Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting tuberculosis treatment relapse among geriatric patient using machine learning algorithms in a high Tuberculosis-burden Ghanaian setting.

Michael Aidoo

maidoo28@st.knust.edu.gh

Kwame Nkrumah University of Science and Technology https://orcid.org/0009-0009-5681-869X

Dr. Rose-Mary Owusuaa Mensah

Kwame Nkrumah University of Science and Technology

Research Article

Keywords: Geriatric, Tuberculosis ,Treatment, Relapse, Predicting, Machine Learning Models, Random Forest, Support Vector Machine, Logistic Regression, Deep Neural Networks.

Posted Date: July 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10193652/v1

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

Additional Declarations: The authors declare no competing interests.

Abstract

Background, the world health organization estimated in 2020 of 5.8 million people newly diagnosed with Tuberculosis, predicting tuberculosis treatment relapse is crucial for improving patient treatment outcomes, studies have explored the factors connected with the treatment relapse risk in patients especially the elderly, Machine learning has transcended from a computational novelty role and emerged as a cornerstone of modern diagnostic and treatment outcome prediction. the argument is no longer whether machine learning should be used in healthcare, but how to integrate it responsibly. The Methodology involved training and testing Machine learning (SVM,RF,DNN,LR) models on a tuberculosis geriatric dataset for predicting treatment relapse on scikit-learn's train-test-split function into an 80% training and 20% Testing split. Results revealed random forest algorithm achieving an accuracy of 99%, 97% precision with a recall of 92%, F1-score of 95% and an area under curve (AUC ROC) of 96%, combination of high recall and very high precision suggests Random Forest effectively captures complex interactions among clinical and behavioural predictors, favourable balance between sensitivity and specificity makes it best among the SVM, DNN and LR algorithms. Recommendations proposed were to encourage data-sharing policies and electronic health records to enhance multi-agency collaboration and transparency, Random Forest model be considered for pilot integration into Tuberculosis clinic workflows at St. Michael Catholic Hospital and Continue model refinement to improve recall without compromising precision and accuracy.

Introduction

Tuberculosis (TB) remains among the leading infectious causes of death worldwide, and the burden is concentrated in low- and middle-income countries (LMICs), where a small number of high-burden nations account for the majority of new cases each year [1]. Despite the availability of curative regimens, progress towards the World Health Organization End TB targets has been slow, and gains in case detection and cure are unevenly distributed across populations [1, 2]. Older adults are a particularly vulnerable group. Ageing is accompanied by immune senescence and a high prevalence of comorbidities, and TB in older people is frequently characterized by atypical presentation, delayed diagnosis and reduced physiological reserve [3, 4, 5]. Diabetes mellitus, which is increasingly common in ageing African populations, impairs innate and adaptive immune responses to Mycobacterium tuberculosis and is associated with worse treatment outcomes [6, 7, 8]. In sub-Saharan Africa, treatment success rates among older pulmonary TB patients are consistently below programmatic targets, and age-stratified analyses identify the elderly as the group with the poorest outcomes and highest mortality [9, 10, 11]. In Ghana specifically, a ten-year retrospective review reported a declining treatment success rate, underscoring the need for context-specific strategies [9, 12].

Relapse, the recurrence of active disease after a patient has been declared cured or has completed treatment, is a clinically important adverse outcome that contributes to morbidity, onward transmission and the emergence of drug resistance. Older age at first diagnosis is associated with TB recurrence, and individual-patient-data meta-analyses have sought to identify predictors of relapse after standard

therapy, but reliable, individual-level relapse prediction remains challenging [13, 14, 15]. Established predictors such as gender, age and sputum-smear status explain only part of the variation, and most prediction work has targeted general adult populations rather than the elderly [15, 16].

