BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

Explainable Machine Learning For Early HIV Detection Using Extra Trees and SHAP Algorithms

Anggi Dewi Nurcahyani[1], Dika Ratu Anisa[2*], Nayla Nurul Azkiya[3]

Informatics[1,2,3]
Universitas Informatika dan Bisnis Indonesia, Bandung, Indonesia[1,2,3]
https://unibi.ac.id/[1,2,3]
dikaratuanisa23@student.unibi.ac.id[2*]

Abstract. Human Immunodeficiency Virus (HIV) remains a global health challenge that requires accurate and reliable early detection approaches. The use of machine learning offers potential in classifying HIV status based on clinical, demographic, and behavioral data. However, the limitations of interpretability in black-box models are an obstacle to clinical application. This study proposes an Explainable Machine Learning approach for early HIV detection by integrating the Extra Trees algorithm and the Shapley Additive exPlanations (SHAP) method. The model was developed using an HIV dataset obtained from the Kaggle platform and processed through standard data preprocessing stages without class balancing. Performance evaluation was conducted using classification metrics, confusion matrices, and learning curves to assess accuracy and learning stability. The results of the experiment show that the Extra Trees model achieved 88% accuracy with strong generalization. SHAP and mean absolute SHAP analyses revealed the dominant features that contributed to the prediction of HIV status consistently at the global and local levels. These findings show that integrating Extra Trees and SHAP produces an HIV early-detection model that is not only competitive in performance but also transparent and clinically relevant, potentially supporting the development of reliable artificial intelligence-based medical decision support systems.

Key words: HIV; early detection; Extra Trees; Explainable Machine Learning; SHAP

1. INTRODUCTION

Human Immunodeficiency Virus (HIV) remains a global health problem that significantly affects the human immune system and has the potential to develop into Acquired Immune Deficiency Syndrome (AIDS). According to the 2024 UNAIDS report, approximately 40.8 million people are living with HIV, with 1.3 million new infections and 630,000 deaths from AIDS[1]. Although various prevention and treatment efforts have shown progress, delays in detecting HIV at an early stage remain a major challenge. Early detection is crucial because it enables the timely initiation of antiretroviral therapy (ART), which reduces morbidity and mortality and lowers the risk of further transmission[2].

As the availability of health data increases, machine learning is increasingly used as an analytical approach to support early HIV detection [3]. Machine learning algorithms can process large-scale data and identify complex patterns and nonlinear relationships between variables that are difficult to capture with conventional statistical methods. Various studies show that this approach can improve the accuracy of HIV status classification using clinical, demographic, and behavioral data[4]. However, most of the models developed remain black boxes, making the decision-making mechanism difficult for medical personnel to understand. This lack of transparency is an obstacle to the widespread adoption of models, especially in clinical practices that demand accountability[5].

In the field of health, the challenges of applying machine learning are not only about achieving high predictive performance but also about ensuring model interpretability [6]. Early HIV detection: understanding the factors that Influence classification results is an important aspect of supporting evidence-based clinical decision-making [7]. Models with high accuracy but inadequate explanation can cause mistrust, misinterpretation, and resistance among healthcare professionals, thereby limiting their use in clinical settings.

Explainable Machine Learning (XML) has emerged as an approach to address this need by providing a mechanism for explaining model decisions. One widely used XML method is Shapley Additive exPlanations (SHAP), which can consistently measure the contribution of each feature to the prediction results [8]. In this study, the Extra Trees algorithm was chosen as the main classification model due to its ability to handle high-dimensional data, reduce overfitting, and produce stable predictions through an ensemble approach[9]. Given the complexity of the Extra

Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

Trees structure, integration with SHAP is a strategic step to improve transparency and understanding of model behavior.

