Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hybrid Statistical Modeling and Machine Learning Ensemblers for Interpretable HIV Viral Load Prediction Among Clients on Antiretroviral Therapy

Muhire Jacob

mrejacob@gmail.com

Central Public Health Laboratories

Jjingo Daudi

Makerere University, IDI-McKinnell Knowledge Centre

Galiwango Ronald

Makerere University, IDI- McKinnell Knowledge Centre

Joyce Nakatumba-Nabende

Makerere University

Aloysious Ssemaganda

Central Public Health Laboratories

Isaac Ssewanyana

Central Public Health Laboratories

Research Article

Keywords: Antiretroviral Therapy, Viral Load, Predictive Modelling, Ensemble Learning, Artificial Intelligence (AI), Explainable AI

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10514528/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hybrid Statistical Modeling and Machine

Learning Ensemblers for Interpretable HIV Viral Load Prediction Among Clients on Antiretroviral Therapy

Muhire Jacob[1*], Jjingo Daudi[2†], Galiwango Ronald[3†], Joyce Nakatumba-Nabende[4], Aloysious Ssemaganda[5†], Isaac Ssewanyana[6†]

[1*] Central Public Health Laboratories, Plot 1062-106 Luzira Butabika Road, Kampala, Uganda.

[2]African Center of Excellence in Bioinformatics, IDI-McKinnell Knowledge Centre, Makerere University, IDI-McKinnell Knowledge Centre, Kampala, P.O Box 22418 Kampala Uganda.

[3]African Center of Excellence in Bioinformatics, Makerere University, IDI-McKinnell Knowledge Centre, Kampala, P.O Box 22418 Kampala Uganda.

[4]School of Computing and Informatics, Uganda, Makerere University, Pool Road, Kampala, P.O Box 7062 Kampala Uganda.

Central Public Health Laboratories, Plot 1062-106 Luzira Butabika Road, Kampala, Uganda.

Central Public Health Laboratories, Plot 1062-106 Luzira Butabika Road, Kampala, Uganda.

*Corresponding author(s). E-mail(s): mrejacob@gmail.com;

Contributing authors: daudics2100@gmail.com;

galiwango.gronald.ronald@gmail.com; joyce.nabende@mak.ac.ug;

assemaganda@gmail.com; sewyisaac@yahoo.co.uk;

†These authors contributed equally to this work.

Abstract

Elimination of HIV transmission requires a multi pronged approach and com-mitment from all stakeholders which includes machine learning engineers. The burden of HIV is highest in sub-Saharan Africa where many countries are in the fight against the disease with a high number of recorded people living with HIV, case in point Uganda with about 1.4 million clients. With the recent advent of Artificial Intelligence, the capabilities in diagnostics have grown in leaps and bounds by leveraging the Machine Learning and Deep Learning. In this study, we present an ensemble learner model that predicts which HIV client's next viral load using HIV viral load laboratory test data. We ably trained and validated our learners on 70% of the data, and tested them on the other 30% for a 5.7 million HIV viral load dataset of ART clients for the period 2019 to 2022 from the Central Public Health Laboratory of Uganda. Our findings showed that the Staking technique outperformed the Bagging, random forest, decision tree, XG Boost, Adaboost, KNN, linear regressors with an R2 score of 0.85, MAE of 0.42, and MSE of 0.41 improving the chances of predicting next visit viral load of Antiretroviral Therapy (ART) clients. For model interpretability, we explored the LIME (Local Interpretable Model-Agnostic Explanations) to provide an insight into the variables of importance during model training. Our study demonstrates the potential of leveraging HIV viral load test data, statistical modeling and machine learning techniques to predict Antiretroviral Therapy clients next visit viral load.

Keywords: Antiretroviral Therapy, Viral Load, Predictive Modelling, Ensemble Learning, Artificial Intelligence (AI), Explainable AI

