



(RESEARCH ARTICLE)



MACHINE LEARNING MODELS FOR EARLY DETECTION AND PUBLIC HEALTH MANAGEMENT OF TUBERCULOSIS

Yahya Khan ¹, Lisa Sullivan ^{2,*}, Danial khan ³, Sumeet Kumar ⁴, Zulqarnain ⁵ and Muhammad Usman ⁶

¹ Infection Prevention and Control Assistant, Care Concern Group.

² Associate Dean, Boston University School of Public Health.

³ Boston University School of Public Health.

⁴ General Medical Officer, Work place Park Hyatt Hadahaa Clinic Maldives.

⁵ Yasir Health City Complex, Pakistan.

⁶ PGR Gastroenterology PIMS P.043000, Pakistan Institute of Medical Sciences Islamabad.

* Corresponding Author

ORCID Details

Yahya Khan: <https://orcid.org/0009-0000-2878-292X>

World Journal of Biology Pharmacy and Health Sciences, 2026, 27(02), 087–098

Publication history: Received on 05 July 2026; revised on 11 August 2026; accepted on 13 August 2026

Article DOI: <https://doi.org/10.30574/wjbphs.2026.27.2.0425>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Background: Tuberculosis (TB) continues to be one of the most pressing health issues worldwide, and in 2024, 10.6 million new cases were reported, resulting in 1.3 million deaths. However, the early and accurate diagnosis is crucial to limit transmission but the existing means are expensive, time consuming, and impractical in resource-limited environments. The objective of the present study was to build a quick and low-cost diagnostic model based on the use of blood biomarkers routinely available in the laboratory. Tuberculosis (TB) continues to be one of the more deadly diseases, particularly in the developing world. This study is aimed at developing a prototype solution where primary signs, symptoms and risk factors of TB would be identified at an early stage and machine learning (ML) predictive algorithms would be applied to them.

Methods: Data for 818 confirmed TB cases and 2,618 healthy controls were analyzed in a retrospective manner. Since the dataset is imbalanced, the ROSE technique was used to balance the training set. Seven machine learning algorithms were trained, and feature selection was done using LASSO regression and forward selection to determine the most predictive variables. SHAP analysis was used to increase interpretability of the model, and the final predictive model was implemented as an interactive Shiny web application to make it easier to be used in the clinic.

Result: The Gradient Boosting Machine (GBM) model demonstrated superior performance on the test set, achieving an area under the curve (AUC) of 0.821, specificity of 85.7%, and sensitivity of 64.9%. SHAP analysis highlighted platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), and platelet distribution width (PDW) as the most influential predictors. Adjusting the classification threshold to 0.24 improved sensitivity to 82.6% while maintaining acceptable specificity (58.9%), underscoring the model's potential utility as a screening tool. The accompanying Shiny application enhances accessibility and practical deployment in clinical settings.

Conclusion: This study presents a robust and interpretable GBM-based diagnostic model leveraging routine hematological parameters to provide a rapid, low-cost TB screening tool suitable for resource-constrained environments. The model's high specificity may reduce unnecessary confirmatory testing, while its core predictors offer biological insights into TB-associated inflammatory processes. The interactive web application facilitates integration into clinical workflows, supporting early detection and improved TB management.

Keywords: Tuberculosis Diagnosis, Machine Learning, Good Health and Well-Being, Optimal ML Model for TB Diagnosis

1 INTRODUCTION

Tuberculosis (TB) is a contagious disease that ranks among the top ten causes of death in the world and is a leading cause of illness and death in many developing countries. It is an infectious disease caused by *Mycobacterium tuberculosis*, which can be transmitted through air by an infected individual when they cough, sneeze, talk, or spit [1]. TB is a pulmonary disease by default, but it is also an extrapulmonary disease when it affects other organs. TB can be latent without any clinical symptoms, or it can be active with clinical symptoms. TB is an ever-growing disease in Pakistan, ranking 5th in the 30 high TB burden countries (World Health Organization, 2024). The burden of TB in Pakistan has shown a decline, but the early identification of suspected cases is crucial in decreasing TB's complications, case burden, transmission, and mortality [2,3]. The rapid development of digital technologies has the potential to improve the discovery of diseases and enhance clinical decision-making. One of the examples to strengthen digital disease detection is Healthcare Informatics (HI). HI is the integration of different domains, such as health, IT, computing, telecommunications, and clinical data with the objective of improving healthcare systems [4]. The U.S. National Library of Medicine describes HI as a field of research aiming to improve healthcare systems by implementing modern tools and technologies through the design and development of new systems in healthcare environments [5]. HI integrates different domains to improve the management of patient data, advance clinical decision-making and the efficiency of healthcare systems [6,7]. In the domain of Healthcare Informatics, Machine Learning (ML), a prominent subset of Artificial Intelligence (AI), has rapidly developed and grown to provide new capabilities in predictive analytics and automated diagnostic processes for diseases. ML affords systems the ability to develop and assess algorithms and learn autonomously from datasets by identifying trends, building models for the future, and yielding decisions and conclusions related to new and future cases, all without the need for extensive and exhaustive coding [8,9]. In the intersection of computational sciences and statistical modeling, ML can demonstrate the different and often abstract covariate relationships in patient populations [10]. For these reasons and many more, the use of ML is being considered more often in the evaluation and construction of processes for the prediction and diagnosis of diseases, as well as the design and evaluation of the clinical risks and treatment options, and clinical support decision making [11].

An obvious synergy between ML and TB (Tuberculosis) screening is the timely detection of at-risk populations, which has the potential to foster prompt diagnostic and therapeutic measures. Traditional TB diagnostic processes may include a clinical evaluation, imaging studies, a variety of laboratory tests, and other procedures; however, the development and implementation of an intelligent screening mechanism using readily accessible and less resource intensive clinical data may support healthcare personnel in the identification and stratification of patients to target additional testing procedures [12,13]. It is the expressed intent of such a mechanism to not substitute clinical or laboratory diagnostic processes; rather, it is an improvement and support for advanced screening processes and the facilitation of clinical decision making. With these objectives in mind, this research focuses on developing an automated Machine Learning-based Tuberculosis (ML-TB) prediction system for the early detection of TB cases based on the identification of major signs and symptoms of TB and associated risk factors [14]. The targeted geographical location of the study is Sindh, Pakistan, where TB remains a significant public-health problem. The author compiled primary data from hospitals based on a self-administered TB screening survey, which was subsequently analyzed using the machine-learning platform in Python. A wide array of machine-learning algorithms was created and tested for their capability to recognize individuals that are at high risk of developing TB [15-16]. The key originality of this work is that it addresses four critical areas. First, the research constructs a formal framework for an ML-based TB screening model with an early detection functionality. Second, the research produces a primary dataset from hospitals based on a formalized TB screening survey. Third, the research assesses the TB-related primary signs, symptoms, and risk factors and analyzes the data using machine-learning constructs in Python. Fourth, the research offers probable findings grounded in the specific context of Sindh, Pakistan, thus providing the local, contextual scholarship regarding technology-supported TB screening services [17,18].

