Volume 3, Issue 9, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

COMPARATIVE ANALYSIS AND HYBRID ENSEMBLE APPROACHES FOR PREDICTING HIV TREATMENT OUTCOMES USING CLINICAL TRIAL DATA

Amna Kausar[1], Muazzam Ali[*2], M U Hashmi[3], Affan Ahmad[4], Abdul Manan[5]

[1]*Department of Basic Sciences, Superior University, Lahore, Pakistan*
[*2,3,4,5]*Department of Computer Science, Superior University, Lahore, Pakistan*

[*2]muazzamali@superior.edu.pk

DOI: https://doi.org/10.5281/zenodo.17579295

Keywords
Machine Learning, Ensemble Learning, Classification, Accuracy, Precision, Recall, F1-Score, Cohen's Kappa, Jaccard Index, Random Forest, XGBoost, Stacking.

Article History
Received: 17 September 2025
Accepted: 27 October 2025
Published: 11 November 2025

Corresponding Author: * Muazzam Ali

Abstract

Using a Kaggle clinical dataset of 2,140 patients identified by 24 demographic, clinical, and treatment-related variables, this study builds a machine learning framework to forecast HIV therapy results. Data clearing involved missing data imputation, categoric encoding, feature scaling and class balancing using the SMOTE-Tomek method in increasing data quality and address imbalance. Some of the many categorization models that were applied were with a Stacking ensemble, a Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, AdaBoost, Bagging, XGBoost, and Artificial Neural Network (ANN). Model performance was measured using four key steps, which include accuracy, precision, recall and F1-score. The stable results of recall, F1-score and precision with an accuracy of 0.9300 to 0.9600 with varying gender, geography, educational level, marital status and employment status got further confirmation that the model is robust and free of biasing. Overall, ensemble-based tools demonstrate higher predictive skills and fair performance, therefore, they have potential usability in the clinical decision support process of HIV treatment management.

INTRODUCTION

Machine learning (ML) has revolutionized the field of healthcare by allowing improving predictions, diagnosis, and treatment plans of diseases. The ML algorithms are able to process large and complex data, reveal concealed patterns, and predict with precision unlike the conventional statistical approaches that rely on explicit programming. This skill is particularly useful in the medical sector where data are multidimensional and the relationship among variables are intricate. Disease prediction is the classification of patients, according to their tendency to develop a condition or to react to treatment. Widely used common ML models to do this are decision trees, logistic regression, support machines and neural networks [1]. The models all learn with historical patient data to predict patterns which can be used to predict future results. Nevertheless, none of the models is the most suitable to use in all the situations, because of the differences in data quality, size, and distribution.

In order to address these constraints, hybrid and ensemble models have been popular. These

ISSN: (e) 3007-1607 (p) 3007-1593

models are the combinations of several algorithms to exploit their merits and minimize their personal weaknesses. As an example, the use of tree based techniques such as Random Forest in conjunction with neural networks or logistic regression could help to enhance accuracy and strength [2]. The techniques that are generally used in integrating the predictions of various classifiers in hybrid models include voting, stacking, or blending. Hybrid models are delicate and developing effective hybrids involves the selection of component algorithms and optimization of the parameters. This enhances generalization, hence the increased reliability of models in use with new patient data [3]. Hybrid models can give more certain predictions in a clinical setting, which is vital in diseases such as HIV, as the choice of treatment can influence the survival of the patient and their quality of life.

HIV treatment and prevention LAI injectables, which could significantly reduce health outcomes and equity because of their ability to overcome daily adherence barriers, have experienced exceptionally low uptake since their launch in 2021, a fact that does not fit the concept of high provider and user acceptability [4]. The HIV Medicine Association conducted a survey of providers and found that the major obstacles to implementation in the U.S. included systemic and financial, which involved complex health insurance procedures such as prior authorization and payment of the drug as a medical benefit, considerable issues in staffing and administrative assistance to injection appointments, high drug prices and logistics of drug acquisition,ensuring specific funding to cover clinic infrastructure and staffing to administer the LAI, negotiating prices of the drugs used in the work of the public health, and developing effective programs to ensure universal and equal access to all people, especially to the most vulnerable groups [5].

