

Prediction of TB treatment outcomes among HIV/TB coinfecting patients in Uganda using routinely collected clinical data: a machine-learning approach

Ambrose Okibure

ambroseokibure@gmail.com

Makerere School of Public Health

Jimmy Patrick Alunyo

Busitema University

Sarah Rachael Akello

Busitema University

Tereza Nyapendi

AIDS Information Centre

Mabel Nakawooya

Ministry of Health

Francis Okello

Busitema University

Noela Regina Akwi Okalany

University of Bergen

Rebecca Nekaka

Busitema University

Ernest Ochepa Ekiro

Makerere School of Public Health

Winfred Jenifer Namuyanja

Makerere School of Public Health

Edward Ikoona

Makerere School of Public Health

Nancy chemutai

Makerere School of Public Health

Azaria Kanyamibwa

Makerere School of Public Health

Ronald Kusolo

Makerere School of Public Health

Noah Kiwanuka

Makerere School of Public Health

Research Article

Keywords: Tuberculosis, HIV co-infection, Machine learning, Treatment outcomes

Posted Date: March 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8870898/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) remains a major public health challenge, particularly in settings with high HIV prevalence. In Uganda, TB/HIV co-infection contributes to suboptimal treatment success rates (TSR), still below the World Health Organisation (WHO) target of $\geq 90\%$. Early identification of patients at risk of poor outcomes is essential to mitigate poor adherence, reduce the risk of multidrug-resistant TB (MDR-TB), and improve health outcomes, including reduced morbidity and mortality as well as transmission. This study applied machine learning (ML) techniques to predict TB treatment outcomes among HIV/TB co-infected patients in Uganda and to identify the most influential predictors of treatment failure using feature importance and partial dependence analyses.

Methods

Data from a retrospective cohort of 5,062 HIV/TB co-infected patients treated in Uganda between 2020 and 2024 were analysed. Machine learning models, including logistic regression, Random Forest, Gradient Boosting Machine (GBM), Support Vector Machine (SVM), AdaBoost, and a stacked ensemble, were developed and evaluated using discrimination, calibration, recall, and accuracy metrics. Feature importance and partial dependence analyses were used to interpret model predictions.

Results

The stacked ensemble and class-balanced logistic regression models achieved the best overall performance (AUC ≈ 0.67 ; accuracy ≈ 0.62). The Random Forest model exhibited the highest discrimination (ROC-AUC = 0.675), while GBM achieved the highest accuracy (0.768) but low sensitivity to treatment failures. Key predictors of treatment success included age, ART status, sex, marital status, TB classification, and treatment model. Treatment success declined progressively with increasing age, particularly beyond 40 years.

Conclusions

The models demonstrated moderate predictive performance and identified key demographic and programmatic predictors of TB treatment outcomes. While not suitable for autonomous clinical decision-making, these models may support risk stratification and targeted patient follow-up.

Trial registration number:

Not Applicable

Background

Globally, tuberculosis (TB) remains among the major causes of morbidity and mortality. Approximately an estimated 10.8 million people fell ill with TB worldwide, including 6.0 million men, 3.6 million women and 1.3 million children in 2023 [1]. The mortality rate for TB disease was disproportionately high, 1.25 million among HIV-negative people compared to 161,000 in HIV-positive people in 2023 [1]. The health target of the United Nations Sustainable Development Goals (SDGs) is to end the TB epidemic by 2030 [2]. The burden of TB is high in the African continent, with over 2.5 million people reported diagnosed with TB in the African region in 2022, accounting for a quarter of new TB cases worldwide. The African continent accounted for over 90% of all global TB cases and 87% of all global co-infection cases. The region contributed more than 400,000 (33%) deaths to the global mortality from the disease in the same year [1].

The World Health Organization (WHO) recommends that early diagnosis and appropriate treatment of tuberculosis (TB) can effectively reduce transmission, avert deaths, and contribute to its elimination. For drug-susceptible TB, a standard six-month course of first-line anti-TB drugs is recommended, while treatment duration may be longer for multidrug-resistant (MDR) or extensively drug-resistant (XDR) TB. Between 2000 and 2020, early diagnosis and treatment of TB were estimated to have averted 66 million deaths [2]. Despite widespread access to TB prevention, diagnosis, treatment and care services, the global treatment success rate was only 86% among HIV-negative people in 2019. For HIV-associated TB, it was even very low at 77%. Sub-Saharan Africa accounts for approximately 90% of global tuberculosis (TB) cases and about 87% of TB/HIV co-infections. The treatment success rate in the region stands at 78.9% and is even lower among patients co-infected with HIV and TB [3]. To address the persistently high TB-related mortality, there is a growing need for interventions aimed at predicting treatment outcomes. Machine learning algorithms offer promising approaches for predicting TB treatment success and improving patient outcomes

In Uganda, by 2023, TB disease was still a big public health challenge, with more than 200 people falling ill with TB disease and approximately 30 losing their lives [4]. Even though 94% of TB/HIV co-infected patients initiated antiretroviral therapy (ART), there was a high mortality rate of 12% [5]. Early detection of TB disease and prompt initiation of treatment with potent anti-TB drug regimens are required to decrease morbidity and mortality from TB disease among HIV-infected patients [6].

Despite implementing these recommendations and improving access to TB prevention, diagnosis, treatment, and care, Uganda still has a very low TSR of 70% among TB/HIV co-infections and 72% among HIV-negative people [7]. Additionally, the country is also ranked among the thirty nations globally grappling with high burdens of TB and TB/HIV co-infection [8]. Annually, an estimated 91,000 individuals in Uganda contract TB, with 32% of these cases occurring in people living with HIV. Additionally, two percent of TB patients face drug-resistant strains that are not responsive to initial medication, and around 15% of TB cases in Uganda involve children under the age of 14 [4]. This Treatment Success Rate is way below the desired WHO target of $\geq 90\%$ [9].