These findings indicate that utilizing clinically pertinent TB indicators in the modeling with ML algorithms can yield excellent predictive outcomes. Early identification of key symptoms and risk factors can help to diagnose suspected TB cases. Accordingly, the ML-TB framework can be used as a TB screening and diagnostic support tool. It enables the clinician to focus on the high-risk population in order to carry out the clinical examination and diagnostic tests. In the long run, this framework can help to improve TB control by facilitating treatment and decreasing the spread of the disease and its associated complications. The following sections of the paper describe relevant literature to support the

research and identify existing gaps in knowledge. The subsequent section contains the proposed ML-TB methodology and modeling, data collection, and preprocessing as well as feature selection, and model evaluation and validation. The section after that focuses on the results of the experiments and the assessments of machine learning models. The last section has the most significant results along with the restrictions, how this research can be used in practice, and future study proposals.

2 MATERIALS AND METHODS

The study was conducted with quantitative approach, cross sectional, and analytical type of research with positivist paradigm and deductive approach. A quasi-structured TB screening questionnaire was used to gather primary data which included patient data, TB related questions, additional clinical data, and feedback. The majority of items were dichotomous (Yes/No) and two were partially open-ended. Pre-testing was used to pre-test the questionnaire prior to data collection and random probability sampling was used. The Institutional Bioethical Committee, University of Sindh, Jamshoro provided ethical clearance (Ref. No. ORIC/SU/311). The total questionnaires were distributed equally in three hospitals of Sindh, Pakistan: Institute of Chest Diseases (TB Sanatorium), Kotri; Liaquat University of Medical and Health Sciences, Jamshoro; and Civil Hospital, Hyderabad. The number of questionnaires returned from these was 950. After excluding 190 incomplete records, were retained for analysis.

The predictive variables were cough, cough duration, cough type, mucus, chest pain, breathing state, body temperature, chills, pulmonary effusion, erythrocyte sedimentation rate (ESR), diet, physique, energy adequacy, smoking, crowding and exposure to TB related risk factors. Data was initially cleaned in MS Excel and later analyzed in Python (Jupyter Notebook) with the help of libraries such as Pandas, NumPy, Scikit-learn, Plotly and Matplotlib. The data preprocessing steps included incomplete record removal, exploratory analysis, categorical encoding, correlation analysis, separation of the predictor variable and the target variable, and feature scaling. Pandas was used to encode categorical variables and Pearson correlation analysis was used to analyze the relationship and possible multicollinearity between the predictors. StandardScaler from Scikit-learn was used to standardize features with different numerical ranges.

The dataset was divided into using Scikit-learn's `train_test_split` function. Five supervised machine learning algorithms were implemented: The model was optimized using grid search and cross-validation was employed to ensure the model was robust. Model performance was assessed. Recall was also highlighted as being of particular importance as it is a marker of the model's capacity to accurately detect suspected cases of TB. Additionally, feature-importance analysis was conducted using Gini importance, permutation importance and SHAP values to identify the most important predictors of TB.

2.1 Predictive Model

The proposed ML-TB predictive model was developed as a supervised binary classification model where the clinical signs, symptoms and risk factors of each participant was used as an input for the model to predict whether an individual was or was not a patient with ML-TB. The predictor vector comprised 16 variables: cough, duration of cough, type of cough, presence of mucus, chest pain, state of breathing, body temperature, chills, pulmonary effusion, erythrocyte sedimentation rate (ESR), diet, physique, energy adequacy, smoking, crowding and TB exposure. Target variable was recoded as a binary outcome with $Y=1$ (TB suspected) and $Y=0$ (not suspected).

For each participant i , the predictive framework can be expressed as:

$$X_i = [x_{i1}, x_{i2} \dots \dots x_{i16}]$$

And

$$Y_i = f_m(X_i, \theta_m)$$

Where X_i represents the set of TB-related predictors, Y_i is the predicted TB status, f_m denotes the selected machine-learning classifier, θ_m and represents the parameters learned during model training. Five classifiers were investigated:

$$m \in \{DT, GNB, LRC, AdaBoost, NN\}$$

Categorical variables were numerically coded and the predictor variables were standardized using StandardScaler. The full data set was then partitioned. The hyperparameters were optimized by the grid search method and the cross-validation method was used to make the model robust.

The overall predictive process can therefore be represented as:

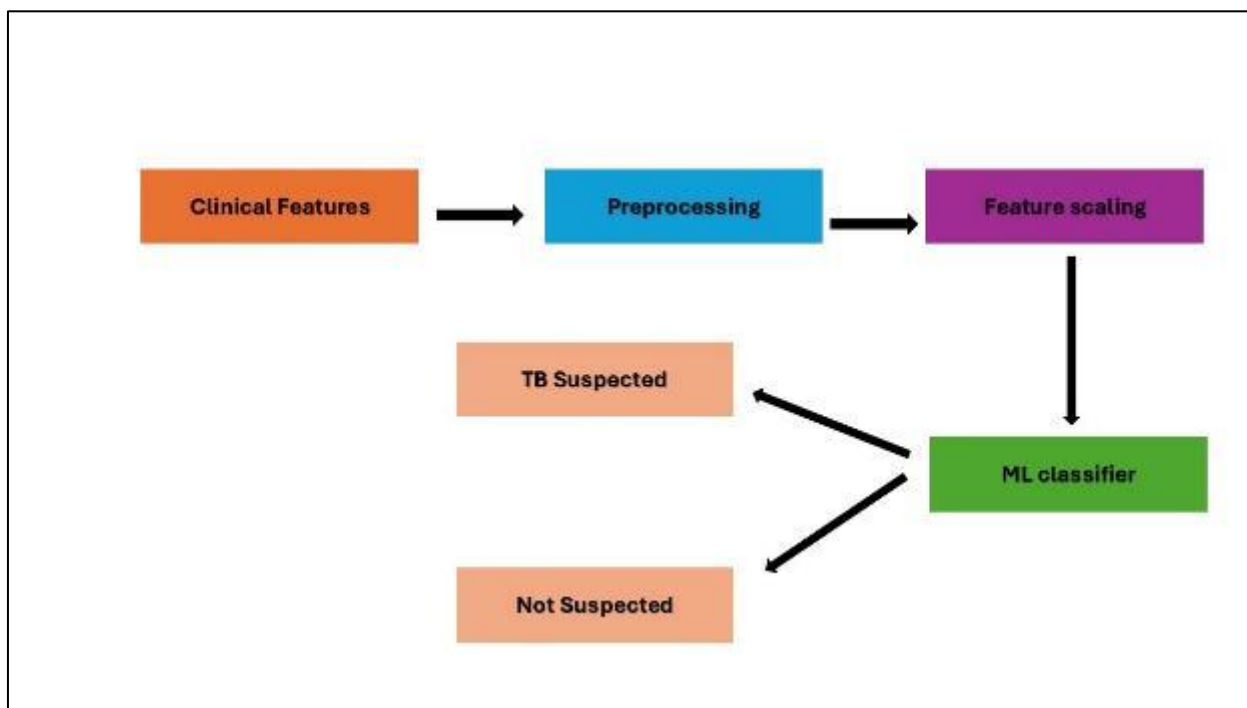


Figure 1: ML-TB predictive model

3 RESULTS

The accuracy, precision, recall, and F1 score were used to assess the predictive ability of the five machine-learning models. The comparative results demonstrated that all the classifiers performed well with relatively high results, but there were differences in the individual evaluation metrics. For easy graphical illustration and interpretation, the numerical results obtained from the original study are reported in separate manner below.

Table 1: Accuracy Performance of the ML-TB Prediction Models

Machine Learning Model	Accuracy (%)
Decision Tree (DT)	92.11
Gaussian Naïve Bayes (GNB)	89.04
Logistic Regression Classifier (LRC)	90.35
AdaBoost	93.42
Neural Network (NN)	92.98

The accuracy of the five machine-learning models' classification results is shown in table 1. AdaBoost had the highest accuracy, meaning that it classified the highest percentage of TB suspected and non-suspected cases. Neural Network had been closely with an accuracy. and Decision Tree had. Logistic Regression achieved an accuracy of and Gaussian Naïve Bayes had the lowest accuracy. However, the accuracy of all the models was very close to or greater than 89% indicating their potential to predict TB.

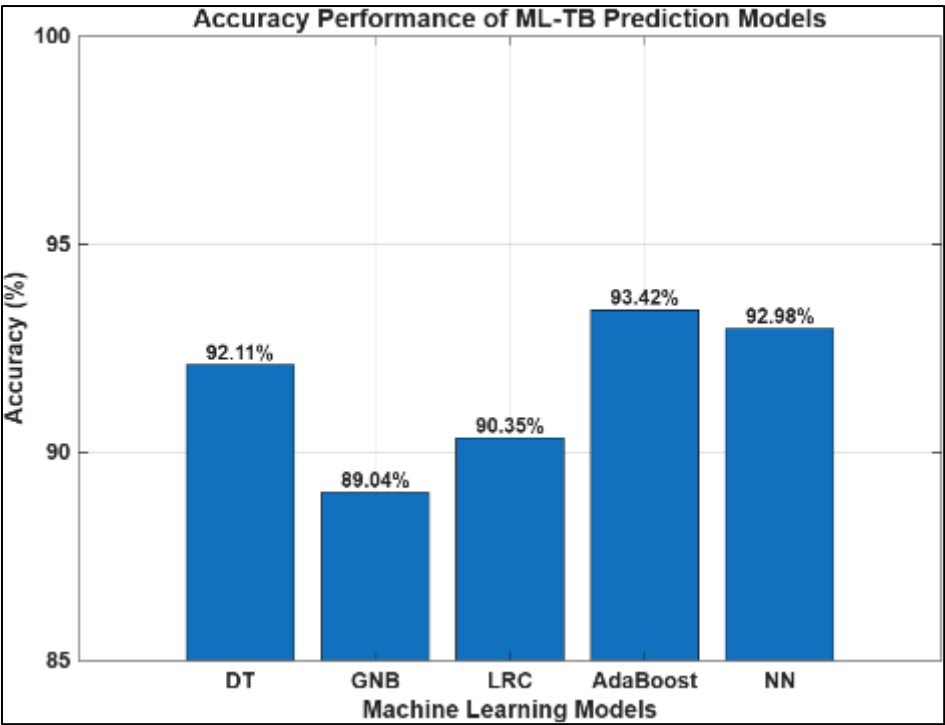


Figure 2: Accuracy performance of ML-TB Prediction Models

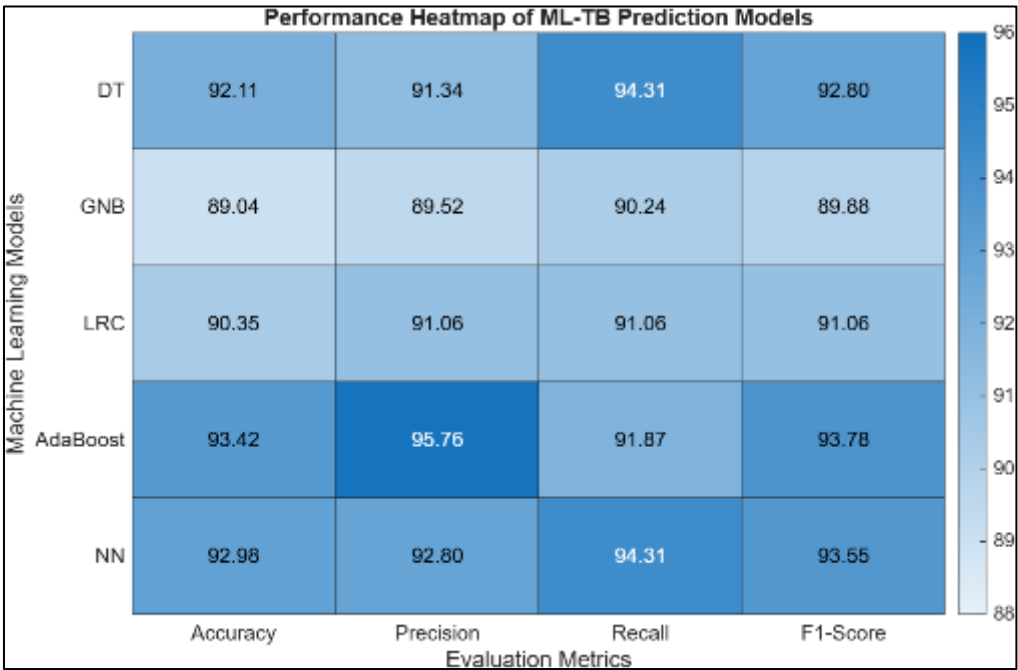


Figure 3: Performance Heatmap of ML-TB prediction Models

Table 2: Precision Performance of the ML-TB Prediction Models

Machine Learning Model	Precision (%)
Decision Tree (DT)	91.34
Gaussian Naïve Bayes (GNB)	89.52
Logistic Regression Classifier (LRC)	91.06
AdaBoost	95.76
Neural Network (NN)	92.80