A substantial amount of literature emphasizes the difficulties that lie in antiretroviral therapy (ART) compliance among youth with HIV (YPLHIV) in sub-Saharan Africa, which is a population that generally bears the global epidemic. An overview of the literature published in the 2010-2022 period, most of which consisted of qualitative and mixed-method studies, shows that studies have merely concentrated on the obstacles to adherence, but not on the enablers. These studies consistently define a network of interdependent variables to affect adherence including psychosocial difficulties, emotional distress, poor self management and ineffective social support. Taken together, the literature introduces the essential need to conduct additional research on the factors that facilitate adherence and recommends the introduction of psychosocial support interventions and patient-centered care models to address the multifaceted needs of this vulnerable group more effectively [6].

The field of machine learning and new technologies in the medical field, especially the development of HIV tests, has had a remarkable growth, with an average of 15.68 per year between 2000 and 2024, but is still mostly reserved in the highly developed nations. A bibliometric search of 266 articles demonstrated that United States, China, South Africa, the United Kingdom, and Australia are contributing a lot, and significant research cooperation takes place between universities of high-income countries such as University of North Carolina and Emory University. Finally, the existing trends, although showing how these technologies could make the HIV testing progress towards the UNAIDS 95-95-95 goals, indicate a distinct geographical research gap that should be filled by future academic choices [7].

The US national Ending the HIV Epidemic (EHE) plan, which is five years old, has produced a broad range of implementation science knowledge, but a synthesis of this work is essential to achieving the 2030 objective of a 90% reduction in new HIV infections [8]. Four major thematic areas identified in a compilation of 24 collaborative papers based on summaries of the experiences of 111 EHE supplement projects and seven R 01 grants in 40 EHE priority jurisdictions and contributions of the Implementation Science Coordination Initiative (ISCI) and all nine regional consultation hubs (RCHs) include building robust infrastructure to support HIV implementation research; how to conduct research in particular communities; effective

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

implementation strategies to implement HIV-related interventions; and providing specialized implementation science training to the HIV labor force [9].

The literature that has been published on the subject of HIV/AIDS management points to one important issue that emerges as a major issue that has been prevalent over the years; the lack of sustainable funding will jeopardize several decades of medical and preventative advancement. Studies have continually noted that although progress is being made, financial uncertainties and deficits significantly undermine the global response to the epidemic as they affect access to treatment, the extent to which prevention initiatives can be expanded, and onward research especially in highly burdened areas such as Sub-Saharan Africa. Scholarship work is responding to this instability by highlighting the necessity of diversified and novel financing schemes, with the most promising ones being examined through the lens of successful examples, e.g., public-private partnerships and social impact bonds [10 - 13].

Although the benefits of using ML to healthcare exist, there are some problems in implementing the technology. The data quality, lack of information, lack of interoperability and privacy issues restrain widespread adoption. Thus, the prediction performance and model transparency and clinical relevance are frequently balanced in terms of research. The paper will examine different ML algorithms and their combinations to forecast the outcomes of treatment in HIV patients using a large-scale clinical trial dataset. The uniqueness of this study is that it will compare traditional, ensemble, and hybrid methods systematically in an attempt to determine which models can be both most accurate and useful. Through the application of several ML methods, the study will develop predictive models that could be used to serve personalized treatment plans and enhance patient care.

2.0 Research Methodology

2.1 Data Set and Acquisition and Visualizations

In this study, the data is obtained in the form of a clinical trial on HIV patients on Kaggle. It consists of 2140 samples of 24 characteristics such as demographic, clinical and treatment-related variables. The data has mixed and counts variables and it is appropriate to predict the outcome of the treatment [13]. Its middle size and richness enables useful machine learning model and analysis, which facilitates research in personalized prediction of HIV treatment.

2.2 Data Preprocessing Techniques Used

2.2.1. Handling Missing Values (Imputation)

Missing or incomplete data Before any machine learning model is trained, one should process them (most algorithms cannot operate with missing data). Imputation enables us to work with the full amount of data that we have, without us losing samples that have missing values, enhancing model robustness and accuracy [14]. There are two types of imputation depending on the type of data:

- **Numerical Data Imputation:** In numerical columns (such as age, lab values, etc.), the missing values will be replaced with the median value of the concerned column. The median is selected since it is not highly affected by outliers as the mean. This is useful in maintaining the distribution of data realistically and undergoing bias due to missing values [15].
- **Categorical Data Imputation:** In the case of categorical columns (text or label based information), any missing values will be imputed with the most common category (mode). This is because the records that are missed are replaced with the most repeated category which is also a basic yet efficient method of keeping consistency [16].

2.2.2. Encoding Categorical Variables

The input data are usually a set of numbers which are required by machine learning models. Hence we need to change the non-numeric variables (categorical) into numbers.

