

Applying Machine Learning for the Prediction of Tuberculosis Among People Living with HIV from Health Facilities in Addis Ababa, Ethiopia

Ayele Tiyou

ayeletiyou@gmail.com

Ministry of Health, Ethiopia

Merga Belina

Addis Ababa University

Yimer Seid

Addis Ababa University

Binyam Haftu

Addis Ababa City Administration Health Bureau

Research Article

Keywords: Tuberculosis (TB), Human Immunodeficiency Virus (HIV), People Living With HIV (PLWHIV), Machine Learning (ML), Deep Learning

Posted Date: August 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10411125/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Tuberculosis (TB) is a major global public health problem and one of the leading causes of death worldwide, particularly in the era of Human Immunodeficiency Virus (HIV). TB is among the leading causes of death for people living with HIV (PLWHIV). Early diagnosis of TB among PLWHIV remains challenging due to non-specific clinical symptoms, coexisting opportunistic infections, and limitations of conventional diagnostic and screening approaches. The objective of this study was to develop and validate a machine learning model to predict TB cases among PLWHIV using smart care data from health facilities.

Methods

The research employs quantitative analysis of secondary data from four health facilities that provided HIV care and treatment to PLWHIV and reported to the Addis Ababa Health Bureau. The study uses an experimental research design applying machine learning and deep learning models.

Results

From four facilities, 16,698 adult patients currently on ART were included; 59% were female, and 85% were married. In this study, the majority of the models performed better when the Hybrid method combined SMOTE with Edited Nearest Neighbors (ENN), and XGBoost had the strongest influence on the predictor of TB occurrence across most evaluation matrices. The XGBoost accuracy was 98%, recall 80%, precision 25%, F1 score 38%, F Beta 30%, MCC 52%, and AUC 96%. ART regimen, taking Tuberculosis Preventive Therapy (TPT), cotrimoxazole prophylaxis, months on ART, patient weight, and functional status are the strongest risk factors for developing TB and for protection against TB development.

Conclusion

Overall, XGBoost outperforms the other models across accuracy, precision, F1 score, and ROC AUC. Using the XGBoost model, the ART regimen, TPT, cotrimoxazole prophylaxis, months on ART, and patient weight were the strongest predictors of whether a patient would be protected against TB or would develop TB.

Introduction

Tuberculosis (TB) is a major global public health problem and one of the leading causes of death worldwide, particularly in the era of Human Immunodeficiency Virus (HIV). TB is among the leading

causes of death for people living with HIV (PLWHIV), accounting for about twenty-five percent of deaths in this group [1]. HIV promotes the progression of latent or recent *Mycobacterium tuberculosis* infection to active disease, increases TB incidence, and may increase susceptibility to TB infection[2]. Despite being preventable and usually curable, TB remains among the top 10 causes of death worldwide, with more than 10 million people falling ill and over 1 million dying from the disease each year [1].

TB is caused by the bacterium *Mycobacterium tuberculosis*, which is spread when people *with TB expel bacteria into the air* (e.g., by coughing). About a quarter of the global population is estimated to have been infected with TB[3]. Without treatment, the death rate from TB disease is high (close to 50%). With the treatments currently recommended by the World Health Organization (WHO) (a course of anti-TB drugs for 4–6 months), about 90% of people with TB can be cured[4]. HIV-related immunosuppression significantly increases the risk of progression from latent *Mycobacterium tuberculosis* infection to active TB, resulting in a substantially higher incidence and mortality rate among PLWHIV compared to the general population [5], [6].

Ethiopia is among the 30 high TB and TB/HIV burden countries globally, with an estimated TB incidence rate of 140/100,000 population (157,000 persons annually), and 21,000 (19/100,000 population) TB deaths in 2018 [7], [8]. TB remains the leading cause of death of people with HIV, accounting for around 40% of acquired immunodeficiency syndrome (AIDS)-related deaths. Both diseases together form a lethal combination, each speeding the other's progress [8].

The TB/HIV Co-infection rate has declined over several years in Ethiopia, ranging from 17% in 2011 to 6% in 2020 [8]. Among people living with HIV who develop TB, both TB treatment and ART for HIV are required to prevent unnecessary deaths from TB and HIV. Since 2019, the global coverage of ART for people living with HIV who were newly diagnosed with TB and reported as a TB case has been maintained at a high level; it reached 91% in 2024, up from 88% in 2023 [1].

TB/HIV co-infection is associated with complex diagnostic and therapeutic challenges and continues to impose a substantial burden on healthcare systems in high-burden countries such as Ethiopia [1]. Although HIV infection significantly increases the risk of developing active tuberculosis, evidence shows that not all people living with HIV (PLWHIV) develop TB, suggesting that HIV status alone does not fully explain TB occurrence [6], [9], [10]. This highlights the role of multiple clinical, immunological, behavioral, and socio-demographic determinants in shaping TB risk among PLWHIV. Understanding these factors within the local context is critical for improving early detection, prevention, and targeted management of TB among PLWHIV in Addis Ababa.

Early diagnosis of TB among PLWHIV remains challenging due to non-specific clinical symptoms, coexisting opportunistic infections, and limitations of conventional diagnostic and screening approaches [11]. These challenges often lead to delayed detection and treatment, increasing the risk of poor clinical outcomes and continued transmission. With the growing availability of large-scale clinical, demographic, and laboratory data, Machine Learning (ML) techniques offer a promising approach for predicting TB occurrence among PLWHIV by identifying complex, non-linear patterns and high-risk profiles that may

not be captured by conventional approaches [12], [13], [14]. Developing accurate ML-based predictive models for TB among PLWHIV can support early risk stratification, targeted screening, and timely intervention, thereby contributing to reduced TB-related morbidity and mortality in this vulnerable population.

