BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Development of a machine learning prediction model for loss to follow-up in HIV care using routine electronic medical records in a low-resource setting

Tamrat Endebu[1*], Girma Taye[1] and Wakgari Deressa[1]

Abstract

Background Despite the global commitment to ending AIDS by 2030, the loss of follow-up (LTFU) in HIV care remains a significant challenge. To address this issue, a data-driven clinical decision tool is crucial for identifying patients at greater risk of LTFU and facilitating personalized and proactive interventions. This study aimed to develop a prediction model to assess the future risk of LTFU in HIV care in Ethiopia.

Methods The study used a retrospective design in which machine learning (ML) methods were applied to the electronic medical records (EMRs) data of adult HIV-positive individuals who were newly enrolled in antiretroviral therapy between July 2019 and April 2024. The data were collected across eight randomly selected high-volume healthcare facilities. Six supervised ML classifiers—J48 decision tree, random forest, K-nearest neighbors, support vector machine, logistic regression, and naïve Bayes—were utilized for training via Weka 3.8.6 software. The performance of each algorithm was evaluated through a 10-fold cross-validation approach. Algorithm performance was compared via the corrected resampled t test ($p < 0.05$), and decision curve analysis (DCA) was used to assess the model's clinical utility.

Results A total of 3,720 individuals' EMR data were analyzed, with 2,575 (69.2%) classified as not LTFU and 1,145 (30.8%) classified as LTFU. On the basis of the ML feature selection process, six strong predictors of LTFU were identified: differentiated service delivery model, adherence, tuberculosis preventive therapy, follow-up period, nutritional status, and address information. The random forest algorithm showed superior performance, with an accuracy of 84.2%, a sensitivity of 82.4%, a specificity of 85.7%, a precision of 83.7%, an F1 score of 83.1%, and an area under the curve of 89.5%. The model demonstrated greater clinical utility, offering greater net benefit than both the 'intervention for all' approach and the 'intervention for none' approach, particularly at threshold probabilities of 10% and above.

Conclusions This study developed a machine learning-based predictive model for assessing the future risk of LTFU in HIV care within low-resource settings. Notably, the model built via the random forest algorithm exhibited

*Correspondence:
Tamrat Endebu
seetame2008@gmail.com

Full list of author information is available at the end of the article

high accuracy and strong discriminative performance, highlighting its positive net benefit for clinical applications. Furthermore, ongoing external validation across diverse populations is important to ensure the model's reliability and generalizability.

Keywords Machine learning, Model development, Prediction tool, Loss to follow-up, HIV, Low-resource setting

Introduction

In 2023, 39.9 million people were living with human immunodeficiency virus (HIV) globally, of which 1.3 million people contracted it and 630,000 died from acquired immunodeficiency syndrome (AIDS) [1]. Despite the global commitment to ending AIDS by 2030, significant barriers persist, particularly in low-resource settings [2]. One of the most critical challenges in reaching this target is the high rate of loss to follow-up (LTFU) among patients enrolled in HIV care [3].

LTFU refers to patients who miss their HIV care or antiretroviral therapy (ART) appointment by more than 28 days from the scheduled date [4]. It is a pervasive problem in low-income settings where healthcare systems are often under resourced and struggle with infrastructural limitations. For example, LTFU rates are alarming in several countries: 23.4% in South Africa [5], 57.4% in Tanzania [6], 27.2% in Kenya [7], and 15.17% in Ethiopia [8]. These figures underscore a systemic failure to retain patients in care, which directly correlates with unsuppressed viral loads and increased morbidity and mortality associated with HIV [9]. Moreover, LTFU contributes not only to poor individual health outcomes but also to ongoing HIV transmission within communities, thereby hampering broader public health initiatives to control the epidemic [10].

Recognizing the risk of LTFU in HIV care—especially during the first five years after initiating ART—is vital [11]. This timeframe is marked by increased vulnerability, with many patients discontinuing treatment [12, 13], which can adversely affect their health outcomes and undermine the overall effectiveness of ART programs [11]. Therefore, an urgent solution is to address this risk and enhance patient retention through innovative approaches that leverage available data sources [14, 15]. Machine learning (ML) offers a promising approach for addressing this challenge by the use of routine electronic medical records (EMRs) to predict which patients are at risk of LTFU. In recent years, the application of ML algorithms to EMRs has gained traction across various medical fields, particularly in predicting patient outcomes and conducting risk assessments [16, 17].

Research has identified a range of risk factors for LTFU by analyzing historical sociodemographic and clinical data within EMRs. Key predictors include sociodemographic variables such as age, sex, and marital status, alongside clinical indicators such as tuberculosis preventive therapy (TPT) [18, 19], differentiated service delivery (DSD) [20], nutritional status [20, 21], adherence to treatment [20, 22], and patient address information [23, 24]. The literature also highlights other risk factors accessible through EMR, including employment status [24, 25], history of missed appointments [19], poor functional status, low CD4 count, and advanced clinical stage [7, 26–28].

While several studies have attempted to predict the risk of loss to follow-up (LTFU) in HIV care, most have been conducted in high-income settings [27, 29–31], limiting their generalizability to low-resource environments. Although some research efforts have emerged from Sub-Saharan Africa—including South Africa [22], Nigeria [32], Tanzania [21], and Ethiopia [20]—these studies often lack comprehensive model performance evaluations. Moreover, few studies have addressed the clinical utility or practical applicability of these models in real-world settings. Thus, this study aimed to develop a machine learning-based prediction model for LTFU in HIV care during the first five years after initiating ART. This study introduces an EMR-based prediction tool that can help clinicians make informed decisions to improve HIV patient retention in care.

Methods

Study settings and participants

The study used a retrospective design in which machine learning methods utilizing EMR data were employed to predict the future risk of loss to follow-up in HIV care in an urban environment in Ethiopia. The estimated HIV prevalence in urban areas of Ethiopia is approximately 3.4% [33], with over 465,457 adult HIV-positive individuals receiving ART [34]. The study included adults aged 15 years and older who tested HIV positive and began ART between July 2019 and April 2024. Patients with incomplete information regarding the outcome variable, as well as those who were transferred in (TI), transferred out (TO), had recorded deaths, or restarted treatment, were excluded from the analysis.