This study aims to develop an HIV early-detection model that not only excels in predictive performance but can also be clearly interpreted. By combining the Extra Trees algorithm and the SHAP method, this study analyzes the main factors contributing to HIV status prediction and their relevance in a clinical context. The contribution of this study lies in applying Explainable Machine Learning to early HIV detection, with an emphasis on model interpretability and reliability. The results of this study are expected to enrich scientific literature and support the development of more transparent and responsible artificial intelligence-based medical decision support systems.

2. RELATED WORK

Proposed a machine learning approach for HIV/AIDS prediction by combining feature selection and data balancing strategies[10]. The study evaluated several algorithms, including Random Forest, Extra Trees, and XGBoost, and showed that tree-based models performed best in handling class imbalance. The evaluation used standard metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, with results showing a significant improvement over approaches without feature selection or data balancing. However, this study did not report the model's interpretability mechanism or an explainable feature contribution analysis, particularly for the Extra Trees algorithm. Hence, the model's decision transparency remains limited despite its high prediction accuracy.

In terms of model interpretability, an Explainable AI framework based on Shapley Additive exPlanations (SHAP) can improve understanding of tree-based models, including Random Forest and Extra Trees. This study emphasizes the importance of local and global explanations in interpreting complex model behavior and identifying nonlinear risk patterns. The study is methodological and does not focus on a specific classification task, so it does not report performance metrics, such as accuracy, in the context of early HIV detection. Nevertheless, its contribution is an important foundation for the development of models that are not only accurate but also systematically explainable.

An alternative approach is demonstrated that uses text mining and machine learning techniques on unstructured clinical records for early HIV detection [11]. This study compares conventional machine learning models with Large Language Models (LLMs) and shows that LLMs provide superior performance in identifying patients suspected of being infected with HIV. The evaluation was conducted on balanced and unbalanced datasets, with results showing improved accuracy and classification capabilities on real-world data. However, numerical accuracy values were not explicitly reported, and the study did not integrate Explainable Machine Learning approaches or evaluate tree-based ensemble models such as Extra Trees.

Based on these three studies, it can be concluded that previous studies generally focused on improving predictive performance or developing interpretability methods separately. To date, no study has explicitly integrated the Extra Trees algorithm with the SHAP method for early HIV detection and evaluated model performance in terms of both accuracy and interpretability. Therefore, this study fills this gap by combining the predictive advantages of Extra Trees with SHAP's ability to explain feature contributions, resulting in an HIV early-detection model that is not only accurate but also transparent and clinically relevant.

3. METHODS

This study uses a data-driven experimental approach to develop an accurate and interpretable HIV early detection model. The study methodology includes systematic stages ranging from data collection, preprocessing, exploratory analysis, modeling, and model performance evaluation. The Extra Trees algorithm was chosen as the main model due to its ability to handle complex medical data and provide high classification performance. To improve transparency and trust in a clinical context, an Explainable Artificial Intelligence approach using SHAP was integrated to explain the contribution of features to the prediction results. This approach ensures that the resulting model is not only effective but also scientifically accountable.