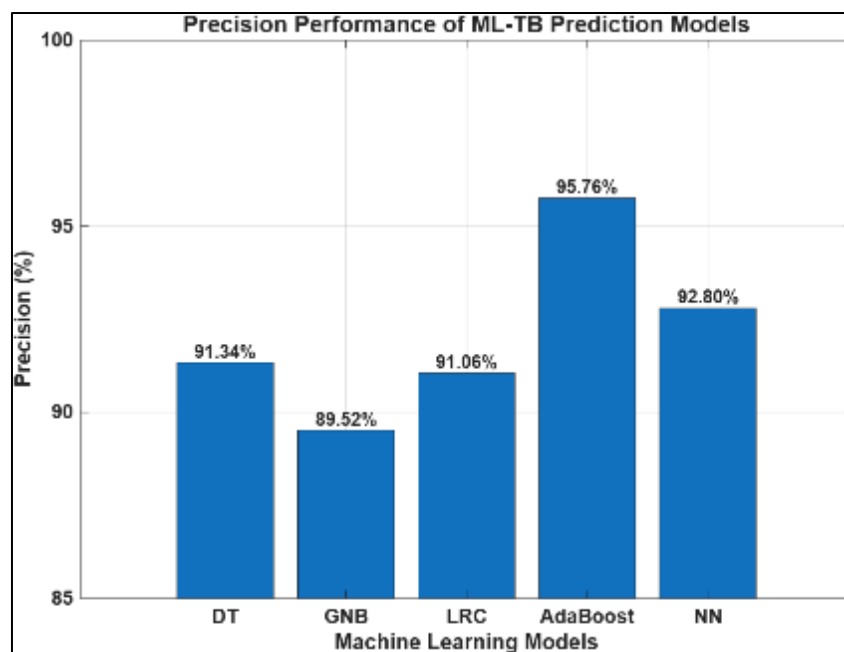


Figure 4: Precision performance of ML-TB prediction models

The precision of the five classifiers is shown in table 2. AdaBoost achieved the highest precision of, meaning a high percentage of the cases which were classified as TB suspected by this model were accurately identified. The second highest precision was achieved by Neural Network, with Decision Tree and Logistic Regression coming after. Gaussian Naïve Bayes had the lowest precision with. The absolute accuracy of AdaBoost is in line with its confusion matrix, which had only five false positive predictions.

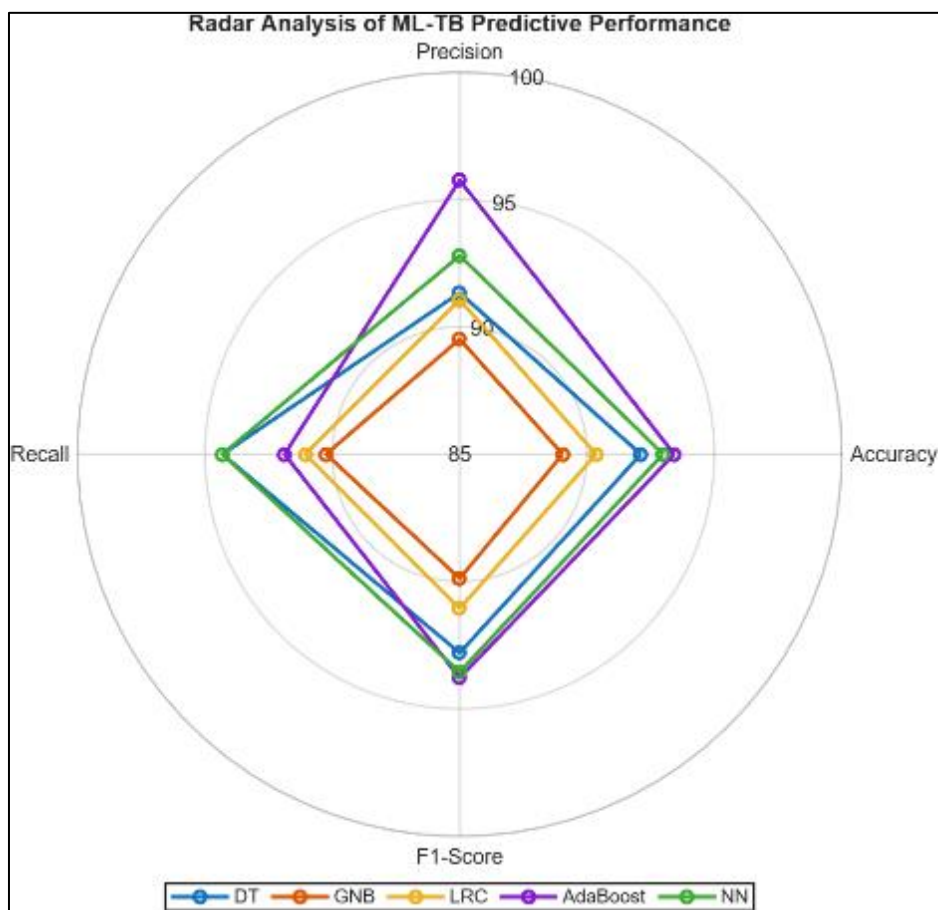
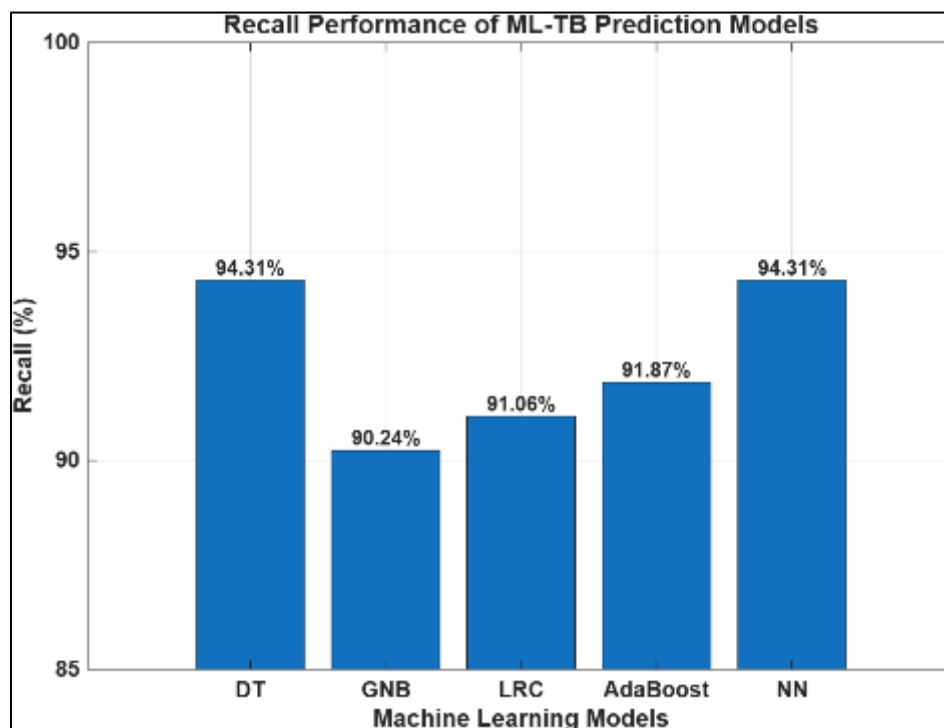