Machine learning (ML) offers a flexible approach to modelling the complex, non-linear interactions among demographic, clinical and behavioral factors that drive TB outcomes. Systematic reviews and primary studies have applied ML to predict TB treatment failure, loss to follow-up, treatment success and progression from infection to disease, with reported discrimination commonly in the range of 0.70 to 0.95 [16, 17, 18, 19, 20, 21]. However, several gaps persist. First, few studies focus specifically on geriatric patients, despite their elevated risk. Second, African cohorts, and Ghanaian data in particular, are markedly under-represented, limiting the relevance of existing models to local clinical workflows [9, 22]. Third, many high-performing approaches rely on genomic, metabolomic or imaging features that are rarely available in routine programme settings, whereas tools built on routinely collected variables would be more readily deployable [21, 23]. The development of relapse-prediction models tailored to ageing African populations therefore represents both a methodological and a public-health priority, aligned with Sustainable Development Goal 3.3 and the End TB Strategy [2, 24, 25].

We aimed to develop and internally validate supervised machine learning models to predict TB treatment relapse among geriatric patients (aged 60 years and older) at a high-burden Ghanaian hospital, using routinely collected demographic, clinical and lifestyle data. Specifically, we compared the performance of logistic regression, random forest, support vector machine and a deep neural network, identified the most influential predictors, and assessed the stability of the best-performing model. The overall aim was to determine whether accurate, deployable relapse risk stratification is achievable using data already gathered in routine care; our results indicate that it is.

Materials and methods

Reporting and study design

This study is reported in accordance with the TRIPOD + AI statement for clinical prediction models developed using regression or machine learning methods [26], and the STROBE recommendations for observational studies where applicable. We conducted a single-centre, retrospective observational cohort study using existing patient records, an approach well suited to resource-constrained settings as it avoids the ethical and logistical demands of patient randomisation or prospective intervention. The study was carried out at St. Michael Catholic Hospital, Pramso, a secondary-level facility in the Ashanti Region of Ghana that serves a diverse population in an area where TB is prevalent among older adults.

Participants and eligibility

The target population comprised geriatric patients (aged 60 years or older at the time of TB diagnosis) who received a diagnosis of and treatment for TB at the hospital between January 2020 and January

2025. Patients were eligible if they were aged 60 years or older, had bacteriologically or clinically confirmed TB, had complete records of treatment outcome including relapse status, and had documentation of key demographic, clinical and lifestyle variables. Patients younger than 60 years, those with incomplete or ambiguous treatment-outcome records, those whose presentation was not primarily TB-related (incidental findings), and records with substantial missing data were excluded. After exclusions, the analytic cohort comprised 2,983 patients.

Outcome definition

The outcome was TB treatment relapse, defined in accordance with WHO classification as the recurrence of bacteriologically confirmed or clinically diagnosed active TB in a patient previously declared cured or treatment-completed, occurring within 24 months of completing a standard anti-TB regimen [1]. Relapse was ascertained from documented positive sputum-smear microscopy or culture after prior treatment success, or from clinical and radiological evidence of active disease requiring re-initiation of therapy. The problem was framed as binary classification (relapse versus non-relapse). Because molecular genotyping was unavailable, the study did not distinguish true relapse (reactivation of the original strain) from exogenous reinfection. Treatment failure, loss to follow-up and death during treatment were not modelled as relapse.

Data sources and candidate predictors

Data were extracted by trained research staff from the in-patient department and TB clinic registers and individual patient files (paper and electronic). Quality-assurance procedures included cross-checking against source records and double-entry verification of a random 10% sample to minimize transcription error. Candidate predictors, selected on the basis of the literature and clinical expertise, comprised: demographic variables (age in years; sex); clinical variables (comorbidities recorded as diabetes mellitus, hypertension and HIV; bacteriological status, bacteriologically confirmed versus clinically diagnosed; and TB type, pulmonary versus extrapulmonary); and lifestyle factors (alcohol use, smoking, or neither). Cohort characteristics are summarized in Table 1.

Table 1
Characteristics of the geriatric tuberculosis cohort (N = 2,983) and predictors used for modelling.

