Leveraging AI for Equitable Healthcare: Predictive Modeling to Address HIV Self-Testing Disparities in Sub-Saharan Africa

Felix Emeka Anyiam (MSc.) [1], Maureen Nokuthula Sibiya (D Tech: Nursing)[2], Olarenwaju Oladimeji (PhD)[3]

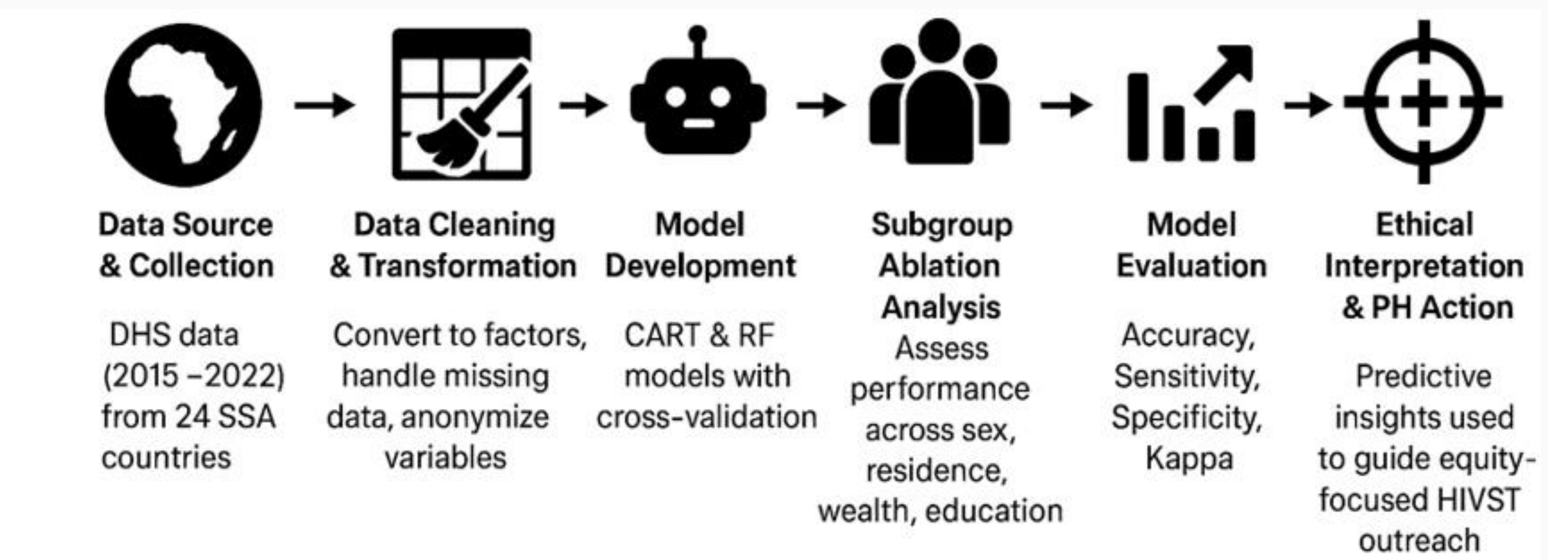
[1] *Faculty of Health Sciences, Durban University of Technology, Durban 4001, South Africa.*
[2] *Vice-Chancellor and Principal's Office, Mangosuthu University of Technology, 511 Mxenge Highway, Umlazi 4031, South Africa*
[3] *Department of Public Health, Sefako Makgatho Health Sciences University, Pretoria 0208, South Africa*

1. Introduction

- HIV self-testing (HIVST) uptake remains low among vulnerable populations (VPs) in Sub-Saharan Africa (SSA) due to socio-economic and demographic inequalities [1].
- VPs—including youth, women, low-income groups, and marginalized communities—face systemic barriers to care [2].
- Socio-demographic data improves intervention planning but poses ethical risks like profiling and stigmatization.
- We applied machine learning (ML) Classification and Regression Tree (CART) and Random Forest (RF) to predict HIVST willingness and **guide equitable, community-level strategies**.
- Our models bridge accessibility gaps while upholding privacy and ethical standards.