Further reductions in the burden of TB disease require improvements in the coverage of TB diagnostic, treatment, and preventive interventions; action on broader determinants that drive new infections or increase the risk of developing disease once infected; and technological breakthroughs, such as a new TB vaccine [1]

Methods

Study design, approach, and data source:

The research used quantitative analysis of secondary data from four public health facilities: Dagmawi Menilik II Hospital, Yekatit 12 Hospital, Zewuditu Memorial Hospital, and Entot 2 Health Center. The study populations were adult PLWHIV currently on ART receiving care and treatment at health facilities in Addis Ababa who were currently on ART for at least 6 months. The study used an experimental research design, applying machine learning and deep learning models to predict TB prevalence cases. The data source was the ART Smart Care database for health facilities.

Study variable:

The dependent Variable for the study was active TB diagnosed status (positive or negative). The independent features include Sociodemographic factors such as Age, Sex, Weight, Height, Marital Status, Educational Level, Occupation, Religion, and Address. Clinical-related factors, including months on ART treatment, duration of HIV infection, WHO stage, Coinfection, Functional status, and nutritional status. Hematologic and immunological factors include recent CD4 count, Viral load status, and viral load. Treatment-related factors like Adherence to treatment, Regimen line, receiving Tuberculosis Preventive Therapy (TPT) initiation, and Cotrimoxazole Preventive Therapy (CPT) use.

Data acquisition:

The health facilities providing HIV care and treatment collect basic and follow-up information with a paper-based patient information capture format, and the data is regularly entered into the Smart Care database by assigned data personnel. After receiving ethical clearance from the Addis Ababa health bureau, the data were extracted from the health facility's database. Based on the metadata review, the data administrators extracted the lists of relevant features in CSV format.

Data pre-process:

Initially, eight features with more than 50% missing data were dropped from the data set (data frame) [15]: religion, occupation, education, Opportunistic Infection (OI), viral load, height, nutrition, and viral count. For the three numerical data, age, and CD4 count, the mean for the specific column and for the ART dose, the median value is computed and replaces missing values.

List-wise or case deletion was applied if the missingness is completely at random (MCAR) for categorical data, that is, the probability of missingness is not related to either the specific value that is supposed to be obtained or the set of observed responses [16]. For categorical data, the mode value of the specific column for Sex, Address, Weight, WHO stage, nutrition, ART regimen, TPT, Marital Status, Adherence, and Functional Status is used to impute missing values.

To enhance data readability, data label encoding was applied to categorical features. Min-max scaling is used to transform the adjusted feature value distribution and reduce the negative effect of marginal values. Then, all unnormalized feature values across the datasets were rescaled to the predefined bounds (0, 1) using min-max scaling.

The data reduction process was applied to improve model efficiency and avoid overfitting. The initial feature selection was guided by domain knowledge to retain the most relevant attributes, and we applied a correlation analysis to identify highly correlated features.

The person correlation analysis among quantitative features revealed that there is a high positive correlation between the number of days since HIV infection was diagnosed and the number of months on ART ($r = 0.81$). Similarly, patient age shows moderate positive correlations with the number of months on ART and with ART dose ($r = 0.65$). The other indicator with moderate positive correlation was the ART dose and number of days since HIV infection ($r = 0.58$) (Fig. 1). Based on the correlation analysis results, we dropped two features, number of days since HIV infection and ART dose, to improve model efficiency during Model building.

Data Balancing:

We have demonstrated various data-balancing techniques and compared their results, including the Synthetic Minority Over-Sampling Technique (SMOTE) with 50%, 25%, and 10% increases, as well as a hybrid method that combines SMOTE with Edited Nearest Neighbors (ENN) (Fig. 2).

Data Splitting:

To facilitate the training, testing, and evaluation of machine learning models, we separate the data, ensuring that the model is trained on one set, validated on another, and tested on a final, independent set. The data was split into 70% for training, 15% for validation, and 15% for testing [17].

Model Selection and Evaluation

In this study, based on similar studies, we have applied eight machine learning models and deep learning models, including Logistic regression (LR), Decision trees (DT), Random forests (RF), Extreme Gradient Boosting (XGClassifier), Support vector machines (SVM), k-nearest neighbors (k-NN), Multilayer Perceptron (MLP), and Convolutional Neural Network (CNN) (Fig. 3).

Models was evaluated and compared using several performance metrics, including accuracy (overall correctness of predictions), precision and recall (the ability to correctly identify positive cases while minimizing false positives and false negatives), F1 score (the harmonic mean of precision and recall, particularly useful for imbalanced datasets) and the Receiver-operating characteristic curve (ROC) curve along with the area under the curve (AUC), which assesses the model's ability to distinguish between classes across various threshold settings and F-Beta metric gives recall more weight than precision through the beta parameter, making it particularly useful for imbalanced classification problems and MCC considers all four confusion matrix elements True Positive (TP), True Negative (TN), False Positive (FP), False Negative(FN) to provide a balanced evaluation.

Result

Exploratory Data Analysis (EDA):

In this study, 16,698 adult patients currently on ART from four public health facilities were included; 58.9% were female, 85.2% were married, and 97.2% lived in Addis Ababa. Based on clinical information, 91.2% of patients were WHO clinical stage I, 47.5% were taking ART regimen 1j, 99.5% classified as good adherence by service providers, 75.6% were working, 91.8% had received TPT treatment, 79.4% received Cotrimoxazole for opportunistic infection (OI), and 98% didn't have TB

Model evaluation

Due to the high imbalance in the data, SMOTE-ENN was applied, and the model performance was evaluated. All results and discussions are based on the SMOTE-ENN result.