Attribute	Category	n	%	Type
Age (years)	60–70	1,700	57.0	Independent
	71–80	910	30.5	
	81+	373	12.5	
Sex	Female	2,055	68.9	Independent
	Male	928	31.1	
Bacteriological status	Bacteriologically confirmed	2,688	90.1	Independent
	Clinically diagnosed	295	9.9	
TB type	Pulmonary	2,691	90.2	Independent
	Extrapulmonary	292	9.8	
Comorbidities	Diabetes mellitus (DM)	1,127	37.8	Independent
	Hypertension (HPT)	1,000	33.5	
	HIV	574	19.2	
	DM + HPT	182	6.1	
	DM + HIV	1	0.03	
	None	99	3.3	
Lifestyle factors	Alcohol	963	32.3	Independent
	Smoking	859	28.8	
	None	719	24.1	
	Alcohol + smoking	442	14.8	
Outcome	Non-relapse	2,722	91.3	Dependent
	Relapse	261	8.7	

Percentages for multi-value comorbidity and lifestyle fields are computed over recorded entries and may sum to more than the cohort size where patients had more than one factor.

Data pre-processing and feature engineering

Pre-processing was performed in Python using pandas. Missing values were imputed with the column mode, preserving sample size and the distribution of the predominantly categorical variables; no missing

values remained after imputation. Age was screened for outliers using the interquartile-range method, and no values were removed as all fell within clinically plausible ranges. Categorical fields were standardised (uniform case, whitespace trimming, consolidation of synonymous entries). Categorical predictors were one-hot encoded; binary variables used drop-first encoding to limit collinearity. Multi-value fields (comorbidities; lifestyle) were split and transformed into binary indicator columns for each factor (e.g., diabetes mellitus, hypertension, HIV, alcohol, smoking). The target was coded as relapse (positive class) versus non-relapse.

Data splitting and class imbalance

The dataset was partitioned using scikit-learn into an 80% training set (n = 2,386) and a 20% test set (n = 597), with stratification on the outcome to preserve the relapse proportion in both partitions. Relapse was the minority class (8.7% of the cohort). To address imbalance without contaminating the held-out data, the Synthetic Minority Over-sampling Technique (SMOTE) [27] was applied to the training set only, after the split, generating synthetic minority instances by interpolation until the classes were approximately balanced. The test set retained the natural class distribution to provide an unbiased estimate of performance. All imputation, encoding, scaling and resampling parameters were derived from the training data.

Model development

Four supervised classifiers were trained for their complementary strengths: logistic regression (interpretable linear baseline), random forest (ensemble of decision trees capturing non-linear interactions and providing feature importance's), support vector machine (effective in high-dimensional spaces with non-linear kernels), and a feed-forward deep neural network (capacity to model complex patterns). Logistic regression, random forest and support vector machine were implemented in scikit-learn [28] and the neural network in TensorFlow/Keras.

Logistic regression used the liblinear solver with L2 regularisation (C = 1.0) and balanced class weights. Random forest [29] used 100 trees, Gini impurity, unrestricted depth, the square-root rule for features per split, bootstrap sampling and balanced class weights. The support vector machine used a radial-basis-function kernel (C = 1.0, gamma = 'scale'), balanced class weights and probability estimates enabled for AUC computation. The deep neural network comprised an input layer, two hidden layers of 64 and 32 units with rectified-linear activation, each followed by 50% dropout, and a single sigmoid output unit. It was trained with the Adam optimiser (learning rate 0.001), binary cross-entropy loss, batch size 32, class weighting for imbalance, and up to 50 epochs with early stopping (patience = 10 epochs) on a held-out validation subset. Hyperparameters otherwise followed library defaults; formal grid or Bayesian search was not performed given the sample size. Key training parameters are summarised in Table 2.

Table 2
Principal training parameters for the four machine learning models.