BIMA
BULLETIN OF INTELLIGENT MACHINES AND ALGORITHMS
Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

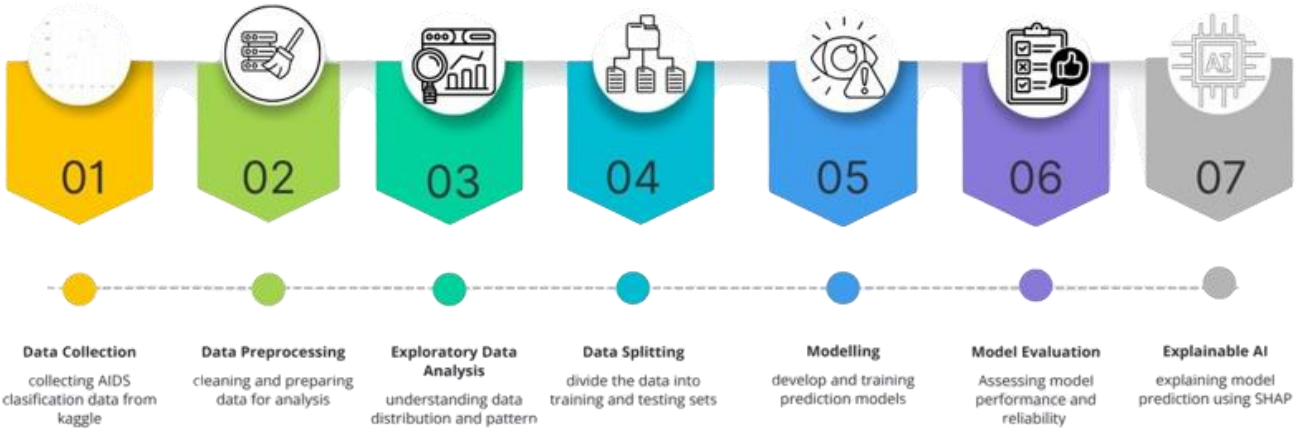

Fig. 1. Proposed Method

The proposed method, as illustrated in Figure 1, is designed to address the main challenges in early HIV detection, particularly those related to prediction accuracy and the interpretability limitations of machine learning models. By integrating the Extra Trees algorithm and the SHAP-based Explainable Machine Learning approach, this method not only optimizes classification performance but also provides a comprehensive understanding of each feature's contribution to the model's decision. This approach enables the quantitative identification of the most influential risk factors, thereby increasing transparency and reliability, and enhancing the model's potential application in a clinical context as a medical decision support system.

3.1. Data Collection.

This study uses an HIV/AIDS classification dataset obtained from the Kaggle platform. The dataset contains clinical, demographic, and behavioral factors relevant to HIV infection status. The data were collected in a structured format and have been anonymized so they are safe to use for the study. This stage aims to provide representative data as a basis for developing an early HIV detection model.

3.2. Data Preprocessing.

The preprocessing stage is performed to improve data quality before modeling. This process includes cleaning data to remove missing values and inconsistencies, maintaining the stability of the model's learning process.

3.3. Exploratory Data Analysis

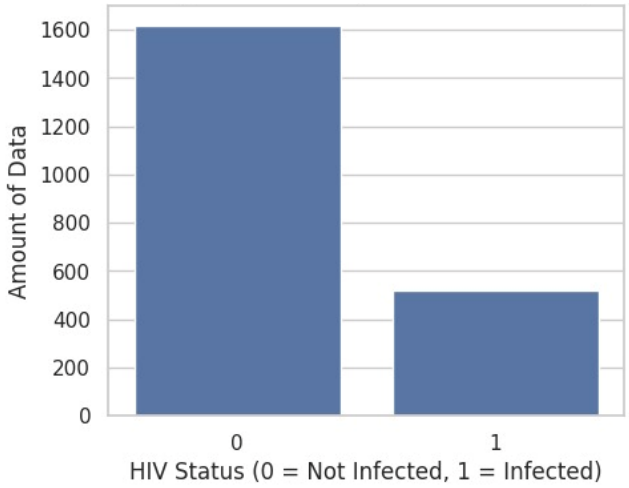

Fig. 2. Target Distribution

Figure 2 shows class imbalance, where the majority of data is in the uninfected class (0) at 75.64%, while the infected class (1) only accounts for 24.36% of the total sample. This condition indicates the need for special handling of class imbalance to prevent the classification model from being biased towards the majority class.

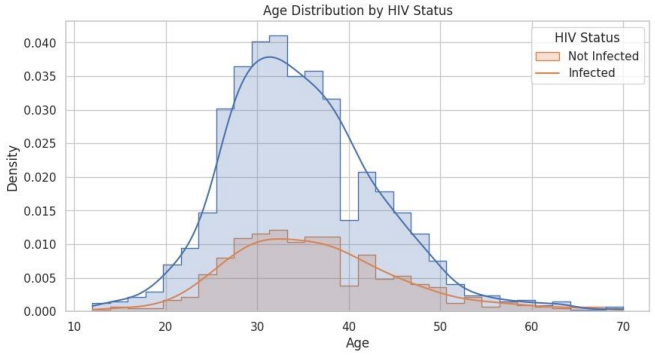

Fig. 3. Age Distribution

Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

Figure 3 shows that HIV-infected individuals are more concentrated in the young adult to productive age range, with peak density around 30–35 years of age. Compared to the uninfected group, this pattern indicates that age could potentially be a distinguishing factor in HIV status classification.