The accuracy performance with SMOTE-ENN ranges from 0.98 to 0.8. The highest accuracy was achieved with the Random Forest, XGBoost, and Decision Tree models, followed by K-NN and MLP. The recall performance with SMOTE-ENN ranges from 0.92 to 0.67. Support vector machine and logistic regression were the leading-performing models in terms of recall, followed by XGBoost with recall scores of 0.92, 0.88, and 0.8, respectively. Random Forest and XGBoost achieve lower false-positive rates, with precision scores of 0.31 and 0.25, respectively. Random Forest (0.43) and XGBoost (0.38) are stronger at balancing precision and recall (F1 score), followed by decision trees (0.36). F-Beta taking beta 2 to give weight for recall results range from 0.3 to 0.55 using SMOTE-ENN, with XGBoost achieving the highest, and the MCC ranges from 0.08 (LR) to 0.43 (XGBoost) (Fig. 5). XGBoost and Random Forest have the highest ROC AUC score of 0.96 and 0.95, respectively, indicating excellent overall discrimination between classes (Figs. 4 and 5).

There is no single model that outperforms across all metrics, but overall, XGBoost outperforms on most evaluation metrics, including ROC AUC, accuracy, F1 score, F-Beta, and MCC. Support vector machine and Logistic regression, followed by XGBoost models, are better in terms of recall and in identifying actual TB-positive cases, with room for improvement in precision to reduce false positives (Fig. 5).

Risk factors of TB using XGBoost

The SHAP analysis identifies the ART regimen as the most influential predictor for developing TB, with a mean absolute SHAP value of 1.4, compared to Adherence, which influences 60 times; followed by taking TPT and cotrimoxazole prophylaxis, with mean SHAP values of 1.2 and 1.1, respectively, compared to Adherence, which influences 53 times and 51 times, respectively. The remaining features, ranked from highest to lowest, are months on ART, patient weight, Functional status, sex, age, and CDC4 count, which are the highest-risk factors for a patient to develop TB or to protect against TB development (Fig. 6).

Discussion

In this study, XGBoost and random forests achieve the highest test accuracy, with 98% and 97%, respectively. In a study conducted in Debre-Markos, Ethiopia, XGBoost correctly identified 82% of all instances as actual positives or actual negatives. A study conducted in China found that XGBoost achieved 84% accuracy. Differences in the source population, training algorithms, data-splitting techniques, included characteristics, and dataset size may all contribute to discrepancies in classification performance. The data used in this study are much higher than those in the Debre Markos and China studies.

In this study, recall level is an important metric, as high false-negative rates among TB patients could cost lives. The support vector machine was the best-performing model in terms of recall for identifying actual positives, achieving 92%. XGBoost achieved the highest recall in the Debre-Markos [18] and China [19] studies, with 84.2% and 71%, respectively. In this study, XGBoost's recall was 80%, comparable to the Debre-Markos study and higher than the China study.

In this study, XGBoost and Random Forest achieved the highest AUC (probability of correctly distinguishing a randomly chosen positive instance from a randomly chosen negative instance), with scores of 96% and 95%, respectively. The AUC score in this study is slightly higher than those in the Debre-Markos [18] and China [19] studies, which are 90% and 93%, respectively. The difference could be explained by the larger sample size, as this study has a much larger sample size.

The precision metrics in this study are lower than those in the Debre Markos [18] and China studies [19]. The better-performing model in this study, with 31% precision, was the random forest, followed by XGBoost at 25%. The study conducted in Debre Markos reported XGBoost with 81% precision, and the China study reported Random Forest and LightGBM with 43% precision. Because the outcome variable is

highly imbalanced, this may explain the studies' low precision, which is often affected by such imbalances.

In addition to F1, we have also applied F-beta, since the F1 score is the harmonic mean of Precision and recall. When precision is low, the F1 score is also affected. While the F-Beta metric gives recall more weight than precision through the beta parameter, making it particularly useful for imbalanced classification problems, MCC is a robust metric that considers all four confusion matrix elements (TP, TN, FP, FN) to provide a balanced evaluation. The best F1 score in this study was Random Forest (0.43), which was much lower than those reported in studies conducted in Debre Markos and China. F-Beta, with a beta value of 2 to emphasize recall over precision, as reducing false negatives is much more important to reduce cases from diagnosis that could lead to mortality. XGBosst has 55%, and MCC for XGBoost was 52%. The MCC score for XGBoost was 43%; one study evaluating business predictive models reported an MCC score of 28.7 for the baseline and 28.5 after applying 5-fold data balancing [20].

In this study, ART regimen, TPT use, cotrimoxazole prophylaxis, months on ART, patient weight, functional status, sex, age, CDC4 count, and WHO stage are the highest-risk factors for developing TB or for protection against TB development. A study conducted in Debre-Markos found that CD4 count, age, duration on ART, sex, WHO clinical stage, address status, Differentiated Service Delivery (DSV) category, and TPT were identified as predictors of tuberculosis incidence among adult HIV patients. The findings of this study, that an ART regimen is associated with TB, align with the global TB report, which indicates that in Africa, a widespread scale-up of antiretroviral therapy (ART) strongly declines the incidence of TB[21]. The second highly associated indicator, taking TPT, is supported by a systematic review of TPT in Ethiopia, which found that PLHIV on IPT, either alone or in combination with ART, were less likely to develop TB than those without IPT [22]. To prevent and reduce the incidence of TB in people living with HIV (PLHIV), the World Health Organization (WHO) recommended the use of isoniazid preventive therapy (IPT) as a mainstay of the “Three I’s” approach [23]. The third variable that has an association in this study was the use of cotrimoxazole prophylaxis, and supported by the Asia study, Multivariate analyses adjusted for viral load (VL), CD4 count, body mass index (BMI) and cART duration showed that CTX reduced the hazard for new TB infection by 28% (HR 0.72, 95% CI | 0.56, 0.93) [18].