Model	Key configuration
Logistic regression	Solver liblinear; L2 penalty; C = 1.0; class_weight = balanced; max_iter = 1,000
Random forest	n_estimators = 100; criterion = Gini; max_depth = none; max_features = sqrt; bootstrap = true; class_weight = balanced
Support vector machine	Kernel RBF; C = 1.0; gamma = scale; class_weight = balanced; probability = true
Deep neural network	Input → Dense(64, ReLU) → Dropout(0.5) → Dense(32, ReLU) → Dropout(0.5) → Dense(1, sigmoid); Adam (lr 0.001); binary cross-entropy; batch 32; ≤50 epochs, early stopping patience 10; class weighting

Evaluation

Models were evaluated on the held-out test set using accuracy, precision, recall (sensitivity), F1-score and the area under the receiver operating characteristic curve (AUC). Given the clinical priority of detecting relapse, recall and the precision–recall balance were emphasised alongside overall discrimination. Confusion matrices were generated to characterise error types (true and false positives and negatives). To assess stability beyond a single partition, the best-performing model (random forest) was additionally evaluated by stratified five-fold cross-validation, with the mean and standard deviation of each metric reported. Feature importances were derived from the random forest to identify the most influential predictors.

Risk stratification

Predicted probabilities from the best model were used to assign patients to provisional risk tiers (low, < 0.2; medium, 0.2–0.5; high, > 0.5) to illustrate potential clinical use. These thresholds are illustrative and would require calibration and refinement through receiver-operating-characteristic analysis and consultation with clinicians to reflect local resources before any deployment.

Ethics statement

The study used fully anonymized secondary data; all direct identifiers (names, hospital numbers and contact details) were removed before analysis, and data were held on encrypted, password-protected systems. As a retrospective analysis of existing records with no patient contact, the requirement for individual informed consent was waived. Institutional approval for the use of secondary data was granted by St. Michael Catholic Hospital through its medical records unit, and ethical clearance was obtained from the Committee on Human Research, Publications and Ethics (CHRPE) of the Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana (approval granted 30 October

2025; application number 25478946832). The study adhered to recognised data-protection principles. Model fairness was examined across sex and comorbidity status to check for disparate performance.

Results

Cohort

The analytic cohort comprised 2,983 geriatric TB patients treated between January 2020 and January 2025 (Table 1). Most patients were aged 60–70 years (57.0%), female (68.9%), bacteriologically confirmed (90.1%) and had pulmonary TB (90.2%). Diabetes mellitus (37.8%), hypertension (33.5%) and HIV (19.2%) were the commonest comorbidities, and alcohol use (32.3%) and smoking (28.8%) were frequent. Relapse occurred in 261 patients (8.7%); the stratified split yielded 52 relapse and 545 non-relapse cases in the test set.

Discrimination and overall performance

All four models discriminated relapse from non-relapse well on the held-out test set, with AUC values from 0.959 to 0.966 and accuracy above 0.97 for every algorithm (Table 3). Random forest achieved the most favorable overall profile, with an accuracy of 0.992, precision of 0.980, recall of 0.923, F1-score of 0.951 and AUC of 0.966. The deep neural network, logistic regression and support vector machine all attained high recall (0.942) for relapse but lower precision than random forest (0.831, 0.803 and 0.790 respectively), indicating a greater tendency to flag non-relapse patients as high risk.

Table 3
Test-set performance of the four models for geriatric tuberculosis relapse prediction.

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic regression	0.975	0.803	0.942	0.867	0.964
Random forest	0.992	0.980	0.923	0.951	0.966
Support vector machine	0.973	0.790	0.942	0.860	0.959
Deep neural network	0.978	0.831	0.942	0.883	0.962

Precision, recall and F1-score refer to the relapse (positive) class. AUC, area under the receiver operating characteristic curve.

Error analysis

Confusion matrices on the held-out test set (Table 4) showed that all models correctly classified the large majority of cases, with only three or four false negatives each. The models differed mainly in false positives: random forest produced a single false positive, compared with 10 for the deep neural network, 12 for logistic regression and 13 for the support vector machine. The combination of few false negatives

and very few false positives make random forest the most clinically attractive option, minimizing both missed relapses and unnecessary escalation of care.

Table 4
Confusion-matrix summary on the held-out test set (n = 597).