TB treatment is typically successful with the standard six-month course of anti-TB drugs. However, treatment outcomes vary widely due to clinical, socio-demographic, environmental, and health-system-related factors. Studies show that treatment success is influenced by patient-related factors (e.g., medication adherence, co-morbidities like HIV), health system factors (e.g., delays in diagnosis, access to healthcare), and environmental factors. However, most previous studies. Few studies in Uganda have applied ML to routine EMR data to predict TB outcomes among HIV/TB co-infected patients, and majority have relied on conventional statistical models, which have limitations in handling complex, high-dimensional data and non-linear relationships among predictors [3, 9]

Machine learning (ML) offers a more robust, data-driven approach to identifying early predictors of treatment failure and MDR-TB progression by handling large, high-dimensional datasets that include patient characteristics, clinical parameters, identifying hidden patterns that conventional models might miss and improving prediction accuracy by considering both linear and non-linear relationships among predictors. Despite growing interest in machine learning for TB outcome prediction, few studies in Uganda have applied these methods to nationally representative EMR data, particularly among HIV/TB co-infected patients. This study addresses this gap by evaluating multiple ML models and identifying key predictors of treatment outcomes.

Methods

Study design

The study utilised a retrospective cohort study design of people with HIV/TB co-infection receiving TB treatment across the country between 2020 and 2024.

Study setting

The study was conducted across 15 Regional Referral Hospitals (RRHs) in Uganda, which serve as key centers for TB diagnosis, treatment, and management, particularly for HIV/TB co-infected patients. These hospitals provide specialized services, including multidrug-resistant TB (MDR-TB) management, Directly Observed Therapy (DOTS), and integration with HIV care through antiretroviral therapy (ART) clinics. They also act as referral centers for lower-level health facilities, such as district hospitals and health center IVs, which initiate TB treatment before referring to complex cases to RRHs. The 15 RRHs include Arua RRH, Fort Portal RRH, Gulu RRH, Hoima RRH, Jinja RRH, Kabale RRH, Lira RRH, Mbale RRH, Mbarara RRH, Moroto RRH, Masaka RRH, Mubende RRH, Naguru RRH (China-Uganda Friendship Hospital), Soroti RRH, Kayunga RRH. TB management at these hospitals follows the Uganda National TB and Leprosy Program (NTLP) guidelines, ensuring standardized patient diagnosis, treatment monitoring, adherence support, and follow-up care.

Study population

The study targeted TB patients diagnosed with HIV or already on HIV care in Regional Referral Hospitals in Uganda. The accessible population were all PLWH diagnosed with TB and started on anti-TB drug regimens in their respective clinics in the country for the period extending from 2020 to 2024. The sample population comprised HIV/TB co-infected patients whose data were randomly selected from the Uganda EMR system and used in this study.

Random selection of patient records was employed to ensure that the data used were representative of the broader national population of HIV/TB co-infected patients. This approach minimized selection bias and enhanced the generalizability of the findings by ensuring that patients from different regions, demographic groups, and treatment centers had an equal chance of inclusion.

Data Description and data sources

The study utilized secondary, routinely collected, anonymized patient data from the Uganda Electronic Medical Records (EMR) system under the Ministry of Health (MOH). Clinical and demographic data: Retrieved from the Uganda EMR system, containing variables on patient demographics, TB disease classification, ART initiation, and health facility characteristics.

Study variables

The primary outcome of the study is TB treatment outcome, categorized as successful (including cure and treatment completion) or unsuccessful (comprising failure, default, death, and not evaluated). Predictor variables will be in categories: socio-demographic (age, sex, residence), patient-related (e.g., initiation of ART, TB disease classification), health system-related (e.g., distance to the facility, facility ownership).

Data preprocessing and exploratory data analysis (EDA)

Data cleaning and handling missing data

Missing data were carefully assessed before deciding whether to exclude or impute values. The extent and pattern of missingness were analyzed using descriptive statistics and correlation matrices to determine whether the missing data were completely random (MCAR), missing at random (MAR), or missing not at random (MNAR). Variables with more than 30% missing values were excluded if they had limited predictive relevance, whereas those with lower levels of missingness underwent appropriate imputation based on their type and missing data pattern.

Categorical variables with missing values that were missing completely at random (MCAR) were imputed using mode imputation. For each categorical variable, the most frequently occurring category (mode) was identified from the observed data, and all missing entries within that variable were replaced with this modal category. This approach assumed that the missing values had the same distribution as

the observed data and helped maintain the original variable structure without introducing bias. For numerical variables, mean or median imputation was applied if the data were normally distributed, whereas K-Nearest Neighbors (KNN) or Multiple Imputation by Chained Equations (MICE) were used for non-normal distributions or MNAR patterns. Sensitivity analyses were performed to compare different imputation methods and ensure minimal bias while maintaining predictive accuracy.

Data transformation and normality checks

Categorical variables (e.g., treatment success/failure, sex, ART status) were encoded and Continuous variables (e.g., age, BMI) were checked for normality using histograms and the Shapiro-Wilk test. If non-normal, transformations (e.g., log transformation) were applied.

Outlier detection and treatment

Outliers were detected using box plots, Z-scores, and interquartile range (IQR) methods and Winsorization were considered for extreme outliers that significantly affect model performance.

Correlation analysis and feature selection

Multicollinearity was assessed using the variance inflation factor (VIF) with a threshold of 5 to identify redundant predictors. A correlation matrix and heatmaps were employed to examine dependencies among variables. Feature selection was performed using recursive feature elimination (RFE) and feature importance scores derived from tree-based models.

Sample size determination

The sample size was determined based on the proportion of unsuccessful TB treatment cases (30%) and on a power analysis to ensure at least 80% statistical power. The final analysis was conducted using data from 5,062 participants, which provided sufficient power for the planned statistical and machine learning analyses.

Machine Learning Models

Model selection criteria

The study evaluated the performance of seven supervised algorithms, logistic regression, random forest, support vector machine and AdaBoost algorithms. Logistic regression offered a straightforward linear approach, while random forest, support vector machine and AdaBoost provide more flexible solutions for complex data. Through this analysis, we aimed at identifying the most effective algorithm for predicting TB treatment outcomes. See Table 1.

