Open Access Full Text Article

ORIGINAL RESEARCH

Artificial Intelligence and Machine Learning Based Prediction of Viral Load and CD4 Status of People Living with HIV (PLWH) on Anti-Retroviral Treatment in Gedeo Zone Public Hospitals

Binyam Tariku Seboka [ID], Delelegn Emwodew Yehualashet [ID], Getanew Aschalew Tesfa [ID]

School of Public Health, Dilla University, Dilla, Ethiopia

Correspondence: Binyam Tariku Seboka, School of public health, Dilla University, P.O Box: 419, Dilla University, Dilla, Ethiopia, Tel +251 920612180, Fax +251 46-331-2568, Email bini555tar@gmail.com

Background: Despite the success made in scaling up HIV treatment activities, there remains a tremendous unmet demand for the monitoring of the disease progression and treatment success, which threatens HIV/AIDS treatment and control. This research presented the assessments of viral load and CD4 classification of adults enrolled in ART care using machine learning algorithms.

Methods: We trained, validated, and tested eight machine learning (ML) classifier algorithms with historical data, including demographics, clinical, and laboratory data. Data were extracted from the ART registry database of Yirgacheffe Primary Hospital and Dilla University Referral Hospital. ML classifiers were trained to predict virological failure (viral load >1000 copies/mL) and poor CD4 (CD4 cell count <200 cells/mL). The model predictive performances were evaluated using accuracy, sensitivity, specificity, precision, f1-score, F-beta scores, and AUC.

Results: The mean age of the sample participants was 41.6 years (SD = 10.9). The experimental results showed that XGB classifier ranked as the best algorithm for viral load prediction in terms of sensitivity (97%), f1-score (96%), AUC (0.99), accuracy (96%), followed by RF. The GB classifier exhibited a better predictive capability in predicting participants with a CD4 cell count <200 cells/mL.

Conclusion: In this study, the XGB and RF models had the highest accuracy and outperformed on various evaluation metrics among the models examined for viral load classification. In the prediction of participants CD4, GB model had the highest accuracy.

Keywords: artificial intelligence, AI, machine learning, ML, anti-retroviral treatment, ART, viral load, CD4 count, HIV

Introduction

Infection with the Human Immunodeficiency Virus Type 1 (HIV-1) continues to pose a hazard to public health globally,[1–3] it is associated with weakening the host's immune system leading to acquired immune deficiency syndrome (AIDS) and high mortality if no therapy is provided. All HIV-positive individuals should receive antiretroviral therapy (ART), according to the World Health Organization (WHO). Thus, this emphasis on HIV medical care helped the 'treatment as a prevention strategy for putting an end to the HIV epidemic by emphasizing the necessity for thorough diagnosis and treatment assistance. As of 2021, 38.4 million people globally were living with HIV/AIDS and roughly 28.7 million people were enrolled in ART.[4] The vast majority of people with HIV were from low- and middle-income countries. Approximately, 53% of people with HIV were from eastern and southern Africa in 2021. In Ethiopia, more than 0.61 million people living with HIV/AIDS, of these 78% to 82% of people with HIV were accessing ART in 2021.[5,6]

There remains a tremendous unmet demand for effective and efficient monitoring and treatment of HIV/AIDS, despite the success made in scaling up HIV treatment activities. Evidence suggests that tests of viral load (a measurement of HIV nucleic acid concentration) and CD4 lymphocyte counts are important tools for effective monitoring of disease

435

progress and treatment response.[7–13] The CD4 cell count is one of the most important indications of immunological function, with lower counts indicating a weaker immune system.[14–16] HIV is a retrovirus that typically infects CD4+ T cells, causing their numbers to diminish with time. Because the number of CD4 cells declines steadily in untreated infected individuals, the CD4 cell count has become an essential indicator for treatment plan selection and ART effectiveness monitoring.[14,16,17] Furthermore, the quantity of CD4+ T cells is a significant signal for judging disease development and determining the patient's prognosis.

However, the long testing interval and the need for advanced laboratory investigations present massive challenges for the effective comparison of treatment outcomes.[7,9–11] As a result, it is extremely difficult for patients and healthcare professionals to predict how the disease will develop or change after starting ART. Furthermore, understanding the conditionally dependent clinical features that drive these outcomes is a major challenge.[10,18–20] The use of predictive modeling to identify patients who are most at risk of treatment failure shows promise. Machine learning (ML) and Artificial Intelligence (AI) algorithms are important in healthcare because they analyze medical data and offer some striking potential in generating useful evidence that can play a significant role in healthcare decision-making.[21,22]

ML algorithms can learn from the huge amount of structured datasets by using a combination of statistical, probabilistic, and optimization techniques.[23–25] Accordingly, ML models have widespread applications in medical research;[18,23,26–32] almost any data type can be used to build predictive models. In particular, ML algorithms had a remarkable potential for creating prediction models that can be very useful in ART treatment.[11,13,30,33–60] ML algorithms employed on data from routine clinical care of HIV patients such as Logistic Regression (LR), Gaussian Naive Bayes (GNB), Decision Trees (DT), K-nearest Neighbor (KNN), eXtreme Gradient Boosting (XGB), Random Forest (RF), and others were effective to model viral load and CD4-related outcomes in ART.[58–62] These make it simple and affordable to identify patients who are at a high risk of treatment failure, and they can be used as a springboard to launch prompt, effective interventions. Several related studies have been conducted to predict viral load and CD4 using ML and other techniques globally. However, there are scant studies in the literature regarding the effective use of ML technologies to support the treatment and management of ART programs, especially in resource-limited settings like Ethiopia.

Therefore, we have employed five standard ML algorithms KNN, DT, GNB, Support Vector Machine (SVM), LR, and three ensemble algorithms Gradient Boosting (GB), XGB, and RF to predict the viral load and CD4 status of adults living with HIV/AIDS and enrolled in ART care in Ethiopia.

Methods and Materials

In this study, we trained, validated, and tested eight ML classifier models using retrospectively collected records of adults living with HIV/AIDS and enrolled in free highly active antiretroviral therapy (HAART) service in the Yirgacheffe Primary Hospital and Dilla University Referral Hospital. The experiment was divided into six steps, which involve data collection or extraction, data pre-processing, data partitioning, data balancing, model building, and model assessment. The study protocol was approved by the Institutional Ethical Review Board of Dilla University, Ethiopia (Reference No: duirb/004/21-10) and was carried out following the Declaration of Helsinki.

Data Acquisition

We extracted the de-identified data from 2907 adults who were 15 years of age or older, using an ART registry database of individuals receiving antiretroviral therapy at the Yirgacheffe Primary Hospital and Dilla University Referral Hospital.

Outcome Measurements

The study's outcomes were evaluated using the Ethiopian general ART framework and the WHO standard for measuring viral load and CD4. A viral load greater than 1000 copies/mL was characterized as virological failure based on two consecutive viral load tests within three months. Individuals' viral loads were thus classed as either: "(0) suppressed (if the viral load was less than 1000 copies/mL), or (1) unsuppressed (if the viral load was greater than or equal to 1000 copies/mL)".[63,64] Similarly, individuals were divided into two groups based on CD4: a CD4 cell concentration greater than or equal to 200 was considered good; or a CD4 cell count <200 cells/mL was categorized as poor or low.[64]

Data Analyses

Python 3 and with libraries such as sklearn library[65] and others were employed to process the data analysis and the experiments. In our study, 2 non-overlapping similar datasets with 2, 907 records of adults with 37 features were created for viral load and CD4 model building. The list of features extracted in the present study along with the targets (outcomes) of the study are provided in Supplementary Table 1. In the respective datasets, we deleted records with missing CD4 or viral load count; hence, 19% and 2.33% of records with missing values of the outcome data (viral load or CD4) were excluded from further analysis from the respective datasets. The final sample size was comprised of 2349 and 2839 individuals in the viral load and CD4 datasets, respectively.

Pre-Processing

The data set was comprised of numeric, binary, and categorical variables. Due to this, the algorithms that would be implemented in ML model building only accept numerical values, and data were pre-processed into the appropriate numerical format and values.[25,66] The data were preprocessed as follows. First were excluded, features which were the indexed value of the unique identity like that of medical record number, blank (null), and features with more than 80% of missing data like those that of pregnancy status where pregnant individuals were less than 0.02%. In the final dataset, all 26 features were similar for both outcomes. Furthermore, one hot encoder and ordinal encoder were employed, this allowed the categorical dataset features to be encoded and turned into numerical values. In the respective datasets, missingness analyses were performed by visual inspection using the VIM (Visualization and Imputation of Missing Values) and the missing values in the data were imputed using the k-Nearest Neighbors (KNN) approach using Euclidean distance metric. In particular, each missing feature was imputed using the average feature value from k = 10 nearest neighbors. As shown in Table 1, features with the higher percentage of missing data were marital status (5.14%), adherence level (3.22%), TB screening status 1), ARV dispensed dose (2.06%), viral load count (0.34%), and WHO stage (0.13%). Further outlier and noise of the dataset were checked, and continuous features were standardized and normalized in the respective datasets.[24,25] The distribution plots are depicted in Supplementary Figure 1.