1 Background

Human immunodeficiency virus (HIV), the main cause of the Acquired Immunodefi-ciency Syndrome (AIDS) or Advanced HIV disease, has affected close to 38 million people across the globe according to WHO statistics. A phylodynamic study showed that the origin of HIV in Uganda dates back to the 1960s and gradually spread to Kampala from south western Uganda in the 1970s with its highest prevalence in early 1990s [1]. Human Immunodeficiency Virus (HIV) is a global concern with an average annual mortality rate of 630,000 from Acquired Immune Deficiency Syndrome (AIDS) related causes as reported by the World Health Organization [5]. The first known cases of HIV in Uganda were recorded in the mid-1980s among expectant mothers in Rakai district, with a prevalence rate of 11%. Due to the uncontrolled spread of the disease from its onset, reaching a peak of 3.3 million infections by 1992, representing 18% of the general population [7]. However, Uganda has done considerable well in reducing HIV prevalence, to the current rate of 5.8%. The assessment also defined adults aged between 15 and 64 years [8]. With the adoption of Antiretroviral therapy (ART) treatment, there have been significant strides in the fight against the HIV/AIDS spread, among the benefits are control of high patient mortality. However, for the treatment options to be effective, adherence is extremely important to prevent the development of antiretroviral treatment failure. Also, routine testing is needed to monitor patient outcomes of the treatment they are on in addition to determining best treatment interventions or detect drug resistance to the treatment they are taking [2]. A viral load test is used to confirm presence of HIV in blood. In the current Ugandan setting, a lot of time is spent between sample collection, testing and eventual receipt of results by patients at health centers which is a lost opportunity in providing patients with intensive adherence counselling before they migrate to other locations or are totally lost to follow-up which in itself will lead to increased spread of the disease among partners who are unaware of their HIV status. Machine learning (ML) is a collection of statis-tical packages that can help us analyze and interpret patterns and structures in data to facilitate learning, synthesis, and decision-making [16][17]. In recent years, machine learning has demonstrated its potential in HIV prevention [11]. ML was employed in the development and validation of an HIV prediction model for identification of candidates for pre-exposure prophylaxis (PReP) [12] With the vast volumes of semantic data, disease risk prediction is now a possibility using client demographic, clinical diagnosis, behavioral, laboratory results, and drug prescriptions. This supports the use of electronic medical records as an input for development of health care prediction models [13] Machine learning was also used in the prediction of Antiretroviral Therapy patient retention and suppression for the setting in South Africa. Adherence to treatment, addressing barriers to care, and early initiation of Antiretroviral Therapy after diagnosis are important factors in managing HIV treatment effectively and pre-venting transmission [16].

A search for previously conducted studies in the field of diagnostics with ensemble learners particularly in the field of HIV treatment. Specifically, a major gap in the reviewed literature shows that many of the derived models are used as single algorithms that may not adequately capture the complex, insights and relationships among variables not showing any weaknesses that could lead to issues around bias or variance of the model training results. Equipped with the HIV viral load test dataset from the Central Public Health Laboratories of Uganda, achieving the aims of the study was made possible. This predictive model is intended to be a decision support tool for healthcare workers in the diagnosis and treatment of HIV clients enrolled on antiretroviral therapies.

2 Methods

The methodology in fig.1 describes how the model was derived; starting with inputting a 4 year HIV Viral load dataset of the period 2019 to 2022 from the Central Public Health Laboratory of Uganda. The patient Antiretroviral Therapy treatment records were studied to gather a wholesome context of HIV key indicators from the laboratory test results. The dataset contained HIV viral load test results of male and female clients either on antiretroviral therapy (ART) performing a routine check or perform-ing a confirmatory test. The dataset contains 5.7 million HIV viral load results of Ugandan ART clients aged 15 to 70 years.

The data was standardized through cleaning and pre-processing to ensure that it is

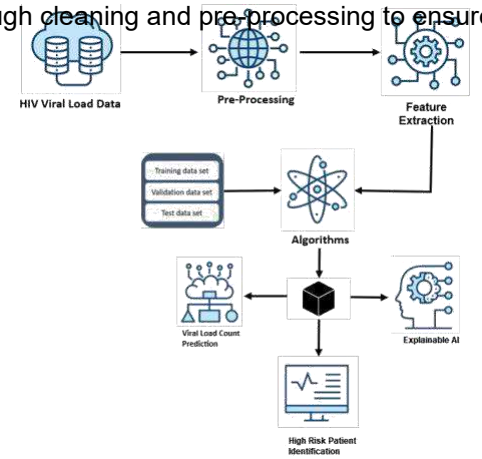


Fig. 1 Methodology

accurate and ready for analysis. This may involve removing incomplete records, handling missing values, standardizing data formats, and encoding categorical variables to ensure that the performance of the models is not negatively impacted.

2.1 Data Pre-processing

The dataset was pre-processed to study the features of interest in the data set and also to clean, organize, impute, normalize and enhance the quality of the variables, to make it more efficient and accurate for analysis. Through data imputation, missingness in variables of interest was addressed. Anonymization of patient identifiers was performed to adhere to privacy and confidentiality through label encoding to convert the unique identifiers to non-traceable numbers.