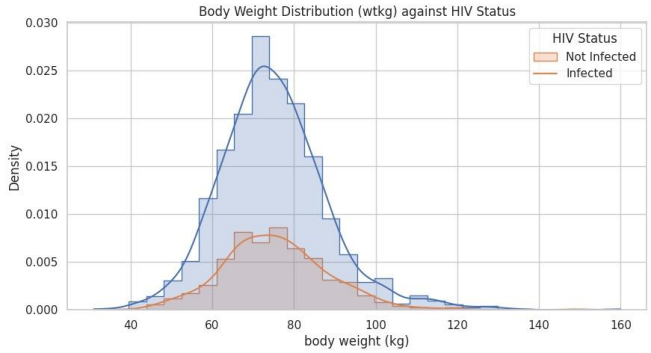

Fig. 4. Body Weight Distribution

Figure 4 shows that HIV-infected individuals tend to have lower body weights, concentrated in the 60–80 kg range. This difference in density patterns indicates that body weight can provide additional signals in distinguishing HIV status, despite overlap between groups.

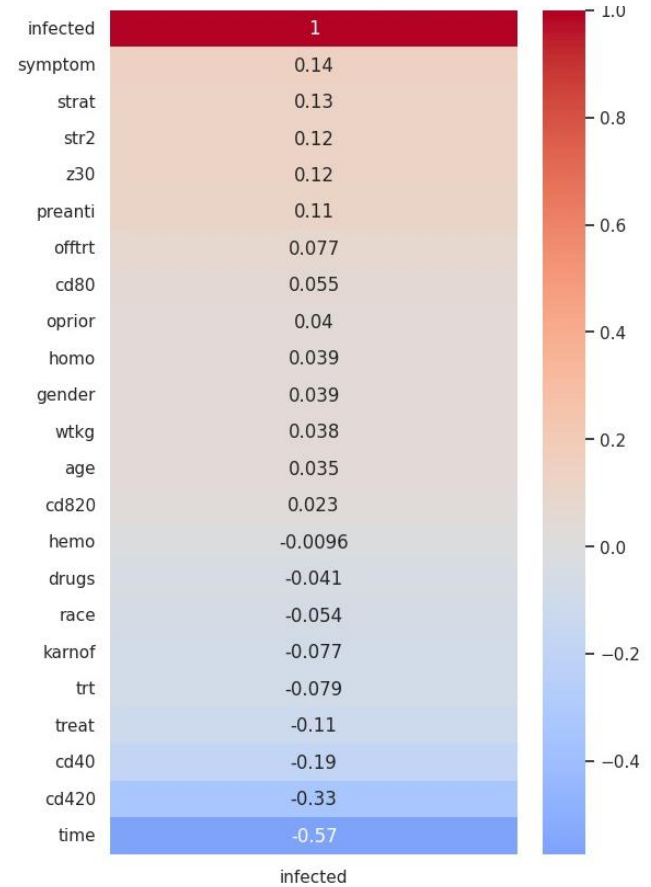

Fig. 5. Feature Correlation with Target (infected)

Figure 5 shows that most features have a weak to moderate relationship with HIV infection status, with the time variable showing the strongest negative correlation (−0.57). Meanwhile, features such as symptom, strat, and preanti have positive correlations, albeit relatively low, indicating that a combination of many factors influences HIV status and is not dominated by a single feature.