Feature Selection

In the feature selection process, we used various feature selection techniques to find the relevant predictive set of features and then take the intersection of the features generated by the different techniques. Accordingly, we employed Pearson correlation, Linear SVM, Lasso, chi-2, Recursive Feature Elimination (RFE) and Logistic Regression, Recursive Feature

Table 1 Characteristics of the Study Participants

Features	Viral Load n=2349 Participants			CD4 Count n=2839 Participants		
	N (%)	Unsuppressed	Missing %	N (%)	< 200 Cells/mL	Missing %
Sex						
Male	838 (35.67)	45 (5.37)	0	994(35.01)	367 (36.92)	0
Female	1511(64.33)	110 (7.28)		1845(64.99)	801 (43.41)	
	X^2= 3.19 P-value 0.007			X^2= 11.246 P-value =0.001		
Marital status						
Single	820(35.91)	46 (5.61)		1014(35.01)	352(34.71)	
Married	1348(57.39)	102(7.57)	0	1610(56.71)	727(45.16)	5.14
Divorced/ Widowed	181(7.71)	7 (3.87)		215(7.57)	89(41.40)	
	X^2= 5.541 P-value = 0.063			X^2= 28.016 P-value =0.000		

(*Continued*)

Table 1 (Continued).

Features	Viral Load n=2349 Participants			CD4 Count n=2839 Participants		
	N (%)	**Unsuppressed**	**Missing %**	**N (%)**	**< 200 Cells/mL**	**Missing %**
Functional status						
Functional	2169(92.34)	2021(93.18)	0	2625(92.46)	1080(41.14)	0
Non functional	180(7.66)	173(96.11)		214(7.54)	88(41.12)	
	X^2= 28.016 P-value = 0.000			X^2= 0.009 P-value =0.005		
Educational status						
Cannot read and write	214(9.11)	18(8.41)		259(9.12)	82(31.66)	
Can read and write	207(8.81)	5(2.42)	0	247(8.70)	104(42.11)	0
Primary school (1–8)	265(11.2)	25(9.43)		591(20.82)	254(42.98)	
High school (9–12)	561(23.88)	39(6.95)		645(22.72)	322(49.92)	
Above secondary	479(20.39)	34(7.10)		784(27.62)	233(29.72)	
College and above	623(26.82)	34(5.46)		313(11.03)	173(55.27)	
	X^2= 15.154 P-value = 0.010			X^2= 99.117 P-value =0.000		
WHO stage						
Stage 1	490(20.86)	32(6.53)		591(20.82)	251(42.47)	
Stage 2	1010(43.0)	64(6.34)	0.01	1174(41.35)	529(45.06)	0.13
Stage 3	739(31.46)	56(7.58)		909(32.02)	356(39.16)	
Stage 4	110(4.68)	3(2.73)		165(5.81)	32(19.39)	
	X^2= 3.940 P-value = 0.268			X^2= 41.568 P-value =0.000		
CD4 at the start of ART						
Less than 100	479(20.39)	26(5.43)	0	575(20.25)	240(41.74)	0
> =100	1870(79.61)	129(6.90)		2264(79.75)	928(40.99)	
	X^2= 1.337 P-value = 0.247			X^2= 0.106 P-value =0.744		
Had opportunistic diseases						
Yes	431(18.35)	27(6.26)	0	498(17.54)	220(44.18)	0
No	1918(81.65)	120(6.67)		2341(82.46)	948(40.50)	
	X^2= 0.095 P-value = 0.007			X^2= 2.298 P-value =0.001		
Adherence level						
Good	1814(77.22)	55(3.03)	2.44	2200(77.49)	903(41.05)	3.22
Fair	394(16.77)	44(11.17)		479(16.87)	187(39.04)	
Poor	141(6.0)	56(39.72)		160(5.64)	70(48.75)	
	X^2= 301.707 P-value = 0.000			X^2= 4.707 P-value =0.095		
Total	**2349**	**155 (6.62)**		**2839**	**1168 (41.14)**	

Elimination (RFE) and Random Forest Classifier, Variance Threshold, recursive feature elimination (RFE), and RF.[25] The selected features were trained among eight state-of-the-art supervised ML algorithms.

Data Balancing

In the respective datasets, the amount of data in outcome classes is significantly imbalanced. As seen in Figure 1, 93.38% of the viral load dataset featured the occurrence of suppressed viral load, and only 6.62% unsuppressed viral load. Similarly, it was observed that the CD4 status distribution was 41.3% for CD4 cell count <200 cells/mL (Figure 2). Therefore, the synthetic minority over-sampling technique (SMOTE) method was utilized in the imbalanced-learn toolbox to balance the dataset to address the class imbalance (https://imbalanced-learn.org/stable/).[67,68]

Data Split and Model-Validation

For machine learning (ML) approaches, a stratified tenfold cross-validation (CV) method was implemented in this study, which partitioned the entire dataset into ten folds. Datasets were split into 70% training set and 30% testing set/train and optimize the models using the training set and the result of cross-validation was registered. Hence, in this research work, nine folds were used to train the model, and then the remaining one fold was used to test the model. This was repeated iteratively until each fold was used once as a test set, thus performing the 10-fold CV. Finally, depending on the average performance, the evaluation classifier was created.

In this work, a stratified tenfold CV was employed for estimation accuracy because of its relatively low-level bias and variety. This method assisted to reduce the deviation in prediction error; increases the use of data for both training and validation, without overfitting or overlap between the test and validation data; and protected against experiment theory
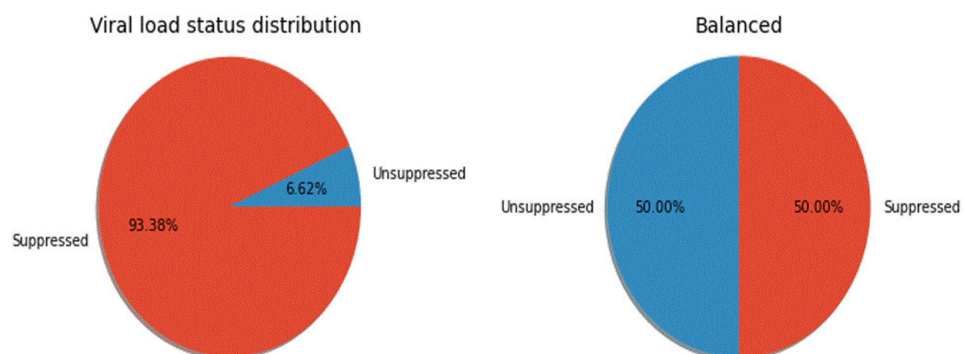

Figure 1 Viral load status distribution among of participants in ART.
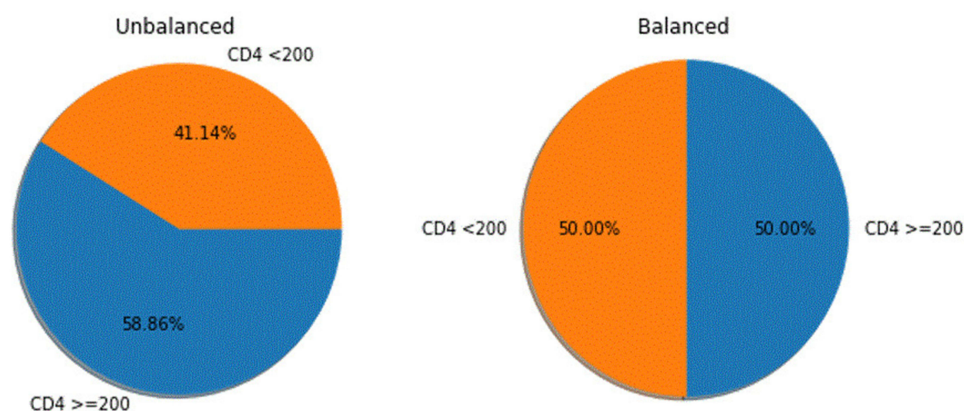

Figure 2 CD4 status distribution among of participants in ART.

proposed by arbitrarily split data. Further, when a CV was used with the stratified sampling method, both the training and test sets were almost the same proportion of the feature with concern as the original dataset. Performing this with the target variable guarantees that the CV outcome was a close approximation of the error function.

Experimental Details

For each cross-validation fold and modeling choice, we employed and trained eight supervised ML algorithms, specifically 5 standard algorithms (KNN, DT, GNB, SVM, LR) and 3 ensemble algorithms (GB, XGB, RF). All eight of our models use a stratified 70–30 split of the data into train-test sets. The split was repeated 100 times, and all presented results were averaged over these repeats. For each 70–30 split, we also conducted a 10-fold cross-validation on the train sets, where the validation loss was used for early stopping.