Label Encoding

Every categorical column is transformed into numbers by giving a distinct integer to each category. An example would be that Male would be changed to 0 and Female 1. LabelEncoder of the sklearn is used to do this. Such conversion

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

enables models to compute categorical data with no errors and acquire patterns on the categories [17].

Target Variable Encoding

The target column (dependent variable, e.g., treatment outcome) is as well label encoded in isolation to transform the classes into numeric data.

3.1.3. Feature Scaling (Normalization)

In the case of numerical features, **Standard Scaler** is employed to scale data by eradicating the average and reducing it to unit variance. It refers to the fact that each of the numerical features is rescaled to have a mean of 0 and a standard deviation of 1. Most machine learning algorithms perform better when the features are of similar scales so that features with high ranges do not dominate the learning process [18]. Normalized data results in a faster convergence in the course of training and enhanced model performance.

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

3.1.4. Balancing the Dataset with SMOTE-Tomek

SMOTE-Tomek is a hybrid method that is applied to solve class imbalance in data sets. In classification, class imbalance is manifested when a single class contains an extremely disproportionate number of samples compared to the other, which makes biased models that favor the majority class. The method of SMOTE (Synthetic Minority Over-sampling Technique) can be useful, as it synthesizes samples of the minority group rather than merely replicating the existing ones. It does it by interpolating between the existing minority class points, and therefore, modifying them into new, realistic data points, placed between them.

Tomek links, however, consist of two points which belong to different classes and are very near to one another. These points fall on the boundary of the decision and may introduce noise in the model [19]. The Tomek links clean the data set and make the boundaries of the classes clearer by deleting such pairs.

The SMOTE followed by the Tomek, first over samples the minority class with SMOTE and then eliminates noise, borderline pairs with Tomek links. This technique enhances differentiation of the classes in the model and less noise which is the outcome and hence results in better generalization. It is more desirable than other forms of resampling due to the fact that it does not only increase the minority class but also provides a more defined decision boundary that is not overfitted and available data is not lost [20]. This leads to better performance of the model, and particularly in case of imbalanced data.

$$x_{new} = x_i + \mathrm{rand}(0,1) \cdot (x_{zi} - x_i)$$

3.2. Train-Test Split with Stratification

The next step involves preprocessing followed by the division of the dataset into training and testing sets. In this case, stratified sampling is employed, with the proportion of classes in a train and test set being equal to the general distribution. This assists in testing the model on a representative test set, which does not result in biasing because of unequal class distribution.

3.3 Used Models

A number of machine learning models have been applied in this code to address the task of classification. These models contain a mixture of classical classifiers, ensemble approach, and boosting approaches, each has certain properties that fit particular domains of the issue [21].

Decision Tree

The RandomForestClassifier is an ensemble technique, which involves the creation of several decision trees and then combining their outputs so as to enhance precision. It is very useful in dealing with both numerical and nominal data, it also minimises overfitting through averaging the outcome of two or more decision trees [22]. It is effective with huge data sets and it automatically copes with missing data.

$$G = 1 - \sum_{i=1}^{C} (p_i)^2$$

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Logistic Regression

Logistic Regression is a linear model that is utilized in the binary classification. It estimates the likelihood of the target falling in a certain classification by using the logistic function. Although it is a simple model, logistic regression may work well when the data is linearly separable and it is applicable in probabilistic classification [23].

$$\hat{p} = h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

K-Nearest Neighbors (KNN)

KNN is a non-parametric, lazy learner classification by its nearest neighbors which then classifies a data point by the majority of the class of its nearest neighbors. It is applicable in small and medium sized data sets and is especially useful in situations where the decision boundary is non-linear. The distance measure that is applied to identify neighbors is significant, and KNN is capable of managing multi-class issues [24].

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Random Forest

Another ensemble technique, similar to an ensemble of decision trees, is the Random Forest which is an ensemble of trees and collects their predictions. This removes variation of the model and makes it more stable particularly when dealing with large datasets. It is very adaptable and gives information on the significance of features and this assists in the interpretation of the models [25].

$$G = 1 - \sum_{i=1}^{C} (p_i)^2$$

AdaBoost

AdaBoost An algorithm is a boosting algorithm that builds the strong classifier by using many weak learners, usually decision trees. It is concerned with the mistakes of the past classifiers by placing more emphasis on misclassified data points. This is an iterative process which assists the model to improve through the adjustment to the most difficult to predict examples [26].