2. Methods

- Demographic and Health Survey (DHS) data from 38 SSA countries (2015–2022) were used to predict HIVST willingness among 594,639 individuals aged 15–49.
- Variables included age, sex, marital status, education, wealth index, residence, and country.
- A **convergent parallel mixed-methods** design combined ML outputs and community interviews.
- CART and RF models were trained using repeated 5-fold cross-validation, 4 folds are used for training, and 1 fold is used for testing.



ML Pipeline for Predicting HIVST Willingness

- **Subgroup ablation analysis** assessed fairness across sex, residence, wealth, and education.
- Data were fully anonymized to protect privacy.

3. Results

- The RF model achieved **98.69%** accuracy (Kappa 0.9653); CART achieved **97.42%** (Kappa 0.9581) (**Figure 2**).
- Key predictors: age (12,318), gender (146.6), wealth index (41.5), urban residence (36.6), Kenya (272.1), Gambia (101.3), marital status "never in union" (558.9), and education level (27.8).
- Younger, urban, wealthier, and more educated individuals showed higher HIVST willingness; rural, low-education, and disadvantaged groups had lower uptake.
- Subgroup ablation confirmed consistent performance: Accuracy 97.2–99.3%, Sensitivity 89.2–98.2%, Specificity ≥99.8%, AUC 0.980–0.996 (**Table 1, Figure 3**).

They demonstrate that the high model accuracy is not primarily driven by correct predictions in only one subgroup.