Conclusions

Overall, we can conclude that XGBoost can be applied to predict TB cases among patients currently on ART, as the models outperform in terms of accuracy, precision, F1 score, F-Beta, ROC AUC, and MCC metric, while also showing better precision and F1 score, with further testing using additional features and high-quality data. The factors found to have a stronger influence on TB development include ART regimen, TPT use, cotrimoxazole prophylaxis, months on ART, patient weight, functional status, sex, age, and CD4 count; these factors are also evident in other similar studies and in WHO recommendations.

Declarations

Ethics approval and consent to participate

The research proposal was reviewed by the Institutional Review Board (IRB) of Addis Ababa University, College of Natural and Computational Sciences, School of Information Science, and approval was granted with Ref. No. EDRE/61/2018/26. The Addis Ababa Health Bureau's ethical review committee reviewed the proposal to grant access to the data and provided a letter for health facilities.

This study used secondary, de-identified data collected through routine HIV care and services. To ensure the confidentiality and privacy of participants, no personal identifiable information—such as names, addresses, or contact details in the dataset for analysis. The study adhered to the ethical principles outlined in the Declaration of Helsinki, including informed consent from the data owners' health facilities, beneficence with maximizing potential benefit for the study population, no harm, ensuring confidentiality of information, and independent ethical review, and complied with all applicable national and institutional guidelines governing research involving human data.

Because the study used secondary analysis of existing data and posed minimal risk to individuals, a waiver of informed consent was sought from the Ethics Review Committee.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request. In addition, the data set was shared with the Addis Ababa University library repository and will be publicly available in accordance with university policy.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific financial support from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

AT conceived and designed the study, drafted the manuscript, and critically reviewed the manuscript. MB and YS contributed significantly to the study design, analysis, data interpretation, manuscript drafting, and critical review. BH contributed to data acquisition and study design. All authors approved and read the final manuscript.

Acknowledgements

The authors thank the Addis Ababa Health Bureau and the health facilities for their ethical review and for sharing the data

Disclaimer

The views, findings, and conclusions represented in this article are those of the authors and do not necessarily represent the official views of the institutions with which they are affiliated.

References

1. *Global tuberculosis report 2025*. Geneva: World Health Organization; 2025. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications/b/81626> (Access Date: Jan 15, 2026).
2. Iliyasu Z, Babashani M. Prevalence and Predictors of Tuberculosis Coinfection among HIV-Seropositive Patients Attending the Aminu Kano Teaching Hospital, Northern Nigeria. *J Epidemiol*. 2009;19(2):81–7. 10.2188/jea.JE20080026.
3. Houben RMGJ, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. *PLOS Med*. Oct. 2016;13(10):e1002152. 10.1371/journal.pmed.1002152.
4. Tiemersma EW, Van Der Werf MJ, Borgdorff MW, Williams BG, Nagelkerke NJD. Natural History of Tuberculosis: Duration and Fatality of Untreated Pulmonary Tuberculosis in HIV Negative Patients: A Systematic Review. *PLoS ONE*. Apr. 2011;6(4):e17601. 10.1371/journal.pone.0017601.
5. World Health Organization. (2023). *Global tuberculosis report 2023*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO. <https://iris.who.int/items/60aa1a1d-04f9-46e2-9897-9836594b8d21>. (Access Date: Jan 15, 2026).
6. Lawn SD, Churchyard G. Epidemiology of HIV-associated tuberculosis. *Curr Opin HIV AIDS*. Jul. 2009;4(4):325–33. 10.1097/COH.0b013e32832c7d61.
7. World Health Organization. (2019). *Global tuberculosis report 2019*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO. <https://www.who.int/publications/i/item/global-tuberculosis-report-2019> (Access Date: Jan 15, 2026).
8. Federal Ministry of Health of Ethiopia (FMOH). National guidelines for TB, DR-TB and Leprosy in Ethiopia - Seven Edition. Addis Ababa: FMOH; 2021.
9. Selwyn PA et al. Mar., A Prospective Study of the Risk of Tuberculosis among Intravenous Drug Users with Human Immunodeficiency Virus Infection, *N. Engl. J. Med.*, vol. 320, no. 9, pp. 545–550, 1989, 10.1056/NEJM198903023200901
10. Corbett EL, et al. The Growing Burden of Tuberculosis: Global Trends and Interactions With the HIV Epidemic. *Arch Intern Med*. May 2003;163:1009. 10.1001/archinte.163.9.1009.
11. Tiberi S, et al. The cursed duet today: Tuberculosis and HIV-coinfection. *Presse Médicale*. Mar. 2017;46(2):e23–39. 10.1016/j.lpm.2017.01.017.
12. Esteva A, et al. A guide to deep learning in healthcare. *Nat Med*. Jan. 2019;25(1):24–9. 10.1038/s41591-018-0316-z.

13. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine, *N. Engl. J. Med.*, vol. 375, no. 13, pp. 1216–1219, Sep. 2016, 10.1056/NEJMp1606181
14. Deo RC. Machine Learning in Medicine, *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, 10.1161/CIRCULATIONAHA.115.001593
15. Van Buuren S. Flexible Imputation of Missing Data, Second Edition, 2nd ed. Second edition. | Boca Raton, Florida: CRC Press, [2019] |: Chapman and Hall/CRC, 2018. 10.1201/9780429492259
16. Khan K, Ahmad W, Amin MN, Ahmad A, Nazar S, Alabdullah AA. Compressive Strength Estimation of Steel-Fiber-Reinforced Concrete and Raw Material Interactions Using Advanced Algorithms. *Polymers*. Jul. 2022;14(15):3065. 10.3390/polym14153065.
17. Ge W, Lalbakhsh P, Isai L, Lensky A, Suominen H. Comparing Deep Learning Models for the Task of Volatility Prediction Using Multivariate Data, 2023, arXiv. 10.48550/ARXIV.2306.12446
18. Tadele DM, Biwota GT, Enyew LM, Tadele MM, Teferi GH. Predicting Tuberculosis Incidence in Adult HIV Patients on ART in Debre Markos, Ethiopia: A Machine Learning Approach, Aug. 28, 2025, *Health Informatics*. 10.1101/2025.08.24.25334330
19. Sun Q, et al. Explainable machine learning for predicting clinical outcomes in HIV/TB co-infection: a comparative retrospective study. *BMC Infect Dis*. Nov. 2025;25(1):1589. 10.1186/s12879-025-11998-w.
20. Sujon KM, et al. Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. *J Big Data*. 2025;12:268. <https://doi.org/10.1186/s40537-025-01313-4>.
21. García-Gonzalo E, Fernández-Muñiz Z, García Nieto P, Bernardo Sánchez A, Menéndez Fernández M. Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers. *Materials*. Jun. 2016;9(7):531. 10.3390/ma9070531.
22. Khurana A, Singh OP. Estimation of Battery Life Cycle & Remaining using Machine Learning. Aug 02. 2024. 10.21203/rs.3.rs-4595067/v1. *In Review*.
23. Chan KY, et al. Deep neural networks in the cloud: Review, applications, challenges and research directions. *Neurocomputing*. Aug. 2023;545:126327. 10.1016/j.neucom.2023.126327.

Figures

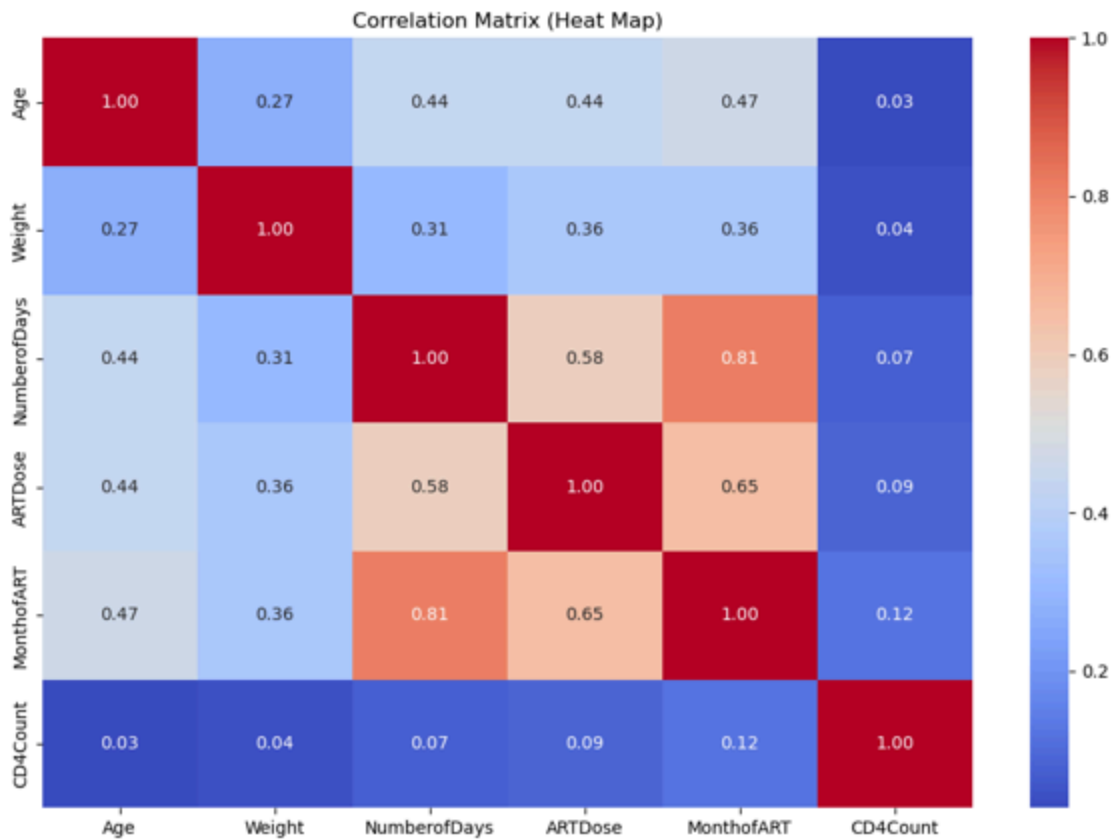


Figure 1

Correlation analysis matrix between study features.