3.4. Data Splitting

The dataset that has undergone preprocessing is split into training and test sets using a stratified sampling scheme with an 80:20 split. This approach maintains consistency in class distribution across both subsets, so that model training and performance evaluation processes better reflect the data conditions.

3.5. Modelling

The modeling stage was carried out using the Extra Trees Classifier, known for its ability to handle high-dimensional data and reduce variance through randomization in feature selection and splitting points. The model was trained on the training data, with adjustments to key hyperparameters to achieve optimal performance. This algorithm was selected for its ability to achieve high accuracy and stable predictions in medical data.

The modeling stage uses the Extra Trees Classifier (also known as Extremely Randomized Trees). This decision tree-based ensemble method introduces strong randomization in feature selection and split points to reduce model variance and improve prediction stability. Mathematically, ensemble predictions are obtained by averaging the predictions of M trees, as shown in Eq. (1).

$$\hat{y}(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x) \qquad (1)$$

Where $T_m(x)$ represents the prediction from tree number m. The randomization mechanism in Extra Trees has been shown to effectively reduce correlation between trees, thereby reducing the risk of overfitting, especially in high-dimensional medical data [12]. To improve model transparency and reliability in a clinical context, Extra Trees is integrated with the SHAP-based Explainable AI method, which calculates the contribution of each feature to the prediction result using Shapley values from game theory, enabling consistent local and global interpretation of model decisions [13].

3.6. Model Evaluation

Model performance was evaluated using classification metrics, including accuracy, precision, recall, F1-score, and further analyzed through a confusion matrix to assess the model's ability to classify infected and uninfected cases accurately. In addition, a learning curve was used to evaluate the model's learning behavior as the amount of training data increased, allowing the balance between bias

Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

and variance to be observed. The evaluation results showed that the Extra Trees model demonstrated stable performance, with high accuracy, and maintained a balance between positive case detection and classification error reduction, indicating the effectiveness of the proposed approach for early HIV detection.

3.7. Explainable AI

To improve the model's transparency and reliability, an Explainable AI approach was applied using SHAP (Shapley Additive exPlanations). SHAP was used to identify the contribution of each feature to the model's predictions, both globally and locally. This approach allows for a more in-depth interpretation of the prediction results. It provides clinically relevant insights, thereby supporting the model's use in medical decision-making.

4. RESULTS AND DISCUSSIONS

The test results show that the Extra Trees model achieves stable classification performance and is highly effective at detecting HIV infection status. The evaluation was conducted by combining key classification metrics, learning curves, confusion matrices, and SHAP-based interpretability analysis. This evaluation approach enables a comprehensive assessment of predictive performance, learning stability, and model decision transparency, ensuring results that emphasize not only accuracy but also clinical relevance and model reliability.

TABLE 1. Classification Metric

Accuracy Value	Extra Trees (0)	Extra Trees (1)
Accuracy	0.88	0.88
Recall	0.97	0.60
Precision	0.89	0.85
F1-Score	0.93	0.70

The classification metrics, as shown in Table 1, indicate that the Extra Trees model achieved an overall accuracy of 88%, indicating the model's ability to classify data well in general. In the uninfected class (0), the very high recall of 97% indicates that the model accurately identifies most uninfected individuals. Meanwhile, in the infected class (1), a recall value of 60% indicates that some positive cases remain undetected. However, the relatively high precision of 85% in the infected class indicates that the model's positive predictions are quite reliable. The F1-score of 70% in the infected class reflects a balance between precision and recall, confirming that the model has strong early-detection potential, though it still needs to improve its sensitivity.