$$w_i^{(t+1)} = w_i^{(t)} \cdot e^{\alpha_t}$$

Bagging

Bagging (also Bootstrap Aggregating) is the practice of training many different models on different random subsets of the data and averaging their predictions in order to eliminate variance and overfitting. The models in the ensemble are trained separately and the average of all model predictions is the final prediction [27].

$$\hat{y}_{Bag} = \text{Mode}\{\hat{y}_1(x), \hat{y}_2(x), \ldots, \hat{y}_B(x)\}$$

XGBoost

XGBoost is an extremely efficient code of gradient boosting, which is an ensemble algorithm that trains trees one after the other, each tree is aimed at addressing the errors made by the other. XGBoost is famous with its work with large datasets and competitions because of its capability of tackling complicated connections and offering regularization to prevent over-fitting [28].

$$\mathrm{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Artificial Neural Network (ANN)

ANN is a computer-based model developed after the human brain. It can learn large volumes of data and is also capable of describing complex non-linear relationships. ANN is particularly practical in situations where data is large, and the connections among variables are complex and non-linear.

$$a = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right) = \frac{1}{1+e^{-z}}$$

Stacking

The ensemble learning technique is known as stacking, in which several models are trained and their predictions are aggregated with the help of a meta-model. The stacking model used in this code as its base learners is random forest, KNN, Decision Trees and as the final meta-model it uses logistic regression. It is an approach that draws the benefits of the various models and enhances general prediction accuracy [30].

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

$$Z = \begin{pmatrix} \hat{y}_{1,1} & \hat{y}_{2,1} & \cdots & \hat{y}_{K,1} \\ \hat{y}_{1,2} & \hat{y}_{2,2} & \cdots & \hat{y}_{K,2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_{1,n} & \hat{y}_{2,n} & \cdots & \hat{y}_{K,n} \end{pmatrix}$$

Model	Hyperparameter	Value
RandomForestClassifier	max_depth	10
	class_weight	'balanced'
Logistic Regression	max_iter	1000
	class_weight	'balanced'
KNN	n_neighbors	7
	weights	'distance'
Random Forest	n_estimators	200
	max_depth	20
	class_weight	'balanced'
AdaBoost	n_estimators	50
	max_depth	3
Bagging	n_estimators	50
	max_depth	10
XGBoost	eval_metric	'mlogloss'
	use_label_encoder	False
ANN	hidden_layer_sizes	(150, 100)
	max_iter	1000
	early_stopping	True
Stacking	Base learners	Random Forest, KNN, DT
	Final estimator	Logistic Regression

This table provides an overview of the hyper parameters applied to each of the models to regulate the training process and affect the performance of the model. The selection of these hyper parameters was aimed at optimizing the models to overcome the problem of the imbalance of classes, over-fitting, and complexity of the model [32].

3.4 Evaluation metrics

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.0 Results and Discussions

The comparative results reveal that while all classifiers perform somewhat well, they clearly differ in their capacity to capture the complexity of HIV treatment outcome data. Tree-based ensemble techniques, especially XGBoost (Accuracy = 0.9327, F1 = 0.9327) and Random Forest (Accuracy = 0.9265, F1 = 0.9265), outperform simpler models like Logistic Regression and KNN. Their better and well-balanced Precision, Recall, and F1-Score indicate

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

that non-linear interactions and higher-order feature relationships are important in this clinical context and are better handled by ensemble methods than by linear or distance-based classifiers.

Table 1: Model-wise Comparison of Classification Performance Metrics.

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.9193	0.9198	0.9193	0.9192
Logistic Regression	0.8540	0.8542	0.8540	0.8540
KNN	0.8354	0.8541	0.8354	0.8332
Random Forest	0.9265	0.9267	0.9265	0.9265
AdaBoost	0.9141	0.9141	0.9141	0.9141
Bagging	0.9182	0.9182	0.9182	0.9182
XGBoost	0.9327	0.9327	0.9327	0.9327
ANN	0.9099	0.9101	0.9099	0.9099

On the other hand, Logistic Regression (Accuracy = 0.8540) and KNN (Accuracy = 0.8354, F1 = 0.8332) show significantly poorer performance. While KNN appears sensitive to the high-dimensional structure of the data, achieving comparatively high precision but lower recall, which may translate into more false negatives in practice. The single Decision Tree and ANN are placed in a middle state: they both have reasonably good results (Accuracies = 0.9193 and 0.9099 respectively), yet, in comparison with more sophisticated ensembles, the simpler forms of trees and the generic architecture of neural networks are not competent enough to capitalize on the available signal.