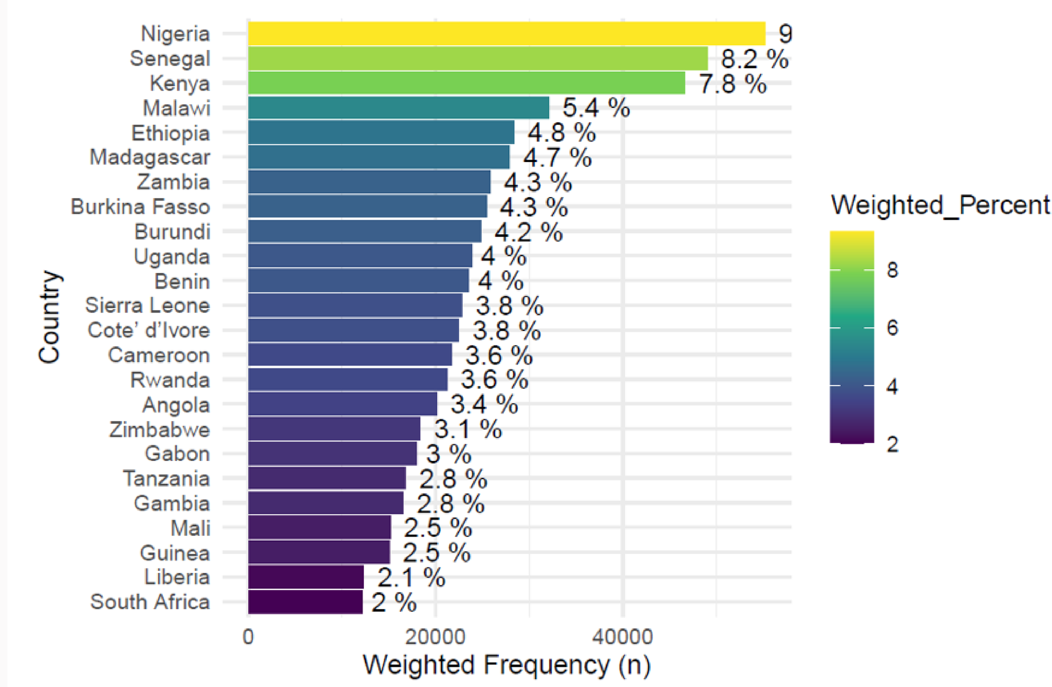


Figure 1: Weighted Frequency and Percent by Country

Table 1: Subgroup-Level Model Performance Summary

Subgroup	Accuracy	Sensitivity	Specificity	Kappa	AUC
Sex					
Male	0.982	0.892	0.999	0.932	0.985
Female	0.993	0.982	0.999	0.985	0.996
Residence					
Rural	0.989	0.942	1.000	0.963	0.989
Urban	0.984	0.955	0.999	0.965	0.995
Wealth Index					
Poorer/Poorest	0.991	0.943	1.000	0.965	0.985
Middle	0.988	0.943	1.000	0.963	0.993
Richer/Richest	0.982	0.954	0.999	0.962	0.994
Education					
No Education	0.993	0.930	1.000	0.960	0.980
Primary	0.991	0.942	1.000	0.965	0.989
Secondary	0.982	0.953	0.999	0.962	0.994
Higher	0.972	0.957	0.998	0.940	0.993