Sample size determination and sampling procedures

To determine the sample size required for developing a prediction model for a binary outcome, several factors must be considered. Key among these are estimating the overall outcome proportion with adequate precision, targeting a small mean absolute prediction error, and establishing a shrinkage factor to minimize optimism in the apparent R^2 Nagelkerke [35]. In Ethiopia, the pooled proportion of patients lost to follow-up was 0.15 [8],

with 30 potential predictor parameters hypothesized. For logistic regression models with this outcome proportion, the maximum R^2 value corresponds to 0.48 [35]. Assuming that the new model would explain 15% of the variability, the anticipated R^2 Nagelkerke value was calculated as $0.15 \times 0.48 = 0.05$. Using Stata with the command "*pmsampsize, type(b) rsquared(0.07) parameters(30) prevalence(0.15)*," the minimum sample size required for developing the new model was 3,706, which included 556 events. Accounting for a 5% attrition rate, the estimated total sample size needed was approximately 3,891.

To select the participants, we first identified 21 high-case-load health facilities in central Ethiopia that had enrolled at least 200 new HIV patients from 2019 to 2023 [36]. We then randomly selected eight facilities: three in Addis Ababa—Zewditu Hospital ($N = 441$), ALERT Hospital ($N = 587$), and Yekatit 12 Hospital ($N = 329$)—and five in nearby Oromia urban areas—Bishoftu Hospital ($N = 827$), Adama Teaching Hospital ($N = 542$), Geda Health Center ($N = 402$), Adama Health Center ($N = 351$), and Asella Hospital ($N = 231$). All eligible patients from these selected facilities were included in the study.

Prediction features

Outcome feature

In this study, LTFU in HIV care was the target feature. LTFU refers to patients who miss their HIV care or ART appointments for 28 days or more from the date of their last scheduled appointment [4]. If the patient was LTFU, the feature was coded as 'Yes,' and if the patient had not been LTFU within the past 5 years, the feature was coded as 'No.'

Predictor features

In this study, we defined the following features used to predict LTFU in accordance with the national consolidated guidelines for comprehensive HIV prevention, care, and treatment [37]. The features identified include demographic information such as sex (male vs. female) and age at enrollment. Address details (green vs. yellow) were categorized as either green or yellow, with green indicating complete and accurate information, including a phone number and a detailed kebele address, while yellow signified incomplete or missing details. The follow-up periods (time 0–12 vs. time 13–60 months) were time intervals after initiating ART. Adherence (good vs. poor) to medication was classified into two categories: good adherence (defined as taking at least 95% of doses) and poor adherence (less than 85%). Additionally, the status of tuberculosis prevention therapy (TPT) status (gold vs. bronze/silver) was recorded as gold, silver, or bronze. Gold indicated the completion of TPT, bronze indicated that TPT had not started, and silver indicated that TPT had started but not completed. The differentiated service delivery (DSD) model category (ASM/3MMD vs. not-enrolled/other DSD forms) was identified for each patient. This included options such as the appointment spacing model, which also referred to those receiving the 6-month multimonth dispensing (MMD) model, 3MMD, and other DSD models such as the advanced disease (ADH) model, key populations model, adolescent model, young people model, and prevention of mother–child transmission (PMTCT) model. Furthermore, we assessed nutritional status (normal vs. undernutrition) on the basis of body weight relative to height. The WHO clinical stage (WHO Stage 1/2 vs. WHO Stage 3/4) was noted, which indicates the severity of disease progression in adults or adolescents with a CD4 count below 200 cells/mm^3. These EMR-based features collectively provide a comprehensive framework for predicting LTFU in patients undergoing ART [37].

Data collection and quality control

The data extraction tool was developed using Ethiopian national HIV care/ART intake and follow-up forms for routine patient care [38]. Each health facility's data manager, under the supervision of two experienced supervisors, extracted deidentified patient data. The research team provided a two-day training session for data collection facilitators, covering the abstraction tool, data management protocols, extraction processes, and confidentiality. A pretest of the extraction tool was conducted at a different ART facility. The facilitators and data managers were blinded to the outcome variable while extracting deidentified data. Prior to extraction, common data quality issues—such as duplication, completeness, consistency, and validation—were addressed via smart-care-ART's data quality assurance features [38].

Statistical analysis and machine learning process

Patient data were extracted from the electronic database in Microsoft Excel and converted to comma-separated values (CSVs) for easier manual preprocessing and machine learning. We followed the model development process in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines [39]. Additionally, we employed an interpretable and transparent machine learning algorithm, which enabled thorough checks and balances [40]. Supervised machine learning methods were employed to develop and validate models via Weka 3.8.6 software, the stable version [41].

Handle missing values

To improve data efficiency, we conducted data preprocessing by handling missing values and transforming the dataset before initiating the machine learning process. Missing data were carefully managed to enhance

the performance and reliability of our predictive models. We excluded features with more than 30% missing data. For example, the data on viral load suppression revealed that 30.8% of the values were missing because the viral load test was not applicable for individuals who had been on ART for less than six months. Instances (cases) with noncritical missing values were removed because of their minimal proportion. We also utilized conditional mean/mode imputation, which fills in missing values on the basis of the conditional relationships between the missing feature and other relevant features [42].

Feature selection

First, we performed preliminary feature selection on the basis of the literature and relevant EMR-accessible features to enhance the clinical applicability of the prediction tool [19]. A multivariable logistic regression analysis was conducted to identify predictors associated with LTFU. Then, via the ML process, we check feature correlation via the correlation attribute evaluator technique and rank features on the basis of their correlation with each other or with the target variable. In addition, we apply the information gain (IG) attribute evaluator in Weka, which ranks features on the basis of their information gain with respect to the target class, to select optimal features. These methods enable feature selection techniques to reduce data dimensionality, address multicollinearity, and improve model performance accuracy and interpretability [43].

Imbalanced data handling

Imbalanced datasets present a significant challenge in machine learning, often leading to the misclassification of instances from minority or infrequently occurring classes as belonging to the majority class [44, 45]. To mitigate this issue, we implemented several techniques designed to handle data imbalance, including the 'class balancer' and synthetic minority oversampling technique (SMOTE). The class balancing technique adjusts the weights assigned to different classes during training. By increasing the weight of the minority class and decreasing the weight of the majority class, this method ensures that the classifier focuses more on the minority class. This adjustment helps improve the model's ability to correctly identify instances from underrepresented classes [46]. SMOTE is another effective strategy that generates synthetic samples for the minority class by interpolating between existing samples. This technique enriches the dataset, making it more balanced and enhancing both the accuracy and fairness of the machine learning model during training [47]. By employing these methods, we were able to create a more equitable training environment for our models, ultimately leading to improved performance and reliability.

