Developing Machine Learning Algorithms to Predict Treatment Interruptions in HIV Care in Uganda

1309

Alex Mirugwe[1*], Solomon Ssevvume[1], Alice Namale[1], Evelyn Akello[1], Jonathan Mpango[1], Solomon Muhumuza[1] , Paul Katongole[1], Rose Baryamutuma[1], Paul Mbaka[2], Enos Sande[3], Arthur G Fitzmaurice[3], Kenneth Musenge[3]

[1]*Makerere University, School of Public Health, Monitoring & Evaluation Technical Support Program, Kampala, Uganda*

[2]*Ministry of Health, Division of Health Information, Kampala, Uganda*

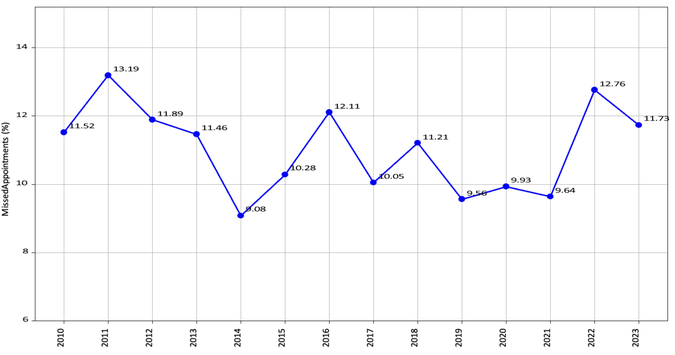
[3]*US Centers for Disease Control and Prevention, Global Health Center, Division of Global HIV & TB, Kampala, Uganda*

BACKGROUND

- Retaining people living with HIV (PLHIV) in care remains a critical challenge in achieving UNAIDS 95-95-95 targets in Uganda
    - Uganda's current status: With 94% of PLHIV knowing their HIV status, 92% receiving ART services, and 84% being virally suppressed
- Loss to follow-up contributes to poor virological suppression, increased transmission rates, and higher mortality
- This initiative explores the use of machine learning to identify high-risk clients before they disengage from care, transforming HIV care delivery into a proactive and targeted system to retain clients in care

METHODS

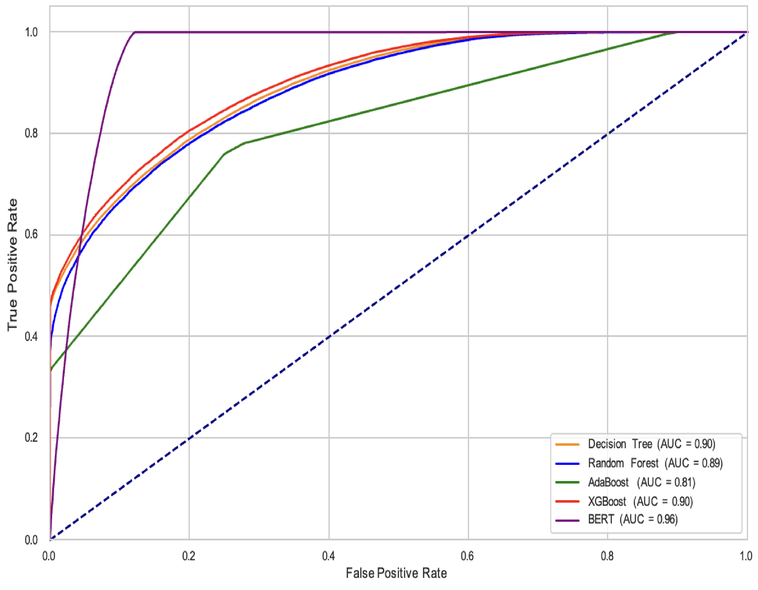
- **Study Design:** Retrospective analysis of longitudinal HIV treatment data from UgandaEMR
- **Dataset:** 66, 206 clients, 1.48 million clinical visits, from 86 facilities (2000 – 2023)
    - 43,132 (65.1%) females and 23,074 (34.9%)
    - **Predictors:** no_IIT_ever, number_visits, no_IIT_in_last12months, treatment_duration, visits_current_regimen, initiation_year, regimen_change, facility_level, entry_point, contact, WHO_stage, regimen_line, age_group, advanced_HIV, pregnancy, season, TB, sex
- **Approach:**
    - Compared traditional machine learning models (Decision Tree, Random Forest, AdaBoost, XGBoost) with deep learning (BERT)
    - Integrated sampling techniques (Undersampling, oversampling, SMOTE) to address data imbalance
    - Evaluated models using accuracy, precision, recall, F1-score, and AUC