Table 1
Machine Learning Algorithms and the rationale for their selection in the model

Algorithm	Reason for Selection
Logistic Regression	Baseline model for comparison, interpretable results
Random Forest	Handles non-linear relationships, robust to outliers
Support Vector Machine (SVM)	Effective for complex, high-dimensional datasets
AdaBoost (Adaptive Boosting)	Enhances weak classifiers, improves prediction accuracy

Model building

The model building methodology involved thorough data preparation, a 70/30 random split into training and test sets, and balanced sampling to ensure equal representation of treatment outcomes. A classifier algorithm was then trained using the training data to establish an optimal predictive model, which was evaluated using the independent test dataset. Python, a freely available programming language, along with TensorFlow libraries, an open-source machine learning framework were used.

Hyperparameter optimization

Hyperparameter tuning was performed to optimize model performance. Random Forest model was fine-tuned by adjusting the number of trees, max depth, and minimum samples per split to balance complexity and accuracy. SVM was optimized by selecting the best kernel type (linear or RBF) and regularization parameter (C). AdaBoost was refined by tuning the learning rate and number of estimators to enhance model stability. Logistic Regression was optimized by adjusting L1 and L2 regularization strength to prevent overfitting. The optimizations were conducted using Grid Search CV and Randomized Search CV for efficient parameter selection. See Table 2.

Table 2
Machine learning algorithms and corresponding hyperparameters tuned during model optimisation

Algorithm	Hyperparameters to Tune
Random Forest	Number of trees, max depth, min samples split
SVM	Kernel type (linear, RBF), regularization parameter (C)
AdaBoost	Learning rate, number of estimators
Logistic Regression	Regularization strength (L1, L2)

Model evaluation

The evaluation of the model's performance was based on accuracy (the proportion of correctly classified observations in the unseen test set), sensitivity (the proportion of known failure outcomes correctly identified by the algorithm), precision (the proportion of predicted failure outcomes that match known

failures in the test set), and specificity (the proportion of known successful outcomes correctly identified). Additionally, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to assess the overall predictive classification performance (The AUC represents the probability that the model will correctly distinguish between a randomly chosen positive case and a randomly chosen negative case). A range of 0.5 to 1.0 on the AUC scale indicates the predictive power of the model, where 0.5 suggests no predictive ability and 1.0 indicates perfect predictive performance. See Table 3.

Table 3
Model evaluation metrics, their definitions, and interpretations

Metric	Definition	Interpretation
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	Measures overall correctness
Precision	$TP / (TP + FP)$	How many predicted positive cases were true positives
Recall (Sensitivity)	$TP / (TP + FN)$	How many actual positive cases were correctly identified
Specificity	$TN / (TN + FP)$	Correct identification of negative cases
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$	Balance between precision and recall
AUC-ROC Score	Area under ROC curve	Discriminative ability of the model

Results

Baseline characteristics of study participants

Table 4 presents the baseline characteristics of patients included in the analysis. We included 5,062 tuberculosis (TB) patients with a mean age of 37.2 years (SD 14.5). Among these, 3,850 (76.1%) achieved treatment success, while 1,212 (23.9%) experienced treatment failure. Men comprised 54.8% of the study population and showed a higher proportion of treatment failure than women (56.4% vs 43.6%). Approximately one-third (34.2%) of the participants were married.

Most patients were classified as having pulmonary bacteriologically confirmed TB (61.1%), followed by pulmonary clinically diagnosed TB (33.7%) and extrapulmonary TB (5.2%). The majority were new TB cases (92.9%), while 5.3% were relapse cases and 1.5% had returned to treatment after loss to follow-up.

The non-digital community-based directly observed treatment (DOT) model was the predominant treatment-support approach reported by 68.8%, followed by facility-based DOT (16.9%) and digital community DOT (14.3%). Treatment failure was most frequent among patients managed under the facility-based DOT model (20.1%).

Regarding ART status, 49.8% of participants were already on ART at baseline, 22.7% were newly initiated during TB treatment, and 27.5% had unknown ART status. Treatment outcomes did not vary substantially across ART categories. See Table 4.

Table 4
Baseline characteristics of patients

Variables	Total	TB Treatment outcome	
		Treatment success	Treatment failure
	n = 5062(%)	n = 3850(%)	n = 1212(%)
Sex			
Female	2289 (45.2)	1760 (45.7)	529 (43.6)
Male	2773 (54.8)	2090 (54.3)	683 (56.4)
Marital status			
Not married	3332 (65.8)	2491 (64.7)	841 (69.4)
Married	1730 (34.2)	1359 (35.3)	371 (30.6)
Age in years mean(sd)	37.2 (14.5)	37.2 (14.4)	37.2 (14.9)
TB Disease Classification (Baseline)			
Extra pulmonary (EP) TB	262 (5.2)	160 (4.2)	102 (8.4)
Pulmonary Clinically Diagnosed (P-CD) TB	1708 (33.7)	1288 (33.5)	420 (34.7)
Pulmonary bacteriologically confirmed (P-BC)TB	3092 (61.1)	2402 (62.4)	690 (56.9)
Type of patient			
New	4704 (92.9)	3581 (93.0)	1123 (92.7)
Relapse	266 (5.3)	205 (5.3)	61 (5.0)
Return after lost to follow up	78 (1.5)	54 (1.4)	24 (2.0)
Treatment after failure	11 (0.2)	7 (0.2)	4 (0.3)
Treatment history Unknown	3 (0.1)	3 (0.1)	0 (0.0)
TB Treatment Model			
Digital Community DOT	726 (14.3)	564 (14.6)	162 (13.4)
Facility DOT	855 (16.9)	611 (15.9)	244 (20.1)

Variables	Total	TB Treatment outcome	
		Treatment success	Treatment failure
	n = 5062(%)	n = 3850(%)	n = 1212(%)
Non-Digital Community DOT	3481 (68.8)	2675 (69.5)	806 (66.5)
DS-TB regimen			
2RHZE/ 10 RH	378 (7.5)	266 (6.9)	112 (9.2)
2RHZE/ 4RH	4684 (92.5)	3584 (93.1)	1100 (90.8)
ART Status			
Already on ART	2521 (49.8)	1908 (49.6)	613 (50.6)
Newly initiated on Art	1149 (22.7)	878 (22.8)	271 (22.4)
Unknown	1392 (27.5)	1064 (27.6)	328 (27.1)

Model performance

Figure 1 presents the comparative performance of machine learning models developed to predict TB treatment outcomes among HIV/TB co-infected patients. Model performance varied across discrimination, recall, and accuracy metrics.

