatening complications. It highlights how machine learning can support clinical decision-making by integrating diverse patient data and generating reliable predictions for improved patient management.

Development of a Predictive Risk Model for Recurrence of Chronic Pulmonary Aspergillosis in Post-Tuberculosis Patients

Authors: Yi Ren, et al. focuses on predicting recurrence of chronic pulmonary aspergillosis (CPA) in post-TB patients using machine learning techniques. A retrospective dataset of 220 patients was analyzed, and LASSO regression identified key predictors such as COPD, age, emphysema, and treatment duration. Multiple algorithms were evaluated, with logistic regression achieving the best predictive performance. The model demonstrated strong calibration and clinical utility, aiding in personalized follow-up strategies. This study underscores the growing role of AI in managing post-TB complications and improving long-term patient outcomes by enabling early detection of recurrence risks.

Machine Learning Algorithms to Predict Treatment Success for Patients with Pulmonary Tuberculosis

Authors: Shaik Ahamed Fayaz, Lakshmanan Babu, Loganathan Paridayal, Mahalingam Vasantha, Palaniyandi Paramasivam, Karuppasamy Sundarakumar, Chinnaiyan Ponnuraja explores the use of machine learning models to predict treatment outcomes in pulmonary tuberculosis patients. By analyzing clinical and demographic variables, the study aims to identify factors influencing treatment success. Various algorithms were tested to improve predictive accuracy and support clinical decision-making. The findings suggest that machine learning can enhance early identification of patients at risk of poor outcomes, enabling targeted interventions. The study contributes to the broader application of AI in TB management and highlights its potential to improve treatment adherence and success rates in high-burden settings.

Predictive Machine Learning Models for Anticipating Loss to Follow-Up in Tuberculosis Patients

Authors: Jingfang Chen, Youli Jiang, Zhihuan Li, Mingshu Zhang, Linlin Liu, Ao Li, Hongzhou Lu, et al.develops predictive models to identify tuberculosis patients at risk of loss to follow-up during treatment. Using large-scale clinical datasets, multiple machine learning techniques were applied to detect patterns associated with treatment discontinuation. The models demonstrated strong predictive capability, enabling early intervention strategies to improve patient retention. By addressing adherence challenges, the research contributes to better treatment outcomes and reduced transmission rates. It highlights the importance of integrating predictive analytics into TB control programs for improved healthcare delivery and patient monitoring.

Post-Treatment Lung Tuberculosis Sequelae: A Clinical-Laboratory Nomogram to Predict Tissue Destruction

This study presents a nomogram-based predictive model to assess the risk of lung tissue destruction following TB treatment. Using clinical, laboratory, and imaging variables, the model identifies key predictors such as inflammatory markers, drug resistance, and delayed treatment initiation. The model achieved high predictive accuracy and demonstrated strong clinical utility. It provides a practical tool for early identification of patients at risk of severe post-TB lung damage, supporting personalized follow-up and intervention strategies. The study emphasizes the need for post-treatment monitoring and highlights the role of predictive tools in improving long-term respiratory outcomes.

Development of a Machine Learning Model for Early Pulmonary Tuberculosis Diagnosis Using Blood Biomarkers

Authors: Liangqiong Chen, et al. introduces a machine learning-based diagnostic model using routine blood biomarkers to detect pulmonary tuberculosis. The study analyzed large datasets and applied multiple algorithms, with Gradient Boosting Machine showing optimal performance. Key biomarkers included platelet-to-lymphocyte ratio and monocyte-to-lymphocyte ratio. The model demonstrated high specificity and reasonable sensitivity, making it suitable for low-resource settings. By enabling rapid and cost-effective diagnosis, the study supports early intervention and improved disease management. It also highlights the potential for integrating such tools into real-time clinical decision support systems.

A Comprehensive Study on Tuberculosis Prediction Models: Integrating Machine Learning into Epidemiological Analysis provides an extensive overview of machine learning approaches applied to tuberculosis prediction at a population level. It integrates epidemiological data with predictive modeling techniques to forecast TB incidence and trends. Various algorithms are compared to evaluate their effectiveness in handling large-scale datasets. The study emphasizes the importance of data visualization and model interpretability in public health decision-making. It demonstrates how predictive analytics can support policy planning, resource allocation, and disease control strategies, offering a broader perspective on TB management beyond individual patient care.