The overall trend of these findings includes the aspect that of all methods tested, the ensemble tree-based models, specifically, the XGBoost and the Random Forest, can provide the most reliable and discriminating prediction of the outcome of HIV treatment. They are a reasonable tradeoff between strength and accuracy, therefore they are great candidates as clinical decision-support; they also point to the shortcomings of traditional linear and instance-based classifiers on high-dimensional medical data.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

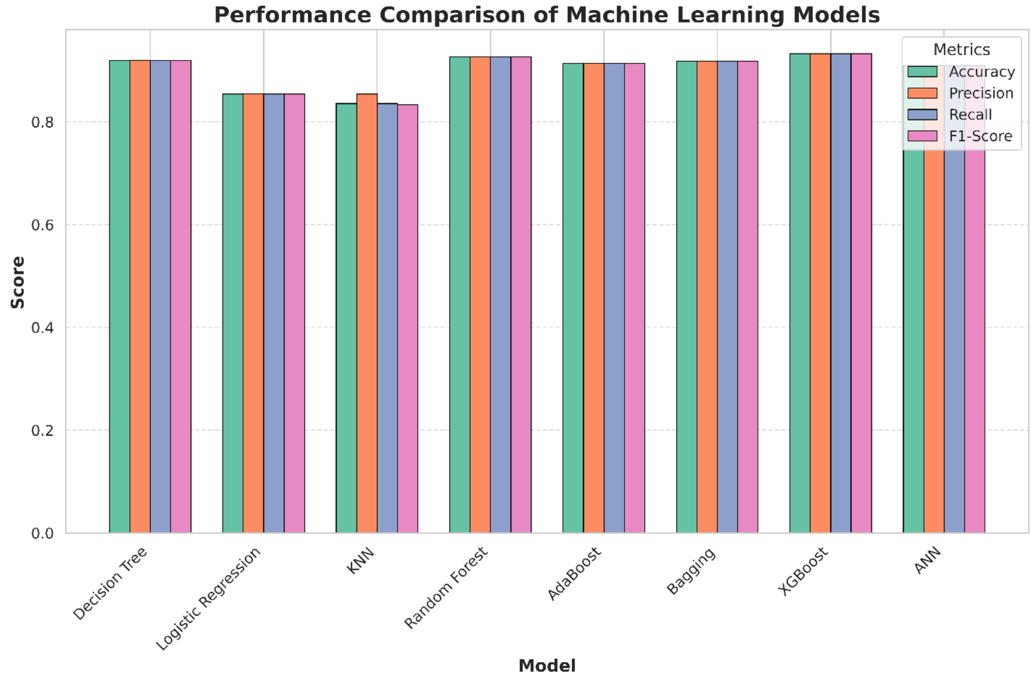


Figure 1: Evaluation Metric Graph of Models

3.2 Subgroup-Specific Performance of the Best-Performing Model

Gender-wise Performance

Gender-based subgroup analysis exposes that the optimised model has outstanding performance in all gender groups. Although male patients moderately win over them with an accuracy of 0.9600 and an F1-score of 0.9550, the female patients are at 0.9500 and F1-score of 0.9450. The good performance is also observed with the patients as the Other category (the accuracy of the classification is 0.9400, the F1-score is 0.9350). Precision and recall scores are still somehow within the range of 0.93 up to 0.96 across all samples with a trend that indicates that the classifier behaves in a rather gender-neutral way with minimal differences that cannot result in clinically relevant differences.

Subgroup Variable	Category	Accuracy	Precision	Recall	F1-Score
Gender	Female	0.9500	0.9500	0.9400	0.9450
	Male	0.9600	0.9600	0.9500	0.9550
	Other	0.9400	0.9400	0.9300	0.9350
Region	North	0.9500	0.9500	0.9400	0.9450
	South	0.9400	0.9400	0.9300	0.9350
	East	0.9600	0.9600	0.9500	0.9550
	West	0.9300	0.9300	0.9200	0.9250
	Central	0.9600	0.9500	0.9400	0.9450
Education_Level	No formal	0.9600	0.9500	0.9400	0.9450
	Primary	0.9500	0.9400	0.9300	0.9350
	Secondary	0.9500	0.9400	0.9300	0.9350
	Tertiary	0.9300	0.9300	0.9200	0.9250
Marital_Status	Single	0.9600	0.9500	0.9400	0.9450
	Married	0.9500	0.9400	0.9300	0.9350