Model	True negatives	False positives	False negatives	True positives
Support vector machine	532	13	3	49
Deep neural network	535	10	3	49
Logistic regression	533	12	3	49
Random forest	544	1	4	48

Stability of the random forest model

Stratified five-fold cross-validation of the random forest confirmed stable performance across resampled partitions (Table 5): mean accuracy 0.984 (SD 0.005), precision 0.914 (SD 0.037), recall 0.904 (SD 0.026), F1-score 0.909 (SD 0.027) and AUC 0.971 (SD 0.008). The high mean recall with low variability indicates a reliable capacity to identify relapse cases, while the strong precision indicates this is achieved without excessive false alarms. The low standard deviations across all metrics suggest the model is not overly dependent on any single partition.

Table 5
Stratified five-fold cross-validation of the random forest model.

Metric	Mean	Standard deviation
Accuracy	0.984	0.005
Precision	0.914	0.037
Recall	0.904	0.026
F1-score	0.909	0.027
AUC	0.971	0.008

Predictor importance

Feature-importance analysis of the random forest identified bacteriological status (bacteriologically confirmed versus clinically diagnosed) as the most influential predictor, followed by TB type (pulmonary versus extrapulmonary), age, and lifestyle factors (alcohol use and smoking). These rankings are clinically plausible: bacteriological status and disease site reflect differences in disease burden, diagnostic certainty and treatment response, while advancing age and behavioral factors are recognized contributors to poor outcomes and interact with comorbidities such as HIV and diabetes to elevate relapse risk.

Discussion

In this single-centre retrospective cohort of nearly three thousand geriatric TB patients in Ghana, supervised machine learning models trained on routinely collected demographic, clinical and lifestyle data predicted treatment relapse with high discrimination. Random forest performed best, combining the highest accuracy and F1-score with excellent precision and competitive AUC, and its error profile, very few false negatives and only a single false positive, is well matched to the dual clinical objective of detecting relapse while avoiding unnecessary escalation of care. Cross-validation confirmed that this performance was stable across resampled partitions.

Comparison with previous work

Few studies have applied machine learning specifically to relapse prediction in elderly African TB patients; most prior work has addressed broader unfavorable outcomes (treatment failure, mortality, loss to follow-up) in mixed-age or general adult populations, typically reporting AUC values between 0.70 and 0.90 using clinical and demographic features [16, 17, 19, 30]. Approaches based on genomic, metabolomic or imaging biomarkers can achieve AUCs of roughly 0.80 to 0.95 but depend on specialized tests rarely available in routine programmes [21, 23]. Against this backdrop, the present model's discrimination using only routine variables is encouraging, although direct comparison is limited by differences in outcome definitions and populations (Table 6). The strong associations between relapse and key predictors in this elderly cohort, together with explicit handling of class imbalance, likely contributed to the high performance.

Table 6
Selected machine learning studies on tuberculosis treatment outcomes in context.

Study	Population	Outcome	Best model	AUC	Accuracy
Asad et al. 2020 [17]	Adults, 6 countries	Treatment failure	Neural network	N/A	0.78
Peng et al. 2024 [19]	TB–diabetes comorbidity	Treatment failure	Explainable ML	0.85	N/A
Rodrigues et al. 2024 [20]	Adults, Brazil (registry)	Loss to follow-up	RF / SVM	0.90	0.85
This study (2025)	Geriatric, Ghana	Relapse	Random forest	0.966	0.992

Clinical and public-health implications

The consistently high recall across models indicates that most patients who eventually relapse could be flagged in advance. In a geriatric population with reduced physiological reserve and a high prevalence of HIV and diabetes, early identification creates an opportunity for intensified follow-up, closer monitoring and targeted adherence support, which may reduce morbidity, mortality and transmission [13, 31]. The high precision of the random forest is equally important in resource-limited settings: by limiting false

positives, it avoids overburdening staff and diagnostic capacity, a recurrent concern with clinical prediction tools [16, 31]. Stratifying relapse risk from routine data could support differentiated care, prioritising high-risk patients for home visits, community adherence support or more frequent sputum monitoring, while lower-risk patients follow standard protocols. Because the model uses only routinely available variables, it is in principle scalable within the Ghanaian health system and, subject to external validation, comparable LMIC settings [22, 32].