Model Evaluation

Different model performance techniques, such as sensitivity, specificity, F1-score, F-beta scores, accuracy, negative predictive value (NPV), and positive predictive value (PPV), which were derived, using the observed data as the gold standard, were used to assess the performances of the given ML models.[25,69] Further, the Receiver Operating Characteristics (ROC) curves were used to express the connection between the model sensitivity and specificity. To assess the effectiveness of classification models and the performance of each class in the prediction model, confusion matrices were also obtained. The significance of predictors for a tree's choice according to their mean decreasing Gini was finally determined based on the most accurate method.

Ethical Approval and Consent to Participate

The Institutional Ethical Review Board of Dilla University in Ethiopia gave its approval for the study protocol, which was carried out following the Declaration of Helsinki's criteria (Reference No. duirb/004/21-10). In this study, the source data was recorded as part of the routine health system prior to the design of this study. The Institutional Ethical Review Board of Dilla University in Ethiopia approved participant's consent waiver. This research was done using these data, and all study participants within the ART database were anonymized for the purpose of de-identification; it did not contain any identifying data of the participants to ensure confidentiality.

Results

Characteristics of the Study Participants

The demographic and clinical-related characteristics of the participants included in the analyses are shown in Table 1. The mean age of the sample adults was 41.6 years (SD = 10.9). In total, 1845 (64.99%) of participants with a CD4 cell count <200 cells/mL were females and 1610 (56.71%) were married. Furthermore among participants with a CD4 cell count <200 cells/mL, 1927 (17.54%) had opportunistic infections. Similarly, the population of participants was divided into two groups based on viral load (viral load <1000 copies/mL or ≥1000 copies/mL) (see Table 1 for detail).

Model Building

After applying various feature selection techniques, 17 and 19 most crucial features were chosen for were chosen as the input for the ML algorithms in the classification of viral load and CD4, respectively. The detailed report on the relative importance of the selected features is shown in Supplementary Figures 1 and 2.

Furthermore, a correlational heatmap was employed to investigate the dependency of included predictors to pass into ML models. The correlation between viral load predictors shown in Figure 3 revealed that the health facility was highly correlated with tpt status and date of regimen change with 0.47 and 0.56, respectively. In addition, as shown in Figure 4 among CD4 predictors, functional status and marital were highly correlated (0.62); and WHO stage and regimen change were highly correlated (0.56). Overall, as it is depicted in Figures 3 and 4, all predictors correlated below (<0.8) indicate that there was no multicollinearity of features.
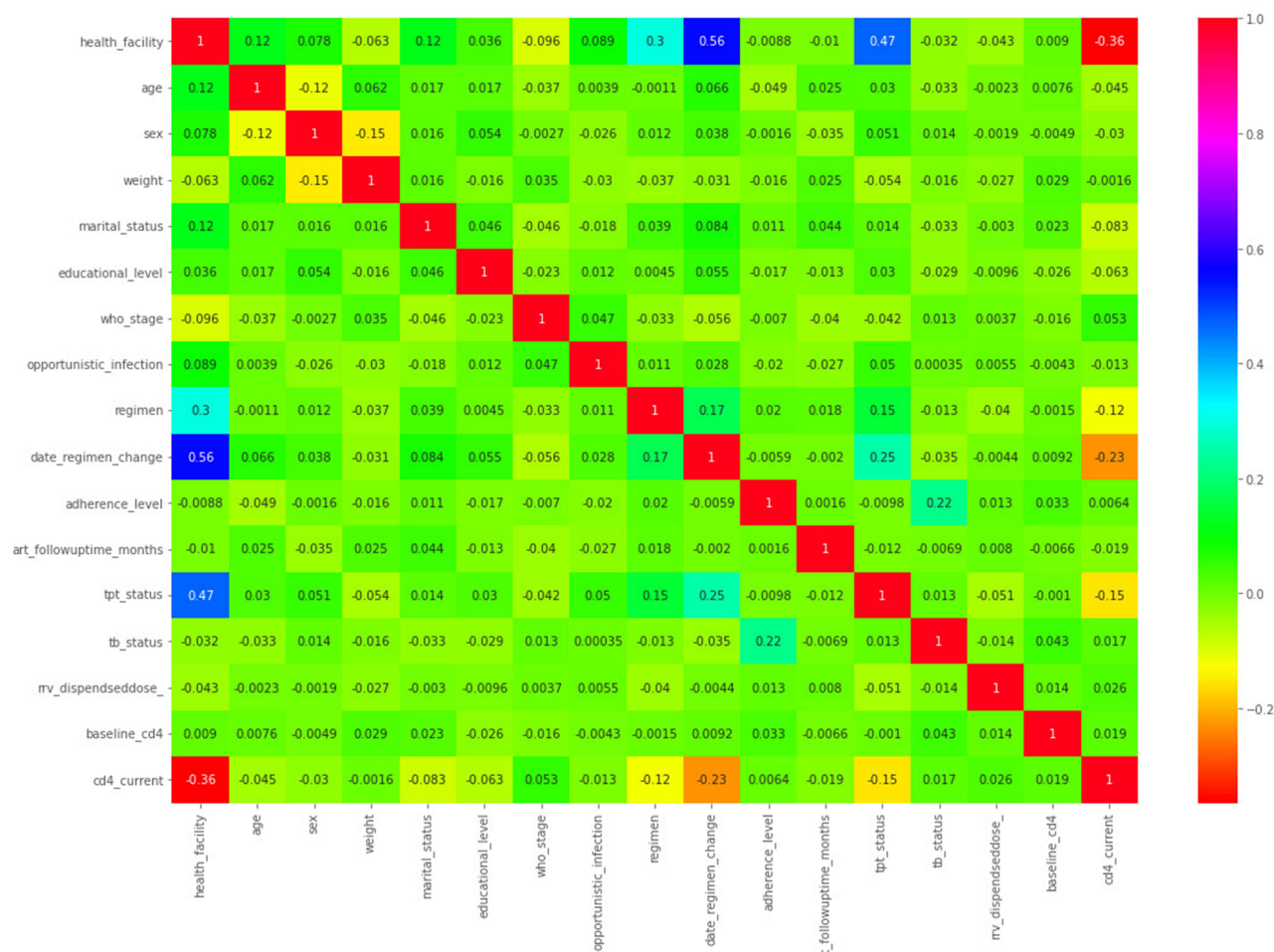
Figure 3 Heat map showing the correlation between features in viral load prediction.

Viral Load Classification Models Evaluation Results

A learning model was developed to distinguish the virological failure of individuals from given adults enrolled in ART, and an ML classifier was developed and validated to predict and separate the virological failure of participants. As shown in Figure 5, the sensitivity ranged from 0.81 (LR) to 0.97 (XGB), while the specificity ranged from 0.73 (GNB) to 0.95 (XGB). Furthermore as shown in Figures 6 and Table 2, the XGB algorithm had an excellent performance considering both the F1 measure and accuracy (>0.96). However, all algorithms had an accuracy of more than 0.78 (Table 2), thereby indicating that they could be used for the classification of virological failure with acceptable accuracy.

Figures 7 and 8 show the performance ROC-AUC of both the XGB and RF algorithms was 0.99. This shows that both models were better at accurately distinguishing suppressed and unsuppressed viral load of participants. In addition, a detailed report on the performance of ROC-AUC ML models was provided in Supplementary Figures 3–8. Overall, based on the findings of various performance evaluation metrics, the results approved that the best and ideal techniques, which have a good performance, were XGB and RF.