Model training and validation

We implemented and evaluated the performance of six machine learning algorithmic classifiers: the J48 decision tree, random forest (RF), k-nearest neighbors (k-NN), support vector machine (SVM), logistic regression (LR), and naïve Bayes (NB) classifiers. Model performance was assessed through tenfold cross-validation, and the performance of the models was evaluated via several binary classification metrics, such as accuracy, sensitivity, specificity, precision, F1 score, and area under the receiver operating characteristic (ROC) curve (AUC) [48]. Accuracy measures the overall correctness of the model, whereas sensitivity (recall) indicates its ability to correctly identify true positives. Specificity reflects the model's capacity to identify true negatives accurately. Additionally, we assessed precision, which measures the accuracy of positive predictions, and the F1 score, which provides a balance between precision and recall. To further enhance our evaluation, we utilized the Matthews correlation coefficient (MCC), which considers all classes in the confusion matrix, as well as the AUC to gauge the model's effectiveness in distinguishing between positive and negative classes [48]. To ensure the reliability of our results, we conducted additional experiments and analyses to compare the performance of the algorithms via the corrected resampled t test (P value < 0.05) [41]. Additionally, we conducted decision curve analysis (DCA) to assess the model's clinical utility. In the DCA, the model was evaluated against two contrasting scenarios: "intervention for all" and "intervention for none" [49].

Association rule mining

Finally, to uncover hidden relationships and identify features that frequently appear together, association rules were mined via the Apriori algorithm. This method was employed to explore and compare the most influential features contributing to the model's predictive performance [50]. The algorithm was initialized with a minimum support threshold of 100%, which was systematically reduced in 5% increments. The iterative process continued until at least ten association rules satisfying a minimum confidence level of 0.9 were generated or until the support threshold reached a lower bound of 10%, whichever occurred first [41]. Notably, this approach aimed to enhance the interpretability and transparency of the machine learning model.

Ethics

This study was approved by the Ethical Review Board of the College of Health Sciences (CHS) at Addis Ababa University (AAU), under reference number 061/23/SPH, on September 20, 2023. The ethics committee waived the requirement for individual informed consent, as the study used deidentified secondary data. All the data were

treated with strict confidentiality and used solely for the purposes of this research. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki.

Results

Patient characteristics

In total, 3720 patients who had newly started ART within the past five years were included in this study. Three-fifths of the patients, 2252 (60.5%), were female. The mean age of the patients was 39 years (± 11.2 SD), with 1,384 (37.2%) between the ages of 15 and 34. With respect to the address information obtained from the EMR system, 548 patients (14.7%) were labeled yellow, indicating that at least one required piece of address information, such as a phone number, kebele, or house number, was missing. On the basis of the categorization of patients by their follow-up periods since initiating ART, 1,355 (36.4%) were in the first 12 months of treatment, whereas 2,365 (63.6%) had been in treatment for 13–60 months.

Table 1 Characteristics of adult HIV patients in Ethiopia, 2019–2024 ($n = 3,720$)

Features	Frequency	Percent (%)
Sex		
Male	1468	39.5
Female	2252	60.5
Age (years)		
15–34	1384	37.2
35+	2336	62.8
Address information		
Green	3172	85.3
Yellow	548	14.7
Follow-up Period (months)		
Time_0–12	1355	36.4
Time_13–60	2365	63.6
TPT Status		
Gold	2576	69.2
Bronze/Silver	1144	30.8
DSD Model		
Not enrolled/Other DSD forms	1201	32.3
ASM/3MMD	2519	67.7
Adherence		
Good	2593	69.7
Poor	1127	30.3
Nutritional Status		
Undernourished	1369	36.8
Normal	2351	63.2
WHO Stage		
Stage 1/2	2372	63.8
Stage 3/4	1348	36.2

Abbreviations: ASM = Appointment Spacing Model; DSD = Differential Service Delivery; MMD = Multimonth Dispensing; TPT = TB Prevention Therapy; WHO = World Health Organization; other DSD models include DSDs such as the Advanced Disease Model, Community-based models, Key Population Model, Adolescents and Young People Model, and Maternity and Child Health Model

One-third of the patients, 1144 (30.8%), were labeled 'bronze/silver,' indicating that they had either not started or not completed TB prevention therapy (TPT). With respect to the DSD model, 2519 patients (67.7%) were enrolled in appointment spacing (ASM) or 3-month multimonth dispensing (MMD), whereas 1201 patients (32.3%) were either not enrolled in any model or were enrolled in other DSD forms, such as the AHD model, adolescent and young DSD, or key population DSD. One-third of the patients, 1127 (30.3%), had poor adherence to their medication, 1369 (36.8%) were undernourished, and 1348 (36.2%) were in WHO advanced clinical stages 3 or 4 (Table 1).

Feature selection

Prior to applying ML-based feature selection, the multivariable logistic regression analysis identified several significant factors associated with LTFU: male sex (adjusted odds ratio (AOR) = 1.71; 95% confidence interval (CI): 1.39–2.09), incomplete address information (yellow) (AOR = 2.60, 95% CI: 2.01–3.37), follow-up period of 0–12 months (AOR = 2.14, 95% CI: 1.75–2.61), TPT status (bronze/silver) (AOR = 2.66, 95% CI: 2.17–3.26), DSD model (not enrolled or other forms) (AOR = 7.78, 95% CI: 6.37–9.50), poor adherence (AOR = 5.01, 95% CI: 4.05–6.18), undernutrition (AOR = 1.92, 95% CI: 1.54–2.39), and WHO stage 1/2 (AOR = 1.36, 95% CI: 1.09–1.70). Patient age was significantly associated with the unadjusted analysis but lost significance in the adjusted model (AOR = 1.21, 95% CI: 0.98–1.48, $P = 0.072$) [Supplementary file 1]. On the basis of the results from the correlation attribute evaluator [Supplementary file 2] and the information gain (IG) ranking, we selected six out of nine features to reduce complexity and improve model efficiency for easier application. The selected features were the DSD model, adherence, TPT status, follow-up period, nutritional status, and address information, which provided the most relevant information for predicting LTFU in HIV care (Fig. 1).

Addressing imbalanced data in machine learning

The original imbalanced data comprised a total of 3,720 individuals, with 2,575 (69.2%) classified as not LTFU and a smaller group of 1,145 individuals (30.8%) classified as LTFU. To address this imbalance, a class balancer was utilized to increase the weight of the minority class while decreasing the weight of the majority class, resulting in both classes being adjusted to an equal weight of 1,860 for the machine learning process. Additionally, the application of SMOTE techniques further balanced the classes, yielding 2,575 individuals classified as not LTFU and 2,290 classified as LTFU (Fig. 2).