2.2 Exploratory Data Analysis (EDA)

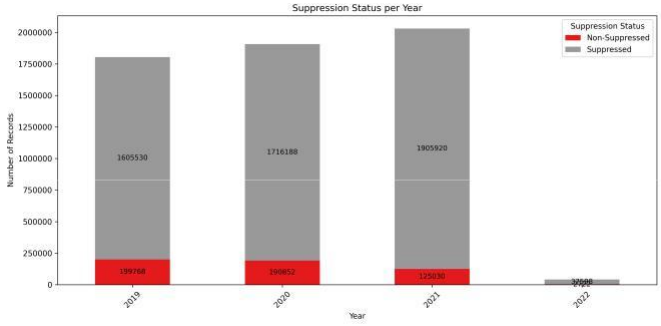


Fig. 2 Suppression and Non-suppression figures

Exploratory Data Analysis (EDA) was performed to analyze and visualize the pre-processed data to gain insights, identify patterns, and understand the relationships between variables. In the stacked bar graph (figure 2), we are shown the distribution of suppressing and non-suppressing clients over the 4-year period, with a higher number of ART clients suppressing compared to the non-suppressing clients for the period 2019 to 2022. It is also visible that the non-suppressing client numbers reduced in numbers from 2019 to 2022 showing an improvement is the HIV program in Uganda; however, this is still a huge number for the country.

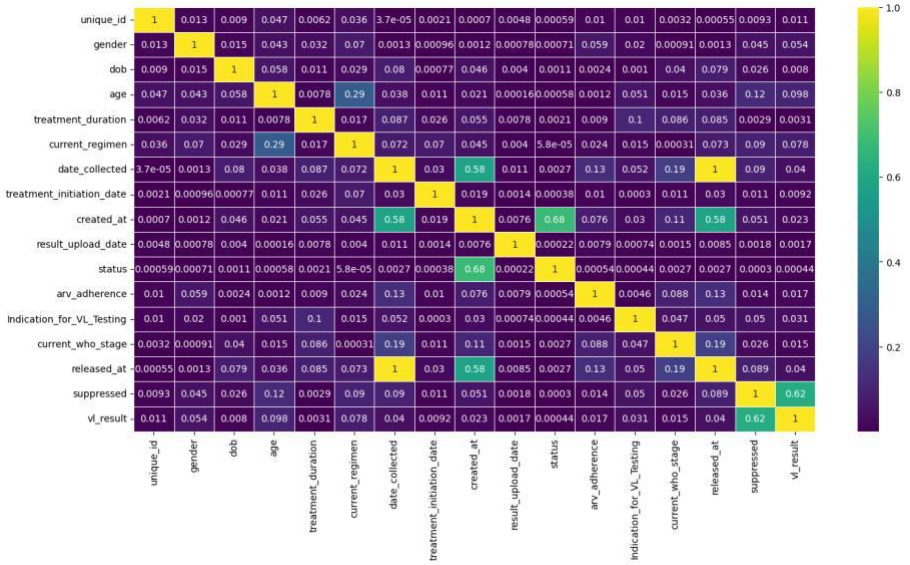


Fig. 3 Correlation Heatmap

A correlation heat map (figure 3) showing the relationships among the variables of the dataset was derived from the dataset. From the heatmap, a number of pairs stood out as highly correlated such as the viral load result and suppression variables which were also considered features of importance for the model training shown in the feature importance bar graph (Figure 4).

In the feature importance diagram; the features that ranked high for the model training include suppression, the unique id, date of results release, date of birth in descending order.

2.3 Model Training and Evaluation

Feature extraction was performed to select or create relevant features from the dataset that are most significant for the analysis or modelling task. Data was split into train, test sets and subjected to the machine algorithms by selecting the model architecture, and tuning hyper-parameters to optimize model performance.