Figures 9 and 10 show the relative importance of features based on XGB and RF models, respectively. Among top five features of great importance for predicting viral load, regimen change, adherence level, level of CD4, and duration of ART follow-up were similarly identified in both algorithms to predict viral load. On the other hand, TB status in XGB (Figure 9) and health facility in the RF (Figure 10) based model were found to be highly predictive.
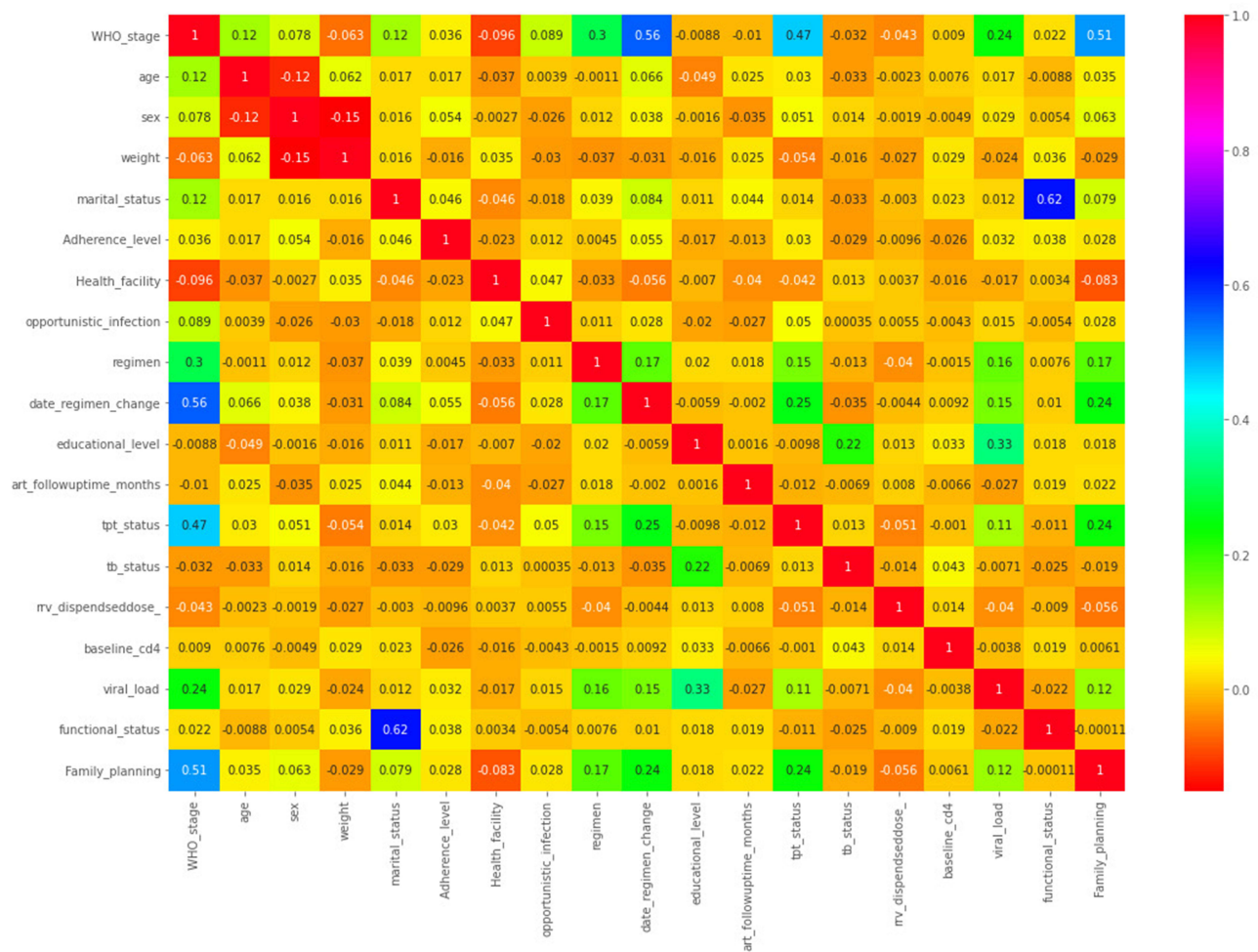
Figure 4 Heat map showing the correlation between features in CD4 prediction.

CD4 Classification Models Evaluation Results

Figure 11 shows the sensitivity, specificity, and AUC of the eight ML algorithms in classifying CD4 of individuals enrolled in ART care from a given adult population enrolled in ART care. The experimental results showed that the GB classifiers yielded better capability with 76.0% sensitivity and 82.0% specificity. Also, the GB algorithm achieved 81.0% precision and 78.0% F1-score shown in Figure 12 and 79% of accuracy (Table 2), indicating that the GB algorithm model had a specific predictive ability.

Furthermore, Figure 13 shows the GB algorithm attained better at accurately distinguishing good and poor CD4 with ROC of 83.0%. The AUC of the remaining algorithms compared was provided in Supplementary Figures 9–15. Overall, GB classifiers were relatively better in the classification of CD4.

Based on the most accurate algorithm (GB), the relative importance of features for classifying the CD4 of participants as good and poor is presented in Figure 14. WHO clinical stage, weight, duration of ART follow-up, age, and regimen change were found to be highly predictive of poor CD4 among participants.

Discussion

The ML models built in this study were used to predict the viral load and CD4 status of adults living with HIV/AIDS and enrolled in ART care in Ethiopia. Accordingly, we trained, validated, and tested five standard ML algorithms KNN, DT, GNB, SVM, LR, and three ensemble algorithms GB, XGB, and RF on a separate dataset for predicting virological failure and poor CD4 of adults enrolled in ART. Our models were able to correctly predict 78% to 96% of viral load un-
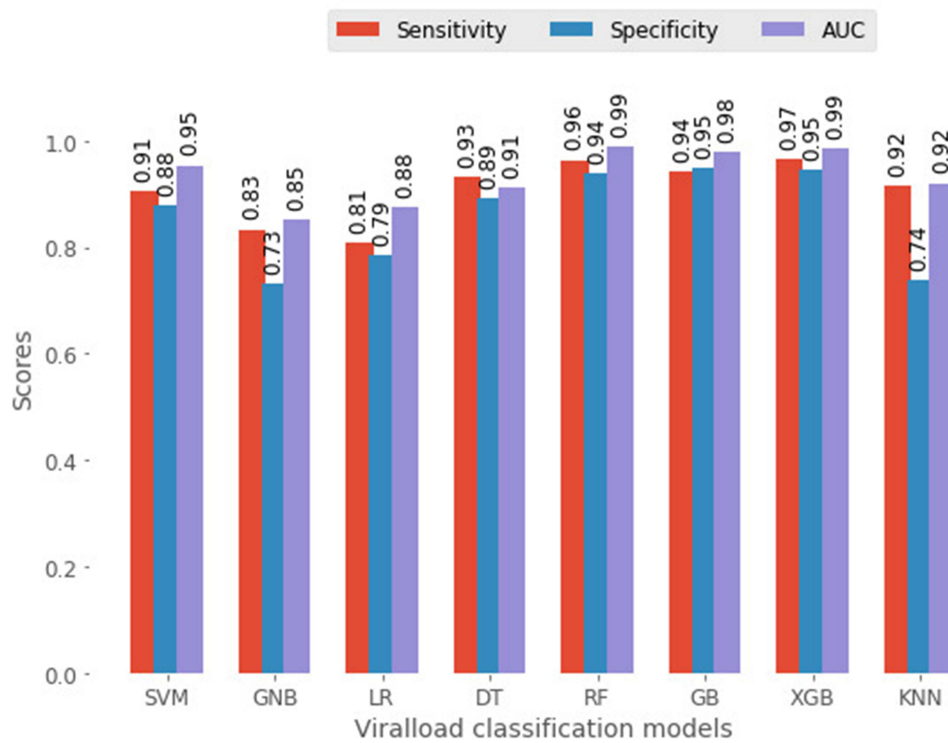
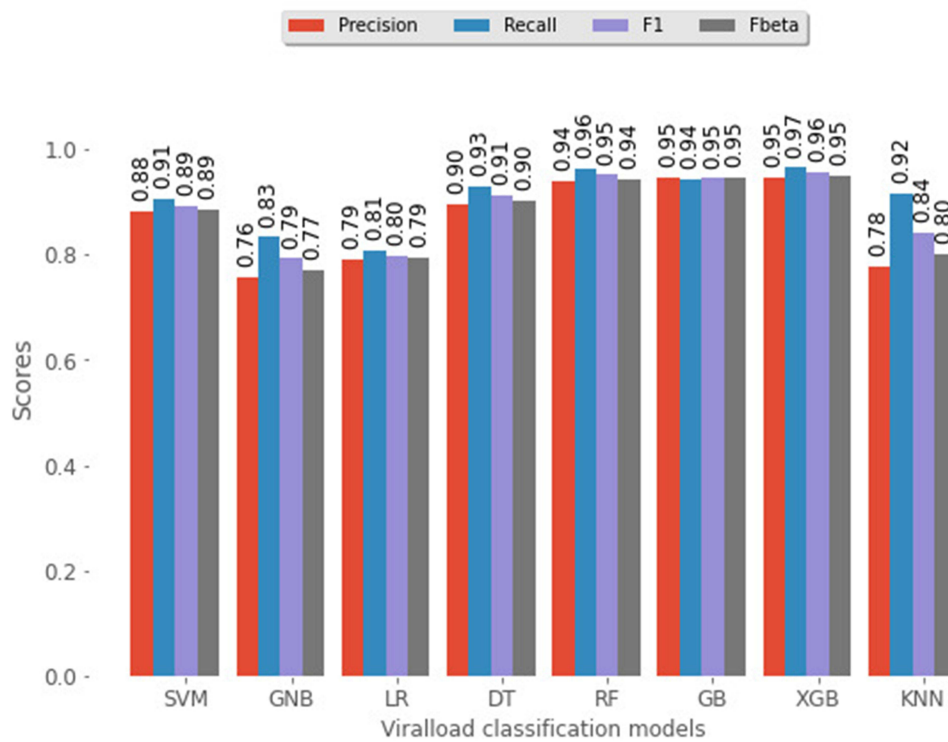**Figure 5** Performance evaluation of eight machine-learning models for viral load prediction.