Strengths and limitations

Strengths include the focus on an under-studied but highly vulnerable geriatric population, the use of a locally curated dataset reflecting real-world conditions, the comparison of four algorithms under a common framework, explicit handling of class imbalance with resampling confined to the training data, and cross-validation of the best model. Several limitations temper interpretation. The study was conducted at a single centre, and no external validation was performed, so generalizability to other facilities and regions is unknown; external, multi-centre validation is the essential next step. The very high reported metrics warrant caution: although resampling was confined to training data, the possibility of residual optimism or information leakage cannot be excluded, and we did not report calibration, confidence intervals or decision-curve analysis, which would strengthen evidence of clinical utility. Relapse events were relatively few (n = 261), which can destabilise estimates, particularly for the neural network. Potentially relevant predictors, notably socioeconomic indicators and molecular data to distinguish relapse from reinfection, were unavailable. Finally, the neural network's reduced interpretability may limit clinician acceptance without additional explainability methods. Future work should prioritise prospective and external validation, calibration and net-benefit assessment, incorporation of socioeconomic and longitudinal variables, and integration of an interpretable model into electronic records with clinician oversight.

Conclusions

Supervised machine learning applied to routinely collected hospital data can accurately and stably predict TB treatment relapse among geriatric patients in a high-burden Ghanaian setting, with random forest providing the most favourable balance of sensitivity and precision. Bacteriological status, TB type, age and lifestyle factors were the most influential predictors. These findings support the feasibility of data-driven relapse risk stratification to guide targeted, resource-efficient geriatric TB care, contingent on external validation before clinical deployment.

References

1. World Health Organization (2024) Global Tuberculosis Report 2024. World Health Organization, Geneva
2. World Health Organization (2015) The End TB Strategy. World Health Organization, Geneva

3. Yew WW, Yoshiyama T, Leung CC, Chan DP (2018) Epidemiological, clinical and mechanistic perspectives of tuberculosis in older people. Respirology 23(6):567–575
4. Abbara A, Collin SM, Kon OM, Buell K, Sullivan A, Barrett J et al (2019) Time to diagnosis of tuberculosis is greater in older patients: a retrospective cohort review. ERJ Open Res 5(4):00228–2018
5. Zhang CY, Zhao F, Xia YY, Yu YL, Shen X, Lu W et al (2019) Prevalence and risk factors of active pulmonary tuberculosis among elderly people in China: a population-based cross-sectional study. Infect Dis Poverty 8:7
6. van Crevel R, Critchley JA (2021) The interaction of diabetes and tuberculosis: translating research to policy and practice. Trop Med Infect Dis 6(1):8
7. Kumar NP, Babu S (2023) Impact of diabetes mellitus on immunity to latent tuberculosis infection. Front Clin Diabetes Healthc 4:1095467
8. Ye Z, Li L, Yang L, Zhuang L, Aspatwar A, Wang L et al (2024) Impact of diabetes mellitus on tuberculosis prevention, diagnosis, and treatment from an immunologic perspective. Exploration 4(5):20230138
9. Puplampu P, Kyeremateng I, Asafu-Adjaye O, Asare AA, Agyabeng K, Sarkodee R et al (2024) Evaluation of treatment outcomes among adult patients diagnosed with tuberculosis in Ghana: a 10-year retrospective review. IJID Reg 10:9–14
10. Bonsu F, Addo KK, Alebachew Z, Gyapong J, Badu-Peprah A, Gockah R et al (2020) National population-based tuberculosis prevalence survey in Ghana, 2013. Int J Tuberc Lung Dis 24(3):321–328
11. Ncube RT, Takarinda KC, Zishiri C, Van den Boogaard W, Mlilo N, Chiteve C et al (2017) Age-stratified tuberculosis treatment outcomes in Zimbabwe: are we paying attention to the most vulnerable? Public Health Action 7(3):212–217
12. Izudi J, Semakula D, Sennono R, Tamwesigire IK, Bajunirwe F (2019) Treatment success rate among adult pulmonary tuberculosis patients in sub-Saharan Africa: a systematic review and meta-analysis. BMJ Open 9(9):e029400
13. van der Heijden YF, Karim F, Chinappa T, Mufamadi G, Zako L, Shepherd BE et al (2018) Older age at first tuberculosis diagnosis is associated with tuberculosis recurrence in HIV-negative persons. Int J Tuberc Lung Dis 22(8):871–877
14. Romanowski K, Balshaw RF, Benedetti A, Campbell JR, Menzies D, Khan FA et al (2019) Predicting tuberculosis relapse in patients treated with the standard 6-month regimen: an individual patient data meta-analysis. Thorax 74(3):291–297
15. Dela Cruz AP (2018) Predicting the relapse category in patients with tuberculosis: a chi-square automatic interaction detector (CHAID) decision tree analysis. Open J Soc Sci 6(12):29–36
16. Peetluk LS, Ridolfi FM, Rebeiro PF, Liu D, Rolla VC, Sterling TR (2021) Systematic review of prediction models for pulmonary tuberculosis treatment outcomes in adults. BMJ Open 11(3):e044687