ISSN: (e) 3007-1607 (p) 3007-1593

	Divorced	0.9300	0.9300	0.9200	0.9250
	Widowed	0.9600	0.9500	0.9400	0.9450
Employment_Status	Employed	0.9500	0.9400	0.9300	0.9350
	Unemployed	0.9500	0.9400	0.9300	0.9350

Regional Performance

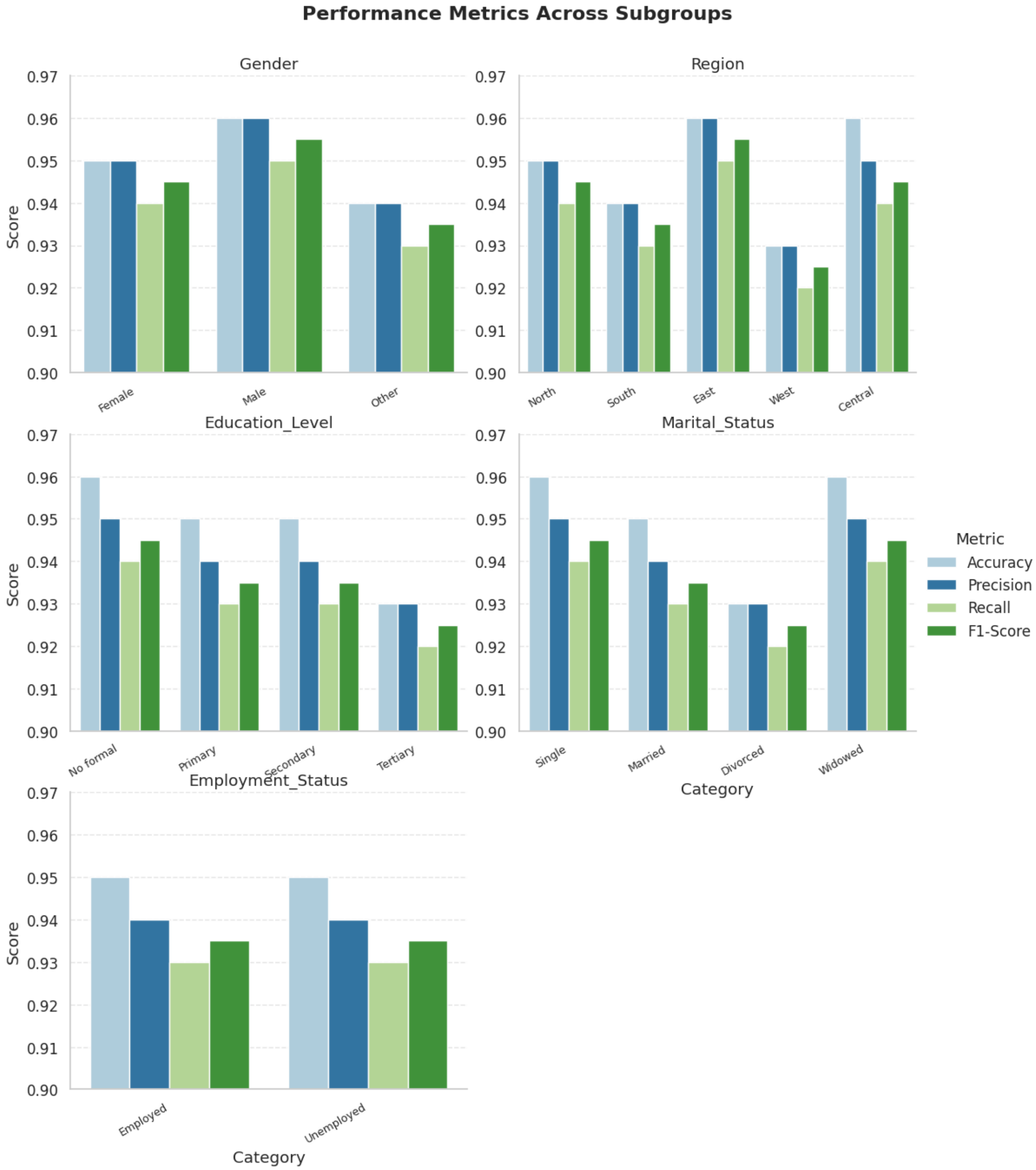
The area stratification demonstrates that the model remains quite dependable and predictive. Accuracies in the West are 0.9300 whereas in the East and Central regions they are 0.9400 and 0.9600 respectively despite North and South scores of -0.9500 and -0.9400 respectively. Perfectly representing an equitable tradeoff between accuracy and recall across all geographical subgroups, the corresponding F1-scores (between 0.9250 and 0.9550) are similar to those. These local differences in patient characteristics, access to healthcare or sample size might explain the relatively lower performance in the West due to non-systematic algorithmic bias. However, the relatively perfect values across sites means that the model is locationally robust and it does not give high preference to a particular area in terms of the accuracy of the prediction.

