Machine Learning Algorithms for HIV/AIDS Prediction using Explainable AI

Nicholas Ho
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
nicholas.ho@binus.ac.id

Anthony Davidson Salim
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
anthony.salim@binus.ac.id

Geoffrey Gohtama
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
geoffrey.gohtama@binus.ac.id

Henry Lucky
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
henry.lucky@binus.ac.id

Derwin Suhartono
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta, Indonesia
dsuhartono@binus.edu

Abstract—Early detection of HIV/AIDS is important in order to reduce transmission rate. By evaluating different machine learning algorithms, this research identifies the best algorithm or model for HIV/AIDS detection based on socio-behavioral data. Even though there are previous research that has shown good results, they still have some challenges to be solved. This study addresses the problems of model interpretability by using an Explainable AI library called SHAP in order to improve model interpretability and transparency. The methodology includes data collection from a public dataset, preprocessing, model training, evaluation using standard metrics (accuracy, precision, recall, F1-score), and interpretability analysis using SHAP. The results show that ensemble models, particularly Random Forest, achieve the highest accuracy at 0.97, outperforming traditional models. SHAP analysis highlights key features such as partner-seeking behavior and drug use as significant predictors. The study concludes that ensemble learning combined with explainability offers a reliable and interpretable approach for HIV/AIDS prediction, thus can help improve the overall result in a medical setting.

Keywords—machine learning, HIV/AIDS, early detection, explainable AI, healthcare

I. INTRODUCTION

Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) are still recognized as crucial global health challenges, affecting millions of people worldwide [1][2]. HIV/AIDS is an infectious disease and the primary modes of HIV transmission are unprotected sexual intercourse, vertical transmission from mother to child during pregnancy, exposure to contaminated hypodermic needles, and transfusions involving infected blood [3]. Despite significant advancements in medical research, HIV/AIDS remains as a chronic disease that affects a lot of people globally. One of the current treatments for HIV/AIDS patients is called antiretroviral therapy (ART) and it has proven to enhance the patient's life expectancy [4][5].

As a result of no definitive cure discovered to treat HIV/AIDS, early detection of this disease is very important. Early detection of this disease is important for proper diagnosis and infection management, ensuring that patients will be treated with a suitable and appropriate treatment. In addition, as HIV/AIDS can be transmitted through several media, early detection can contribute to prevent the transmission of HIV/AIDS disease [6].

Following the advancement of technology, the healthcare sector starts to utilize artificial intelligence, specifically machine learning in order to help doctors to improve predictive accuracy, diagnostic efficiency, and clinical decision-making capabilities. In addition, machine learning algorithms can help doctors to identify early detection of a disease [7]. Numerous studies utilize machine learning in order to predict or detect disease, starting from common disease to chronic disease, such as influenza [8], diabetes [9], heart disease [10], and cancer [11]. These applications highlight its growing importance in healthcare, particularly for disease prediction and detection [12]. Moreover, machine learning continues to expand toward tackling more complex conditions, such as HIV/AIDS.

Recent studies have successfully applied machine learning to predict HIV/AIDS, including research focused on detecting HIV infection among men who have sex with men (MSM) [13][14], such as a Random Forest model accurately predicting HIV among high-risk groups in Zhejiang, China [13], and a Decision Tree model aiding in adolescent HIV testing predictions in Ethiopia [15]. Another study explored HIV risk prediction with Explainable AI (XAI) [16], but only discussed theoretical roles and potential benefits without developing models or reporting experimental results. To address the challenge of model interpretability, XAI methods like SHAP (SHapley Additive exPlanations) are used to

improve transparency by showing feature contributions to predictions, making results more interpretable for medical professionals; although other methods like LIME exist, SHAP is favored for its consistency and strong theoretical foundation. Despite these advances, interpretability remains a challenge and a key area for further research. Therefore, this study aims to compare multiple machine learning algorithms for HIV/AIDS prediction using socio-behavioral data, identify the most accurate model, and enhance transparency with SHAP to evaluate interpretability and highlight influential predictors, ultimately supporting more informed decision-making in clinical and public health settings.