Figure 5: Radar Analysis of ML-TB prediction models

Table 3: Recall Performance of the ML-TB Prediction Models

Machine Learning Model	Recall (%)
Decision Tree (DT)	94.31
Gaussian Naïve Bayes (GNB)	90.24
Logistic Regression Classifier (LRC)	91.06
AdaBoost	91.87
Neural Network (NN)	94.31

**Figure 6: Performance of ML-TB prediction models**

The recall or sensitivity of the models that are implemented are summarized in Table 3. Decision Tree and Neural Network had the best recall of (94.31%), meaning that they were more successful in capturing the actual TB suspected cases. AdaBoost performed best with. Logistic Regression with and Gaussian Naïve Bayes with were the other two best-performing classifiers. Recall is especially important when screening for TB as increasing sensitivity decreases the chance of potentially suspected cases being wrongly identified as negative. For this reason, the study revealed that the model of DT and NN are better if sensitivity is emphasized.

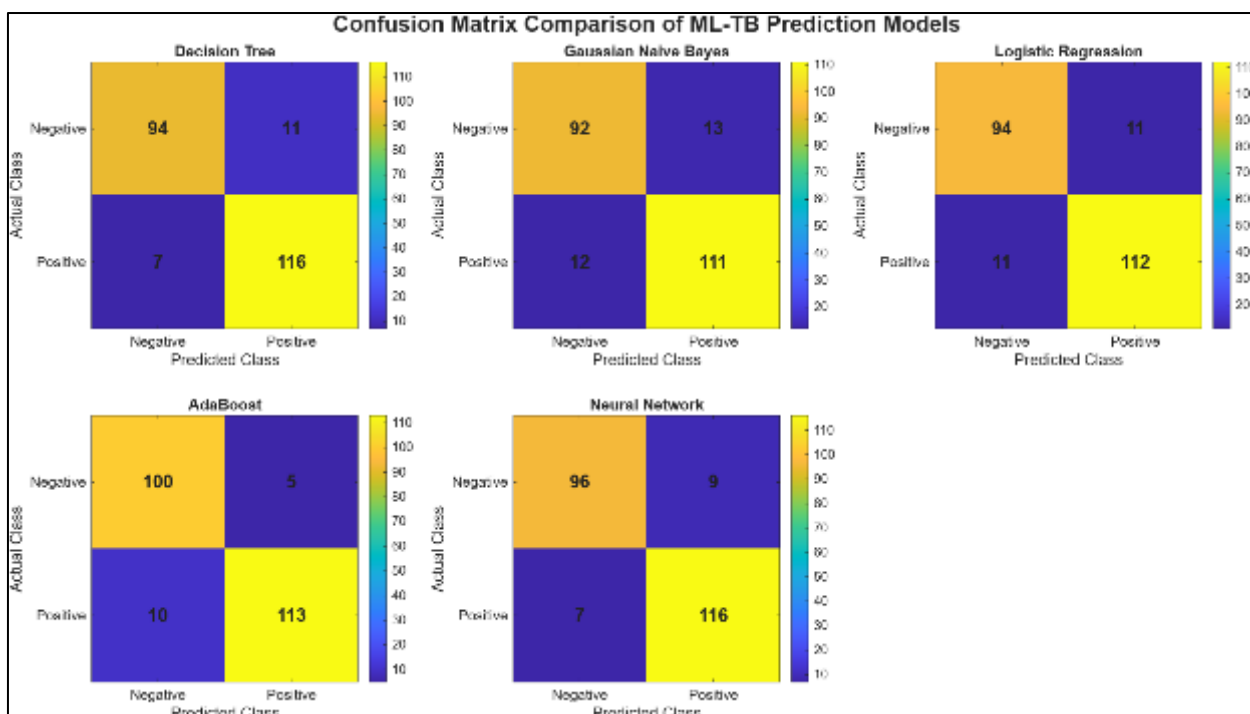


Figure 7: Confusion matrix comparison of ML-TB prediction models

Table 4: F1-Score Performance of the ML-TB Prediction Models

Machine Learning Model	F1-Score (%)
Decision Tree (DT)	92.80
Gaussian Naïve Bayes (GNB)	89.88
Logistic Regression Classifier (LRC)	91.06
AdaBoost	93.78
Neural Network (NN)	93.55