Figure 6 Precision, Recall, F1, and F-beta scores based evaluation of eight machine-learning models for viral load prediction.

Table 2 Comparison of Viral Load and CD4 Classification Models Using Accuracy and AUC

Algorithms	Comparison Method	Viral Load Models	CD4 Models
K Nearest Neighbours (KNN)	Accuracy (%)	83.0	71.0
	AUC	0.92	0.76
Support Vector Machine (SVM)	Accuracy (%)	89.0	76.0
	AUC	0.95	0.79
Logistic Regression (LR)	Accuracy (%)	80.0	75.0
	AUC	0.88	0.79
Decision Tree (DT)	Accuracy (%)	91.0	65.0
	AUC	0.91	0.65
Gaussian Naive Bayes (GNB)	Accuracy (%)	78.0	72.0
	AUC	0.85	0.75
Random Forest (RF)	Accuracy (%)	95.0	77.0
	AUC	0.99	0.83
Gradient Boosting (GB)	Accuracy (%)	95.0	79.0
	AUC	0.98	0.83
eXtreme Gradient Boosting (XGB)	Accuracy (%)	96.0	76.0
	AUC	0.99	0.81

Abbreviation: AUC, area under curve.

suppression (or suppression) and 79% in CD4 classification. These results suggest some important implications for the application of supervised ML algorithm methodology in clinical practice.

When we evaluated the performance of eight ML models to predict the viral load un-suppression (or suppression) of individuals enrolled in ART, we found that the XGB and RF models had better performance than the other ML models tested. AUC of XGB and RF models in this study was 0.99, and the accuracy was 0.96 and 0.95, respectively. Different studies have been evaluating the application of ML techniques in predicting viral load of individuals with HIV/AIDS.[33,36,44] This is somewhat higher than Maskew et al, which reports that ML based predictive modeling achieves an AUC of 0.76.[39]
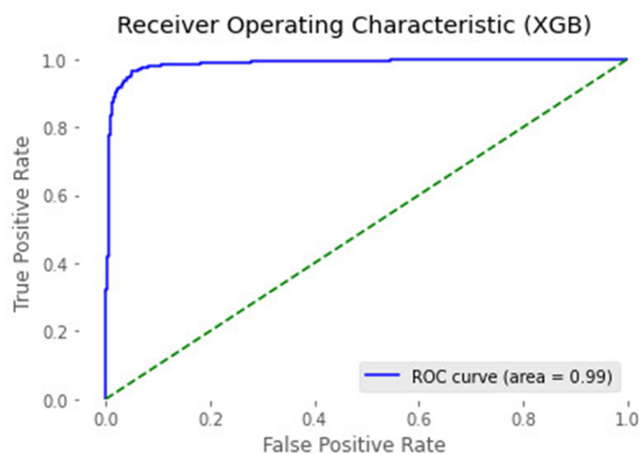

Figure 7 Receiver operating characteristic curves of eXtreme Gradient Boosting (XGBoost) model for viral load prediction.
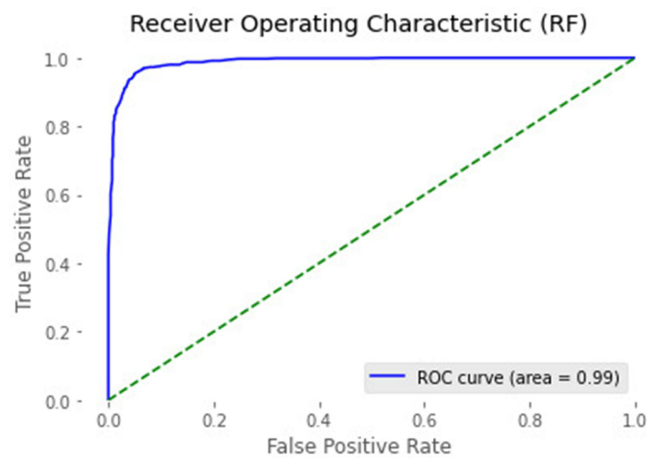
Figure 8 Receiver operating characteristic curves of Random Forest (RF) model for viral load prediction.
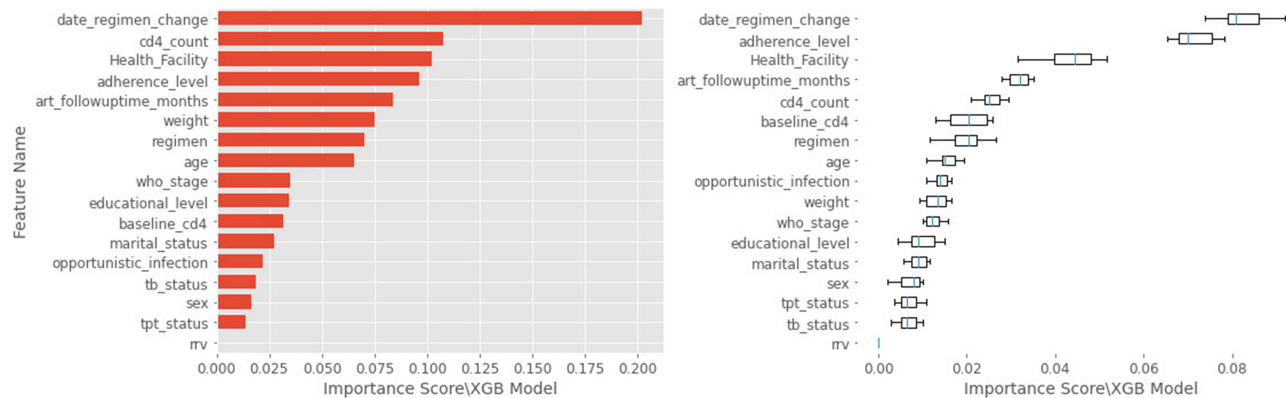

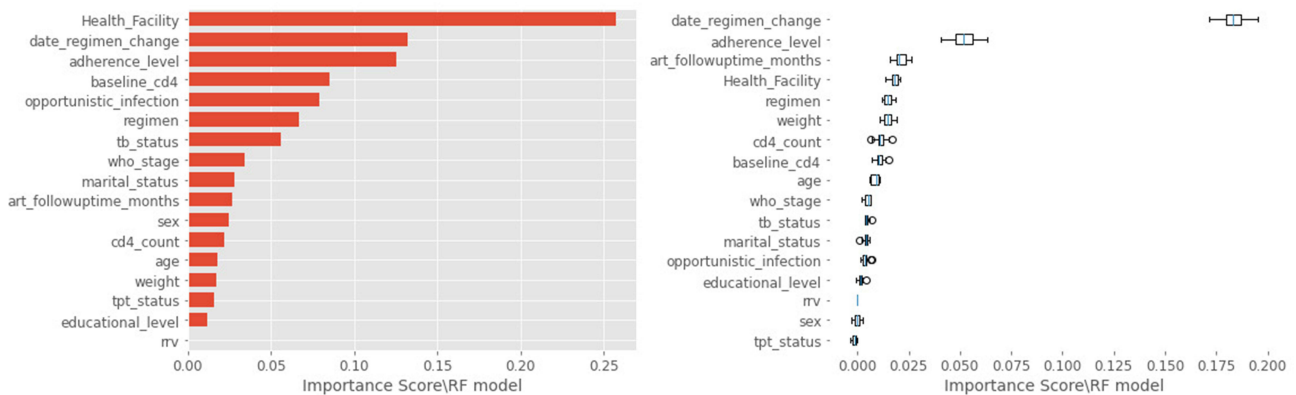**Figure 9** Predictors of viral load in the eXtreme Gradient Boosting (XGB) model.

Figure 10 Predictors of viral load in the Random Forest (RF) model.