II. RELATED WORKS

In the last few years, there has been a lot of interest in the use of artificial intelligence in healthcare, especially in the area of machine learning for disease detection or prediction. Specifically, machine learning has become increasingly important for improving predictive accuracy, diagnostic efficiency, and decision-making capabilities in complex healthcare scenarios, such as HIV/AIDS detection. Several studies have explored how machine learning can enhance healthcare outcomes compared to conventional statistical methods. For example, a research using machine learning models to analyze data from multiple surveys demonstrated that these models outperformed traditional statistical approaches in HIV/AIDS risk prediction [17][18].

There have been multiple studies on high-risk populations, including men who have sex with men in China [13][14] and adolescents in Ethiopia [15], that tested various classifiers such as Logistic Regression, SVM, Decision Trees, and Random Forest. From these studies, tree-based ensemble methods, particularly Random Forest and Gradient Boosting, consistently achieved the highest accuracy, while traditional models performed less effectively. Similar findings can be seen in socio-behavioral risk prediction studies in East and Southern Africa [20] and clinical classification tasks [19], emphasizing the strength of ensemble methods for HIV/AIDS prediction.

Moving to the integration of Explainable AI, there is a paper that explores the role of machine learning in HIV risk prediction with the integration of Explainable AI [16]. This paper focused on providing insights about the theoretical role, potential benefits of machine learning in HIV risk prediction, and the importance of implementing Explainable AI, such as SHAP to enhance the model interpretability. The gap of this paper is that it does not specifically conduct model development and experiments to predict or detect HIV/AIDS.

These existing studies collectively showed the growing role of machine learning in HIV/AIDS prediction. They demonstrated the effectiveness of different models in various contexts, from clinical diagnosis to behavioral risk assessment. However, despite the promising results, challenges such as model interpretability remain areas for further research. Addressing these problems could be important in advancing the real-world application of machine learning for HIV/AIDS detection and prevention.

In order to address these gaps, future experiments need to focus on comparing different machine learning models for HIV/AIDS detection or prediction with the integration of Explainable AI, such as SHAP to evaluate and enhance model interpretability.

III. METHODOLOGY

This study aims to compare the performance of different machine learning models (Logistic Regression, Naive Bayes, Decision Tree, Random Forest, and Support Vector Machine) in order to detect HIV/AIDS infection. This study also integrates Explainable AI, such as SHAP to enhance the model interpretability.
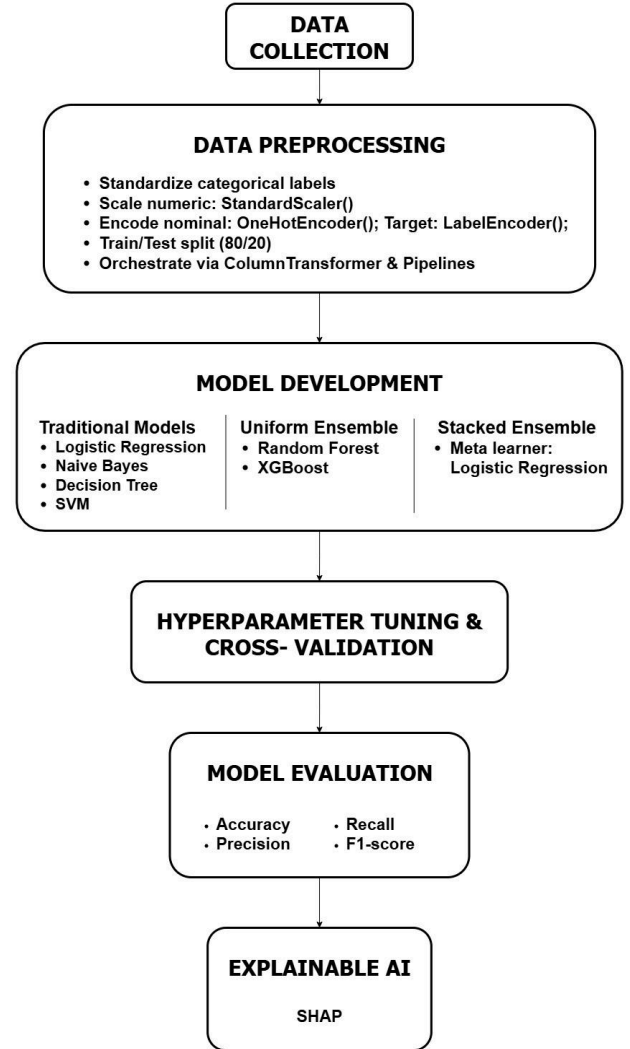