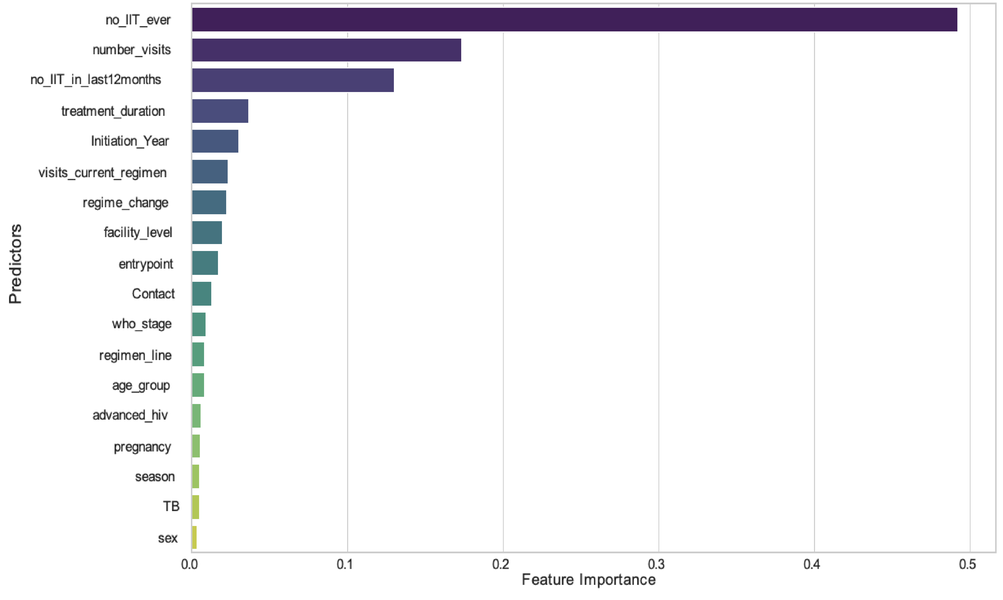


Proactive approaches leveraging ***machine learning techniques*** can effectively identify HIV clients at the ***highest risk of dropping out of care***. These predictive models enable targeted interventions that improve retention

RESULTS

- **BERT Model:** AUC: 0.96, Accuracy: 94.8%, Precision: 97.1%, Recall: 100%, F1-Score: 94.2%.
- **XGBoost with Undersampling**: AUC: 0.90, Accuracy: 80.7%, Precision: 97.1%, Recall: 80.8%, F1-Score: 88.2%





- **Key Predictors:** Historical treatment adherence (e.g., no_IIT_ever, number_visits, no_IIT_in_last12months), visit frequency, treatment duration, and regimen details
    - Longitudinal treatment history emerged as the most significant factor, while demographic variables like sex and age group were less influential
- **Missed Appointment Trends:** Persistent rate of 10.65% annually highlights the need for intervention.

CONCLUSIONS

- This study highlights the transformative potential of integrating machine learning into HIV care systems to enhance retention and achieve sustainable viral suppression
- By proactively identifying clients at the highest risk of disengaging from care, data-driven strategies address long-standing challenges in HIV care delivery
- The integration of the BERT model into UgandaEMR will enable clinicians to routinely identify high-risk clients and provide targeted interventions, significantly improving retention and care outcomes

Acknowledgment:

We acknowledge all health facilities collecting and entering data into UgandaEMR and appreciate the support of PEPFAR Implementing Partners in strengthening HIV care and treatment.