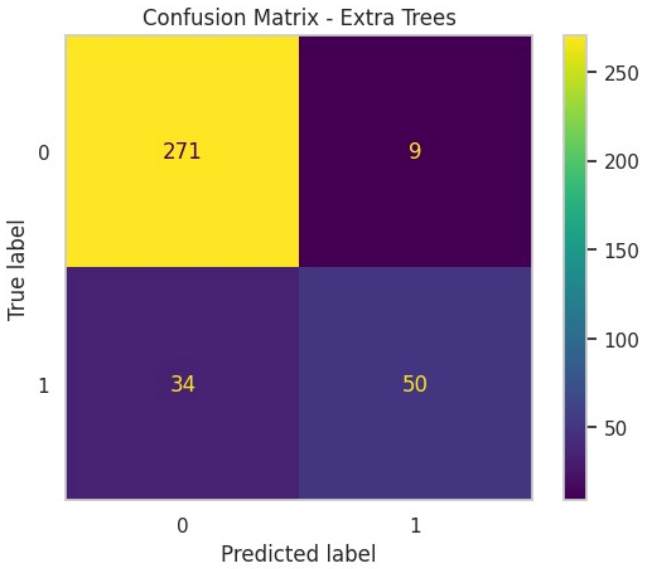

Fig. 6. Confusion Matrix

Figure 6 is a confusion matrix showing that the model correctly classified the majority of uninfected data (271 samples), with very few false positives. However, there were still 34 false negative cases in the infected class, indicating that the model did not detect some positive cases. These findings confirm the trade-off between high specificity and sensitivity, which can still be improved, an important aspect in the context of early HIV detection.

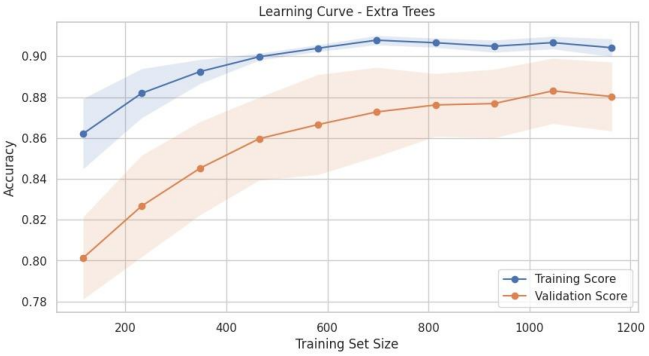

Fig. 7. Learning Curve

Figure 7 is a learning curve graph showing that accuracy on the training and validation data increases as the training data size increases, then stabilizes. The difference between training and validation scores is relatively small at larger data sizes, indicating that the model generalizes well and does not overfit. This pattern shows that Extra Trees can effectively leverage additional data to improve predictive performance.

Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

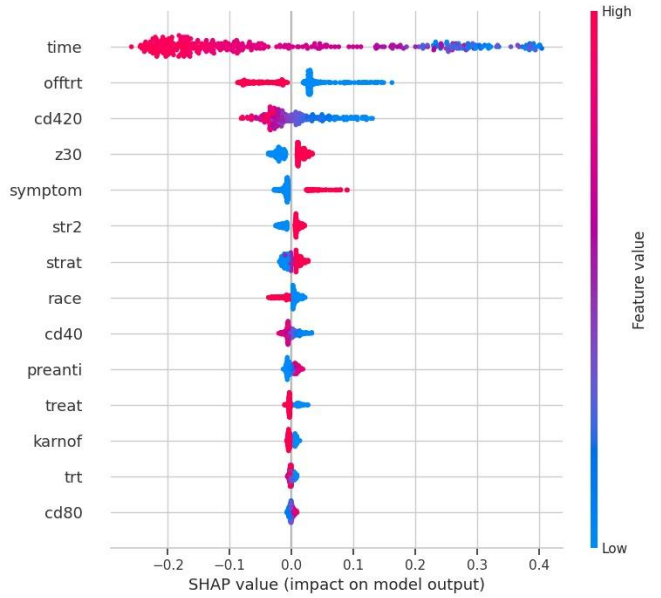

Fig. 8. SHAP Summary Plot

Figure 8 is a SHAP summary plot showing that the time feature has the greatest impact on the model output, followed by offtrt, cd420, and z30. The wide distribution of SHAP values on these features indicates a significant and consistent Influence on the model's decision. The plot shows that variations in feature values increase or decrease the probability of predicting HIV status, providing an in-depth understanding of the direction and magnitude of each feature's Influence.