Fig.1. Proposed Methodology

A. Data Collection

The dataset used for this study is obtained from Kaggle [21]. This dataset contains 698 samples. Each data contains 10 features which are:

- **Age**: Numerical representation of the patient's age.

- **Marital Status:** Whether the patient is married, unmarried, divorced, etc.

- **STD:** The patient's history of sexually transmitted diseases (Yes/No).

- **Educational Background:** The patient's highest level of education.

- **HIV Test in Past Year:** Whether the patient took an HIV test in the last year (Yes/No).

- **AIDS Education:** Whether the patient has received education about AIDS (Yes/No).

- **Places of Seeking Sex Partners:** Locations where the patient seeks sexual partners.

- **Sexual Orientation:** The patient's sexual preference (e.g., heterosexual, homosexual, bisexual, etc).

- **Drug-Taking:** Whether the patient has a history of drug use (Yes/No).

- **Result:** The HIV/AIDS status (Positive/Negative).

This dataset is suitable for this research because the dataset provides relevant patients' socio-behavioral features that are commonly linked with HIV/AIDS. As a result, the machine learning models will be able to use these to accurately find the pattern and correlations to HIV/AIDS.

B. Data Preprocessing

The dataset underwent several preprocessing steps to ensure data quality and prepare it for model training. First, inconsistencies in categorical values were corrected by standardizing labels and fixing typographical errors across features. Next, the dataset was split, with an 80:20 train–test ratio. This ratio was chosen to have an adequate amount of training set for model learning while keeping enough test data to provide a reliable and unbiased evaluation. Stratification was not required because the dataset had a balanced target distribution. Then, numerical attributes were scaled using StandardScaler to normalize feature ranges. Nominal categorical features were encoded using OneHotEncoder to avoid introducing artificial ordinal relationships. For target encoding, LabelEncoder was applied to transform the result into binary labels. These transformations were organized into pipelines using ColumnTransformer, allowing consistent preprocessing during both training and testing. This structured approach ensured the dataset was clean, consistent, and in a standardized numerical format suitable for both traditional and ensemble machine learning models.

C. Hyperparameter Tuning and Cross-Validation

To optimize model performance, we applied hyperparameter tuning using GridSearchCV from *scikit-learn*, which performs a search over specified parameter grids combined with 5-fold cross-validation. For example, for Random Forest we tuned the number of estimators, maximum tree depth, and minimum samples per split, while for XGBoost we adjusted the learning rate, maximum depth, and subsampling ratio. This approach ensured that the selected parameters generalize well across unseen data, reducing the risk of overfitting. The choice of 5-fold cross-validation provided a balance between computational efficiency and reliable performance estimation.

This is how each model was hyperparameter tuned:

1. **Logistic Regression:** Tuned over C (0.01, 0.1, 1, 10), penalty (l2), and solver (lbfgs).

2. **Naive Bayes (GaussianNB):** Tuned var_smoothing (1e-9, 1e-8, 1e-7).

3. **Decision Tree:** Tuned max_depth (None, 10, 20),

min_samples_split (2, 5), and min_samples_leaf (1, 2).

4. **Support Vector Machine (SVM):** Tuned C (0.1, 1, 10), kernel (rbf, linear), and gamma (scale, auto).

5. **Random Forest:** Tuned n_estimators (100, 200), max_depth (None, 10, 20), min_samples_split (2, 5), and min_samples_leaf (1, 2).