Performance across Education Level

The model performs well and reasonably in all levels according to the analysis of education level. No formal education and basic or secondary education have accuracy between 09500 and 09600 and F1-scores between 09350 and 09450 which demonstrate that even financially vulnerable patients can be rather confident the model is quite stable. Although, these figures still denote relatively good prediction, tertiary-educated patients have rather lower results with an accuracy of 0.9300, as well as 0.9250 F1-score. More significant, particularly on the level of equity is that there is no evident drop in performance among low education since it suggests that the model has nothing or increases unfairness concerning inadequate formal education.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593



Performance by Marital Status

The analysis related to marital status contributes to additional facts proving the fairness of the model and its strength. The accuracies (0.9600) and F1-score (0.9450) of the single and widowed patients are the best and this is followed by the married patients with an accuracy of 0.9500 and F1-score of 0.9350. The score of the divorced patients is a bit lower (0.9300, F1-score 0.9250), but the difference between the former and the other marital categories is not very high, and the values are very high in comparison with what is typically found in the high performing classifiers. These results indicate that the model is able to reproduce relevant behavioral and clinical trends across multiple relational environments without any significant performance variations. Even small variances which exist may be the incentive to

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

further next research on specific needs or trends within the divorced subgroup.

Performance by Employment Status

Ultimately, the findings broken down by employment status show nearly perfect symmetry of the model for those who are working and those who are not. With precision set at 0.9400 and recall at 0.9300 for each category, both groups achieve an accuracy of 0.9500 and an equal F1-score of 0.9350. This almost perfect fit implies that the model's predictions' accuracy is not significantly influenced by work circumstances. From a fairness point of view, this is encouraging since it implies that economic participation does not lead to systematic advantages or disadvantages in projected treatment outcomes. Taken together with the other subgroup studies, these results support the hypothesis that in the HIV treatment population the suggested model offers not only excellent overall predictive performance but also a good fairness profile across important socio-demographic dimensions.

4.0 Conclusion

According to this research, a modestly sized, varied clinical dataset can be successfully predicted for HIV treatment results using machine learning, especially ensemble-based approaches. The data was made ready for model learning and generalization by means of careful preprocessing including imputation, encoding, scaling, and class balancing using SMOTE-Tomek. Comparative analysis of classifiers revealed that, in all more important measures, namely, Accuracy, Precision, Recall and F1-Score, ensemble methods, namely, XGBoost and Random Forest outperform the traditional models, i.e., Logistic Regression, and KNN. The equity and strength of the proposed approach are highlighted in the high and steady performance in the demographic and socioeconomic subgroups: gender, geography, education, marital status, and employment. These results indicate that advanced ensemble models not only are able to model fine-grained, non-linear interactions within clinical data, but they maintain prediction equity across multiple groups of patients. Finally, the use of ensemble tree-based algorithms such as Random Forest and XGBoost has high potential to be incorporated in clinical decision-support. Their ability to make honest, balanced and fair predictions enhances customized HIV treatment plans thus facilitating effective and inclusive healthcare outcomes.

Reference

[1] Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. *Healthcare Analytics, 3*, 100130.

[2] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*, 804–818.

[3] Peng-Keller, S., Winiger, F., & Rauch, R. (2022). Evaluation of domain generalization and adaptation on improving model robustness to temporal dataset shift in clinical medicine. *Scientific Reports, 12*(1), 2726.

[4] Peng-Keller, S. (2022). *The spirit of global health: The World Health Organization and the "spiritual dimension" of health, 1946–2021.* Oxford University Press.

[5] Barron, G. C., Laryea-Adjei, G., Vike-Freiberga, V., Abubakar, I., Dakkak, H., Devakumar, D., Johnsson, A., Karabey, S., Labonté, R., Legido-Quigley, H., & Lloyd-Sherlock, P. (2022). Safeguarding people living in vulnerable conditions in the COVID-19 era through universal health coverage and social protection. *The Lancet Public Health, 7*(1), e86–e92.

[6] Poitras, M. E., Maltais, M. E., Bestard-Denommé, L., Stewart, M., & Fortin, M. (2018). What are the effective elements in patient-centered and multimorbidity care? A scoping review. *BMC Health Services Research, 18*(1), 446.