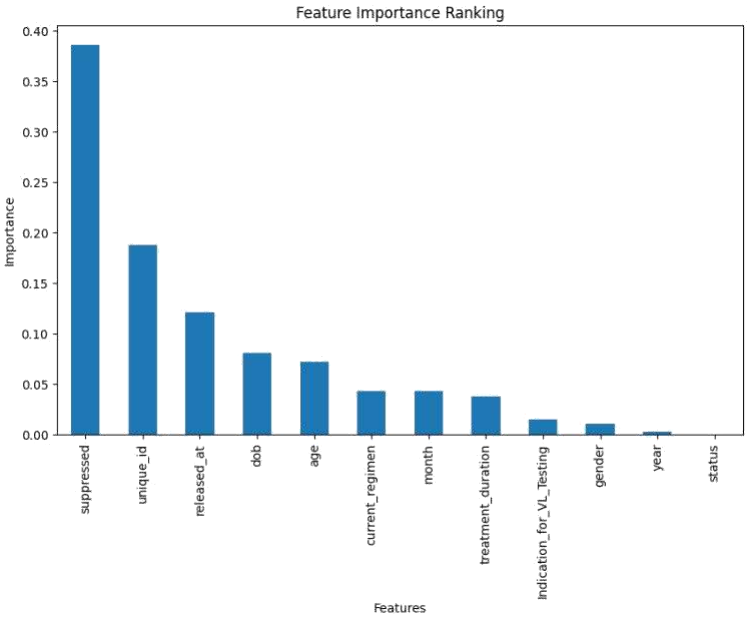


Fig. 4 Feature Importance

2.4 Ensemble Learning Techniques

Depending on whether the task is a classification or regression in nature, ensemble learning involves training of two or more machine learning algorithms. The algorithms that are applied during the ensemble learning model development are referred as “base models”, “base learners” or “weak learners”. The base models are usually designed with a single algorithm or by the combination of several algorithms. Ensemble learning typically involves these three approaches Bagging (bootstrap-aggregating), Boosting or Stacking (Blending) techniques to introduce variability in the base models during training [19].

Bagging, also known as bootstrap aggregation, is an ensemble learning technique that combines the positive attributes of bootstrapping and aggregation to produce a balanced model that improves the predictive performance of the machine learning model [20].

Boosting trains a sequence of models where each model is subjected through training on a weighted training set. Weights are assigned based on the errors of the previous models in the sequence. The main concept behind sequential training is to have each model correct the errors of its predecessor. This continues until the predefined number of trained models or some other criteria are met.

Stacking uses the predictions of base models that are provided as input to a meta-model (or meta-learner). The meta-model then takes the results of the predictions from the base models to arrive at a final prediction.

3 Results

3.1 Model Selection and Optimization

From the 5.7 million HIV Viral Load dataset that was for 2019 to 2023, the inclusion criteria of considering Antiretroviral Therapy clients was atleast 3 previous viral load results during the period under consideration which resulted in 1 million records that met the criteria. By splitting the dataset into a ratio of 70:20:10, where 70% of the data was trained, 20% was tested during model development and 10% for model validation. The data was subjected to ensemble learners by selecting the model architecture, and tuning hyper-parameters to optimize model performance for the best scores. The ensemble learner algorithms explored in this study comprised of Random Forest, KNN, Decision Tree, Linear, ADA Boost, Gradient Boost and XGBoost Regressors. During the model training, feature importance was examined and revealed as shown in figure 4, that the suppression variable was the highest ranked during the training.

The evaluation metrics of Root Mean Squared Error(R2 score), Mean Absolute Error (MAE) and Mean Squared Error (MSE) were derived to compare the performance of the 9 algorithms which included 2 ensemble learners namely; Stacking, Bagging, Random Forest, Decision tree, XGBoost, Gradient Boost, AdaBoost, KNN and Linear Regressors that were explored in the study to make an informed choice of the most suitable model based on metric evaluation. The results of the models derived as shown in Table 1, indicate that the Stacking Regressor performed best with an R2 score of 0.85, MSE of 0.41 and MAE of 0.42

To achieve satisfactory metric scores prior to model training, the data was subjected to statistical modeling that was implemented in the following ways. Before training the model with the guidance of the heatmap, highly correlated features were eliminated while focusing on features of highest importance like suppression and the target feature of viral load. During the model training, there were discoveries of skewed results indicating that the dataset had imbalances in the records and this called for application of log transformation. Log transformation is an essential approach when dealing with skewed datasets where it improves model performance and visualization when implemented on the target feature or dataset. Additionally, random sampling was per-formed on the data upon identification of data artifacts that were introducing outliers in the results and causing high error rates in the trained model. With the application of these data modeling approaches, the results were satisfactory as shown in table 1. In figure 5, we observe the blue line representing the reported viral load and yellow markers represent the next visit viral load. With this visualization, it clearly shows that majority ART clients with high viral load count are non-suppressing and require immediate intervention measures to keep them from negative effects such as developing drug resistance or other comorbidities.