6. **XGBoost:** Tuned n_estimators (100, 200), max_depth (3, 6, 10), learning_rate (0.01, 0.1, 0.2), and subsample (0.8, 1.0).

7. **Stacked Ensemble:** Tuned final estimator LogisticRegression parameters (C [0.01, 0.1, 1, 10], penalty [l2]) and passthrough (True, False).

In this study, a stacked ensemble was constructed by using the best-performing base models (Logistic Regression, Naive Bayes, Decision Tree, and SVM) as inputs with Logistic Regression as the meta-learner, enabling the model to combine the strengths of different algorithms and achieve better overall generalization.

D. Model Evaluation

The performance of machine learning models could be evaluated using a set of evaluation or performance metrics. The typical evaluation metrics that are used for evaluating the performance of classification models are confusion matrix, accuracy, precision, recall, and f1-score.

These evaluation metrics were chosen because they provide a comprehensive assessment of classification performance. The confusion matrix allows visualization of true vs. false predictions, while accuracy gives an overall measure of correctness. Since accuracy alone can be misleading in healthcare applications, precision and recall are included to balance the concern of false positives and false negatives, and the F1-score is used to combine both into a single, balanced metric.

E. Explainable AI Integration

This study will apply an Explainable AI technique called SHAP (SHapley Additive exPlanations) to see how the machine learning algorithms predict HIV/AIDS. By doing so, it will provide transparency and give reasoning behind the results. SHAP will measure how much a feature contributes to the overall result.

The result from using SHAP will then be used to determine which one plays the most significant role. These visuals will also help understand how different features affect the model's decisions and how they interact with each other. This will be important in supporting the research by helping medical professionals recognize which are the most significant clinical indicators for HIV/AIDS diagnosis.

IV. RESULT AND DISCUSSION

A. Performance Evaluation

In order to compare the machine learning models in predicting HIV status, we tested three groups of models, which are Base Models, Uniform Ensemble, and Stacked Ensemble. The predictions were evaluated using

classification reports on the test set.

TABLE I. Result of ML Model

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.82	0.93	0.72	0.81
Gaussian Naive Bayes	0.76	0.79	0.75	0.77
Decision Tree	0.94	0.91	0.97	0.94
SVM	0.94	0.91	0.97	0.94
Random Forest	0.97	0.96	0.99	0.97
XGBoost	0.96	0.94	1.00	0.97
Stacked Ensemble	0.95	0.94	0.97	0.95

1. Base Models

For the base models, four traditional machine learning models were trained using a pipeline and hyperparameter tuning.

The Logistic Regression model achieved 82% accuracy. It had strong recall (0.94) for negative cases. However, the recall (0.72) for positives was lower. This indicates it might underpredict positive cases.

Gaussian Naive Bayes produced 76% accuracy, which is the lowest among the base models. This can be caused due to the strong assumption of feature independence, which rarely holds true in real-world health data. Even though it had reasonably balanced performance, it struggled to match the precision and recall of the other classifiers.

Decision Tree scored 94% accuracy. This model got high F1-scores for both classes (0.93 for negative, 0.94 for positive). However, this model has a higher tendency to overfit. This means that while the model performs well on the training and test sets, it might generalize poorly to new data.

Support Vector Machine (SVM) also reached 94% accuracy. The results were identical to the Decision Tree. The SVM used an RBF (Radial Basis Function) kernel, which is effective in handling non-linear relationships between features. Compared to Decision Trees, SVMs are less prone to overfitting and often generalize better.

2. Ensemble Learning

Two uniform ensemble models were trained using hyperparameter tuning. The Random Forest model reached the highest individual model performance at 97% accuracy, with excellent precision and recall across both classes. XGBoost got slightly worse results than Random Forest, with 96% accuracy. However, it achieved perfect precision (1.00) for class 0 and perfect recall (1.00)

for class 1.