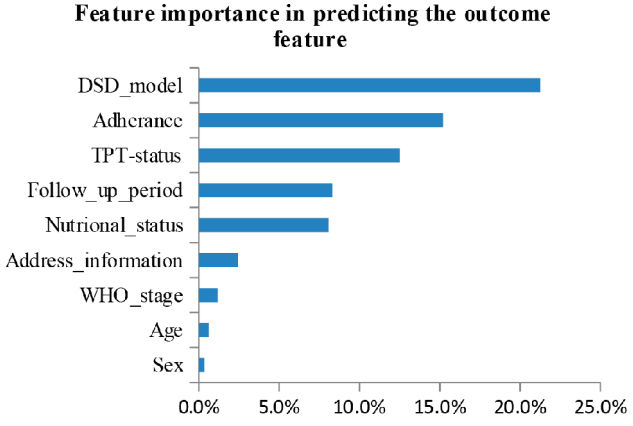

Fig. 1 Information gain (IG) of features for predicting loss to follow-up in HIV care. Abbreviations: DSD = Differentiated Service Delivery, TPT = TB Prevention Therapy, WHO = World Health Organization

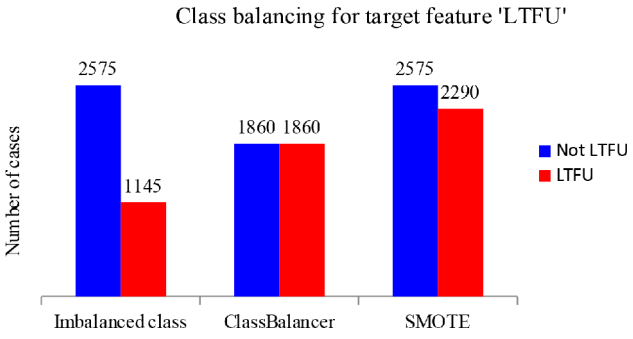

Fig. 2 Class distribution after applying class balancing techniques to the target feature, addressing the original imbalanced data. Abbreviations: LTFU = Loss to Follow-Up, SMOTE = Synthetic Minority Oversampling Technique

Model training and evaluation

We trained the model via six distinct algorithms—RF, J48, K-NN, SVM, LR, and naïve Bayes—and internally validated it via 10-fold cross-validation, keeping all the hyperparameters at their default settings. The performance analysis in Table 2 highlights the impact of class balancing on machine learning algorithms applied to imbalanced data, with notable improvements in sensitivity and slight changes in accuracy and AUC. For example, RF shows a stable accuracy of approximately 84% across all methods (84.8% for imbalanced data, 84.1% with class balancing, and 84.2% with SMOTE). However, the sensitivity increases from 68.3% with imbalanced data to 82.4% with SMOTE, indicating a significant improvement in detecting minority classes. The AUC for RF increases slightly from 89.1% (imbalanced) to 89.5% with SMOTE. Similarly, the sensitivity of J48 increases from 66.8% (imbalanced) to 82.5% with SMOTE, although the accuracy decreases slightly from 85.2 to 83.9%. K-NN experiences a sensitivity increase from 66.8 to 82.4% with SMOTE, whereas its accuracy remains stable at approximately 84%. Although SVM achieves a slight decline in accuracy from 85.0 to 81.9%, it benefits from improved sensitivity, increasing from 66.3 to 80.0% with SMOTE. LR decreases the accuracy from 84.6 to 81.7%, with the sensitivity slightly lower at 78.1% when SMOTE is used; however, the AUC remains constant at 88.5%. Naïve Bayes, while improving the sensitivity from 70.9 to 75.9% with SMOTE, maintains a constant AUC of 88.3% (Table 2). Therefore, SMOTE was selected as the preferred method for balancing because of its ability to enhance minority class detection without compromising overall model performance.

Table 2 Performance of ML algorithms on original imbalanced data vs. data balanced with class balancing and SMOTE

ML algorithm	Comparison metrics	Original imbalanced classes (%)	Class balancer (%)	SMOTE (%)
RF	Accuracy	84.8	84.1	84.2
	Sensitivity	68.3	82.3	82.4
	AUC	89.1	89.0	89.5
J48	Accuracy	85.2	83.7	83.9
	Sensitivity	66.8	82.2	82.5
	AUC	87.2	87.8	88.0
K-NN	Accuracy	84.9	84.0	84.2
	Sensitivity	66.8	82.3	82.4
	AUC	89.1	89.1	89.5
SVM	Accuracy	85.0	81.8	81.9
	Sensitivity	66.3	80.1	80.0
	AUC	79.8	81.8	81.8
LR	Accuracy	84.6	81.1	81.7
	Sensitivity	68.2	80.3	78.1
	AUC	88.6	88.5	88.5
Naïve Bayes	Accuracy	83.9	81.2	81.8
	Sensitivity	70.9	79.3	75.9
	AUC	88.3	88.3	88.3

Abbreviations: AUC - area under the curve, J48 is a decision tree algorithm based on the C4.5 algorithm (J48 is its implementation in Weka). K-NN = k-nearest neighbors, LR = logistic regression, ML = machine learning, RF = random forest, SMOTE = synthetic minority oversampling technique, SVM = support vector machine

Furthermore, we conducted robust experiments comparing six algorithms via the corrected paired t test with a p value < 0.05. Both RF and KNN outperform the other algorithms, achieving the highest values across the most important metrics, establishing it as the best algorithm for predicting LTFU in HIV care (Fig. 3) [Supplementary file 3]. On the basis of further considerations, we chose RF for its practical advantages over KNN, such as better handling of large datasets, noise resilience, automatic feature importance, and scalability [51].