The stacked ensemble model demonstrated the strongest overall performance, with a recall for treatment failure of 0.60, ROC-AUC of 0.669, and overall accuracy of 0.618. Logistic regression models with class balancing (L1/Lasso and L2/Ridge) achieved similar performance, detecting 58–59% of failures, with ROC-AUC values around 0.65 and accuracy between 0.61 and 0.62.

The tuned Random Forest model yielded the highest ROC-AUC (0.675) among individual algorithms, maintaining balanced recall for treatment failures (0.58) and successes (0.67). The tuned Gradient Boosting Machine (GBM) achieved the highest accuracy (0.768) and near-perfect recall for successes (0.99), but its sensitivity to failures was low (recall = 0.07). Baseline Random Forest and Gradient Boosting models with SMOTE achieved accuracies of 0.709 and 0.715, respectively, but recalled ≤ 0.32 of failures.

The Support Vector Machine (SVM) model exhibited the weakest discrimination (ROC-AUC = 0.466). AdaBoost with SMOTE achieved moderate accuracy (0.672) and recall for successes (0.77) but detected only 35% of failures. See Fig. 1.

Model calibration

Figure 2 shows the calibration and precision-recall performance of the evaluated models. Across models, ROC-AUC values during cross-validation ranged from 0.637 to 0.655, reflecting fair discrimination. Precision-recall AUC (PR-AUC) values were consistently high (≥ 0.836), indicating that model predictions of treatment failure were generally reliable when made.

The tuned Random Forest achieved the highest PR-AUC (0.848), although calibration plots indicated slight overconfidence. Logistic regression models and the stacked ensemble exhibited moderate calibration with minor deviation from the ideal line at higher predicted probabilities. The tuned GBM model achieved the best calibration performance (Brier score = 0.172), although this improvement did not enhance failure detection. See Fig. 2

Feature importance analysis

Figure 3 presents the relative importance of predictors of TB treatment outcomes derived from the Random Forest model. Age was the most influential predictor, accounting for nearly half of the model's predictive strength. Other key demographic and clinical predictors included sex (male), marital status (married), ART status (unknown or newly initiated), TB classification (pulmonary bacteriologically confirmed and clinically diagnosed), and treatment model (community vs facility DOT). Programmatic variables, such as hospital care and treatment regimen (2RHZE/4RH), also contributed moderately to model predictions. Logistic regression analysis further identified facility-specific and clinical predictors associated with treatment success. Facilities including Mbale (OR = 2.82), Naguru (OR = 2.47), Hoima (OR = 2.25), Gulu (OR = 2.13), and Lira (OR = 1.98) Regional Referral Hospitals exhibited higher odds of treatment success. In contrast, Soroti (OR = 0.49), Moroto (OR = 0.49), Masaka (OR = 0.57), and St. Kizito Matany (OR = 0.63) hospitals had lower odds. Patients with pulmonary bacteriologically confirmed TB (OR = 2.15) and clinically diagnosed TB (OR = 1.64) were more likely to achieve treatment success, whereas those classified as "treatment after failure" (OR = 0.51) or managed under facility-based DOT (OR = 0.68) had lower odds. See Fig. 3.

Comparison of feature importance across models

Figure 4 compares normalised feature importance across Random Forest, GBM, and Logistic Regression models. Age consistently emerged as the most influential predictor across all models. Random Forest and GBM models additionally emphasised sex (male), ART status (unknown or newly initiated), marital status (married), X-ray result (unknown), and treatment model (non-digital community DOT).

The Logistic Regression model identified male sex (OR = 0.33), unknown ART status (OR = 0.40), and newly initiated ART (OR = 0.37) as factors associated with lower odds of treatment success. Patients under the non-digital community DOT model (OR = 0.43) also exhibited poorer outcomes. See Fig. 4.

Partial dependence analysis

Figure 5 illustrates the partial dependence of age on the predicted probability of TB treatment success. The relationship was non-linear, with younger patients (< 20 years) demonstrating the highest predicted probabilities of treatment success. The effect plateaued during early adulthood and declined progressively beyond 40 years, reaching the lowest probabilities among patients aged 70 years or older. See Fig. 5.

Discussions

The study evaluated the predictive capacity of several machine learning algorithms, including logistic regression, Random Forest, Gradient Boosting Machine (GBM), Support Vector Machine (SVM), AdaBoost, and a stacked ensemble modelling forecasting tuberculosis (TB) treatment outcomes among HIV/TB co-infected patients. Model performance was evaluated across key metrics of discrimination, recall, accuracy, and calibration to determine both their predictive power and clinical applicability. Furthermore, feature importance analysis was used to identify the most influential predictors of treatment outcomes, while partial dependence analysis illustrated the marginal effects of specific predictors, particularly age, on treatment success probabilities.

These findings indicate that the stacked ensemble and class-balanced logistic regression models achieved the most favourable balance between discrimination, calibration, and recall for treatment failures, supporting their application in programmatic contexts where accurate identification of patients at risk of poor outcomes is critical. The Random Forest model achieved the highest ROC-AUC, indicating strong discriminative ability, whereas the GBM attained the highest overall accuracy but was less sensitive to treatment failures. Across all models, key predictors of TB treatment outcome included age, ART status, sex, marital status, TB classification, and treatment modality. The partial dependence analysis further revealed a clear decline in treatment success with increasing age, particularly beyond 40 years, highlighting the clinical relevance of demographic and programmatic factors in predicting TB treatment outcomes.