Compared to other models, their ability to model complex, non-linear relationships and interaction between socio-behavioral features are superior compared to traditional models. Both algorithms are tree-based ensembles that evaluate various splits and combinations allowing them to capture subtle patterns in the data. Random forest mitigates overfitting data by aggregating predictions from multiple randomized decision trees, while XGBoost improves generalization through sequential boosting and regularization.

Another method used is the stacked ensemble. The stacked ensemble used the tuned base models as its input. It used Logistic Regression as the meta learner. It achieved 95% accuracy and outperforms all of the base models. It produced balanced precision and recall.

B. SHAP Interpretability Analysis

For transparency and explainability, we implemented SHAP to explain the feature contributions. The analysis was performed on the tuned Random Forest model. This is because it achieved the highest accuracy out of all the models.The Beeswarm, Bar, and Waterfall plot were chosen for the visualizations.
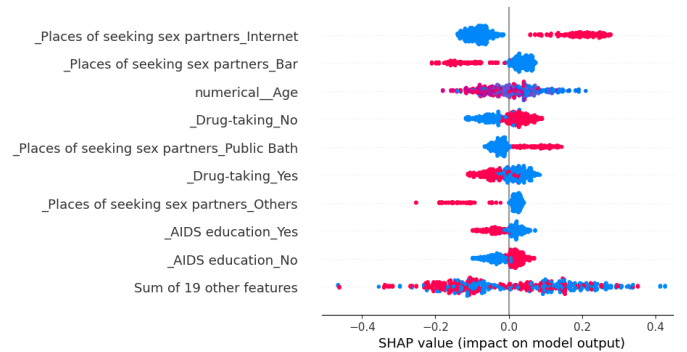

Fig.2. SHAP Beeswarm Plot on Random Forest Model

The Beeswarm Plot ranks features by their overall impact on model output. It shows how each feature's value affects the prediction across all samples. Beeswarm is suitable for SHAP as it provides predictions, directions, variability and relationship to feature within the entire dataset [22]. Each point represents a single prediction. The x-axis shows the SHAP value which is the feature's effect on the model output, and the color indicates the feature value (red = high, blue = low). From Fig. 2, we can see that most of the features show a clear separation between high and low values in terms of their impact. The only feature with a noticeably mixed pattern is "Age," where both high and low values are spread across positive and negative SHAP values. This suggests the model's interpretation of age varies more depending on other interacting features.
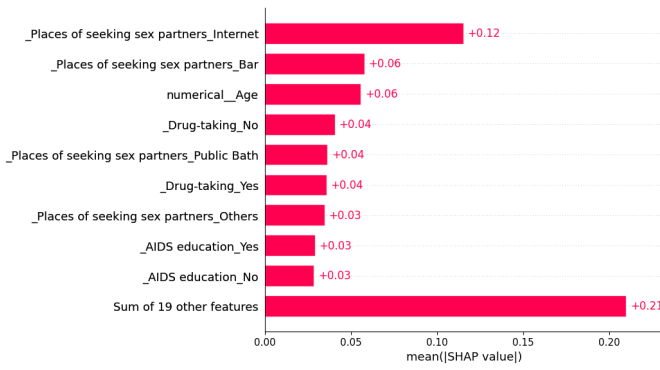
Fig.3. SHAP Absolute Mean Value for 9 highest impact on Random Forest Model
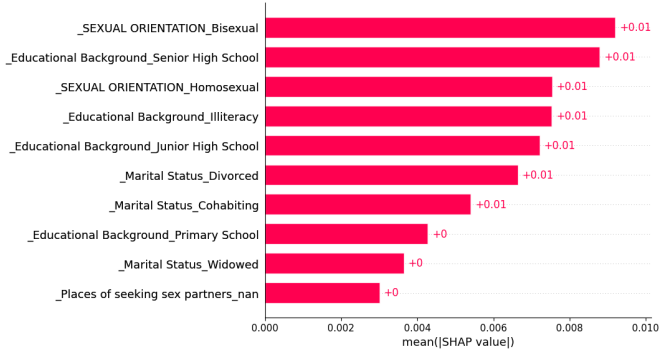