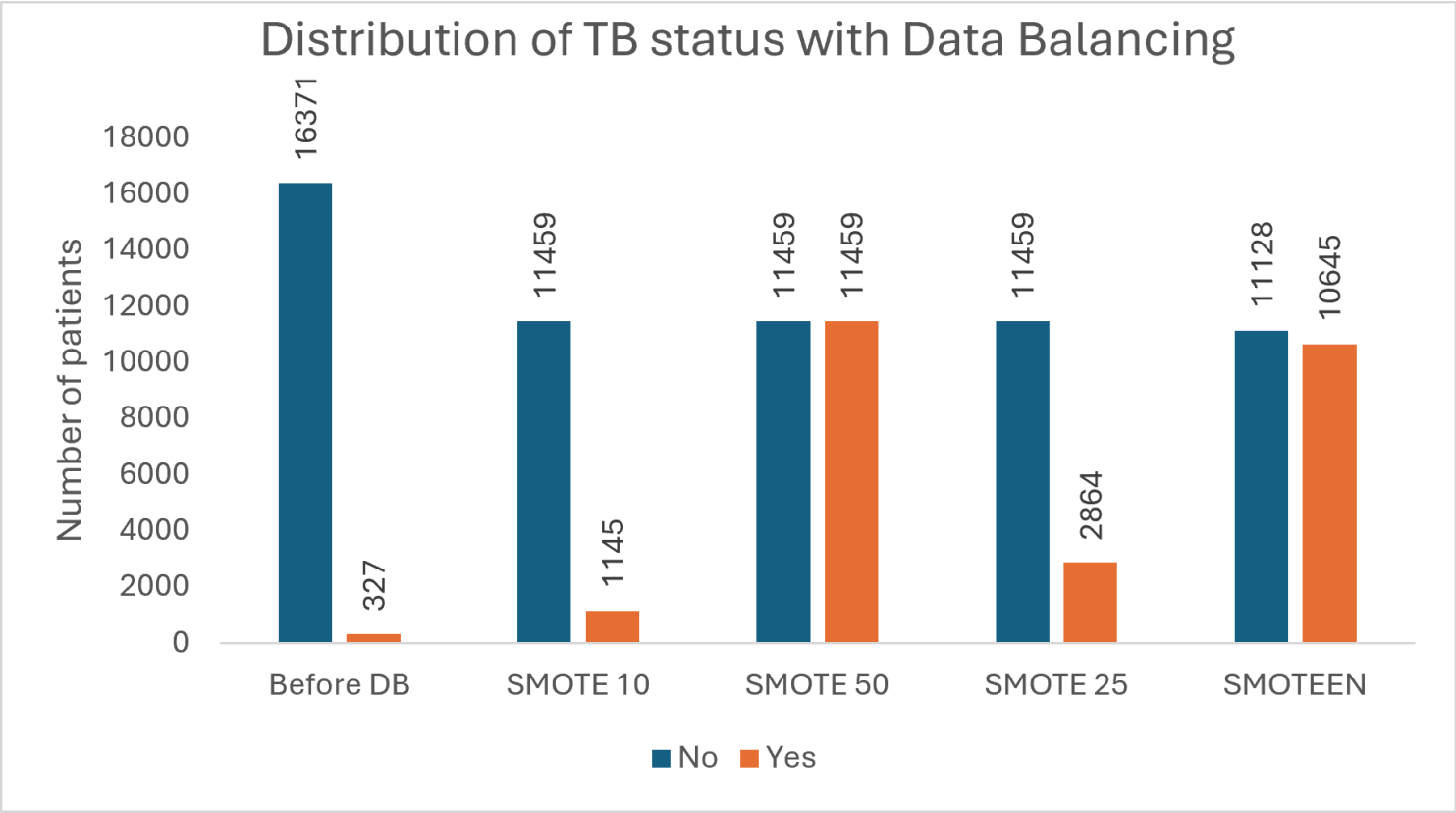


Figure 2

Distribution of TB cases before and after the data balance technique was applied

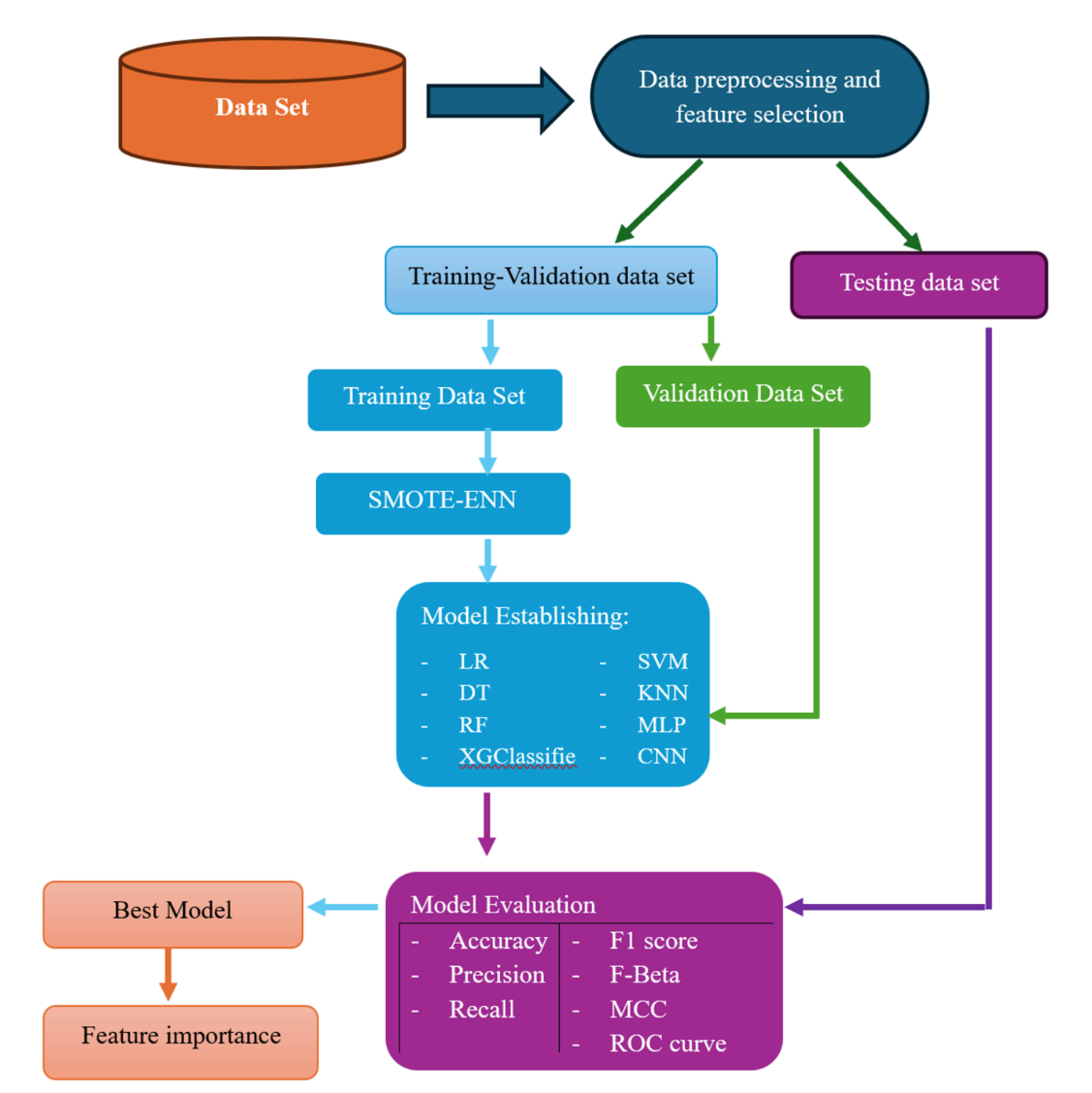


Figure 3