Recently, Kamal et al conducted a study employing RF-based prediction of virologic outcomes in Switzerland, which achieved an AUC of 0.77.[38] Meanwhile, our accuracy finding is in agreement with a study of Bisaso et al, which reports a predictive modeling accuracy of 92.9%.[36] Overall, our XGB model shows good predictive performance in various diagnostic metrics, such as accuracy, precision, sensitivity, and F1 score. This is in line with evidences which shows XGB model had similar performance for predicting the CD4/CD8 ratio and HIV/AIDS prognosis.[40–42,44,45]
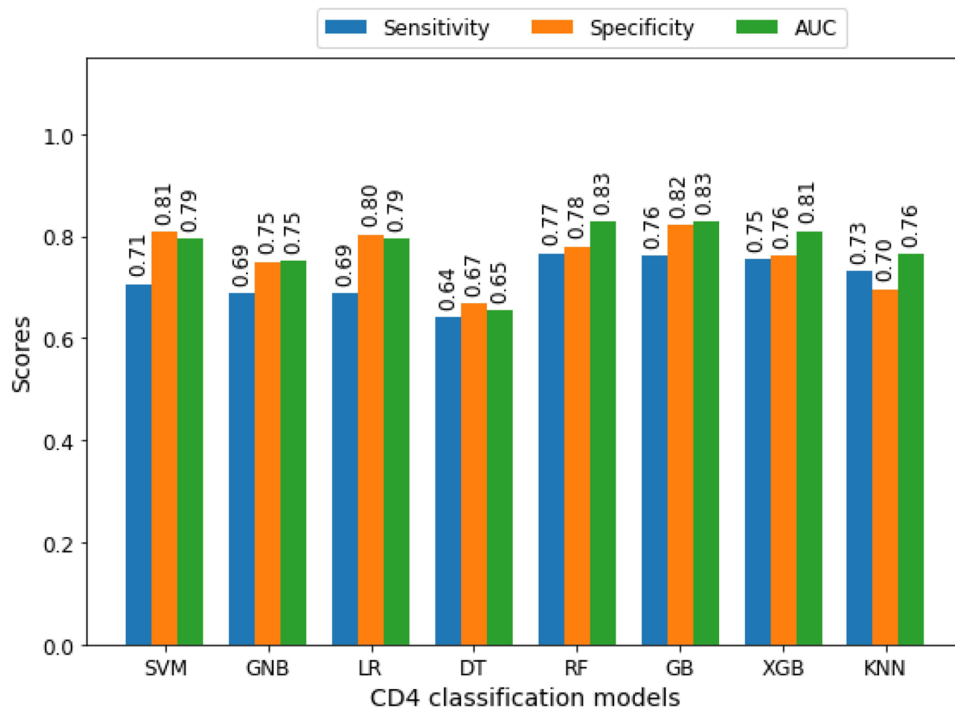
Figure 11 Performance evaluation of eight machine-learning models for CD4 prediction.
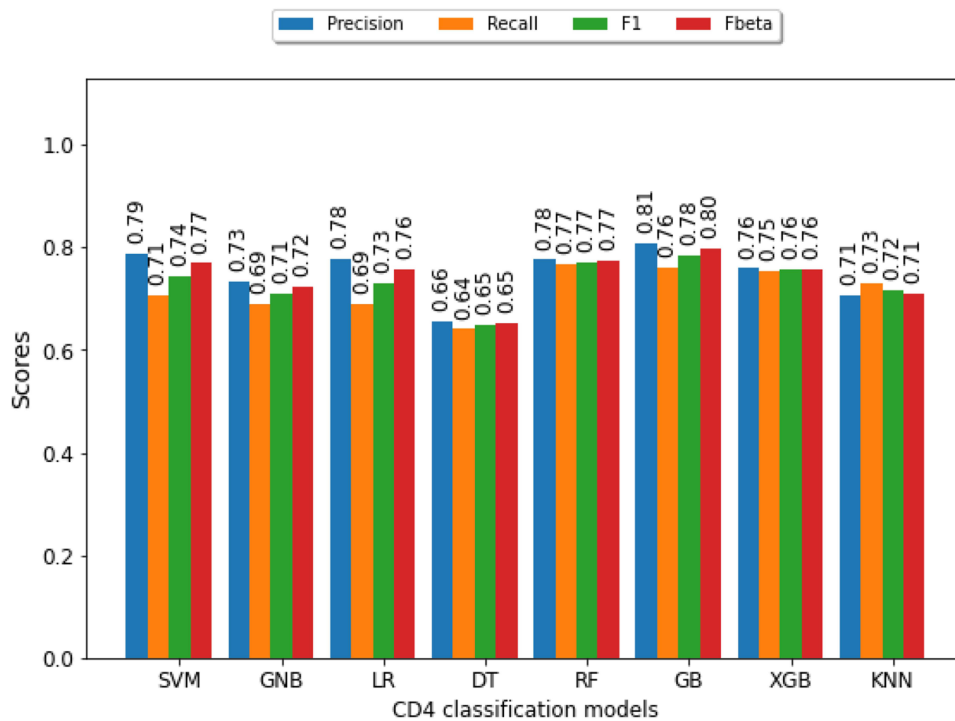

Figure 12 Precision, Recall, F1, and F-beta scores based evaluation of eight machine-learning models for CD4 prediction.

Regarding CD4 prediction, our model developed by the GB yielded better performance with (0.79 accuracies, ROC-AUC = 0.83), in predicting low or poor CD4 among other models developed by LR, SVM, KNN, DT, GNB, XGB, and RF. This accuracy is lower than a study done in Ethiopia which reported 83.5% to 99.8%, across J48, Decision tree, Neural
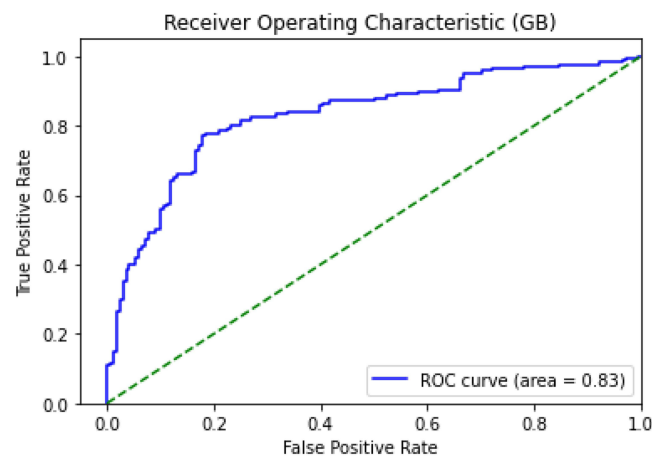
Figure 13 Receiver operating characteristic curves of Gradient Boosting (GB) model for CD4 prediction.
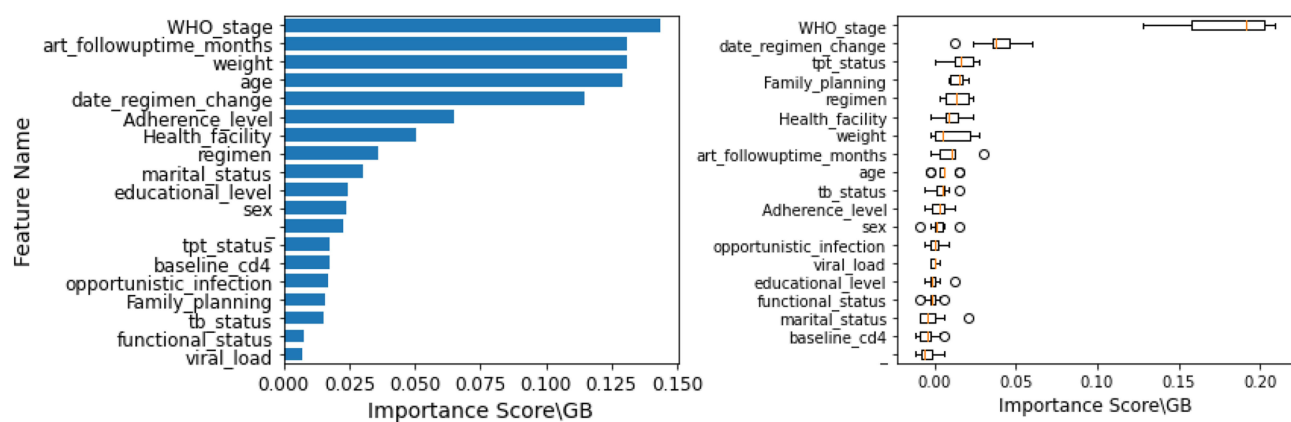

Figure 14 Predictors of CD4 in the Gradient Boosting (GB) model.

Network, and RF algorithms in predicting CD4.[59] Similarly, it is lower than 83% returned in South Africa.[54] Regarding other model's performance metrics (specificity = 0.82%) in this study were in line with a study from Yunnan, China.[70]

Using the XGB and RF models, we assessed the importance of predictors that contribute to better performance in virological failure prediction. In this study, regimen change, adherence level, level of CD4 and duration of ART follow-up, and TB status were significant in predicting viral load. In our study, findings indicated that the level of adherence in ART care was the most important feature in predicting viral load failure. This is in line with findings of previous ML findings.[38,39,52] Furthermore, duration on ART was an important feature for predicting virological failure. Participants who had a longer duration of stay on ART had a higher risk of developing virological failure; this is consistent with the previous studies,[38,39] which reported longer ratio of follow-up increases the chances of developing virological failure. Our finding indicated participants who had low CD4 counts had higher chance of virological failure. The low of CD4 count in particular is a well-known predictor of ART treatment outcome that has been strongly associated with viral load failure.[38,46,49] Similarly, opportunistic infection was found to have the most crucial impact in predicting viral load failure among individuals living with HIV/AIDS in the literature.[36,39,44]

This study determined the significant influencing factors for low or poor CD4 status using the relevance values of features for the GB classifier. Accordingly, WHO clinical stage, weight, duration of ART follow-up, age, and regimen change were predictive of CD4 status in this study. We found that the WHO clinical stage of participants was most crucial in predicting CD4, which is consistent with the previous findings.[62] The age and weight of the HIV/AIDS participants were also the most important features in predicting CD4.[11,49,59,60] The latter characteristics indicated

duration of ART follow-up heightened and possibly played a role in the low level of CD4. Several studies obtain the same results for the individual's duration of ART follow-up.[42,49,58,60,69]

The results of this work should be interpreted as an illustration of the presented statistical methods in a realistic setting. The practical conclusions should be seen in light of the data limitations. Analysis was restricted to routinely collected retrospective data describing the demographics and clinical status of participants. There may be considerable reservation with unexplored behavioral information, such as smoking, alcohol consumption, exercise, and dietary habits. Hence, unexplored and missing information may also contribute to viral load and CD4 status, the impact of these factors is unknown and needs to be examined. Furthermore, data we used in this study were routinely collected on two-centers and it requires external validation. Fear of generalization should be alleviated to a large and diverse datasets to include more variables and improve the performance of ML models. This study used a supervised ML algorithm, feature selection methods to select the best subset features to develop the models. It is better to see the difference in performance results using unsupervised or deep learning algorithm models.