Random forest algorithm

A random forest model using 10-fold cross-validation took 0.26 s to build, employing bagging with 100 iterations and a base learner [Supplementary file 4]. It demonstrated an accuracy of 84.2%, a sensitivity of 82.4%,

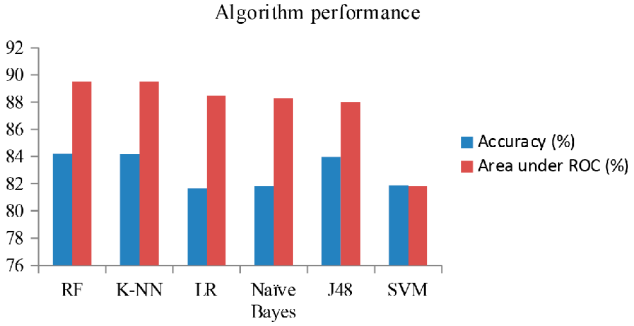

Fig. 3 Comparison of different ML algorithms after applying SMOTE to balance the data. Abbreviations: J48 is a decision tree algorithm based on the C4.5 algorithm. K-NN = k-nearest neighbors, LR = logistic regression, RF = random forest, ROC = receiver operating characteristic curve, SVM = support vector machine

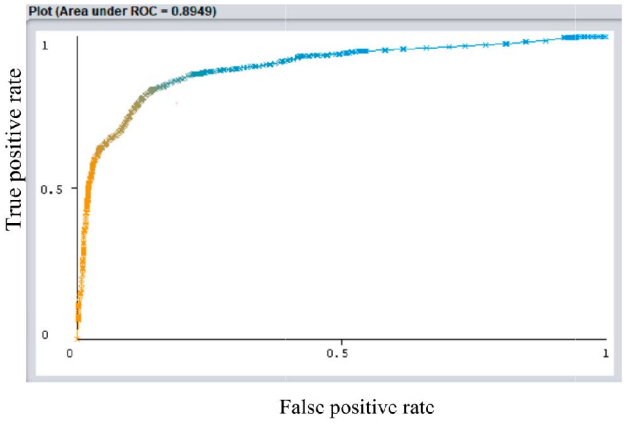

Fig. 4 ROC curve and AUC of the random forest algorithm for predicting LTFU in HIV Care, Ethiopia

a specificity of 85.7%, a precision of 83.7%, an F1 score of 83.1%, an MCC of 68.3%, and an area under the PRC (precision-recall curve) of 88.7% (Table 3). Figure 4 shows the AUC for the random forest classifier, indicating a model performance of 89.5% in distinguishing between the true positive rate (sensitivity) and the false positive rate (1-specificity) across all thresholds (Fig. 4).

Clinical utility of the model

We conducted a decision curve analysis to determine the clinical effectiveness of a prediction model aimed at assessing the risk of LTFU in HIV care. In Fig. 3, decision curve analysis (DCA) illustrates the optimal thresholds at which our model achieves the best balance between benefit and harm, enhancing our understanding of when to intervene in cases of LTFU. Notably, at thresholds of 10% and above, the model demonstrates a greater net benefit than both the "intervention for all" and "intervention for none" strategies do (Fig. 5).

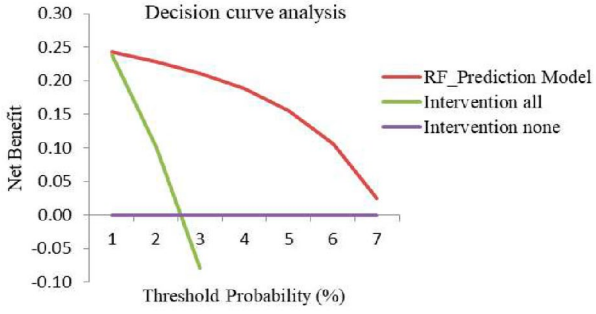

Fig. 5 Decision curve analysis (DCA) assessing the clinical utility of the model for predicting LTFU in HIV care, Ethiopia

Association rule results

Association rules were mined via the Apriori algorithm to find relationships or patterns between features in a dataset and compare the most significant features. A total of ten association rules were identified, each with a confidence level exceeding 90% and a minimum support of 0.2. The rules indicate that TPT status, the DSD model, adherence, and the follow-up period are strongly associated with LTFU. For example, in Rule 1, if the TPT status is *bronze/silver* and the DSD model is *not enrolled/other DSD forms*, then the class (LTFU) is likely to be '*Yes*' with 93% confidence and a strong association (lift = 1.98). Rule 2: When the DSD model is *not enrolled/other DSD forms* and *adherence is poor*, the class is predicted to be '*Yes*' with 92% confidence and a strong association (lift = 1.96). Rule 6: When the DSD model is *not enrolled/other DSD forms* and the follow-up period is between *0 and 12 months*, the class is predicted to be '*Yes*' with 90% confidence and a strong association (lift = 1.92) [Supplementary file 5].

Table 3 Performance of the random forest algorithm in predicting LTFU in HIV care, Ethiopia

ML algorithm	Class	Sensitivity (%)	Specificity (%)	Precision (%)	F1 score (%)	MCC (%)	Area under ROC (%)	Area under PRC (%)
RF	Not LTFU	85.7	82.4	84.6	85.2	68.3	89.5	87.5
	LTFU	82.4	85.7	83.7	83.1	68.3	89.5	88.7
	Weighted average	84.2	84.0	84.2	84.2	68.3	89.5	88.0

Abbreviations: LTFU = loss to follow-up, MCC = Matthews correlation coefficient, ML = machine learning, RF = random forest. ROC = receiver operating characteristic curve, PRC = precision-recall curve

Discussion

In this study, a prediction model was developed to estimate the five-year risk of LTFU in HIV care after ART initiation via machine learning algorithms trained on routine electronic medical records. The dataset used for model development revealed a 30.8% prevalence of LTFU in HIV care, which is consistent with findings from other low-resource settings where ML-based prediction models were developed for patient disengagement, such as a 27% LTFU rate reported in Nigeria and 23% in Mozambique [52]. Similarly, a study conducted in Ethiopia developed a prediction model using data with a 25.7% prevalence of LTFU [20]. However, the prevalence of LTFU in the current study was higher than that reported from South Africa, where a prediction model was developed using data with a 10.5% prevalence of LTFU in HIV care [22]. The higher LTFU incidence in our study might be attributed to the inclusion of patients who tested HIV positive and newly began ART within the past five years, a period during which LTFU rates tend to be higher. Additionally, the data were collected from urban settings and high-caseload facilities in central Ethiopia, including the capital, Addis Ababa, where patient LTFU may be more prevalent.

In this study, the machine learning–based feature selection process identified six locally relevant and operationally defined predictors of LTFU: the differentiated service delivery (DSD) model, adherence level, tuberculosis preventive therapy (TPT) status, follow-up period, nutritional status, and address information. These predictors were consistent with findings from other studies that predict LTFU in HIV care. For example, in a previous similar study conducted in Ethiopia, factors such as the appointment spacing model (ASM) for DSD, TPT status, adherence level, and nutritional status were used to develop a prediction model for LTFU [20]. In a similar study conducted in South Africa, the duration of follow-up on ART [22, 53] was used, whereas in Tanzania, body weight and WHO clinical stage were utilized to predict the risk of disengagement from HIV care [21]. Patient address information was also a factor for follow-up efforts and interventions to re-engage patients who may have become LTFUs. Having or lacking complete and detailed information, including a phone number and a precise kebele address, makes it a valuable feature in predictive modeling [23, 24]. Other predictors previously reported in the literature, such as age, sex, and WHO stage, did not emerge as significant features in our model selection process. However, these factors were important predictors in models developed during earlier times, suggesting that when historical information is lacking, they can still serve as valuable indicators of LTFU in HIV care.