The Efficacy of Machine Learning Models in Lung Cancer Risk Prediction with Explainability

Authors: Refat Khan Pathan, Israt Jahan Shorna, Md. Sayem Hossain, Mayeen Uddin Khandaker, Huda I. Almohammed, Zuhal Y. Hamd evaluates machine learning models for predicting lung cancer risk while emphasizing explainability. Various algorithms were applied to clinical datasets, and interpretability techniques such as SHAP were used to understand model predictions. The research demonstrates how explainable AI can enhance clinician trust and decision-making. Although focused on lung cancer, the methodologies are highly relevant to post-TB lung disease prediction. The study highlights the importance of transparent models in healthcare, ensuring that predictions are not only accurate but also clinically interpretable and actionable.

Prediction of Long-Term Mortality Using Machine Learning in Interstitial Lung Disease Patients

Authors: D. Sun, Y. Wang, Q. Liu, applies machine learning techniques to predict long-term mortality in patients with interstitial lung disease. By analyzing clinical and demographic variables, the study develops models capable of identifying high-risk patients. The findings demonstrate the effectiveness of machine learning in handling complex, longitudinal healthcare data. Although not specific to TB, the study provides valuable insights into chronic lung disease prognosis and risk stratification. It reinforces the potential of predictive analytics in improving patient management and guiding clinical decision-making in respiratory diseases.

Predictive Analysis of Tuberculosis Treatment Outcomes Using Machine Learning: A Large-Scale Study

Authors: SeshaSai Nath Chinagudaba, Darshan Gera, Krishna Kiran Vamsi Dasu, Uma Shankar S, Kiran K, Anil Singarajpure, Shivayogappa U, Somashekar N, Vineet Kumar Chadda, Sharath B N study uses machine learning to predict TB treatment outcomes using national-level data. With over 500,000 patient records, the research applies advanced algorithms to generate accurate risk predictions. The model achieved high recall and AUC scores, demonstrating strong performance in identifying patients at risk of poor outcomes. The study also incorporates natural language processing to enhance predictive capability. It highlights the scalability of machine learning in public health and its potential to improve TB control programs through data-driven decision-making and personalized patient care.

III. PROPOSED METHOD

The proposed system, titled Post-TB Lung Risk Forecaster, is a real-time clinical decision support system designed to predict the likelihood of long-term pulmonary complications in patients who have completed tuberculosis (TB) treatment. The system aims to assist healthcare professionals in early identification of high-risk individuals, enabling timely intervention and improved disease management.

The architecture of the system consists of four primary modules: data acquisition, data preprocessing, predictive modeling, and user interface. The data acquisition module collects patient-specific information from electronic health records (EHRs), including demographic details (age, gender), clinical history (smoking status, comorbidities), treatment outcomes, radiological findings (chest X-rays or CT scans), and pulmonary function test results such as spirometry. This comprehensive dataset ensures a holistic assessment of each patient’s health status.In the preprocessing stage, the collected data undergo cleaning, normalization, and transformation to handle missing values and inconsistencies. Feature selection techniques are applied to identify the most relevant variables influencing post-TB lung complications. This step enhances model performance and reduces computational complexity.

The core of the system lies in the predictive modeling module, where machine learning algorithms such as Random Forest, Gradient Boosting, and Logistic Regression are employed. These models are trained on longitudinal datasets to capture patterns associated with adverse pulmonary outcomes, including fibrosis, chronic airflow obstruction, and recurrent infections. The system generates a personalized risk score for each patient, categorizing them into low, moderate, or high-risk groups. Model performance is evaluated using metrics such as accuracy, precision, recall, and area under the ROC curve (AUC) to ensure reliability and robustness.

The final module is a user-friendly interface designed for clinicians. It provides real-time risk predictions along with visual summaries and key contributing factors, enabling transparent and explainable decision-making. The interface can be integrated into existing hospital information systems, allowing seamless access during routine clinical workflows.Additionally, the system incorporates continuous learning capabilities, where new patient data can be used to update and refine the model over time. This ensures adaptability to evolving clinical patterns and population characteristics.

Overall, the Post-TB Lung Risk Forecaster offers a proactive approach to post-TB care by combining advanced analytics with practical clinical utility. It supports personalized treatment planning, improves patient monitoring, and contributes to reducing the long-term burden of respiratory complications.