3.2 ML Model Selection Accountability

The inner workings of ML models should be understood by consumers of their results in order for trust AI systems to grow. An algorithm that performs the task of providing that understanding is LIME (Local Interpretable Model-Agnostic Explanations). It

Metric/Algorithm	R-Squared Score	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Stacking Regressor	0.85	0.42	0.41
Random Forest Regressor	0.84	0.48	0.45
Bagging Regressor	0.82	0.5	0.52
Decision Tree Regressor	0.77	0.33	0.64
XGBoost Regressor	0.63	0.84	1.06
Gradient Boost Regressor	0.61	0.87	1.12
Linear Regressor	0.53	0.99	1.32
Adaboost Regressor	0.44	1.12	1.58
KNN Regressor	0.12	1.24	2.49

Table 1 ML Model & Ensemble Learner Results

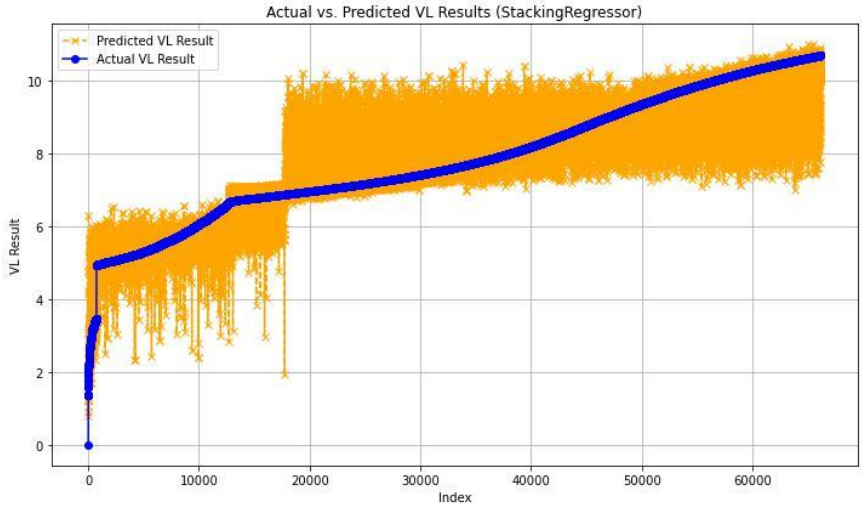


Fig. 5 Actual vs Predicted HIV viral load

provides a general framework for deciphering black boxes, explains the ”why” behind forecasts or suggestions made by AI and also addresses explainability concerns.

The LIME explainer takes the observation to be explained and the model prediction that is supposed to be interpreted. Then it generates new samples around the observation to be explained and produces the predictions of those new samples using the model.

The results from Random Forest Regressor with LIME shown in figure 6, shows the suppression variable is of most importance overall in the prediction outcome for an Antiretroviral Therapy client’s viral load prediction. For the decision tree regressor (figure 7), we see that the suppression variable is of most importance in the prediction of the ART client next viral load.

In figure 8, the GBoost LIME Explainer shows that suppression variable is an important during the prediction through the high ranking indicated in the diagram.