In this study, among the six ML algorithms tested—RF, J48, K-NN, SVM, LR, and naïve Bayes—RF outperformed the other algorithms in predicting the risk of LTFU in HIV care. It was selected for its high accuracy (84.2%), sensitivity (82.4%), and AUC (89.5%). These findings were comparable with findings from several other studies that have employed machine learning techniques to address similar challenges in HIV care settings. A study from South Africa revealed that RF models were among the top performers for predicting patient retention, achieving predictive power (AUC = 0.69) [22], although this value was slightly lower than the findings of the current study. In Nigeria, the 'Data—FI' initiative uses machine learning to predict LTFU among ART clients, achieving over 70% accuracy and demonstrating the potential of RF models in HIV care settings [32]. A study conducted in Mozambique applied various machine learning algorithms and selected a random forest, which achieved an AUC of 0.65, to predict LTFU among ART clients, even though its performance was lower than that of the current study [52]. The predictive performance of the current study exceeded that of a similar study conducted in Tanzania, which utilized machine learning with routine EMR indicators and achieved an accuracy of 75.2% and a sensitivity of 54.7% [21]. Similarly, the current model outperformed a previous study conducted in Ethiopia, which achieved an AUROC of 85.9%, a maximum sensitivity of 72.07%, and a specificity of 83.49% [20]. The improved performance in the current study may be attributed to several factors, including a larger sample size, the incorporation of diverse potential predictors, and the use of robust machine learning algorithms such as random forest [51]. Furthermore, the model underwent rigorous internal evaluation via a 10-fold cross-validation approach, which helps mitigate overfitting and provides a more generalized estimate of its predictive power [54].

We evaluated the clinical utility of the model through decision curve analysis (DCA). At thresholds of 10% or higher, the model demonstrated a greater net benefit than did strategies that either intervene with all patients or none. The DCA in the current study was also consistent with broader research trends advocating for machine learning's role in enhancing clinical decision-making and patient management across various healthcare domains. For example, a study on machine learning algorithms, including random forests, showed that DCA effectively assessed model performance in predicting surgical outcomes, highlighting the clinical value of predictive models for decision-making in healthcare [55]. Often, a clinically relevant range (e.g., 5–30%) of thresholds ensures that the analysis aligns with practical decision-making contexts and reflects patient preferences and clinical guidelines [56, 57]. A study conducted elsewhere reported that multidomain prediction models outperformed single-domain models in terms of net benefit when DCA was used, particularly at treatment threshold

probabilities above 10% [58]. This aligns with the current findings, as both studies emphasize the importance of tailored interventions on the basis of predictive analytics rather than a one-size-fits-all approach. Thus, the DCA in this study underscores the importance of assessing predictive models for targeted interventions.

One limitation of this study is the categorization of continuous variables, such as age and follow-up period. This approach can result in a loss of information that may be critical for understanding the relationships within the data. Additionally, we combined several subcategories within the predictor feature "differentiated service delivery (DSD) model." While these subcategories may appear insignificant on their own, their merging into broader categories could overlook important nuances. Finally, the developed model was validated solely with internal data, which restricts its external validity, which may affect the model's generalizability to new observations and diverse populations, potentially reducing its applicability in broader contexts.

Conclusions

In this study, a machine learning prediction model was developed to assess the future risk of LTFU within five years of initiating antiretroviral therapy in a low-resource setting. A model was built using six predictors of LTFU: the DSD model, adherence, TPT status, follow-up period, nutritional status, and address information. Notably, the model built via the random forest algorithm demonstrated high accuracy and strong discriminative performance, highlighting its potential clinical utility through a positive net benefit. Future research should focus on external validation across diverse populations to ensure its generalizability and effectiveness.

Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ASM	Appointment Spacing Model
AUC	Area Under the Receiver Operating Characteristic Curve
CSV	Comma Separated Values
DCA	Decision Curve Analysis
DSD	Differentiated Service Delivery
EMR	Electronic Medical Records
HIV	Human Immunodeficiency Virus
J48	A decision tree algorithm based on the C4.5 algorithm
K-NN	K-Nearest Neighbors
LR	Logistic Regression
LTFU	Loss to Follow Up
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MMD	Multimonth Dispensing
PRC	Precision-Recall Curve
RF	Random Forest
ROC	Receiver Operating Characteristic Curve
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
TPT	TB Prevention Therapy
UNAIDS	Joint United Nations Programme on HIV/AIDS
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-025-03030-7.

Supplementary Material 1

Acknowledgements

We sincerely thank the team leaders of the ART care and treatment units from all the health facilities where data was collected, especially Mrs. Assiya Jeylan Hussen, the ART team leader at Adama Hospital Medical College.

Author contributions

TE: conceptualizing, designing, facilitating data collection, analyzing data, interpreting the results, and drafting of the manuscript; WD: designing and supervising the overall research process, including the data collection process, data analysis, result interpretation, and critical revision of the manuscript; GT: designing, supervising the data collection process, data analysis, result interpretation, and critical revision of the manuscript; All authors approved the final manuscript, including the authorship list.

Funding

This work was made possible by the financial support of the Doris Duke Charitable Foundation (DDCF) under grant number 2017187. The DDCF's mission is to enhance people's quality of life by funding performing arts, environmental conservation, medical research, and child well-being, as well as preserving Doris Duke's cultural and environmental legacy. The funder was not involved in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Data availability

The data are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethical Review Committee of the College of Health Sciences, Addis Ababa University (Approval No. 061/23/SPH, September 20, 2023). The ethics committee waived the requirement for individual informed consent, as the study used deidentified secondary data. All the data were treated with strict confidentiality and used solely for the purposes of this research. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Epidemiology and Biostatistics, School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

Received: 14 October 2024 / Accepted: 12 May 2025
Published online: 19 May 2025

References

1. UNAIDS epidemiological estimates. 2023. [Internet]. Available from: https://www.unaids.org/en/resources/documents/2024/global-aids-update-2024
2. UNAIDS. Political declaration on HIV and AIDS: Ending inequalities and getting on track to end AIDS by 2030. [Internet]. Available from: https://www.unaids.org/en/resources/documents/2021/2021_political-declaration-on-hiv-and-aids
3. PEPFAR Ethiopia (PEPFAR-E). Ethiopia Country Operational Plan COP2020/FY2021 Strategic Direction Summary March 23, 2020.