Fig.4. SHAP Absolute Mean Value for 10 lowest impact on Random Forest Model

The Bar Plot (Mean Absolute SHAP Values) shows a global feature importance ranking by averaging the absolute SHAP values across all samples. The Bar Plot shows a clear and easy to identify on which features have the biggest impact on the model prediction [23]. The Fig. 3 shows the 9 most influential features, with "Places of seeking sex partners_Internet" (+0.12), "Places of seeking sex partners_Bar", and the individual's age having the highest contribution. Other features like drug-taking behavior, use of public baths, and AIDS education moderately influence the model's predictions. We can also see from Fig. 4 that the features with the lowest impact are sexual orientation, various educational backgrounds, and marital statuses like divorced or widowed. These features have minimal contribution on the model, each contributing less than +0.01 on average.
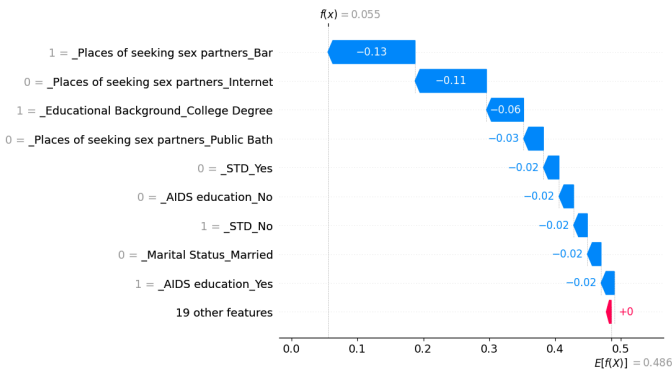

Fig.5. SHAP Waterfall Plot on Random Forest Model

The Waterfall Plot shows how an individual prediction was formed from base values and SHAP value

contributions [24]. The baseline prediction or the average value of the model is 0.486. As seen on Fig. 5, for this individual prediction, the final model output is 0.055. This means that it has a low probability of being HIV-positive. Each bar in the graph represents a feature's impact on the prediction. Key features that reduced the predicted risk include seeking sex partners at bars and on the internet, having a college degree, and AIDS education.

Keep in mind that SHAP assumes feature independence when attributing contributions, which means it evaluates each feature without impact on the other features as if it was unrelated. In this socio-behavioral dataset, features often correlate with one another which may slightly affect the accuracy of individual SHAP value interpretation.

C. Comparative SHAP Analysis (Random Forest vs. Logistic Regression)

In order to further enhance and understand the interpretability discussion, SHAP analysis is also conducted on the Logistic Regression model.
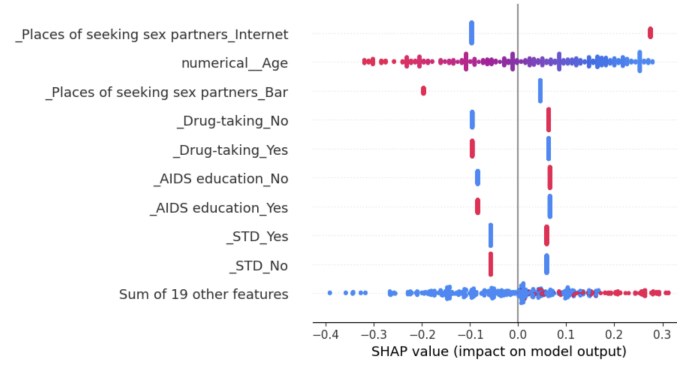