Study flow diagram of data preparation, model establishment, and evaluation

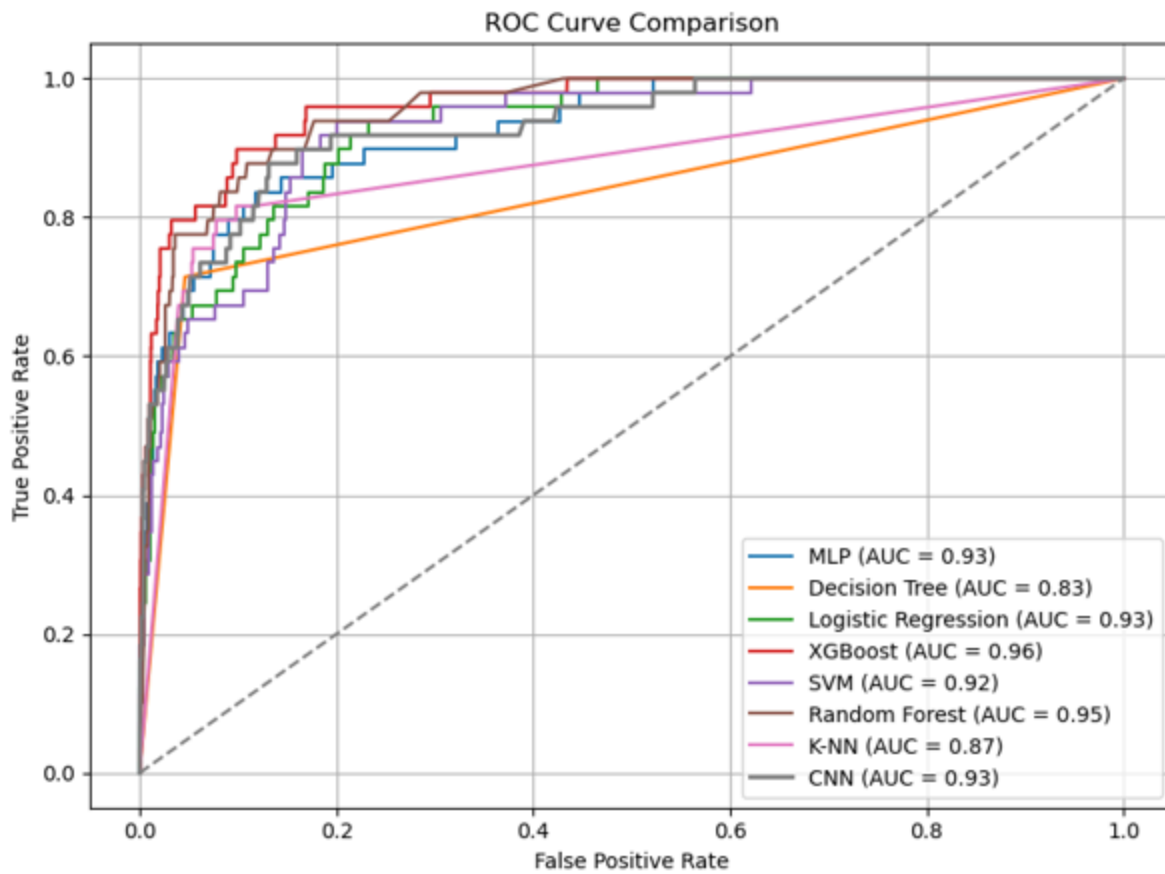


Figure 4

The performance comparison of the machine learning models using Receiver Operating Characteristic (ROC) curves