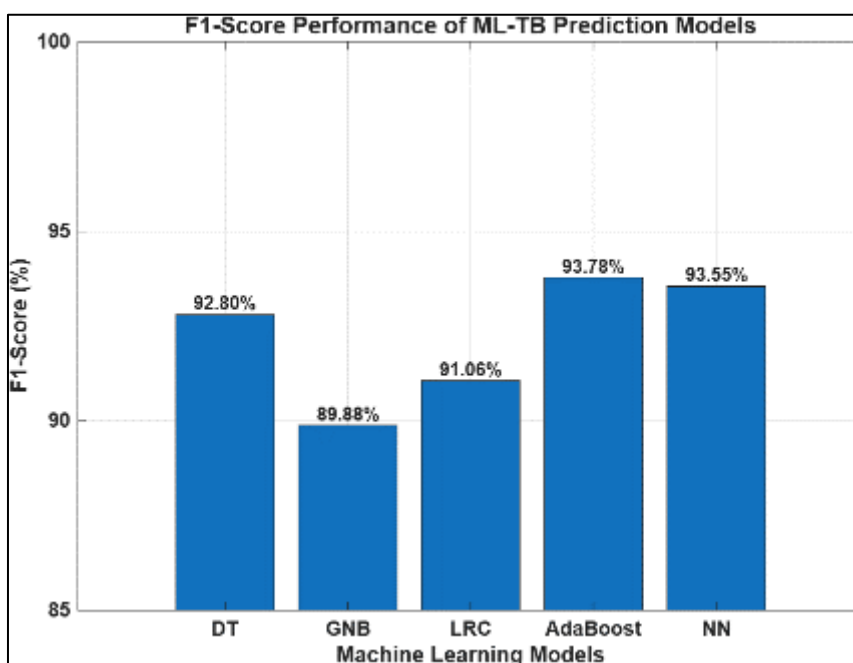


Figure 8: F1 Score performance of ML-TB prediction models

The balance between precision and recall is shown in the F1-score in Table 4. AdaBoost had the best overall performance on F1 score with a value of 0.8, indicating a best balance between correctly detecting TB suspected cases and reducing the number of false positive predictions. Neural Network had a very similar F1-score of, while Decision Tree had. Logistic Regression resulted in a score of and Gaussian Naïve Bayes a lowest F1 score. The findings further validate the overall excellent performance of AdaBoost and verified the classification capacity of NN and DT as good as AdaBoost.

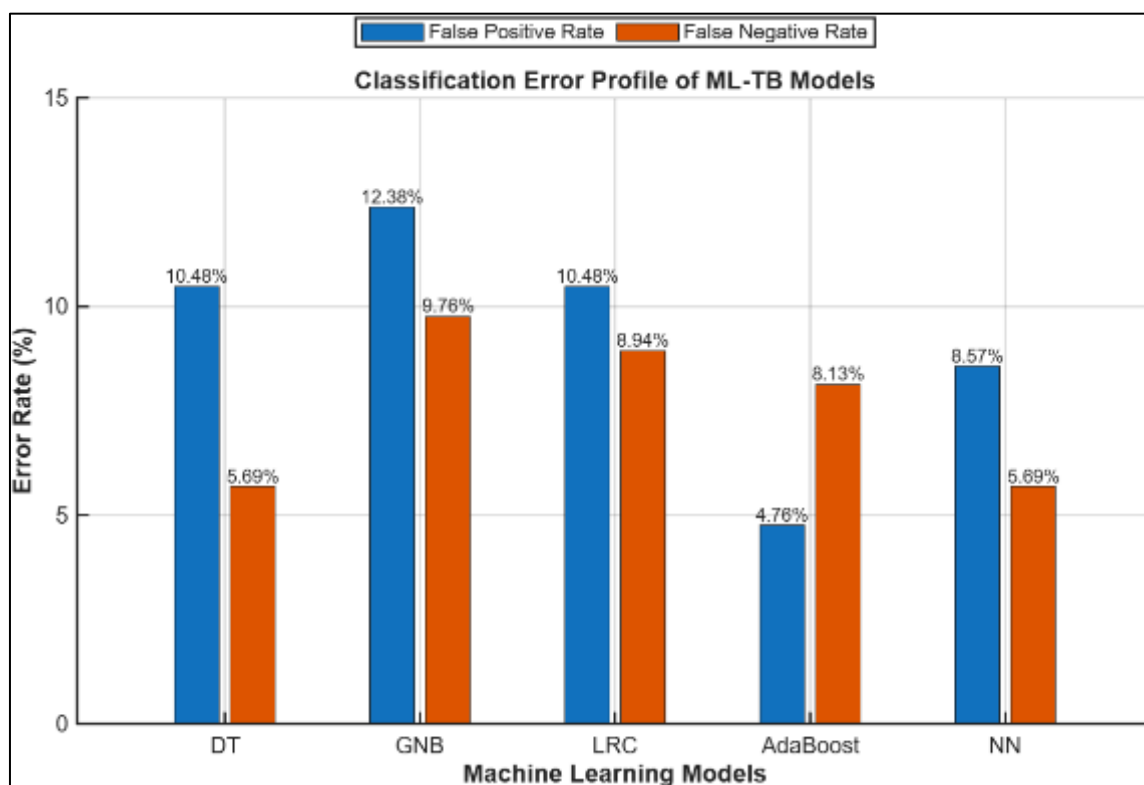


Figure 9: Classification error profile of ML-TB models

4 DISCUSSION

In the present study, the potential of five supervised machine learning algorithms namely Decision Tree (DT), Gaussian Naïve Bayes (GNB), Logistic Regression Classifier (LRC), Adaptive Boosting (AdaBoost) and Neural Network (NN) for early detection of suspected tuberculosis cases based on primary clinical signs, symptoms and risk factors were evaluated. The overall performances of all five models were also relatively good with accuracies ranging from 89.04% to 93.42%. AdaBoost resulted in the best overall accuracy (93.42%), best precision (95.76%) and best F1-score (93.78%) and NN had the best recall (94.31%). The results show that no single classifier performed best in each evaluation measure and it is concluded that when choosing classifiers, the clinical goal must be taken into account and not just the accuracy of the classifier. The best overall classification accuracy was obtained by AdaBoost out of the classifiers evaluated. The accuracy (93.42%) means that it correctly classified a large proportion of both suspected and non-suspected TB cases and precision (95.76%) means that it had a high likelihood of giving a correct classification for non-TB cases. This can be seen from the confusion matrix used, where AdaBoost produced 100 T.N. (True Negative), 113 T.P. (True Positive) and only five F.P (False Positive) and 10 F.N (False Negative). DT had 11 false positives and seven false negatives, NN had nine false positives and seven false negatives, LRC had 11 false positives and 11 false negatives and GNB had 13 false positives and 12 false negatives. AdaBoost exhibited a positive relationship between accuracy of classification and false positive control [19-21].

The additional confusion-matrix analyses based on these reported values further elucidate the behaviour of the classifiers. The estimated specificity of AdaBoost was 95.24% and the false positive rate was 4.76%, showing that it has a good performance in discriminating among the non-suspected individuals. The balanced accuracy of its derived model was **93.55%** and was also maximum with this model compared to the other five models. NN achieved a balanced accuracy of approximately 92.87%, followed by DT at 91.92%, LRC at 90.29%, and GNB at 88.93%. These values are secondary values that have been calculated from the confusion matrices, and are not additional metrics explicitly reported by the original authors. However, they agree with the result that AdaBoost gives a good tradeoff between the two measures. However, if the test is meant for screening purposes, the most important factor is that of recall or