Conclusion

In order to predict viral load and CD4 status of individuals in ART, this study compared the effectiveness of eight ML classifier models. In this study, the XGB and RF models had the highest accuracy and outperformed on various evaluation metrics among the models examined for viral load classification. In the prediction of participants CD4, GB model had the highest accuracy.

Therefore, ML-based predictive models, particularly the XGB, RF, and GB algorithm, potentially facilitate identifying the patients who are at high risk of virological failure and poor CD4 to inform proper interventions by the clinicians.

Abbreviations

AIDS, Acquired Immunodeficiency Syndrome; ART, Anti-Retroviral Treatment; DT, Decision Trees; XGB, eXtreme Gradient Boosting; GB, Gradient Boosting; GNB, Gaussian Naive Bayes; HAART, Highly Active Anti-Retroviral Treatment; HIV, Human Immunodeficiency Virus; LR, Logistic Regression; KNN, K-nearest Neighbor; RF, Random Forest; SVM, Support Vector Machine; WHO, World Health Organization.

Data Sharing Statement

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

Disclosure

The authors declare that they have no competing interests.

References

1. UNAIDS. Global HIV and AIDS statistics - 2020 fact sheet; 2020. Available from: https://www.unaids.org/en/resources/fact-sheet#:~:text=GLOBAL%20HIV%20STATISTICS&text=38.0%20million%20%5B31.6%20million–44.5,AIDS-related%20illnesses%20in%202019. Accessed January 27, 2023.
2. STATISTICS, G.H.A.A. Global information and education on HIV and AIDS; 2019. Available from: https://www.avert.org/global-hiv-and-aids-statistics. Accessed January 27, 2023.
3. World Health Organization. HIV/AIDS; 2020.
4. Statistics, G. Global HIV/AIDS statistics; 2021. Available from: https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics. Accessed January 27, 2023.

5. Assefa Y, Hill PS, Van Damme W, et al. Leaving no one behind, lessons from implementation of policies for universal HIV treatment to universal health coverage. *Global Health*. 2020;16(1):17. doi:10.1186/s12992-020-00549-4

6. Barnabas G, Sibhatu MK, Berhane Y. Antiretroviral therapy program in Ethiopia benefits from virology treatment monitoring. *Ethiop J Health Sci*. 2017;27(Suppl 1):1–2. doi:10.4314/ejhs.v27i1.1S

7. Hamid A. The impact of viral load monitoring and CD4 in patient taking anti-retroviral treatment at kicukiro health center. *World J Pharm Res*. 2017;1897–1908. doi:10.20959/wjpr20178-9114

8. Mazrouee S, Little SJ, Wertheim JO. Incorporating metadata in HIV transmission network reconstruction, A machine learning feasibility assessment. *PLoS Comput Biol*. 2021;17(9):e1009336. doi:10.1371/journal.pcbi.1009336

9. Pimentel V, Pingarilho M, Alves D, et al. Molecular epidemiology of HIV-1 infected migrants followed up in Portugal, trends between 2001–2017. *Viruses*. 2020;12:3. doi:10.3390/v12030268

10. Qian Y, Wu Z, Chen C, et al. Detection of HIV-1 viral load in tears of HIV/AIDS patients. *Infection*. 2020;48(6):929–933. doi:10.1007/s15010-020-01508-2

11. Sharma R, Pai C, Kar H. A retrospective analysis of discordant CD4 and viral load responses in HIV patients on anti-retroviral therapy. *Int J Sci Res Publ*. 2013;3:1–3.

12. Shoko C, Chikobvu D. A superiority of viral load over CD4 cell count when predicting mortality in HIV patients on therapy. *BMC Infect Dis*. 2019;19(1):169. doi:10.1186/s12879-019-3781-1

13. Stockman J, Friedman J, Sundberg J, et al. Predictive analytics using machine learning to identify ART clients at health system level at greatest risk of treatment interruption in Mozambique and Nigeria. *J Acquir Immune Defic Syndr*. 2022;90:2. doi:10.1097/QAI.0000000000002947

14. Aavani P, Allen LJS. The role of CD4 T cells in immune system activation and viral reproduction in a simple model for HIV infection. *Appl Math Model*. 2019;75:210–222. doi:10.1016/j.apm.2019.05.028

15. Søgaard OS. Deciphering the association between HIV-specific immunity and immune reconstitution. *EBioMedicine*. 2021;67:103350. doi:10.1016/j.ebiom.2021.103350

16. Migueles SA, Connors M. The role of CD4+ and CD8+ T cells in controlling HIV infection. *Curr Infect Dis Rep*. 2002;4(5):461–467. doi:10.1007/s11908-002-0014-2

17. Tripiciano A, Picconi O, Moretti S, et al. Anti-Tat immunity defines CD4(+) T-cell dynamics in people living with HIV on long-term cART. *EBioMedicine*. 2021;66:103306. doi:10.1016/j.ebiom.2021.103306

18. Edelman EJ, Rentsch CT, Justice AC. Polypharmacy in HIV, recent insights and future directions. *Curr Opin HIV AIDS*. 2020;15(2):126–133. doi:10.1097/COH.0000000000000608

19. Takahashi N, Ardeshir A, Holder GE, et al. Comparison of predictors for terminal disease progression in simian immunodeficiency virus/simian-HIV-infected rhesus macaques. *Aids*. 2021;35(7):1021–1029. doi:10.1097/QAD.0000000000002874

20. Tu W, Johnson E, Fujiwara E, et al. Predictive variables for peripheral neuropathy in treated HIV type 1 infection revealed by machine learning. *Aids*. 2021;35(11):1785–1793. doi:10.1097/QAD.0000000000002955

21. Javaid M, Haleem A, Pratap Singh R, et al. Significance of machine learning in healthcare, Features, pillars and applications. *Int J Intell Netwrk*. 2022;3:58–73. doi:10.1016/j.ijin.2022.05.002

22. Secinaro S, Calandra D, Secinaro A, et al. The role of artificial intelligence in healthcare, a structured literature review. *BMC Med Inform Decis Mak*. 2021;21(1):125. doi:10.1186/s12911-021-01488-9

23. Witten I. *Data Mining Practical Machine Learning Tools and Techniques*. 2nd ed. Amsterdam, Boston, Heidelberg, London, New york, Oxford, Paris, San diego, San Francisco, Singapore, Sydney, Tokyo: Elsevier; 2010.

24. Singh P. *Chapter 5 - Diagnosing of Disease Using Machine Learning, in Machine Learning and the Internet of Medical Things in Healthcare*. Academic Press; 2021:89–111.

25. Erickson BJ. Basic artificial intelligence techniques, machine learning and deep learning. *Radiol Clin North Am*. 2021;59(6):933–940. doi:10.1016/j.rcl.2021.06.004

26. Capitaine L, Genuer R, Thiébaut R. Random forests for high-dimensional longitudinal data. *Stat Methods Med Res*. 2021;30(1):166–184. doi:10.1177/0962280220946080

27. Duke ER, Williamson BD, Borate B, et al. CMV viral load kinetics as surrogate endpoints after allogeneic transplantation. *J Clin Invest*. 2021;131:1. doi:10.1172/JCI133960

28. Jamal S, Nikolić N, Mildner M, et al. Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis. *Sci Rep*. 2020;10:1. doi:10.1038/s41598-019-56847-4

29. Kuhn T, Kaufmann T, Doan NT, et al. An augmented aging process in brain white matter in HIV. *Hum Brain Mapp*. 2018;39(6):2532–2540. doi:10.1002/hbm.24019

30. Zarei H, Kamyad AV, Heydari AA. Fuzzy modeling and control of HIV infection. *Comput Math Methods Med*. 2012;2012:893474. doi:10.1155/2012/893474

31. Sajda P. Machine learning for detection and diagnosis of disease. *Biomed Eng*. 2006;8:537–565.

32. Abirami N, Kamalakannan T, Muthukumaravel A. A study on analysis of various data mining classification techniques on healthcare data. *Int J Emerging Technol Adv Eng*. 2013;3(7):604–607.