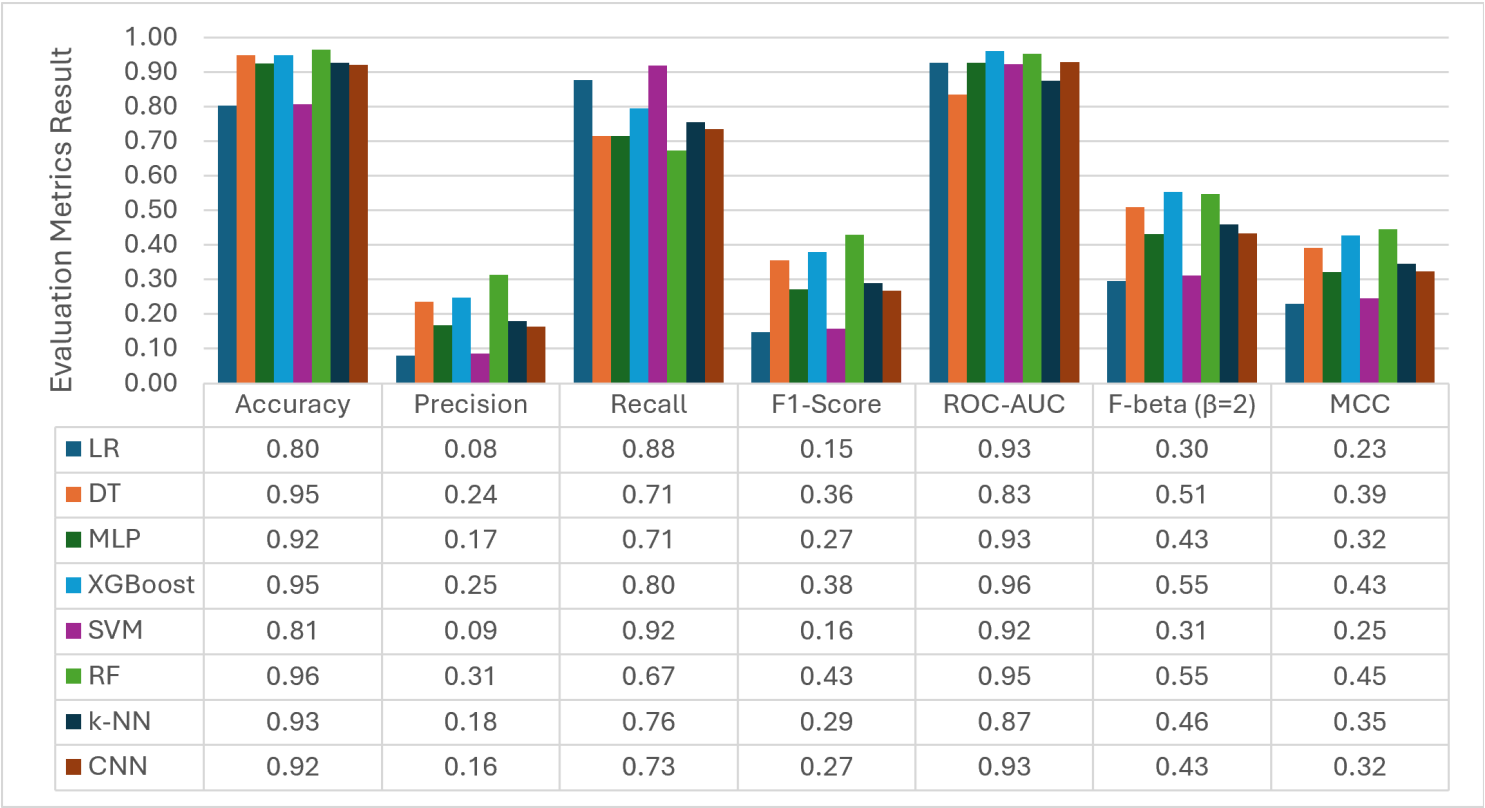


Figure 5

Evaluation metrics comparison for respective models with SMOTE-EEN

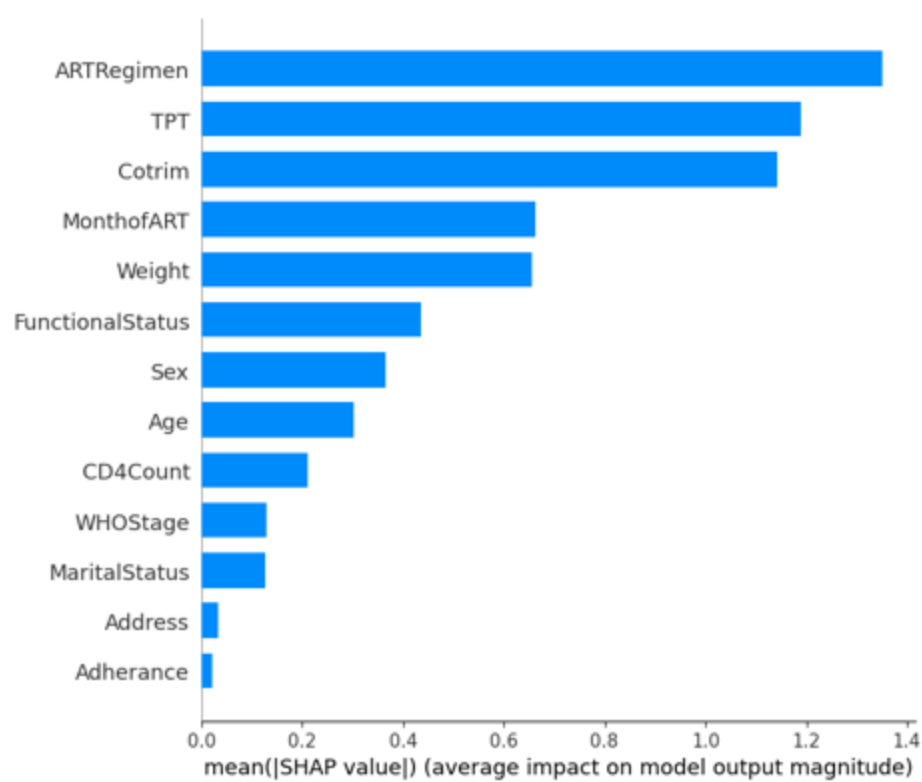


Figure 6