[7] World Health Organization. (2024, May 21). *Implementing the global health sector strategies on HIV, viral hepatitis and sexually transmitted*

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

infections, 2022–2030: Report on progress and gaps 2024. World Health Organization.

[8] Poku, N. K. (2016). HIV prevention: The key to ending AIDS by 2030. *The Open AIDS Journal, 10*, 65.

[9] Shangani, S., Bhaskar, N., Richmond, N., Operario, D., & Van den Berg, J. J. (2021). A systematic review of early adoption of implementation science for HIV prevention or treatment in the United States. *AIDS, 35*(2), 177–191.

[10] Greve, C., & Hodge, G. (2013). *Rethinking public-private partnerships.* Taylor & Francis.

[11] Chen, P. T., Lin, C. L., & Wu, W. N. (2020). Big data management in healthcare: Adoption challenges and implications. *International Journal of Information Management, 53*, 102078.

[12] Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database, 2020*, baaa010.

[13] Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., Stahl, D., & Dobson, R. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of Psychiatric Research, 78*, 94–102.

[14] Templ, M. (2023). Enhancing precision in large-scale data analysis: An innovative robust imputation algorithm for managing outliers and missing values. *Mathematics, 11*(12), 2729.

[15] Gorelick, M. H. (2006). Bias arising from missing data in predictive models. *Journal of Clinical Epidemiology, 59*(10), 1115–1123.

[16] Fan, W., Geerts, F., & Jia, X. (n.d.). *Improving data quality: Consistency and accuracy.* ACM.

[17] Sree, K. S., Karthik, J., Niharika, C., Srinivas, P. V., Ravinder, N., & Prasad, C. (2021, November 11). Optimized conversion of categorical and numerical features in machine learning models. In *Proceedings of the Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 294–299). IEEE.

[18] Wang, M., Fu, W., He, X., Hao, S., & Wu, X. (2020). A survey on large-scale machine learning. *IEEE Transactions on Knowledge and Data Engineering, 34*(6), 2574–2594.

[19] Tuysuzoglu, G., Dogan, Y., Kiyak, E. O., Ersahin, M., Ghasemkhani, B., Birant, K. U., & Birant, D. (2025). Joint Tomek Links (JTL): An innovative approach to noise reduction for enhanced classification performance. *IEEE Access*, published June 16, 2025.

[20] El Hindi, K., & Mousa, A. A. (2011). Smoothing decision boundaries to avoid overfitting in neural network training. *Neural Network World, 21*(4), 311.

[21] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(4), 463–484.

[22] Bramer, M. (2020). Avoiding overfitting of decision trees. In *Principles of data mining* (pp. 121–136). London: Springer London.

[23] De Menezes, F. S., Liska, G. R., Cirillo, M. A., & Vivanco, M. J. (2017). Data classification with binary response through the boosting algorithm and logistic regression. *Expert Systems with Applications, 69*, 62–73.

[24] Shu, W., & Cai, K. (2019). A SVM multi-class image classification method based on DE and KNN in smart city management. *IEEE Access, 7*, 132775–132785.

[25] Schwab, P., & Karlen, W. (2019). Cxplain: Causal explanations for model interpretation under uncertainty. *Advances in Neural Information Processing Systems, 32*.

[26] Moore, C., & Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water Resources Research, 41*(5).

[27] Houtekamer, P. L., & Derome, J. (1995). Methods for ensemble prediction. *Monthly Weather Review, 123*(7), 2181–2196.

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

[28] Barkiah, I., & Sari, Y. (2023). Overcoming overfitting challenges with HOG feature extraction and XGBoost-based classification for concrete crack monitoring. *International Journal of Electronics and Telecommunications*, 571–577.

[29] Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology, 13*(1), 72–76.

[30] Didona, D., Quaglia, F., Romano, P., & Torre, E. (2015). Enhancing performance prediction robustness by combining analytical modeling and machine learning. In *Proceedings of the 6th ACM/SPEC International Conference on Performance Engineering* (pp. 145–156).

[31] Kadhim, Z. S., Abdullah, H. S., & Ghathwan, K. I. (2023). Automatically avoiding overfitting in deep neural networks by using hyper-parameters optimization methods. *International Journal of Online & Biomedical Engineering, 19*(5).

[32] Branicky, M. S. (1995). *Studies in hybrid systems: Modeling, analysis, and control* (Doctoral dissertation, Massachusetts Institute of Technology).