Model performance and predictive utility

The study revealed that the stacked ensemble and class-balanced logistic regression models provided the most balanced and clinically useful predictions of TB treatment outcomes. These models were able to handle the imbalance between treatment success and failure effectively, achieving moderate discrimination and calibration. Their performance reflects the ability of ensemble approaches to combine complementary strengths of multiple algorithms, while regularized logistic models remain robust in managing collinearity and mixed data types. Comparable studies have reported similar findings, where ensemble and logistic regression models achieved an optimal balance between interpretability and accuracy in TB outcome prediction [10, 11]. This consistency supports the

application of these models in real-world TB programs, where interpretability and operational feasibility are as crucial as accuracy.

Discrimination of individual models

The tuned Random Forest exhibited the highest ROC-AUC among single models, while the tuned GBM yielded the highest overall accuracy but failed to adequately detect treatment failures. The Random Forest's superior discrimination may be attributed to its ability to capture complex, non-linear interactions among heterogeneous predictors. In contrast, the GBM's overemphasis on optimising overall accuracy led to overfitting to the dominant outcome class, treatment success, resulting in poor sensitivity to failures. This trade-off between accuracy and recall has been consistently documented in previous studies [12, 13], highlighting the need for careful tuning of machine learning models when dealing with imbalanced clinical datasets.

We found that age was the most influential predictor of TB treatment outcomes. The partial dependence analysis indicated a non-linear relationship between age and the probability of treatment success, with younger patients achieving higher success rates than older adults, particularly those aged 40 years and above, showing declining probabilities. This trend is biologically plausible, as older patients often have weakened immunity, higher prevalence of comorbidities, and greater treatment fatigue, which collectively undermine adherence and clinical response. Studies conducted in Uganda and elsewhere corroborate this finding, demonstrating that older age is associated with increased risk of poor TB outcomes[14, 15]. These observations emphasize the need for age-sensitive treatment support and follow-up strategies.

We also found that the ART status of these patients significantly influenced treatment outcomes. Patients who were newly initiated on ART or had unknown ART status were more likely to experience treatment failure compared to those already established on ART at baseline. This finding likely reflects late initiation of HIV care, poor linkage between TB and ART services, or incomplete treatment documentation. Unknown ART status also signals potential programmatic weaknesses in data recording or continuity of care. Similar associations have been reported in Ethiopia and Uganda, where delayed or interrupted ART initiation was a major contributor to unfavourable TB outcomes [16]. Strengthening ART–TB integration and ensuring early initiation of ART remain key programmatic priorities.

The study also found that male patients were more likely to experience treatment failure compared to female patients. This difference could be attributed to gendered disparities in health-seeking behaviour, adherence, and social support. Men are often diagnosed later, face occupational mobility challenges, and may have weaker support systems during treatment. Consistent with this, [17] and [14] observed that male patients had lower treatment success and higher default rates in Uganda. These findings underscore the importance of incorporating gender-sensitive approaches into TB programs, including flexible clinic schedules and targeted counselling for men.

In addition, marital status was also associated with treatment outcomes, with unmarried individuals experiencing higher failure rates. Married participants likely benefit from stronger social and emotional

support, enhancing adherence and clinic attendance. Similar associations between marital status and TB treatment success have been reported in sub-Saharan Africa [3], emphasizing the role of social support structures in sustaining adherence.

Patients with pulmonary bacteriologically confirmed or clinically diagnosed TB achieved better outcomes compared with those categorized under retreatment or treatment-after-failure groups. This likely reflects the benefits of early detection and diagnostic certainty, which facilitate appropriate treatment initiation and monitoring. In contrast, retreatment and treatment-after-failure categories typically include patients with prior drug resistance, adherence difficulties, or advanced disease. Comparable findings were reported by [3] and [16], who observed that diagnostic category and treatment history were strong predictors of TB treatment success in African settings. These results reinforce the importance of early diagnosis and differentiated care models for high-risk or previously treated patients.

The facility-based DOT model was associated with higher rates of treatment failure compared with community-based DOT models. This could be due to limited patient follow-up in busy facility settings and the convenience and social support offered by community-based models. Community DOT allows treatment supervision closer to patients' homes, reducing transportation barriers and improving adherence. Similar findings were observed in systematic reviews by [18] and Ugandan programmatic studies by [19], which showed superior treatment success under community-based approaches. However, [20] reported reduced success rates among previously treated patients managed under digital DOT, suggesting that the effectiveness of DOT interventions depends on context, implementation fidelity, and patient characteristics. Our results highlight the need to refine digital and facility-based DOT strategies to ensure optimal adherence support and supervision.

Model calibration and reliability

Model calibration and precision, recall analyses demonstrated fair discrimination and high precision for predicted failures, suggesting that when the models predicted a failure, they did so with high reliability. However, some models, particularly GBM, showed reduced recall, indicating difficulty in identifying all failure cases. This challenge arises from the inherent class imbalance in programmatic datasets, where treatment success overwhelmingly dominates. Recent studies emphasize that in clinical prediction, calibration metrics such as the Brier score and PR-AUC are as important as ROC-AUC for assessing real-world utility [11, 13]. The fair calibration observed in this study suggests that the models provide reasonably accurate probabilities that could be integrated into decision-support systems to identify high-risk patients.

Strengths and Limitations

This study benefited from a large, nationally representative large dataset that enhanced the reliability and generalizability of model findings. The application of multiple machine learning algorithms allowed for robust performance comparisons, while inclusion of calibration and interpretability analyses

strengthened clinical relevance. Feature importance and partial dependence analyses further added explanatory value by identifying key drivers of treatment outcomes.

Nevertheless, several limitations should be acknowledged. The retrospective design limited control over unmeasured confounders such as nutritional status, adherence behaviours, and socioeconomic conditions. Missing data, especially for ART status, may have introduced bias in model training. The imbalance between treatment success and failure restricted the models' sensitivity to failures, even after applying balancing techniques. Additionally, lack of external validation constrains the generalizability of findings to other settings.

Conclusions

This study demonstrates that machine learning models, particularly ensemble and regularised logistic regression approaches, can moderately predict TB treatment outcomes among HIV/TB co-infected patients in Uganda. The models achieved fair discrimination and calibration, highlighting their potential utility in programmatic risk assessment, although their performance remains insufficient for autonomous clinical decision-making.