Fig. 6. SHAP Beeswarm Plot on Logistic Regression Model

The SHAP Beeswarm plot for the Logistic Regression model in Fig. 6 indicates that the top contributing features remain consistent with those in Random Forest, particularly "Seeking sex partners in the internet", "Age", and "Seeking sex partners in the Bar". However, the distribution of SHAP values here is notably narrower, reflecting the linear or base model's reduced capacity to capture complex non-linear interactions.
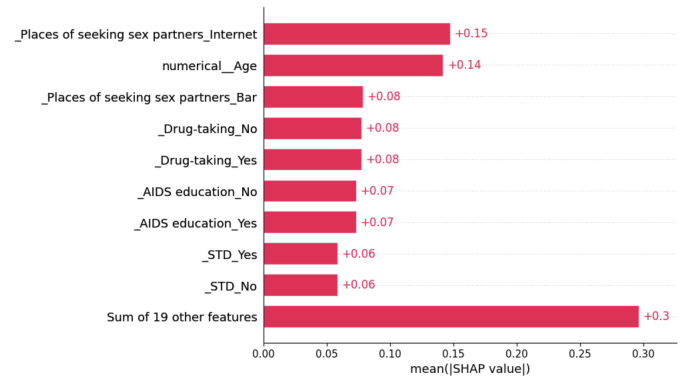

Fig. 7. SHAP Absolute Mean Value for 9 highest impact on Logistic Regression Model

The mean absolute SHAP value plot for Logistic Regression in Fig. 7 confirms that "Seeking sex partners in the Internet", "Age", and "Seeking sex partners in the Bar"

as the top three features that influence the model's decision. However, it shows a closer gap between subsequent features such as "No Drug Taking", "Yes Drug Taking" and "AIDS Education". This suggests that the linear or base model spreads its importance more evenly, whereas Random Forest concentrates it on the most discriminative predictors.

Overall, the analysis and comparison highlights that while both models identify similar high-impact features, Random Forest captures complex feature interactions and non-linear effects, resulting in higher SHAP value ranges and stronger predictive performance. On the other hand, Logistic Regression provides a simpler and more interpretable global trend but sacrifices the nuanced decision boundaries that boosted Random Forest's performance.

V. CONCLUSION

The results of the machine learning models show that the models with the highest predictive performance are Random Forest and XGBoost, achieving accuracy of 97% and 96% respectively. On the other hand, Logistic Regression and Gaussian Bayes, as part of the traditional methods, did not perform as effectively as the uniform ensemble. Even Stack Ensemble, which is a generally powerful model, could not perform better than uniform ensemble methods such as Random Forest and XGBoost in this case, achieving accuracy of 95%. The most possible reason is that the Uniform Ensemble methods were already advanced enough to process the data efficiently, leaving less room for improvement through stacking. SHAP analysis shows the key features that influence the results of the Random Forest model, which enhances both reliability and interpretability in detecting HIV status using Beeswarm, Bar and Waterfall plot. The key features that shape the model's predictions are partner-seeking behavior, education level, and drug use, which supports the accuracy in the real-world applications. While the proposed approach achieved high accuracy, the relatively small dataset size may limit the generalizability of the findings. Future work will focus on expanding the dataset and incorporating additional real-world samples or medical dataset. In addition, exploring other Explainable AI such as LIME may affect the interpretability of the model and potentially yield different insights.

REFERENCES

[1] E. Kumah, D. S. Boakye, R. Boateng, and E. Agyei, "Advancing the global fight against HIV/Aids: Strategies, barriers, and the road to eradication," *Ann. Glob. Health*, vol. 89, no. 1, 2023, doi: 10.5334/aogh.4277.

[2] R. D. Govender, M. J. Hashim, M. A. Khan, H. Mustafa, and G. Khan, "Global epidemiology of HIV/AIDS: A resurgence in North America and Europe," *J. Epidemiol. Glob. Health*, vol. 11, no. 3, pp. 296–301, 2021, doi: 10.2991/jegh.k.210621.001.

[3] P. K. Gupta and A. Saxena, "HIV/AIDS: Current updates on the disease, treatment and prevention," *Proc. Natl. Acad. Sci. India Sect. B*, vol. 91, no. 3, pp. 495–510, 2021, doi: 10.1007/s40011-021-01237-y.

[4] S. G. Deeks *et al.*, "Research priorities for an HIV cure: International AIDS Society Global Scientific Strategy 2021," *Nat. Med.*, vol. 27, no. 12, pp. 2085–2098, 2021, doi: 10.1038/s41591-021-01590-5.