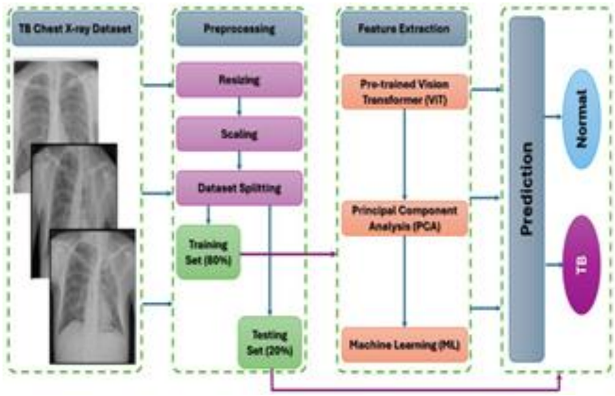


Fig.1.System Architecture

The Post-TB Lung Risk Forecaster is structured into several functional modules, each designed to ensure efficient data handling, accurate prediction, and seamless clinical usability. These modules work collaboratively to deliver real-time risk assessment for post-tuberculosis lung complications.

The first module is the Data Acquisition Module, which is responsible for collecting patient-related information from multiple sources such as electronic health records (EHRs), laboratory systems, and radiology databases. It gathers comprehensive data including demographic details (age, gender), clinical history (smoking habits, prior illnesses), TB treatment outcomes, imaging results, and pulmonary function test values. This module ensures that diverse and relevant datasets are captured for meaningful analysis.

The second module is the Data Preprocessing Module, which prepares raw data for analysis. It involves data cleaning to remove inconsistencies, handling missing values through imputation techniques, and normalization to standardize data formats. Feature selection methods are applied to identify the most significant variables affecting post-TB lung disease risk. This module enhances the quality of input data, thereby improving the efficiency and accuracy of the predictive models.

The third module is the Predictive Modeling Module, which forms the core of the system. It utilizes machine learning algorithms such as Random Forest, Logistic Regression, and Gradient Boosting to analyze patterns in the processed data. These models are trained using historical and longitudinal datasets to predict the likelihood of complications like fibrosis, airflow obstruction, and recurrent infections. The module generates a risk score for each patient and classifies them into different risk categories, supporting early clinical decision-making.

The fourth module is the Risk Visualization and User Interface Module, designed to present the prediction results in an accessible and interpretable manner. It provides clinicians with real-time dashboards, graphical summaries, and key contributing factors influencing the risk score. This ensures transparency and supports evidence-based decisions at the point of care.

Finally, the Model Update and Feedback Module enables continuous system improvement. It incorporates new patient data and clinician feedback to retrain and optimize the predictive models over time. This adaptive capability ensures that the system remains relevant and accurate in changing clinical environments.

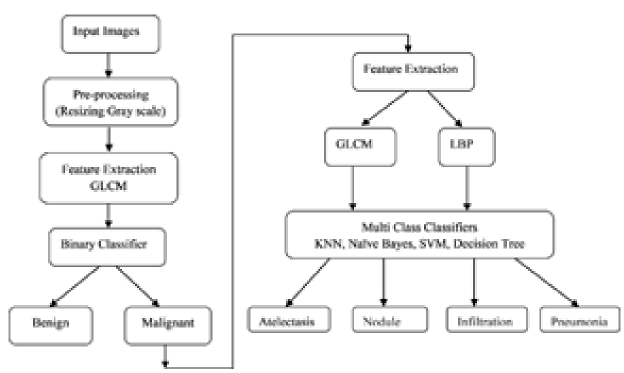


Fig.2.Methodology workflow of the A Real-Time Clinical Decision Support System

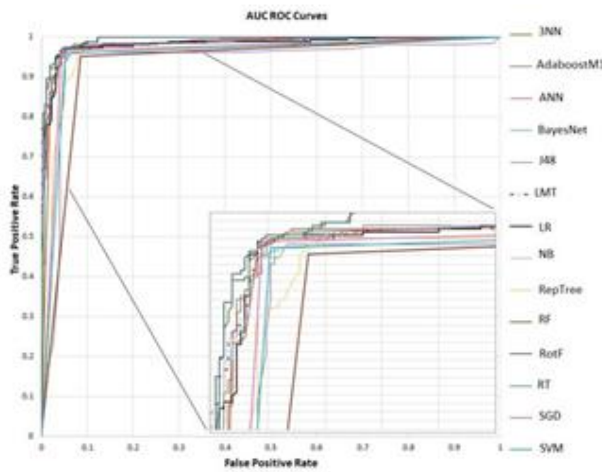
Overall Working Flow of the Proposed System:
The Post-TB Lung Risk Forecaster operates through a well-defined workflow that efficiently converts patient data into meaningful clinical insights for predicting long-term respiratory complications after tuberculosis treatment. The system follows a step-by-step pipeline to ensure accuracy, reliability, and ease of use in clinical settings.

The process begins with the data collection stage, where patient information is obtained from various sources, including electronic health records, laboratory systems, radiology reports, and pulmonary function tests. This data typically includes demographic details, medical history, TB treatment outcomes, imaging findings, and spirometry measurements. All collected information is then fed into the system for further analysis.