33. Chen S, Owolabi Y, Dulin M, et al. Applying a machine learning modelling framework to predict delayed linkage to care in patients newly diagnosed with HIV in Mecklenburg County, North Carolina, USA. *Aids*. 2021;35(Suppl 1):S29–s38. doi:10.1097/QAD.0000000000002830

34. Ekpenyong ME, Etebong PI, Jackson TC. Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy. *Heliyon*. 2019;5(7):e02080. doi:10.1016/j.heliyon.2019.e02080

35. Benitez AE, Musinguzi N, Bangsberg DR, et al. Super learner analysis of real-time electronically monitored adherence to antiretroviral therapy under constrained optimization and comparison to non-differentiated care approaches for persons living with HIV in rural Uganda. *J Int AIDS Soc*. 2020;23(3):e25467. doi:10.1002/jia2.25467

36. Bisaso K, Karungi SA, Kiragga A, et al. A comparative study of logistic regression based machine learning techniques for prediction of early virological suppression in antiretroviral initiating HIV patients. *BMC Med Inform Decis Mak*. 2018;18:1. doi:10.1186/s12911-018-0659-x

37. Chakraborty S, Roy M, Chatterjee S, et al. Detection of HIV-1 progression phases from transcriptional profiles in ex vivo CD4+ and CD8+ T cells using meta-heuristic supported artificial neural network. *Multimed Tools Appl*. 2022;81(11):15103–15126. doi:10.1007/s11042-022-12534-7

38. Kamal S, Urata J, Cavassini M, et al. Random forest machine learning algorithm predicts virologic outcomes among HIV infected adults in Lausanne, Switzerland using electronically monitored combined antiretroviral treatment adherence. *AIDS Care*. 2021;33(4):530–536. doi:10.1080/09540121.2020.1751045

39. Maskew M, Sharpey-Schafer K, De Voux L, 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. doi:10.1038/s41598-022-16062-0

40. Murnane PM, Ayieko J, Vittinghoff E, et al. Machine learning algorithms using routinely collected data do not adequately predict viremia to inform targeted services in postpartum women living with HIV. *J Acquir Immune Defic Syndr*. 2021;88(5):439–447. doi:10.1097/QAI.0000000000002800

41. Paul RH, Cho KS, Belden AC, et al. Machine-learning classification of neurocognitive performance in children with perinatal HIV initiating de novo antiretroviral therapy. *Aids*. 2020;34(5):737–748. doi:10.1097/QAD.0000000000002471

42. Paul RH, Cho KS, Luckett P, et al. Machine learning analysis reveals novel neuroimaging and clinical signatures of frailty in HIV. *J Acquir Immune Defic Syndr*. 2020;84(4):414–421. doi:10.1097/QAI.0000000000002360

43. Peng X, Zhu B. Different features identified by machine learning associated with the HIV compartmentalization in semen. *Infect Genet Evol*. 2022;98:105224. doi:10.1016/j.meegid.2022.105224

44. Petersen ML, LeDell E, Schwab J, et al. Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring. *J Acquir Immune Defic Syndr*. 2015;69(1):109–118. doi:10.1097/QAI.0000000000000548

45. Shi M, Lin J, Wei W, et al. Machine learning-based in-hospital mortality prediction of HIV/AIDS patients with Talaromyces marneffei infection in Guangxi, China. *PLoS Negl Trop Dis*. 2022;16(5):e0010388. doi:10.1371/journal.pntd.0010388

46. Wang D, Larder B, Revell A, et al. A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy. *Artif Intell Med*. 2009;47(1):63–74. doi:10.1016/j.artmed.2009.05.002

47. Yang X, Zhang J, Chen S, et al. Utilizing electronic health record data to understand comorbidity burden among people living with HIV, a machine learning approach. *Aids*. 2021;35(Suppl 1):S39–s51. doi:10.1097/QAD.0000000000002736

48. Zhang J, Olatosi B, Yang X, et al. Studying patterns and predictors of HIV viral suppression using A big data approach, a research protocol. *BMC Infect Dis*. 2022;22(1):122. doi:10.1186/s12879-022-07047-5

49. Li B, Li M, Song Y, et al. Construction of machine learning models to predict changes in immune function using clinical monitoring indices in HIV/AIDS patients after 9.9-years of antiretroviral therapy in Yunnan. *Front Cell Infect Microbiol*. 2022;12:867737. doi:10.3389/fcimb.2022.867737

50. Weissman S, Yang X, Zhang J, et al. Using a machine learning approach to explore predictors of healthcare visits as missed opportunities for HIV diagnosis. *Aids*. 2021;35(Suppl 1):S7–s18. doi:10.1097/QAD.0000000000002735

51. Pulliam L, Liston M, Sun B, et al. Using neuronal extracellular vesicles and machine learning to predict cognitive deficits in HIV. *J Neurovirol*. 2020;26(6):880–887. doi:10.1007/s13365-020-00877-6

52. Soogun AO, Kharsany ABM, Zewotir T, et al. Identifying potential factors associated with high HIV viral load in KwaZulu-Natal, South Africa using multiple correspondence analysis and random forest analysis. *BMC Med Res Methodol*. 2022;22(1):174. doi:10.1186/s12874-022-01625-6

53. Ioannidis JP, Goedert JJ, McQueen PG, et al. Comparison of viral load and human leukocyte antigen statistical and neural network predictive models for the rate of HIV-1 disease progression across two cohorts of homosexual men. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1999;20(2):129–136. doi:10.1097/00042560-199902010-00004

54. Yashik Singh MM. Support vector machines to forecast changes in CD4 count of HIV-1 positive patients. *Sci Res Essays*. 2010;5(17):2384–2390.

55. Madigan EA, Miklos Zrinyi OLC, Zrinyi M. Workforce analysis using data mining and linear regression to understand HIV/AIDS prevalence patterns. *Hum Resour Health*. 2008;6:1–6. doi:10.1186/1478-4491-6-2

56. HEALTH, I.G. Machine Learning for predicting default from HIV services in Mozambique; 2022.

57. Kuteesa R, Bisaso GTA, Susan A, Kiragga A, Castelnuovo B. Karungi, Agnes Kiragga, Barbara Castelnuovo A survey of machine learning applications in HIV clinical research and care. *Comput Biol Med*. 2017;91:366–371. doi:10.1016/j.compbiomed.2017.11.001

58. Chala TD. Data mining technology enabled anti retroviral therapy (ART) for HIV positive patients in Gondar University Hospital, Ethiopia. *Bioinformation*. 2019;15(11):790–798. doi:10.6026/97320630015790

59. Kebede M, Zegeye DT, Zeleke BM. Predicting CD4 count changes among patients on antiretroviral treatment, Application of data mining techniques. *Comput Methods Programs Biomed*. 2017;152:149–157. doi:10.1016/j.cmpb.2017.09.017

60. Nemomsa G, Azath M. Designing a predictive model for antiretroviral regimen at the antiretroviral therapy center in Chiro Hospital, Ethiopia. *J Healthc Eng*. 2021;2021:1161923. doi:10.1155/2021/1161923

61. Sibanda W, Pretorius P. A review of applications of neural networks in the modeling of HIV epidemic. *Int J Comput Appl*. 2012;44:16.

62. Romero-Rodríguez D, Ramírez C, Imaz-Rosshandler I, et al. Machine learning-selected variables associated with CD4 T cell recovery under antiretroviral therapy in very advanced HIV infection. *Transl Med Commun*. 2020;5:1. doi:10.1186/s41231-020-00058-x

63. Federal minstry of health (FMOH). National guidelines for comprehensive HIV prevention, care and treatment; 2017.

64. World Health Organization. *Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring, Recommendations for a Public Health Approach, WHO, Editor*. Geneva: World Health Organization; 2021.

65. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn, machine learning in python. *J Mach Learn Res*. 2012;12:1.

66. Kulkarni A, Chong D, Batarseh FA. *5 - Foundations of Data Imbalance and Solutions for a Data Democracy, in Data Democracy*. Batarseh FA, Yang R, Editors. Academic Press; 2020:83–106.

67. Chawla N, Bowyer KW, Hall LO, et al. SMOTE, synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357. doi:10.1613/jair.953

68. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn, A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 2017;18(1):559–563.

69. Huang C, Li S-X, Caraballo C, et al. Performance metrics for the comparative analysis of clinical risk prediction models employing machine learning. *Circ Cardiovasc Qual Outcomes*. 2021;14(10):e007526. doi:10.1161/CIRCOUTCOMES.120.007526

70. Li B, Li M, Song Y, et al. Construction of machine learning models to predict changes in immune function using clinical monitoring indices in HIV/AIDS patients after 9.9-years of antiretroviral therapy in Yunnan, China. *Front Cell Infect Microbiol*. 2022;12:1.