sensitivity; a test that returns a negative may result in a person who may be infected not being called for further testing. Both DT and NN had the highest recall (94.31%) with only seven false-negative predictions. The false negative rates of these were around 5.69% which was better than GNB, LRC and AdaBoost. This discovery also sheds light on why DT and NN were appealing options in the original study when it was important to maximize identification of potential TB cases. From a practical standpoint, if the primary goal is to minimize the number of false-positive classifications with an emphasis on achieving high overall predictive balance, then AdaBoost is likely to be preferred; if sensitivity is more important than DT and NN are likely to be considered. This is especially applicable for screening systems where there are differing outcomes for missed cases and unnecessary referrals. These trade-offs continue to be reflected in the F1 score results. AdaBoost recorded the highest F1-score of 93.78% followed by NN with 93.55% and DT with 92.80%. The difference between the F1-score of AdaBoost and NN is small, meaning that both models kept a good precision/recall balance when detecting the suspected cases and minimizing the possibility of false positive results. In the extra comparative analysis performed in this study when all four principal parameters (accuracy, precision, recall, and F1-score) are taken together, the means of the performance scores were around 93.71%, 93.41%, 92.64%, 90.88% and 89.67% for the AdaBoost, NN, DT, LRC, and GNB, respectively. The descriptive mean should not be considered as a substitute for the standard description metrics of the models, but it gives a compact picture of the overall relative performance of the models [21-24].

The learning characteristics of the models may also help to explain the performance differences between the models. AdaBoost uses the idea of sequential boosting, i.e., giving more weight to the observations that are hard to classify, and this is likely to be the reason for its generally high accuracy. The high recall and F1-score of the Neural Network, on the other hand, might be due to the ability of the algorithm to learn the complex nonlinear relationships among multiple signs and risk factors. Although relatively simple, Decision Tree was able to outperform other models, suggesting that interpretable models can be effective for symptom-based TB screening. While the accuracy, precision, recall and F1-score of GNB were the lowest of the classifiers assessed, they were still close to 90%. Its relative performance may be explained by the fact that it is based on the independence assumptions that may not hold true for clinical variables that are correlated among each other. This is in line with the correlation analysis showed in the study. There was a strong correlation between cough type and cough duration (0.73), and a correlation of 0.68 between cough and cough duration, and several other TB related variables were associated with each other. Moderate relationships were also found between ESR and mucus and pulmonary effusion and body temperature. The inter-relationships between the different TB-related manifestations indicate that no single manifestation is independent of the others which might hinder methods that rely on more feature-independence assumptions, but might support algorithms that can model interactions or even non-linear patterns [25]. Clinically meaningful insight into the process of prediction can also be gained from the feature importance assessment. The study used the Gini importance, permutation importance, and SHAP analysis and found the predictors of suspected TB that had the greatest impact to be persistent cough and weight loss. Thus, the ability to obtain readily available clinical manifestations and risk factors that include sufficient predictive information to support preliminary screening is supported by this result, which is in line with the central premise of the proposed ML-TB framework. It is important to note that such a model should be used as a screening-decision support tool rather than a diagnostic tool. Suspected TB cases should undergo a proper clinical evaluation and testing.

The results obtained are also broadly similar to previous machine learning studies of TB. Yahiaoui, Er and Yumusak (2017) reported 96.68% accuracy when using an SVM based approach; Shahaboddin et al (2014) reported 99.14% accuracy using a hybrid approach that combines artificial immune systems, genetic algorithms and neural networks, which was more computationally intensive. Previous research on neural network and artificial immune systems has been around 90% accurate while Alcantara et al. (2017) reached an 89.6% accuracy with a deep convolutional neural network and mobile health technology. The remaining image-based studies mentioned in the paper had an accuracy of around 84.6% to 98.7%. However, caution should be taken when comparing the results of the aforementioned studies directly owing to the significant differences in the sample sizes, input features, imaging modalities, patient population, algorithms and validation procedures used. The importance of this study is less that the universal superiority was claimed and more that the symptom and risk-factor based data, which is easily available, would have competitive predictive performance in the specific context of Sindh, Pakistan. The other added value of the study is the fact that it has used data from the primary hospital and not only international datasets or chest radiographs, which are available in public. After discarding the incomplete observations, 760 complete observations were left in the final dataset, from three hospitals in Sindh. There was also a relatively balanced distribution of the target distribution of around 48% TB-suspected and 52% non-suspected cases. This gives development of models a valuable local context, especially as the authors state that Pakistan and especially Sindh is an underrepresented context in previous TB research based on ML models. However, the predictive power seen in this data set would need to be replicated in independent hospitals and geographically separate populations before it could be expanded to a wider population.

Statistically significant differences between classifiers in terms of accuracy and precision scores ($p < 0.05$) were also reported and confidence interval was calculated to evaluate robustness of classifiers in the study. The reported results do not include details of statistical comparison procedure; however, numerical confidence intervals are provided. Thus, these statistical results are to be taken with a grain of salt. Future research needs to report the exact statistical test used,

confidence intervals, repeated cross-validation results and uncertainty around each of the performance estimates in order to provide better evidence that observed differences in performances between classifiers are not attributable to sampling variation. There are a few caveats to note as well. The methodological limitations mentioned in the original paper are the relatively limited feature set and the lack of important evaluation measures like the receiver operating characteristic area under the curve (ROC-AUC) and calibration measures. Furthermore, while grid search and cross validation were employed, these are not fully described in the reported paper, nor are the results of the final optimized hyperparameters, neural-network architecture or random seed. There was no report on external validation. Restrictions from these factors reduce the extent to which the models can be reproduced and the ability to infer the accuracy of the models if they are applied to entirely new clinical cohorts. The proposed framework should thus be further tested in large, multi-center studies, and in a more varied patient population in future studies. Beyond accuracy, precision, recall, and F1-score, there is a need to incorporate other metrics into evaluation such as ROC-AUC, specificity, negative predictive value, measures of calibration, confidence intervals, and clinical metrics relevant to the clinical decision. There is a need for external and prospective validation prior to clinical use, especially. The paper also suggests expansion of extrapulmonary TB, increasing the sample size, using automated machine-learning techniques, reinforcement learning and finally the creation of a practical implementation of the ML-TB diagnostic support system. In general, the results show that machine-learning models using primary signs, symptoms, and risk factors have a lot of potential to assist in early TB detection. AdaBoost had the best overall predictive balance with NN and DT having the highest sensitivity. These results, therefore, highlight the need to choose the model based on the clinical goal: AdaBoost could be beneficial if overall classification accuracy and the reduction of false positive (FP) results are most important; DT and NN could be beneficial if the reduction of missed suspected cases (MS) is the most important goal. If validated externally, having larger datasets, a thorough evaluation of calibration and incorporation into the clinical workflow, these ML-based systems may offer valuable assistance with early detection of TB, especially in resource-limited and TB-endemic environments.