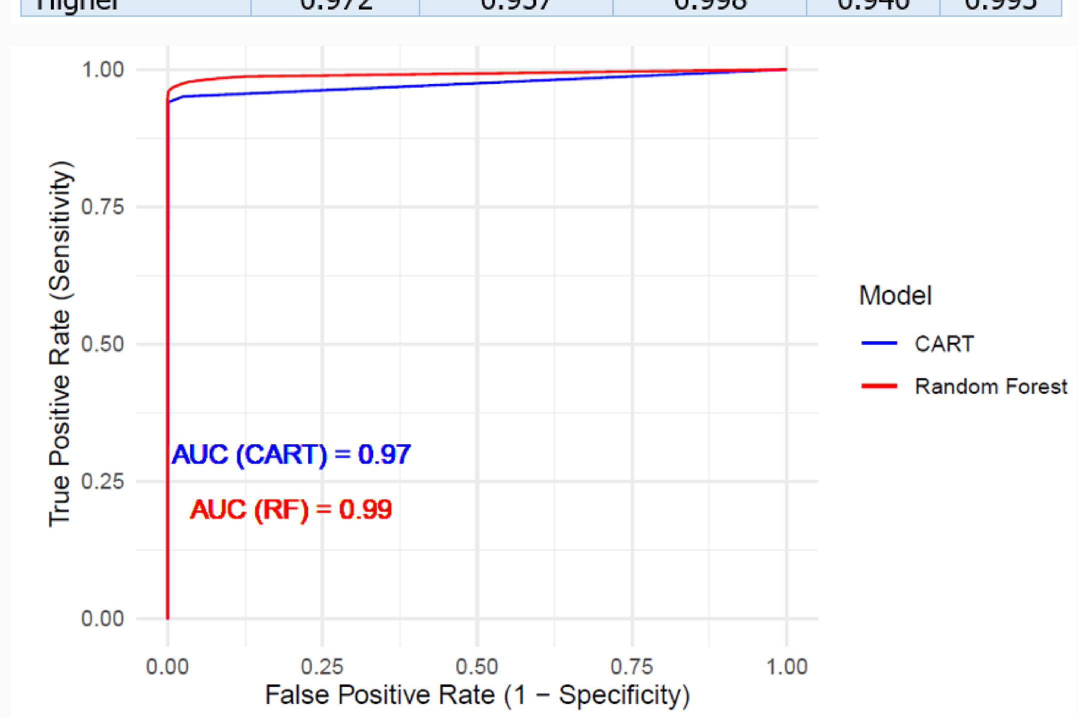


Figure 2. Comparison of AUC-ROC Curves for CART and RF Models

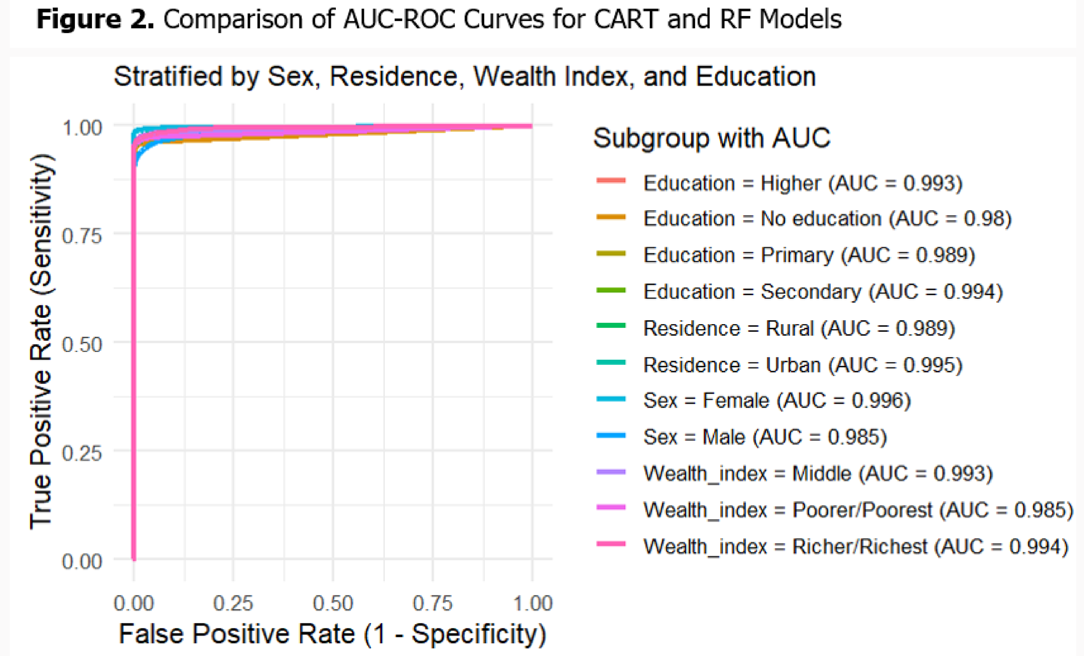


Figure 3: ROC Curves by Subgroup

4. Discussion

- Findings highlight the urgent need for equity-focused interventions to improve HIVST uptake in SSA.
- ML models revealed marked socio-demographic disparities, with lower uptake among rural, less-educated, and economically disadvantaged groups, reflecting deep-rooted structural inequities, which aligned with existing literature [3].
- Interventions should expand test kit access, use community messengers, and address misinformation.
- Subgroup ablation confirmed consistent model performance across demographics.
- Ethical risks were minimized through anonymization and population-level analysis, promoting responsible, inclusive AI for public health planning.

5. Conclusion

- This study demonstrates that ethical machine learning can identify HIVST disparities and guide targeted interventions.
- Integrating outputs into community health efforts, through NGOs, mHealth platforms, and outreach workers, can enable equitable kit distribution, stigma reduction, and resource linkage to advance health equity across SSA.

6. References

1. Anyiam FE, Sibiya MN, Oladimeji O. Factors influencing the acceptability and uptake of HIV self-testing among priority populations in Sub-Saharan Africa: a scoping review. Public Health Rev. 2025;46:1608140. doi:10.3389/phrs.2025.1608140.
2. Daniels SI, Cheng H, Gray C, Kim B, Stave CD, Midboe AM. A scoping review of implementation of health-focused interventions in vulnerable populations. Transl Behav Med 2022;12(9):935–944.
3. Marko JG, Neagu CD, Anand PB. Examining inclusivity: the use of AI and diverse populations in health and social care: a systematic review. BMC Medical Informatics and Decision Making. 2025;25(1):57.

7. Acknowledgments

I sincerely thank my PhD supervisors, Professors M.N. Sibiya and O. Oladimeji, for their invaluable guidance. I also appreciate the support of Durban University of Technology and thank the SAIL 2025 Organizers for the opportunity to share this research internationally.