4. Standard Operating. Procedures (SOP) for comprehensive HIV/AIDS prevention, treatment, care and support services. Oromia National Regional State Health Bureau; Sept 2018.
5. Mberi MN, Kuonza LR, Dube NM, Nattey C, Manda S, Summers R. Deter_ minants of loss to follow-up in patients on antiretroviral treatment, South Africa, 2004–2012: a cohort study. BMC Health Serv Res. 2015;15(1):1–11.
6. Makunde WH, Francis F, Mmbando BP, Kamugisha ML, Rutta AM, Mandara CI, Msangeni HA. Lost to follow up and clinical outcomes of HIV adult patients on antiretroviral therapy in care and treatment centers in Tanga City, Northeastern Tanzania. Tanzan J Health Res. 2012;14(4):250–6.
7. Wekesa P, McLigeyo A, Owuor K, Mwangi J, Nganga E, Masamaro K. Factors associated with 36-month loss to follow-up and mortality outcomes among HIV-infected adults on antiretroviral therapy in central Kenya. BMC Public Health. 2020;20(1):328.
8. Abebe Moges N, Olubukola A, Micheal O, Berhane Y. HIV patients retention and attrition in care and their determinants in Ethiopia: a systematic review and meta-analysis. BMC Infect Dis. 2020;20(1):439.
9. Mugavero MJ, Westfall AO, Cole SR, Geng EH, Crane HM, Kitahata MM, et al. Beyond core indicators of retention in HIV care: missed clinic visits are independently associated with all-cause mortality. Clini Infect Dis. 2014;59(10):1471-9.
10. Zürcher K, Mooser A, Anderegg N, Tymejczyk O, Couvillon MJ, Nash D, et al. Outcomes of HIV-positive patients lost to follow-up in African treatment programmes. Trop Med Int Health TM IH. 2017;22(4):375–87.
11. Verguet S, Lim SS, Murray CJL, Gakidou E, Salomon JA. Incorporating loss to follow-up in estimates of survival among HIV-infected individuals in sub-Saharan Africa enrolled in antiretroviral therapy programs. J Infect Dis. 2013;207(1):72–9.
12. Flores D, Leblanc N, Barroso J. Enroling and retaining human immunodeficiency virus (HIV) patients in their care: A metasynthesis of qualitative studies. Int J Nurs Stud. 2016;62:126–36.
13. Hendricks L, Eshun-Wilson I, Rohwer A. A mega-aggregation framework synthesis of the barriers and facilitators to linkage, adherence to ART and retention in care among people living with HIV. Syst Reviews. 2021;10(1). Article 54. https://doi.org/10.1186/s13643-021-01582-z.
14. Villanueva M, Miceli J, Speers S, Nichols L, Carroll C, Jenkins H, et al. Advancing data to care strategies for persons with HIV using an innovative reconciliation process. PLoS One. 2022;17(5):e0267903.
15. Budhwani H, Kiszla BM, Hightow-Weidman LB. Adapting digital health interventions for the evolving HIV landscape: Examples to support prevention and treatment research. Curr Opin HIV AIDS. 2022;17(2):112–8.
16. Schwartz JT, Gao M, Geng EA, Mody KS, Mikhail CM, Cho SK. Applications of machine learning using electronic medical records in spine surgery. Neurospine. 2019;16(4):643–53.
17. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, et al. Machine learning applications to clinical decision support in neurosurgery: An artificial intelligence augmented systematic review. Neurosurg Rev. 2020;43(5):1235–53.
18. Gezae KE, Abebe HT, Gebretsadik LG. Incidence and predictors of LTFU among adults with TB/HIV coinfection in two governmental hospitals, Mekelle, Ethiopia, 2009–2016: survival model approach. BMC Infect Dis. 2019;19(1):107.
19. Kebede HK, Mwanri L, Ward P, Gesesew HA. Predictors of lost to follow up from antiretroviral therapy among adults in sub-Saharan Africa: A systematic review and meta-analysis. Infect Dis Poverty. 2021;10(1):33.
20. Fentie DT, Kassa GM, Tiruneh SA, Muche AA. Development and validation of a risk prediction model for lost to follow-up among adults on active antiretroviral therapy in Ethiopia: A retrospective follow-up study. BMC Infect Dis. 2022;22(1):727.
21. Fahey CA, Wei L, Njau PF, Shabani S, Kwilasa S, Maokola W, et al. Machine learning with routine electronic medical record data to identify people at high risk of disengagement from HIV care in Tanzania. PLOS Glob Public Health. 2022;2(9):e0000720.
22. Maskew M, Sharpey-Schafer K, De Voux L, Crompton T, Bor J, Rennick M, et al. Applying machine learning and predictive modeling to retention and viral suppression in South African HIV treatment cohorts. Sci Rep. 2022;12(1):12715.
23. Opio D, Semitala FC, Kakeeto A, Sendaula E, Okimat P, Nakafeero B, et al. Loss to follow-up and associated factors among adult people living with HIV at public health facilities in Wakiso district, Uganda: a retrospective cohort study. BMC Health Serv Res. 2019;19(1):628.
24. Kiwanuka J, Mukulu Waila J, Muhindo Kahungu M, Kitonsa J, Kiwanuka N. Determinants of loss to follow-up among HIV positive patients receiving antiretroviral therapy in a test and treat setting: A retrospective cohort study in Masaka, Uganda. PLoS One. 2020;15(4):e0217606.
25. Telayneh AT, Tesfa M, Woyraw W, Temesgen H, Alamirew NM, Haile D, et al. Time to lost to follow-up and its predictors among adult patients receiving antiretroviral therapy retrospective follow-up study Amhara Northwest Ethiopia. Sci Rep. 2022;12(1):2916.
26. Mussini C, Lorenzini P, Cozzi-Lepri A, Mammone A, Guaraldi G, Marchetti G, et al. Determinants of loss to care and risk of clinical progression in PLWH who are re-engaged in care after a temporary loss. Sci Rep. 2021;11(1):9632.
27. Pence BW, Bengtson AM, Boswell S, Christopoulos KA, Crane HM, Geng E, et al. Who will show? Predicting missed visits among patients in routine HIV primary care in the united States. AIDS Behav. 2019;23(2):418–26.
28. Tweya H, Oboho IK, Gugsa ST, Phiri S, Rambiki E, Banda R et al. Loss to follow-up before and after initiation of antiretroviral therapy in HIV facilities in Lilongwe, Malawi. Beck EJ, editor. PLOS One. 2018;13(1):e0188488.
29. Pettit AC, Bian A, Schember CO, Rebeiro PF, Keruly JC, Mayer KH, et al. Development and validation of a multivariable prediction model for missed HIV health care provider visits in a large US clinical cohort. Open Forum Infect Dis. 2021;8(7):ofab130.
30. Ramachandran A, Kumar A, Koenig H, De Unanue A, Sung C, Walsh J, et al. Predictive analytics for retention in care in an urban HIV clinic. Sci Rep. 2020;10(1):6421.
31. Woodward B, Person A, Rebeiro P, Kheshti A, Raffanti S, Pettit A. Risk prediction tool for medical appointment attendance among HIV-Infected persons with unsuppressed viremia. AIDS Patient Care STDs. 2015;29(5):240–7.
32. Palladium. Predicting Loss-to-Follow-Up among HIV/AIDS clients in Nigeria: Report on the retrospective application of machine learning. Washington, DC, USA: Data.FI; 2021. https://pdf.usaid.gov/pdf_docs/PA00X7ZG.pdf.
33. Ethiopia Demographic and Health Survey 2016.
34. MOH, Ethiopia. National_guidelines_for_comprehensive_hiv_prevention_care_and_treatment_-_February_2022_pocket_guide.pdf. https://hivpreventioncoalition.unaids.org/sites/default/files/attachments/national_guidelines_for_comprehensive_hiv_prevention_care_and_treatment_-_February_2022_pocket_guide.pdf
35. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ [Internet]. 2020 Mar 18 [cited 2021 Nov 11];m441. Available from: https://www.bmj.com/lookup/doi/https://doi.org/10.1136/bmj.m441
36. Addis Ababa Health Bureau. Department of HIV/AIDS prevention and Control program unit report of 2021.
37. FMOH. National Consolidated Guidelines for Comprehensive HIV Prevention, Care and Treatment. 2018. https://www.afro.who.int/publications/national-consolidated-guidelines-comprehensive-hiv-prevention-care-and-treatment
38. Ethiopian FMOH. SmartCare-ART Module Participant Manual Ver1.1; May 2019.
39. Patzer RE, Kaji AH, Fong Y. TRIPOD reporting guidelines for diagnostic and prognostic studies. JAMA Surg [Internet]. 2021 Jul 1 [cited 2022 Dec 14];156(7):675–6. Available from: https://doi.org/10.1001/jamasurg.2021.0537
40. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. Nat Med [Internet]. 2022;28(10):1996–9. Available from: https://doi.org/10.1038/s41591-022-01961-6
41. Eibe Frank MA, Hall. and Ian H. Witten. The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Fourth Edition, 2016.
42. Raheem E. Missing Data Imputation: A Practical Guide. In: Mitra AK, editor. Statistical Approaches for Epidemiology: From Concept to Application [Internet]. Cham: Springer International Publishing; 2024. pp. 293–316. Available from: https://doi.org/10.1007/978-3-031-41784-9_18
43. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. Neurocomputing. 2018;300:70–9.
44. Michelucci U. Unbalanced Datasets and Machine Learning Metrics. In: Michelucci U, editor. Fundamental Mathematical Concepts for Machine Learning in Science [Internet]. Cham: Springer International Publishing; 2024. pp. 185–212. Available from: https://doi.org/10.1007/978-3-031-56431-4_8
45. Chen W, Yang K, Yu Z, Shi Y, Chen CLP. A survey on imbalanced learning: Latest research, applications and future directions. Artif Intell Rev. 2024;57(6):137.
46. Abhishek K, Abdelaziz DM. Machine Learning for Imbalanced Data: Tackle imbalanced datasets using machine learning and deep learning techniques. 2023.