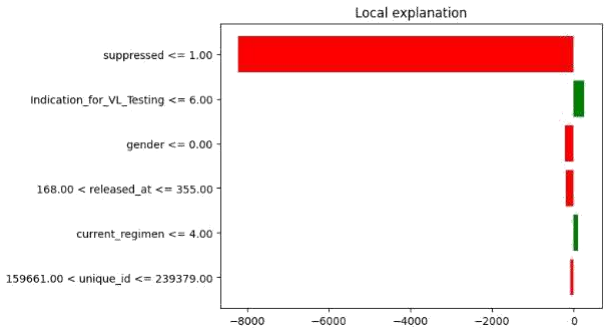


Fig. 6 RF LIME Explainer

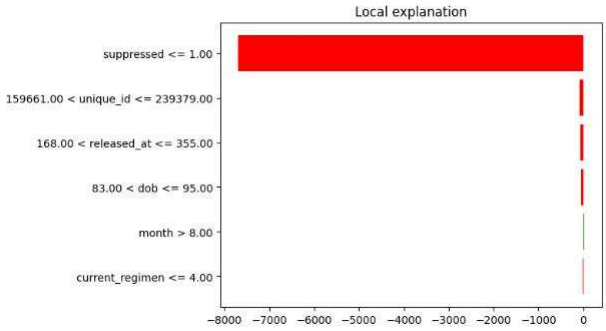


Fig. 7 DT LIME Explainer

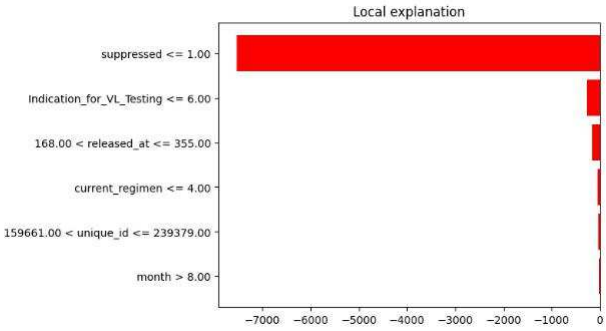


Fig. 8 GBoost LIME Explainer

For the KNN regressor in figure 9, the date of birth, treatment duration, current regimen, date of result release, patient unique identifier and age are important variables for its prediction of the viral load result.

For the linear regressor in figure 10, the year and suppression including the indication for viral load testing are the important variables for the prediction.

In figure 11, we see from the XGBoost regressor that the suppression and date of birth variables are of highest importance in the prediction of the model. The XGBoost

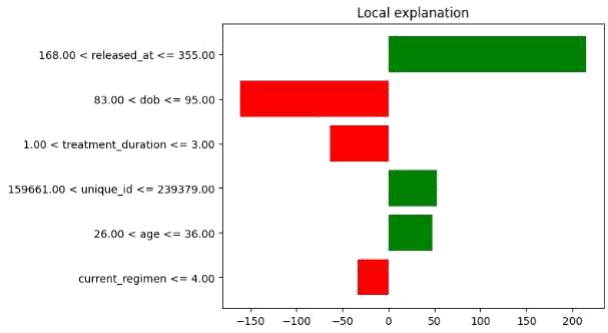


Fig. 9 KNN LIME Explainer

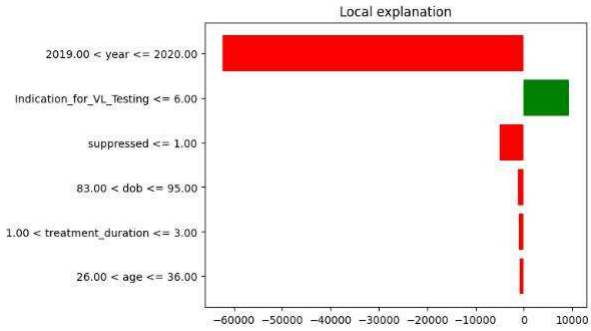


Fig. 10 Linear Regressor LIME Explainer

regression model also provided the best prediction and therefore guides the variables to consider for HIV viral load prediction.

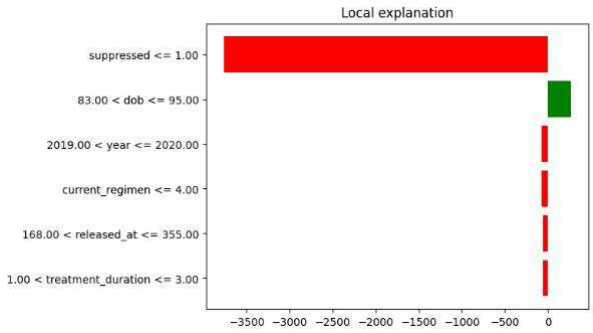


Fig. 11 XGB Regressor LIME Explainer

In figure 12, the ADA Boost Regressor shows the suppression variable is of most importance overall in the prediction outcome of the model however, the result from the model training does not rank higher than the XGBoost result.

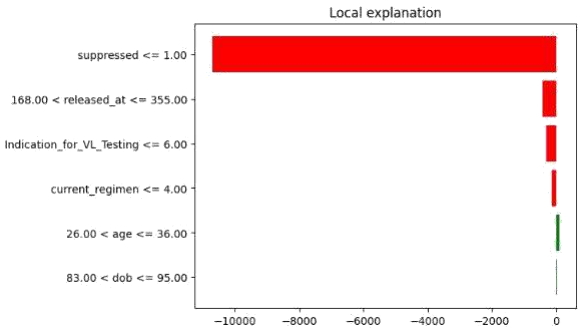


Fig. 12 ADA Boost Regressor LIME Explainer

4 Discussion

In this work, the potential of utilizing HIV viral load laboratory test data with the ensemble learners for the prediction of next visit HIV viral load result was demonstrated. The derived models provided results using secondary laboratory test data from the Central Public Health Laboratories, Uganda.