5 CONCLUSION

In this study, it has been shown that the supervised machine learning methods can be used to perform early screening of tuberculosis based on easily available clinical signs, symptoms and risk factors. This data set consisted of 760 full patient records from three hospitals in Sindh, Pakistan and five machine learning classifiers were developed and compared: Decision Tree (DT), Gaussian Naïve Bayes (GNB), Logistic Regression Classifier (LRC), Adaptive Boosting (AdaBoost) and Neural Network (NN). The results showed that the data on symptoms and risk factors can be informative and useful for prediction purpose and can help health-care workers identify people who may need more investigation for suspected TB. The classification results of all five models were good, ranging from 89.04% to 93.42% accuracy. AdaBoost demonstrated the best overall performance, achieving the highest accuracy (93.42%), precision (95.76%), and F1-score (93.78%). DT and NN had the highest recall of 94.31% which showed a better ability to correctly identify suspected TB cases. These results suggest that choosing a model for use should be guided by the clinical application. In screening applications, where minimizing missed suspected cases is important, DT and NN might be better choices when a high overall predictive balance and reduction of false positive classifications are not important, while AdaBoost can be useful. Primary clinical manifestations, such as persistent cough and weight loss were also found to be influential predictors of TB in the study. This study is relevant to the context of Pakistan, specifically in the province of Sindh, where the data from the hospital was obtained from the local population and not only from international public datasets or complex imaging modalities. The potential of the proposed ML-TB framework for supporting early TB screening in high burden and resource-limited healthcare settings is thus low cost and easily accessible. The developed system needs to be considered a clinical decision support tool, screening tool at a preliminary level, and not as a substitute for confirmatory TB diagnosis. People who were suspected of having the disease would still need clinical evaluation, laboratory, imaging, or other validated diagnostic tests.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] Alcantara, M. F., Cao, Y., Liu, C., Liu, B., Brunette, M., Zhang, N., ... & Curioso, W. H. (2017). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú. *Smart Health*, 1, 66-76.
- [2] Dongardive, J., Xavier, A., Jain, K., & Abraham, S. (2011, July). Classification and rule-based approach to diagnose pulmonary tuberculosis. In *International Conference on Advances in Computing and Communications* (pp. 328-339). Berlin, Heidelberg: Springer Berlin Heidelberg.

- [3] Santos, A. D., Pereira, B. D. B., Seixas, J. D., Mello, F. C. Q., & Kritski, A. L. (2007). Neural networks: an application for predicting smear negative pulmonary tuberculosis. In *Advances in statistical methods for the health sciences: applications to cancer and AIDS studies, genome sequence analysis, and survival analysis* (pp. 275-287). Boston, MA: Birkhäuser Boston.
- [4] El-Solh, A. A., Hsiao, C. B., Goodnough, S., Serghani, J., & Grant, B. J. (1999). Predicting active pulmonary tuberculosis using an artificial neural network. *Chest*, 116(4), 968-973.
- [5] Elveren, E., & Yumuşak, N. (2011). Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. *Journal of medical systems*, 35(3), 329-332.
- [6] Er, O., Temurtas, F., & Tanrıku, A. Ç. (2010). Tuberculosis disease diagnosis using artificial neural networks. *Journal of medical systems*, 34(3), 299-302.
- [7] Er, O., Yumusak, N., & Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, 37(12), 7648-7655.
- [8] Er, O., Yumusak, N., & Temurtas, F. (2012). Diagnosis of chest diseases using artificial immune system. *Expert Systems with Applications*, 39(2), 1862-1868.
- [9] Holzinger, A. (2016). Machine learning for health informatics. In *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges* (pp. 1-24). Cham: Springer International Publishing.
- [10] Hwang, S., Kim, H. E., Jeong, J., & Kim, H. J. (2016, March). A novel approach for tuberculosis screening based on deep convolutional neural networks. In *Medical imaging 2016: computer-aided diagnosis* (Vol. 9785, pp. 750-757). SPIE.
- [11] Islam, S. S., Haque, M. S., Miah, M. S. U., Sarwar, T. B., & Nugraha, R. (2022). Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study. *PeerJ Computer Science*, 8, e898.
- [12] Karmani, P., Chandio, A. A., Karmani, V., Korejo, I. A., & Chandio, M. S. (2020). Taxonomy on healthcare system based on machine learning approaches: tuberculosis disease diagnosis.
- [13] Karmani, P., Chandio, A. A., Korejo, I. A., & Chandio, M. S. (2018, October). A review of machine learning for healthcare informatics specifically tuberculosis disease diagnostics. In *International conference on intelligent technologies and applications* (pp. 50-61). Singapore: Springer Singapore.
- [14] Kumar, P., & Clark, M. L. (2012). *Kumar and Clark's clinical medicine E-Book*. Elsevier health sciences.
- [15] Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2), 574-582.
- [16] Li, Y., Wang, B., Wen, L., Li, H., He, F., Wu, J., ... & Hou, D. (2023). Machine learning and radiomics for the prediction of multidrug resistance in cavitary pulmonary tuberculosis: a multicentre study. *European Radiology*, 33(1), 391-400.
- [17] Lopes, U. K., & Valiati, J. F. (2017). Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in biology and medicine*, 89, 135-143.
- [18] Omisore, M. O., Samuel, O. W., & Atajeromavwo, E. J. (2017). A Genetic-Neuro-Fuzzy inferential model for diagnosis of tuberculosis. *Applied Computing and Informatics*, 13(1), 27-37.
- [19] Parry, D. (2014). Health informatics. In *Springer Handbook of Bio-/Neuroinformatics* (pp. 555-564). Berlin, Heidelberg: Springer Berlin Heidelberg.
- [20] Shamshirband, S., Hessam, S., Javidnia, H., Amiribesheli, M., Vahdat, S., Petković, D., ... & Kiah, M. L. M. (2014). Tuberculosis disease diagnosis using artificial immune recognition system. *International journal of medical sciences*, 11(5), 508.
- [21] Tasci, E., Uluturk, C., & Ugur, A. (2021). A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Computing and Applications*, 33(22), 15541-15555.
- [22] The World Health Organization. 2024a. Tuberculosis profile: Afghanistan. Retrieved July 10, 2022.
- [23] The World Health Organization. 2024b. World health organization—Eastern Mediterranean region. Retrieved April 07, 2018.
- [24] Yahiaoui, A., Er, O., & YUMUŞAK, N. (2017). A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines. *BIOMEDICAL RESEARCH-INDIA*, 28(9).
- [25] Zhang, B., Dong, X., Hu, Y., Jiang, X., & Li, G. (2023). Classification and prediction of spinal disease based on the SMOTE-RFE-XGBoost model. *PeerJ Computer Science*, 9, e1280.