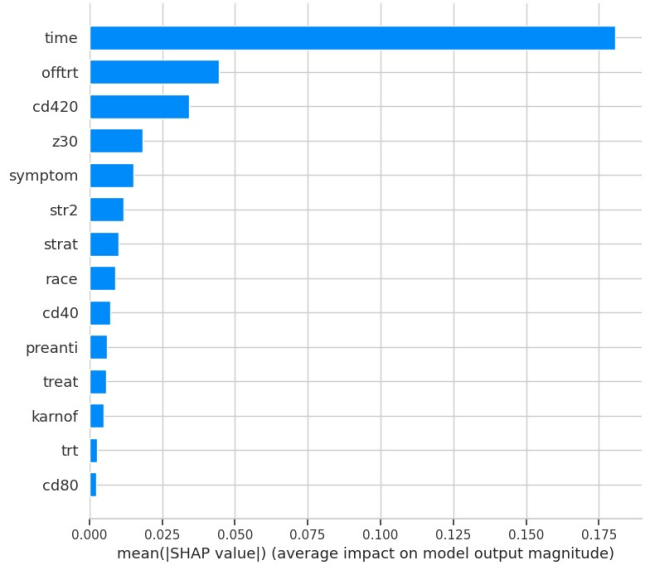

Fig. 9. Mean Absolute SHAP

Figure 9 is a mean absolute SHAP graph that confirms the global contribution of features to model prediction. The time feature ranks highest, indicating that temporal factors are the main determinants in HIV status classification. Other clinical features, such as offtrt and cd420, also have significant contributions, while other features show relatively smaller impacts. This order reinforces the consistency between SHAP's global and local interpretations.

Overall, the results of this study show that the Extra Trees approach, when integrated with SHAP-based Explainable AI, is not only capable of achieving good predictive performance but also provides clinically relevant explanations. Although the model shows high specificity and good learning stability, improving sensitivity to the infected class remains an area for development. Further study could focus on class weight adjustment, threshold tuning, or external validation across more diverse clinical datasets to enhance the model's practical impact in supporting early HIV detection.

The results of this study confirm that the Extra Trees algorithm is an effective approach for early HIV detection in complex medical data. High accuracy and a stable learning curve indicate that the model balances bias and variance well. These findings reinforce the idea that strong randomization in Extra Trees improves model stability and generalization.

However, the lower recall in the infected class indicates a risk of false negatives, with important implications in a clinical context. Although the model shows high specificity and can minimize errors in the uninfected class, increased sensitivity to the infected class is still needed for the system to be more effective as an early HIV screening tool.

The integration of Explainable Machine Learning using SHAP provides significant added value to model reliability and transparency. The identification of dominant features such as time, offtrt, and cd420 shows that the model's decisions are consistent with relevant clinical factors. SHAP's ability to explain feature contributions at the individual and population levels increases confidence in prediction results and opens up opportunities for model adoption in medical decision support systems.

Although the results obtained are promising, this study still has limitations, including the use of a single dataset and the lack of external validation. Further study could focus on optimizing model sensitivity through class-weight adjustment or threshold tuning, as well as on cross-population evaluation to improve the generalization and clinical impact of the proposed approach.

5. CONCLUSIONS

This study proposes an HIV early-detection approach based on Explainable Machine Learning, using the Extra Trees algorithm integrated with the SHAP method to improve model transparency and reliability. The experimental results show that the Extra Trees model achieved an accuracy of 88%, indicating strong overall classification performance. Learning curve analysis shows stable learning performance and strong generalization, while a confusion matrix indicates that the model accurately identifies 97% of uninfected cases. However, sensitivity to the infected class remains at 60%, indicating that several

Vol. 1 No. 2 January 2025
E-ISSN: 3123-5115
DOI: 10.65780/bima.v1i2.8

positive cases have not been detected, which is an important challenge in a clinical context.