From the various illustrations of the LIME explainers of the algorithms considered for this study, and from the review of the evaluation metrics where the Stacking Regressor model performed best, and can be adopted in model deployment for decision support in a clinical or laboratory setting for ART client health outcome prediction.

Due to resource constraints of compute power, and finance; next steps would be deployment of the model in a real-world clinical setting to support decision making and capture of clinical variables on Antiretroviral Therapy clients. With additional clinical variables about ART clients, additional patterns can be discovered that would open other avenues for future studies in addition to improving decision making for health care workers.

4.1 Limitations

In spite of our approach's merits in using machine learning to predict HIV viral load for ART clients, the study had some limitations. Among the limitations experienced during the study was the unavailability of comprehensive datasets capturing both clinical and laboratory testing data about ART clients. This is an issue experienced in low resource settings due to low adoption of digital or information systems for patent management. The data available comprises mainly of laboratory HIV viral load test results of ART clients and a few programmatic features captured at health facility level. Addition of features such as socio-economic status of a client like there employment status, income level, education level, marital status, and access to pyscho-social support among others would provide a better picture for model predictions.

5 Conclusion

In this work, we demonstrated the potential of utilizing HIV viral load laboratory test data with the ensemble learners for the prediction of future viral load result. The derived models provided results from analysis was conducted using secondary laboratory test data. Due to resource constraints of time and finances, we otherwise would carry out the model deployment on a platform with high volume of ART clients. Utilization of additional variables on Antiretroviral Therapy clients such as clinical information would be beneficial and an avenue for further studies. With these additional variables, other patterns would be discovered that would present opportunities and improve decision making for health care workers.

Declarations

Ethics approval and consent to participate

Ethical approval for this study was obtained from the Research and Ethics Committee (REC) of the Uganda National Health Laboratory Services (UNHLS), Uganda (Approval No. **UNHLS-2023-71**). The study involved secondary analysis of routinely collected, de-identified patient data and was determined to pose no more than minimal risk to participants. The requirement for informed consent to participate was waived by the Research and Ethics Committee of the Uganda National Health Laboratory Services because the study used anonymized retrospective data with no direct participant contact and no personally identifiable information.

All methods were carried out in accordance with the ethical principles of the Declaration of Helsinki and all applicable institutional guidelines and national regulations governing research involving human participants and health data.

Consent for publication

Not applicable. This manuscript does not contain any identifiable individual person's data, images, or other information requiring consent for publication.

Availability of data and materials

The datasets analyzed during the current study are not publicly available because they contain sensitive health information owned by the Uganda National Health Laboratory Services. De-identified data may be made available from the corresponding author upon reasonable request and with permission from the Central Public Health Laboratories, subject to applicable ethical and institutional approvals.

Competing interests

The authors declare that they have no competing interests.

Funding

This work received support from the Research, Innovations, Services and Engagement (RISE) Grant, College of Computing and Information Sciences, Makerere University, which supported the ethical approval process for the study. The funders had no role in the study design, data collection, data analysis, interpretation of results, manuscript preparation, or the decision to submit the manuscript for publication.

Authors' contributions

MJ conceived and designed the study, performed the experiments, analyzed the data, and drafted the original manuscript. AS, JD, JNN, and RG supervised the scientific and methodological aspects of the study and contributed to manuscript review and editing. All authors read and approved the final manuscript.

Acknowledgements

The authors sincerely acknowledge the staff of Central Public Health Laboratories, for their support in data acquisition, curation, validation, and scientific guidance throughout this study.

The authors also acknowledge the Research, Innovations, Services and Engagement (RISE) Grant at the College of Computing and Information Sciences, Makerere University, for supporting the ethical approval process, and the Africa Centre of Excellence in Bioinformatics and Data Intensive Sciences (ACE), Uganda, for providing mentorship and access to high-performance computing infrastructure used for model development and training.