[5] E. U. Alum, D. E. Uti, O. P. C. Ugwu, and B. N. Alum, "Toward a cure—advancing HIV/AIDs treatment modalities beyond antiretroviral therapy: A review," *Medicine (United States)*, vol. 103, no. 27, 2024, doi: 10.1097/MD.0000000000038768.

[6] N. E. Rosenberg, C. D. Pilcher, M. P. Busch, and M. S. Cohen, "How can we better identify early HIV infections?" *Curr. Opin. HIV AIDS*, vol. 10, no. 1, pp. 61–68, 2015, doi: 10.1097/COH.0000000000000121.

[7] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," *Int. J. Intell. Netw.*, vol. 3, pp. 58–73, 2022, doi: 10.1016/j.ijin.2022.05.002.

[8] S. K. Hung *et al.*, "Developing and validating clinical features-based machine learning algorithms to predict influenza infection in influenza-like illness patients," *Biomed. J.*, vol. 46, no. 5, 2023, doi: 10.1016/j.bj.2022.09.002.

[9] F. A. Khaleel and A. M. Al-Bakry, "Diagnosis of diabetes using machine learning algorithms," *Mater. Today Proc.*, vol. 80, pp. 3200–3203, 2023, doi: 10.1016/j.matpr.2021.07.196.

[10] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," *SN Comput. Sci.*, vol. 1, no. 6, 2020, doi: 10.1007/s42979-020-00365-y.

[11] Y. Li *et al.*, "Machine learning for lung cancer diagnosis, treatment, and prognosis," *Genomics Proteomics Bioinform.*, vol. 20, no. 5, pp. 850–866, 2022, doi: 10.1016/j.gpb.2022.11.003.

[12] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 10–19, 2021, doi: 10.38094/jastt20179.

[13] J. He *et al.*, "Application of machine learning algorithms in predicting HIV infection among men who have sex with men: Model development and validation," *Front. Public Health*, vol. 10, 2022, doi: 10.3389/fpubh.2022.967681.

[14] B. Lin, J. Liu, K. Li, and X. Zhong, "Predicting the risk of HIV infection and sexually transmitted diseases among men who have sex with men: Cross-sectional study using multiple machine learning approaches," *J. Med. Internet Res.*, vol. 27, 2025, doi: 10.1016/j.ijin.2022.05.002.

[15] M. S. Alie and Y. Negesse, "Machine learning prediction of adolescent HIV testing services in Ethiopia," *Front. Public Health*, vol. 12, 2024, doi: 10.3389/fpubh.2024.1341279.

[16] J. Fieggen *et al.*, "The role of machine learning in HIV risk prediction," *Front. Reprod. Health*, vol. 4, 2022, doi: 10.3389/frph.2022.1062387.

[17] M. Jaiteh *et al.*, "The application of machine learning algorithms to predict HIV testing in repeated adult population–based surveys in South Africa: Protocol for a multiwave cross-sectional analysis," *JMIR Res. Protoc.*, vol. 14, 2025, doi: 10.2196/59916.

[18] A. Abade *et al.*, "A comparative analysis of classical and machine learning methods for forecasting TB/HIV co-infection," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-69580-4.

[19] G. Airlangga, "Performance evaluation of machine learning models for HIV/AIDS classification," *J. Inf. Syst. Res.*, vol. 6, no. 2, 2025, doi: 10.47065/josh.v6i2.6790.

[20] E. Orel *et al.*, "Prediction of HIV status based on socio-behavioural characteristics in East and Southern Africa," *PLoS ONE*, vol. 17, no. 3, 2022, doi: 10.1371/journal.pone.0264429.

[21] HIV prediction. (2023, May 4). *Kaggle*. https://www.kaggle.com/datasets/ishigamisenku10/hiv-predicti

[22] Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888. doi: https://arxiv.org/abs/1802.03888

[23] Bifarin, O. O. (2023). Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. Plos one, 18(5), e0284315. doi: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0284315a

[24] Ponce‐Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clinical and Translational Science, 17(11), e70056. doi:https://ascpt.onlinelibrary.wiley.com/doi/full/10.1111/cts.70056