During the data preprocessing phase, the raw data is refined to improve its quality and usability. This involves handling missing values through suitable imputation methods, eliminating inconsistencies, and detecting outliers. The data is then standardized and normalized to ensure uniformity across all inputs. Additionally, feature selection techniques are used to identify the most important variables that influence post-TB lung disease risk.

The refined data is subsequently processed in the predictive modeling phase, where trained machine learning algorithms evaluate patterns and relationships within the dataset. Based on this analysis, the system generates an individualized risk score for each patient, indicating their probability of developing complications such as lung fibrosis, airflow obstruction, or recurrent infections. Patients are then grouped into different risk levels to support clinical prioritization.

The outcomes are presented via a user-friendly interface, allowing clinicians to access real-time predictions, visual summaries, and key influencing factors. This facilitates informed and timely decision-making. A continuous feedback mechanism enables the system to learn from new data and clinical outcomes, ensuring ongoing improvement, adaptability, and sustained effectiveness in real-world healthcare environments.



$$P(y=1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}}$$

Fig.3.Performance Evaluation of A Real-Time Clinical Decision

Support System

This equation represents the logistic regression model used to estimate the probability of a patient developing post-TB lung complications. The output is a value between 0 and 1, interpreted as the risk probability. Here, x_1,x_2,...,x_ndenote input features such as age, smoking status, lung function values, and imaging findings, while β_0,β_1,...,β_nare model coefficients learned during training. The sigmoid function ensures the output remains bounded within a probabilistic range. This equation is particularly useful for binary classification, such as predicting whether a patient is at high risk or not. It is widely used due to its interpretability and efficiency in clinical decision support systems.

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x)$$

This equation defines the prediction mechanism of the Random Forest algorithm, where the final output y ̂is obtained by averaging predictions from multiple decision trees. Each h_t (x)represents an individual decision tree trained on a subset of the data, and Tis the total number of trees in the forest. By combining multiple models, Random Forest reduces overfitting and improves prediction accuracy. In the context of post-TB risk forecasting, each tree evaluates patient features such as clinical history and test results, and the aggregated output provides a robust risk estimate. This ensemble approach enhances stability and handles complex, nonlinear relationships effectively in healthcare datasets.

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x)$$

This equation represents the Gradient Boosting model, where the final prediction F(x)is constructed as a weighted sum of multiple weak learners h_m (x). Each model is trained sequentially to correct the errors of the previous one, with γ_mrepresenting the weight assigned to each learner. This iterative approach allows the model to gradually improve performance by focusing on difficult cases. In the proposed system, Gradient Boosting helps capture subtle patterns in patient data, such as interactions between clinical variables and imaging results. It is particularly effective for structured medical datasets and contributes to highly accurate risk prediction for post-TB lung complications.

V. CONCLUSION

The Post-TB Lung Risk Forecaster presents an effective approach to addressing the often-overlooked burden of post-tuberculosis lung disease by enabling early prediction of long-term respiratory complications. By integrating clinical, demographic, and diagnostic data with advanced machine learning techniques, the system provides accurate and personalized risk assessments for patients who have completed TB treatment. Its real-time clinical decision support capability allows healthcare providers to make informed decisions, implement timely interventions, and design tailored follow-up strategies.

The modular architecture and user-friendly interface ensure seamless integration into existing healthcare systems, enhancing usability and efficiency. Additionally, the system's adaptability through continuous learning improves its long-term relevance and performance. Overall, this approach shifts the focus from reactive treatment to proactive care, ultimately contributing to better patient outcomes, reduced healthcare burden, and improved management of post-TB lung disease in diverse clinical settings.

VI. FUTURE WORK

The proposed Post-TB Lung Risk Forecaster offers a strong foundation for predictive and preventive healthcare; however, several enhancements can be explored in future work to improve its effectiveness and scalability. One important direction is the integration of advanced deep learning techniques, particularly for analyzing radiological images such as chest X-rays and CT scans. This would enable automated feature extraction and improve prediction accuracy.

Another potential improvement involves incorporating real-time data from wearable devices and remote monitoring systems to track patient respiratory parameters continuously. This would allow dynamic risk assessment and early detection of deterioration. Expanding the dataset to include multi-center and geographically diverse populations can further enhance model generalizability and robustness.Additionally, implementing explainable AI (XAI) techniques will improve transparency, helping clinicians better understand the reasoning behind predictions and increasing trust in the system.