SHAP integration enables global and local interpretations of model decisions by identifying key features that contribute to HIV status predictions, including temporal factors and specific clinical indicators. This approach increases model transparency and supports confidence in prediction results as a medical decision support system. Overall, the results show that the combination of Extra Trees and Explainable AI is an effective and promising approach for early HIV detection. Further study is recommended to improve sensitivity to the infected class and perform external validation on cross-population clinical datasets to strengthen the practical impact of this approach in supporting public health efforts.

REFERENCES

[1] "Global HIV & AIDS statistics — Fact sheet," UNAIDS, [Online]. Available: https://www.unaids.org/en/resources/fact-sheet

[2] "Long-Term Benefits from Early Antiretroviral Therapy Initiation in HIV Infection," *NEJM Evidence*, vol. 2, no. 3, Feb. 2023, doi: 10.1056/EVIDoa2200302.

[3] Y. Xiang, J. Du, K. Fujimoto, F. Li, J. Schneider, and C. Tao, "Application of artificial intelligence and machine learning for HIV prevention interventions," Jan. 01, 2022, *Elsevier Ltd*. doi: 10.1016/S2352-3018(21)00247-2.

[4] S. U. Nisa, A. Mahmood, F. S. Ujager, and M. Malik, "HIV/AIDS predictive model using random forest based on socio-demographical, biological and behavioral data," *Egyptian Informatics Journal*, vol. 24, no. 1, pp. 107–115, Mar. 2023, doi: 10.1016/j.eij.2022.12.005.

[5] N. Khan, M. Nauman, A. S. Almadhor, N. Akhtar, A. Alghuried, and A. Alhudhaif, "Guaranteeing Correctness in Black-Box Machine Learning: A Fusion of Explainable AI and Formal Methods for Healthcare Decision-Making," *IEEE Access*, vol. 12, pp. 90299–90316, 2024, doi: 10.1109/ACCESS.2024.3420415.

[6] R. ElShawi, Y. Sherif, M. Al-Mallah, and S. Sakr, "Interpretability in healthcare: A comparative study of local machine learning interpretability techniques," *Comput. Intell.*, vol. 37, no. 4, pp. 1633–1650, Nov. 2021, doi: 10.1111/coin.12410.

[7] L. Famiglini, A. Campagner, M. Barandas, G. A. La Maida, E. Gallazzi, and F. Cabitza, "Evidence-based XAI: An empirical approach to design more effective and explainable decision support systems," *Comput. Biol. Med.*, vol. 170, Mar. 2024, doi: 10.1016/j.compbiomed.2024.108042.

[8] R. Zhou and T. Hu, "Evolutionary approaches to explainable machine learning," Jun. 2023, doi: 10.1007/978-981-99-3814-8_16.

[9] M. Yousefi, V. Oskoei, H. R. Esmaeli, and M. Baziar, "An innovative combination of extra trees within adaboost for accurate prediction of agricultural water quality indices," *Results in Engineering*, vol. 24, Dec. 2024, doi: 10.1016/j.rineng.2024.103534.

[10] A. Mizwar, A. Rahim, P. Hartato, A. Ridwan, and F. Asharudin, "Machine Learning-Based Approach for HIV/AIDS Prediction: Feature Selection and Data Balancing Strategy," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[11] R. Morales-Sánchez, S. Montalvo, A. Riaño, R. Martínez, and M. Velasco, "Early diagnosis of HIV cases by means of text mining and machine learning models on clinical notes," *Comput. Biol. Med.*, vol. 179, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108830.

[12] S. M. Lundberg et al., "From Local Explanations to Global Understanding with Explainable AI for Trees," Nature Machine Intelligence, 2020.

[13] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python—Tree-Based Ensemble Updates," Journal of Machine Learning Research, pembaruan praktik ensemble pohon, 2021–2023