References

[1] G. Yebra, “Analysis of the history and spread of HIV-1 in Uganda using phylodynamics.” Journal of General Virology, vol. 96, no. 7, pp. 1890-1898, 2015, doi: 10.1099/vir.0.000107

[2] C. Watera et al., ”HIV drug resistance among adults initiating antiretrovi-ral therapy in Uganda,” Journal of Antimicrobial Chemotherapy, vol. 76, no. 9, pp. 2407–2414, May 2021. Accessed: Feb. 16, 2023. [Online]. Available: https://doi.org/10.1093/jac/dkab159

[3] I. J. Collins et al., ”Incidence of switching to second-line antiretroviral therapy and associated factors in children with HIV: An international cohort collaboration,” The Lancet HIV, vol. 6, no. 2, Feb. 2019, Art. no. e105-e115. Accessed: Feb. 16, 2023. [Online]. Available: https://doi.org/10.1016/s2352-3018(18)30319-9

[4] HIV and AIDS Prevention and Control Act, 2015. Available: https://ulii.org/akn/ug/act/2015/1/eng@2015-02-13

[5] ”Estimated number of people (all ages) living with HIV.” World Health Organi-zation (WHO). https://www.who.int/data/gho/data/indicators/indicatordetails/ GHO/estimated-number-of-people–living-with-hiv (accessed Feb. 10, 2023).

[6] ”Population, total - uganda — data.” World Bank Open Data — Data. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=UG (accessed Feb. 16, 2023).

[7] R. Mafigiri et al., “HIV prevalence and uptake of HIV/AIDS services among youths (15–24 Years) in fishing and neighboring communities of Kasensero, Rakai Dis-trict, South Western Uganda,” BMC Public Health, vol. 17, no. 1, Mar. 2017, doi: https://doi.org/10.1186/s12889-017-4166-2.

[8] “Uganda summary sheet 2020-2021 - PHIA project.” PHIA Project. https://phia.icap.columbia.edu/uganda-summary-sheet-2020-2021/ (accessed Feb. 16, 2023).

[9] “Consolidated Guidelines for the Prevention and Treatment of HIV in Uganda.” Accessed: Jul. 17, 2023. [Online]. Available: https://platform.who.int/docs/default-source/mca-documents/policydocuments/

guideline/UGA-RH-43-01-GUIDELINE-2018-eng-Final-Uganda- HIV-Guidelines.pdf

[10] J Liu, T.; Shafer, R.Web Resources for HIV type 1 Genotypic-Resistance Test Interpretation. Clin. Infect. Dis. 2006, 42, 1608–1618. [Google Scholar] [CrossRef] [Green Version]

[11] C. K. Mutai, P. E. McSharry, I. Ngaruye, and E. Musabanganji, “Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa,” BMC Medical Research Methodology, vol. 21, no. 1, Jul. 2021. Accessed: Feb. 16, 2023. [Online]. Available: https://doi.org/10.1186/s12874-021-01346-2

[12] K. Patil, S. Pawar, P. Sandhyan, and J. Kundale, “Multiple Disease Prognostication Based On Symptoms Using Machine Learning Techniques,” ITM Web of Conferences, vol. 44, p. 03008, 2022, doi: 10.1051/itmconf/20224403008.

[13] J. L. Marcus, W. C. Sewell, L. B. Balzer, and D. S. Krakower, “Artificial Intelli-gence and Machine Learning for HIV Prevention: Emerging Approaches to Ending the Epidemic,” Current HIV/AIDS Reports, vol. 17, no. 3, pp. 171–179, Apr. 2020, doi: https://doi.org/10.1007/s11904-020-00490-6.

[14] D. S. Krakower et al., “Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a mod-elling study,” The Lancet HIV, vol. 6, no. 10, pp. e696–e704, Oct. 2019, doi: https://doi.org/10.1016/s2352-3018(19)30139-0.

[15] C. Comito, D. Falcone, and A. Forestiero, “AI-Driven Clinical Decision Support: Enhancing Disease Diagnosis Exploiting Patients Similarity,” IEEE Access, vol. 10, pp. 6878–6888, 2022, doi: https://doi.org/10.1109/access.2022.3142100.

[16] S. Climer, “A machine-learning evaluation of biomarkers designed for the future of precision medicine,” Jul. 2023, doi: https://doi.org/10.1101/2023.07.09.23292430.

[17] L. Blassel, A. Zhukova, C. J. Villabona-Arenas, K. E. Atkins, S. Hu e, and O. Gascuel, “Drug resistance mutations in HIV: new bioinformatics approaches and challenges,” Current Opinion in Virology, vol. 51, pp. 56–64, Dec. 2021, doi: https://doi.org/10.1016/j.coviro.2021.09.009.

[18] M. Maskew et al.,“Machine learning to predict retention and viral suppression in South African HIV treatment cohorts,”Feb. 2021, doi: https://doi.org/10.1101/2021.02.03.21251100.

[19] https://en.wikipedia.org/wiki/Ensemble _learning

[20] http://www.baeldung.com/cs/bagging-boosting-stacking-ml-ensemble-models