Key predictors of treatment success identified across models included age, ART status, sex, marital status, TB classification, and treatment model. Among these, age and ART status were the most influential variables, reflecting the complex interplay between biological, clinical, and behavioural factors in determining treatment outcomes. The findings also revealed that older adults, males, and patients newly initiated or with unknown ART status were at higher risk of treatment failure. These insights highlighted the importance of integrating data-driven predictive models into routine TB and HIV program management to enable early identification and targeted support for patients at risk of poor outcomes. Overall, this study adds to the existing body of evidence on the application of machine learning in TB treatment monitoring and offers a structured framework for the development of predictive tools designed to strengthen existing clinical and surveillance systems, particularly in resource-limited settings.

Recommendations

Based on the study findings, several recommendations are proposed to advance both research and practice. For researchers, future investigations should prioritize prospective and external validation of the best-performing models using independent datasets from diverse geographical regions in Uganda. Such validation will enhance the generalizability and robustness of predictive models, allowing them to be effectively adapted to various programmatic contexts. Researchers should also consider including additional clinical, behavioural, and socioeconomic variables, such as HIV viral load, adherence patterns, nutritional indicators, and income levels, in future model development to improve predictive accuracy and interpretability.

For the Ministry of Health (MOH) and the National TB and Leprosy Program (NTLP), integration of predictive models into electronic medical record systems and digital health dashboards within national TB control programs should be prioritized. Embedding predictive analytics in these systems would enable real-time identification of high-risk patients, facilitate data-driven decision-making, and strengthen monitoring and evaluation of treatment outcomes.

For health workers and program implementers, TB control programs should strengthen community-based and digital directly observed treatment (DOT) models to improve adherence and continuity of care. Tailored follow-up strategies, such as home-based visits and mobile health interventions, should be expanded to support patients who face structural or socioeconomic barriers to treatment completion.

Finally, for policymakers and program managers, targeted interventions should focus on vulnerable subgroups, particularly older adults, men, and patients newly initiated or with unknown ART status. Differentiated adherence counselling, personalised treatment support, and closer clinical monitoring for these groups can reduce treatment failure, enhance survival, and contribute to Uganda's broader TB and HIV control goals.

Abbreviations

AUC

Area Under the Curve

ART

Antiretroviral Therapy

DOT

Directly Observed Therapy

DS

TB–Drug–Susceptible Tuberculosis

EDA

Exploratory Data Analysis

EMR

Electronic Medical Records

EP

Extra–Pulmonary

F1

Score–Harmonic Mean of Precision and Recall

GBM

Gradient Boosting Machine

HIV

Human Immunodeficiency Virus

IQR

Interquartile Range

KNN
K-Nearest Neighbors
L1
Lasso Regularization
L2
Ridge Regularization
MAR
Missing At Random
MCAR
Missing Completely At Random
MDR
TB-Multidrug-Resistant Tuberculosis
MICE
Multiple Imputation by Chained Equations
ML
Machine Learning
MNAR
Missing Not At Random
MOH
Ministry of Health
NTLP
National TB and Leprosy Program
PLWH
People Living With HIV
PR
AUC-Precision-Recall Area Under the Curve
RBF
Radial Basis Function
REC
Research Ethics Committee
RFE
Recursive Feature Elimination
ROC
Receiver Operating Characteristic
ROC
AUC-Area Under the Receiver Operating Characteristic Curve
RRH
Regional Referral Hospital
SD
Standard Deviation

SDGs
Sustainable Development Goals
SMOTE
Synthetic Minority Oversampling Technique
SVM
Support Vector Machine
TB
Tuberculosis
TN
True Negative
TP
True Positive
TSR
Treatment Success Rate
VIF
Variance Inflation Factor
WHO
World Health Organization
XDR
TB–Extensively Drug–Resistant Tuberculosis

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Makerere University School of Public Health Research Ethics Committee, Approval No. **542**. This study involved a secondary analysis of routinely collected, de-identified clinical data, and no direct interaction with human participants occurred. All study procedures were conducted in accordance with applicable ethical guidelines and regulations for research involving human participants. A waiver of informed consent was granted by the Research Ethics Committee due to the use of de-identified data. Data confidentiality and participant privacy were strictly maintained throughout all stages of the research process.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

There was no funding for this study

Authors' contributions

AO conceptualised and designed the study. JPA participated in data analysis, report writing, and drafting of the manuscript. AO and FO conducted data analysis. MN, EOE, EI, NC, and AK contributed to data acquisition, data management, and preliminary analyses. JPA, SRA, TN, RN, FO, and NRAO provided methodological and subject-matter expertise and critically reviewed and revised the manuscript for important intellectual content. RK supervised and reviewed the manuscript. NK provided overall supervision and reviewed the manuscript for scientific soundness. All authors read and approved the final version of the manuscript.

Acknowledgements

The authors gratefully acknowledge the Uganda Ministry of Health, Division of Health Information, for providing access to the data used in this study.

References

1. WHO, Tuberculosis. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis> (2025).
2. World Health Organization. The End Strategy TB. End TB Strateg. 2015;53:1689–99.
3. Teferi MY, El-Khatib Z, Boltana MT et al. Tuberculosis treatment outcome and predictors in africa: A systematic review and meta-analysis. Int J Environ Res Public Health; 18. Epub ahead of print 2021. 10.3390/ijerph182010678
4. WHO. World Tuberculosis Day. 2025: Uniting to End TB in Uganda. <https://www.afro.who.int/countries/uganda/news/world-tuberculosis-day-2025-uniting-end-tb-uganda> (2023).
5. MOH. *Uganda National TB and Leprosy Program Republic of Uganda Ministry of Health*. 2020.
6. Ali SA, Mavundla TR, Fantu R et al. Outcomes of TB treatment in HIV co-infected TB patients in Ethiopia: a cross-sectional analytic study. BMC Infect Dis 2016; 1–9.
7. MoH. Acceleration of HIV prevention in Uganda: A road map towards zero new infections by 2030.
8. WHO. *Global tuberculosis report, 2020*. 2021.
9. Izudi J, Tamwesigire IK, Bajunirwe F. Explaining the successes and failures of tuberculosis treatment programs; a tale of two regions in rural eastern Uganda. 2019; 2: 1–10.