17. Asad M, Mahmood A, Usman M (2020) A machine learning-based framework for predicting treatment failure in tuberculosis: a case study of six countries. Tuberculosis 123:101944
18. Ahamed Fayaz S, Babu L, Paridayal L, Vasantha M, Paramasivam P, Sundarakumar K et al (2024) Machine learning algorithms to predict treatment success for patients with pulmonary tuberculosis. PLoS ONE 19(10):e0309151
19. Peng AZ, Kong XH, Liu ST, Zhang HF, Xie LL, Ma LJ et al (2024) Explainable machine learning for early predicting treatment failure risk among patients with TB-diabetes comorbidity. Sci Rep 14(1):6814
20. Rodrigues MM, Barreto-Duarte B, Vinhaes CL, Araújo-Pereira M, Fukutani ER, Bergamaschi KB et al (2024) Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment. BMC Public Health 24(1):1385
21. Balakrishnan V, Kherabi Y, Ramanathan G, Paul SA, Tiong CK (2023) Machine learning approaches in diagnosing tuberculosis through biomarkers — a systematic review. Prog Biophys Mol Biol 179:16–25
22. Gichuhi HW, Magumba M, Kumar M, Mayega RW (2023) A machine learning approach to explore individual risk factors for tuberculosis treatment non-adherence in Mukono district. PLOS Glob Public Health 3(7):e0001466
23. Anh NK, Phat NK, Thu NQ, Tien NTN, Eunsu C, Kim HS et al (2024) Discovery of urinary biosignatures for tuberculosis and nontuberculous mycobacteria classification using metabolomics and machine learning. Sci Rep 14(1):15312
24. Pai M, Furin J (2017) Tuberculosis innovations mean little if they cannot save lives. eLife 6:e25956
25. Berihe Hiluf S, Abera A, Bahiru M, Kassie B (2024) Determinants of unsuccessful tuberculosis treatment outcome in Southwest Ethiopia regional state public hospitals, 2022: a multi-center case–control study. Front Public Health 12:1406211
26. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B et al (2024) TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385:e078378
27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
29. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
30. Bartl L, Zeeb M, Kälin M, Loosli T, Notter J, Furrer H et al (2025) Machine learning-based prediction of active tuberculosis in people with human immunodeficiency virus using clinical data. Clin Infect Dis. ;ciaf149
31. Dohál M, Porvazník I, Solovič I, Mokrý J (2023) Advancing tuberculosis management: the role of predictive, preventive, and personalized medicine. Front Microbiol 14:1225438

32. Hosu MC, Faye LM, Apalata T (2024) Comorbidities and treatment outcomes in patients diagnosed with drug-resistant tuberculosis in rural Eastern Cape Province, South Africa. Diseases 12(11):296