Integration with mobile health applications could also make the system more accessible in remote and resource-limited settings.Finally, prospective clinical validation and deployment in real-world healthcare environments are essential to evaluate system performance, usability, and impact on patient outcomes, ensuring its readiness for large-scale adoption.

REFERENCES

1. X. Wei, N. M. Norsuddin, H. A. Hamid, M. I. Azmi, G. Zhang and J. Tian, "Accuracy of Machine Learning in Identifying Drug Resistance in Tuberculosis: A Systematic Review and Meta-Analysis," Health Sci. Rep., 2025.
2. M. Singh, G. V. Pujar, S. A. Kumar, M. Bhagyalalitha, H. S. Akshatha, B. Abuhaija, A. R. Alsoud, L. Abualigah, N. M. Beeraka and A. H. Gandomi, "Evolution of Machine Learning in Tuberculosis Diagnosis: A Review," Electronics, vol. 11, 2022.

3. M. Cabanillas-Lazo, C. Quispe-Vicuña, M. Pascual-Guevara, J. Barja-Ore, M. E. Guerrero, A. Munive-Degregori and F. Mayta-Tovalino, "Bibliometric Analyses of AI Applications in Tuberculosis," Indian J. Med. Res., 2022.
4. [4] R. Rahman, A. Khandakar, M. A. Kadir, et al., "Reliable Tuberculosis Detection Using Chest X-Ray with Deep Learning," IEEE Access, vol. 8, 2020.
5. A. Wong, J. R. H. Lee, H. Rahmat-Khah, A. Sabri, A. Alaref and H. Liu, "TB-Net: A Deep CNN for Tuberculosis Detection," Frontiers in AI, 2022.
6. R. R. Shettigar and Padmanayana, "Efficient Deep Learning Models for Automated TB Diagnosis," in Lecture Notes in Networks and Systems, Springer, 2025.
7. H. M. Sunny, "Machine Learning-Based Tuberculosis Detection Using C4.5 Decision Tree," IJISAE, vol. 13, 2025.
8. D. Hindustani, S. Hindustani and P. Nguyen, "Machine Learning for Tuberculosis Detection Using Deep Models," arXiv, 2025.
9. S. Sathitratanacheewin and K. Pongpirul, "Deep Learning for TB Chest X-Ray Classification," arXiv, 2018.
10. [10] A. T., S. Natarajan and K. N. B. Murthy, "A Data Mining Approach for TB Diagnosis," arXiv, 2011.
11. A. A. Septiandri, Aditiawarman, R. Tjiong, E. Burhan and A. Shankar, "Cost-Sensitive ML for TB Screening," arXiv, 2020.
12. World Health Organization, "Global Tuberculosis Report," WHO, 2021.
13. J. Melendez, B. van Ginneken, P. Maduskar, R. H. Philipsen, H. Ayles and C. I. Sanchez, "CAD for Tuberculosis Detection," IEEE Trans. Med. Imaging, 2016.
14. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep CNNs," NIPS, 2012.
15. K. Gasmi, I. B. Ltaifa, G. Lejeune, H. Alshammari, L. B. Ammar and M. A. Mahmood, "Deep Neural Networks for Medical Question Answering," Cybernetics & Systems, 2021.
16. N. Walia, H. Singh, S. K. Tiwari and A. Sharma, "Decision Support System for TB Diagnosis," Int. J. Soft Computing, 2015.
17. C. S. Poornimadevi and H. Sulochana, "Automatic Detection of TB Using Image Processing," in Proc. IEEE WiSPNET, 2016.
18. R. Hooda, A. Mittal and S. Sofat, "Automated TB Classification Using Deep Learning," Multimedia Tools Appl., 2019.
19. Y. Arzhaeva, L. Hogeweg, P. de Jong, M. Viergever and B. van Ginneken, "CAD of TB Using Dissimilarity-Based Classification," MICCAI, 2009.
20. X. B. Wei, N. M. Norsuddin, M. I. Azmi and H. A. Hamid, "AI in TB Imaging: Bibliometric Analysis," Radiography, 2026.
21. T. M. Mitchell, Machine Learning, McGraw-Hill, 1997.
22. R. C. Gonzalez and R. E. Woods, Digital Image Processing, Prentice Hall, 2008.
23. Y. Zhou, Medical Imaging: Principles and Applications, IntechOpen, 2019.
24. M. Ullah, M. Bari, A. Ahmed and S. Naveed, "Lung Disease Detection Using Image Processing: A Review," Mehran Univ. Res. J., 2019.
25. C. Dicente Cid, V. Liauchuk, V. Kovalev and H. Müller, "ImageCLEF Tuberculosis Challenge Overview," CLEF Working Notes, 2017.