47. Chawla N, Bowyer K, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Oversampling Technique. ArXiv [Internet]. 2002;abs/1106.1813. Available from: https://api.semanticscholar.org/CorpusID:1554582
48. Berrar D. Performance Measures for Binary Classification. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C, editors. Encyclopedia of Bioinformatics and Computational Biology [Internet]. Oxford: Academic Press; 2019. pp. 546–60. Available from: https://www.sciencedirect.com/science/article/pii/B9780128096338203518
49. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol. 2018;74(6):796–804.
50. Nisbet R, Miner G, Yale K. Chapter 7 - Basic Algorithms for Data Mining: A Brief Overview. In: Nisbet R, Miner G, Yale K, editors. Handbook of Statistical Analysis and Data Mining Applications (Second Edition) [Internet]. Boston: Academic Press; 2018. pp. 121–47. Available from: https://www.sciencedirect.com/science/article/pii/B9780124166325000074
51. Schonlau M, Zou RY. The random forest algorithm for statistical learning. Stata J. 2020;20(1):3–29.
52. Stockman J, Friedman J, Sundberg J, Harris E, Bailey L. Predictive analytics using machine learning to identify ART clients at health system level at greatest risk of treatment interruption in Mozambique and Nigeria. JAIDS J Acquir Immune Defic Syndr [Internet]. 2022;90(2). Available from: https://journals.lww.com/jaids/fulltext/2022/06010/predictive_analytics_using_machine_learning_to.6.aspx
53. Esra RT, Carstens J, Estill J, Stoch R, Le Roux S, Mabuto T, et al. Historical visit attendance as predictor of treatment interruption in South African HIV patients: Extension of a validated machine learning model. PLOS Glob Public Health. 2023;3(7):e0002105.
54. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nat Methods. 2016;13(9):703–4.
55. Ul Banna H, Zanabli A, McMillan B, Lehmann M, Gupta S, Gerbo M, et al. Evaluation of machine learning algorithms for trabeculectomy outcome prediction in patients with glaucoma. Sci Rep. 2022;12.
56. Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X et al. Decision curve analysis: a technical note. Ann Transl Med Vol 6 No 15 August 08 2018 Ann Transl Med Focus Cardiovasc Dis Cardiometabolic Risk Adv Underst Pathophysiol Public Health Burd Clin Care [Internet]. 2018 [cited 2018 Jan 1]; Available from: https://atm.amegroups.org/article/view/20389
57. Andrew Vickers. Statistical Thinking - Seven Common Errors in Decision Curve Analysis [Internet]. 2023 [cited 2024 Sep 29]. Available from: https://www.fharrell.com/post/edca/
58. Huber M, Schober P, Petersen S, Luedi MM. Decision curve analysis confirms higher clinical utility of multidomain versus single-domain prediction models in patients with open abdomen treatment for peritonitis. BMC Med Inf Decis Mak. 2023;23(1):63.

Publisher's note