10. Gichuhi HW, Magumba M, Kumar M, et al. A machine learning approach to explore individual risk factors for tuberculosis treatment non-adherence in Mukono district. *PLOS Glob Public Heal.* 2023;3:1–20.
11. Hosu MC, Faye LM, Apalata T. Optimizing Drug-Resistant Tuberculosis Treatment Outcomes in a High HIV-Burden Setting: A Study of Sputum Conversion and Regimen Efficacy in Rural South Africa. *Pathogens*; 14. Epub ahead of print 2025. 10.3390/pathogens14050441
12. Wang K, Wang Z, Li Z et al. Oriented Object Detection in Optical Remote Sensing Images using Deep Learning: A Survey. Epub ahead of print 2023. 10.1007/s10462-025-11256-0
13. Zhang L, Yu T, Zheng G, et al. Using machine learning to predict selenium content in crops: Implications for soil health and agricultural land utilization in longevity regions. *Sci Total Environ.* 2025;964:178520.
14. Kirenga BJ, Ssengooba W, Muwonge C, et al. Tuberculosis risk factors among tuberculosis patients in Kampala, Uganda: Implications for tuberculosis control. *BMC Public Health.* 2015;15:1–7.
15. Omara G, Bwayo D, Mukunya D, et al. Tuberculosis treatment success rate and its predictors among TB HIV co-infected patients in East and North Eastern Uganda. *Sci Rep.* 2025;15:5532.
16. Omara G, Bwayo D, Mukunya D, et al. Tuberculosis treatment success rate and its predictors among TB HIV co-infected patients in East and North Eastern Uganda. *Sci Rep.* 2025;15:5532.
17. Baluku JB, Mukasa D, Bongomin F, et al. Gender differences among patients with drug resistant tuberculosis and HIV co-infection in Uganda: a countrywide retrospective cohort study. *BMC Infect Dis.* 2021;21:1–11.
18. Wright DM, Reid N, Montgomery WI, et al. Herd-level bovine tuberculosis risk factors: Assessing the role of low-level badger population disturbance. *Sci Rep.* 2015;5:1–11.
19. Makabayi-Mugabe R, Musaaazi J, Zawedde-Muyanja S, et al. Community-based directly observed therapy is effective and results in better treatment outcomes for patients with multi-drug resistant tuberculosis in Uganda. *BMC Health Serv Res.* 2023;23:1–12.
20. Izudi J, Okello G, Bajunirwe F. Low treatment success rate among previously treated persons with drug-susceptible pulmonary tuberculosis in Kampala, Uganda. *J Clin Tuberc Other Mycobact Dis.* 2023;32:100375.

Figures

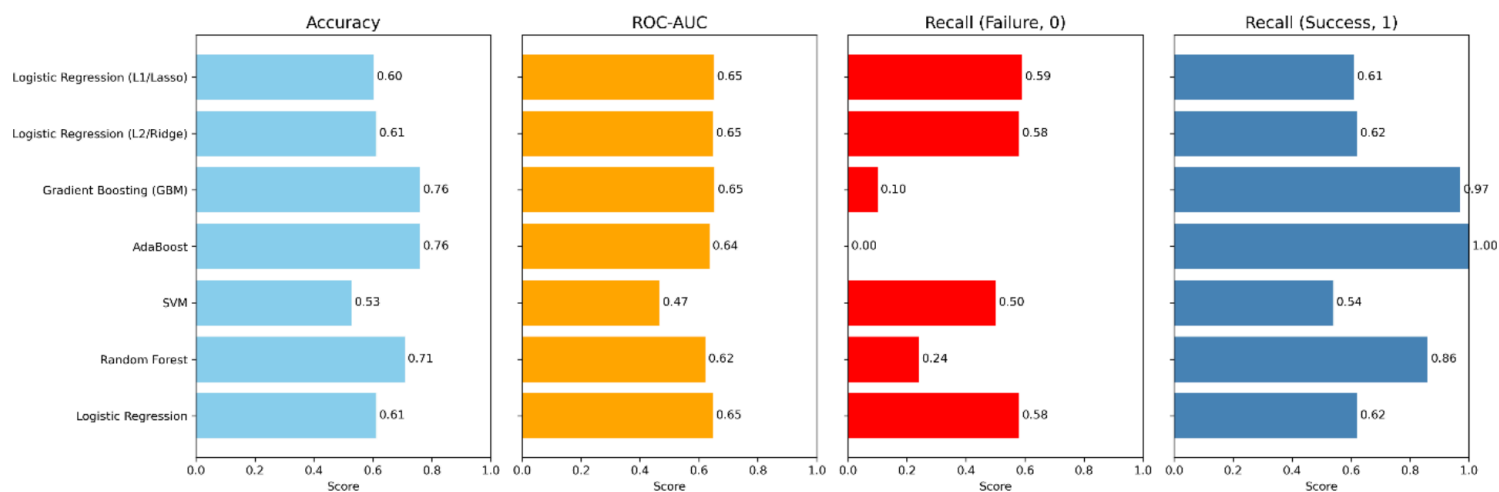


Figure 1

Performance of Machine Learning Models for TB Treatment Outcomes

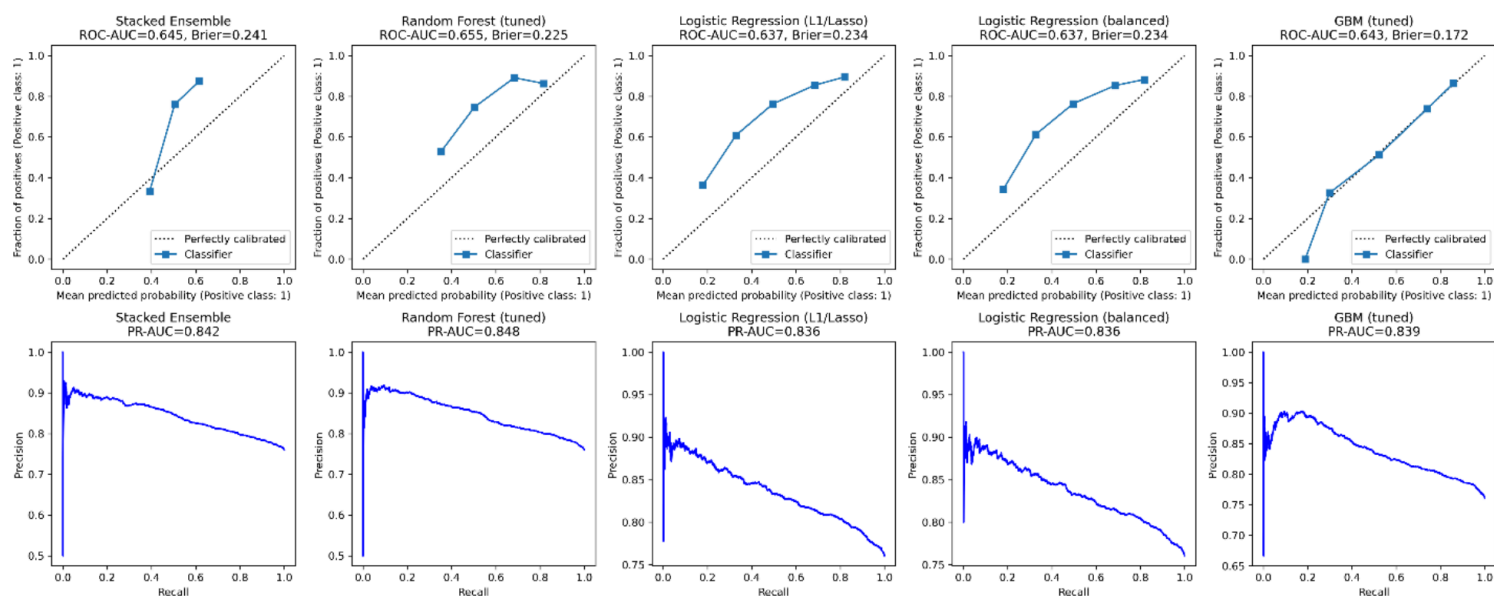


Figure 2

Calibration and precision-recall analysis of machine learning models.

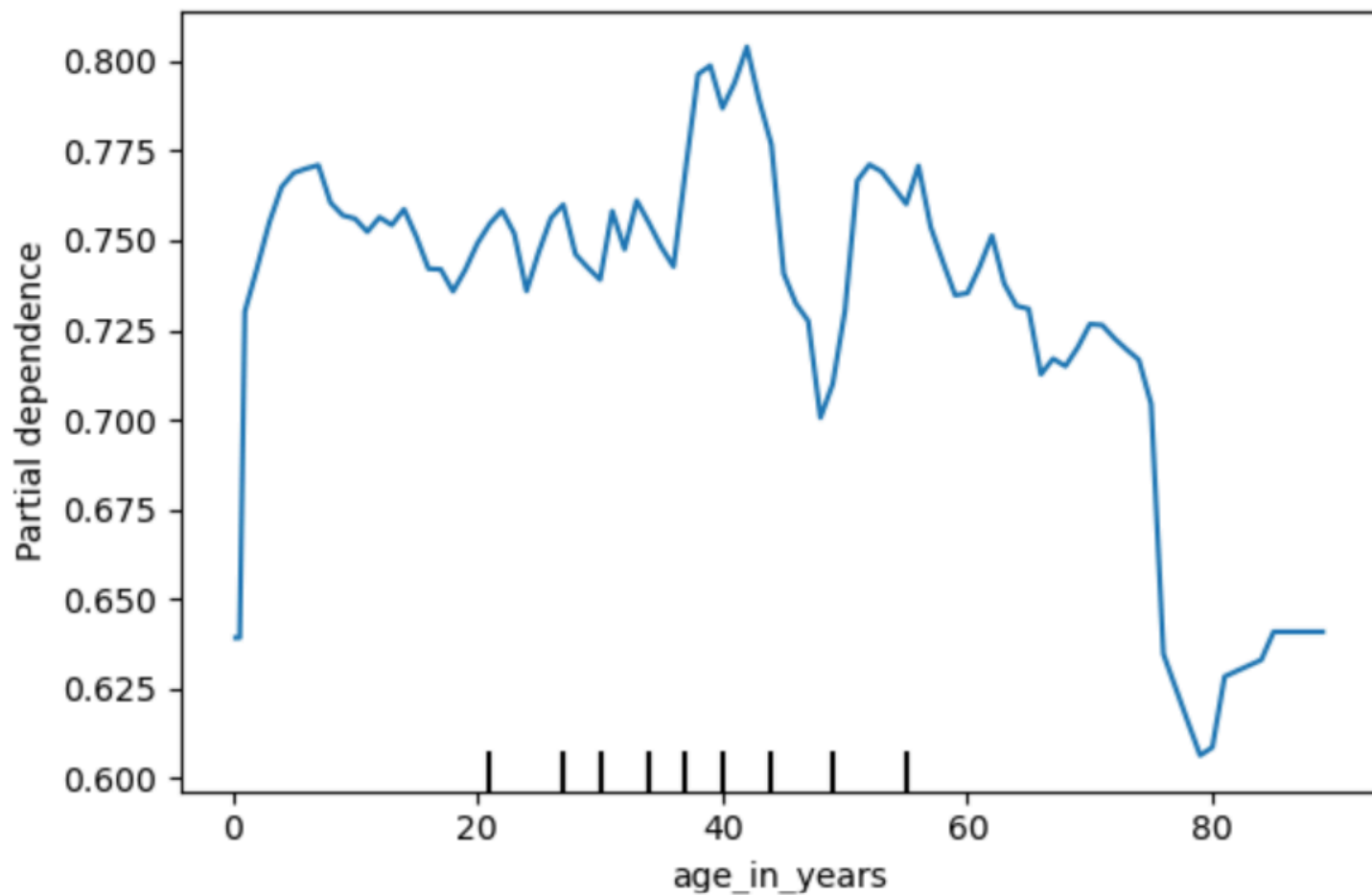


Figure 5

Partial dependence plot of age and predicted probability of TB treatment success