Using Machine Learning to Optimize HIV Risk Prediction and Case Identification in Eastern Kenya

Presenter: Dr. Peter Kyalo

Authors: Felix Otieno Makori[1], John Kioko[1], Patrick Eshikumo[1], Peter Kyalo[1], Sam Wafula[1], Emily Koech[1]

[1]**Center for International Health, Education and Biosecurity (CIHEB)-Kenya**

Disclaimer

The findings and conclusions in this abstract are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention and other funding institutions.

CROI 2026

Outline

- Background
- Methodology
- Results
- Conclusion and Recommendations
- Acknowledgement

Background

- **UNAIDS 95–95–95 Goals**: Reduce HIV incidence by 2030, focusing on improved diagnosis, prevention, treatment and viral suppression
- **Achieving and sustaining** the UNAIDs 95-95-95 goals requires continuous identification of People living with HIV in an environment of diminishing resources
- **Limitation:** Current HIV testing eligibility screening tools lack precision in risk categorization, which limits their effectiveness in identification of high-risk individuals for provision of HIV prevention interventions
- **Machine Learning (ML)**: has the potential to improve HIV screening accuracy and refine risk categorization of patients, addressing the shortcomings of existing eligibility screening tools

Objectives

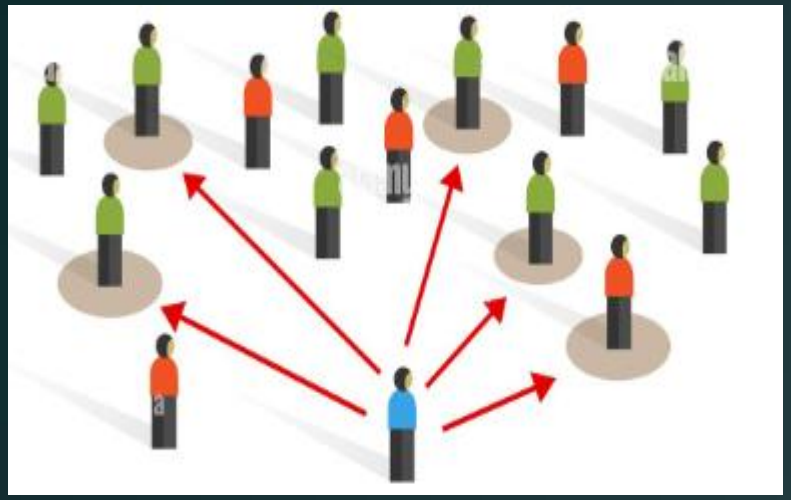
- To evaluate the potential of ML to enhance HIV case identification.

- To address the challenge of routine eligibility screening for HIV in risk categorization and enhance a focused HIV testing approach that prioritizes people at greatest risk.

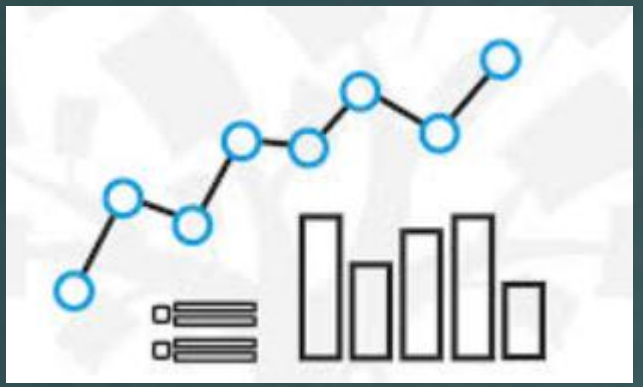
Methodology

- **Study design:** Retrospective analysis (Jan 2024–Mar 2025)
- **Setting:** 58 health facilities, Lower Eastern Kenya
- **Data source:** Kenya EMR system with ML integration





- **ML risk model inputs:**
  Age, sex, STI history, prior HIV testing, number of sexual partners, condom use, partner HIV status
- **Risk classification:**
  Weighted indicators → cumulative ML score → **Low, Moderate, High, Very High risk**

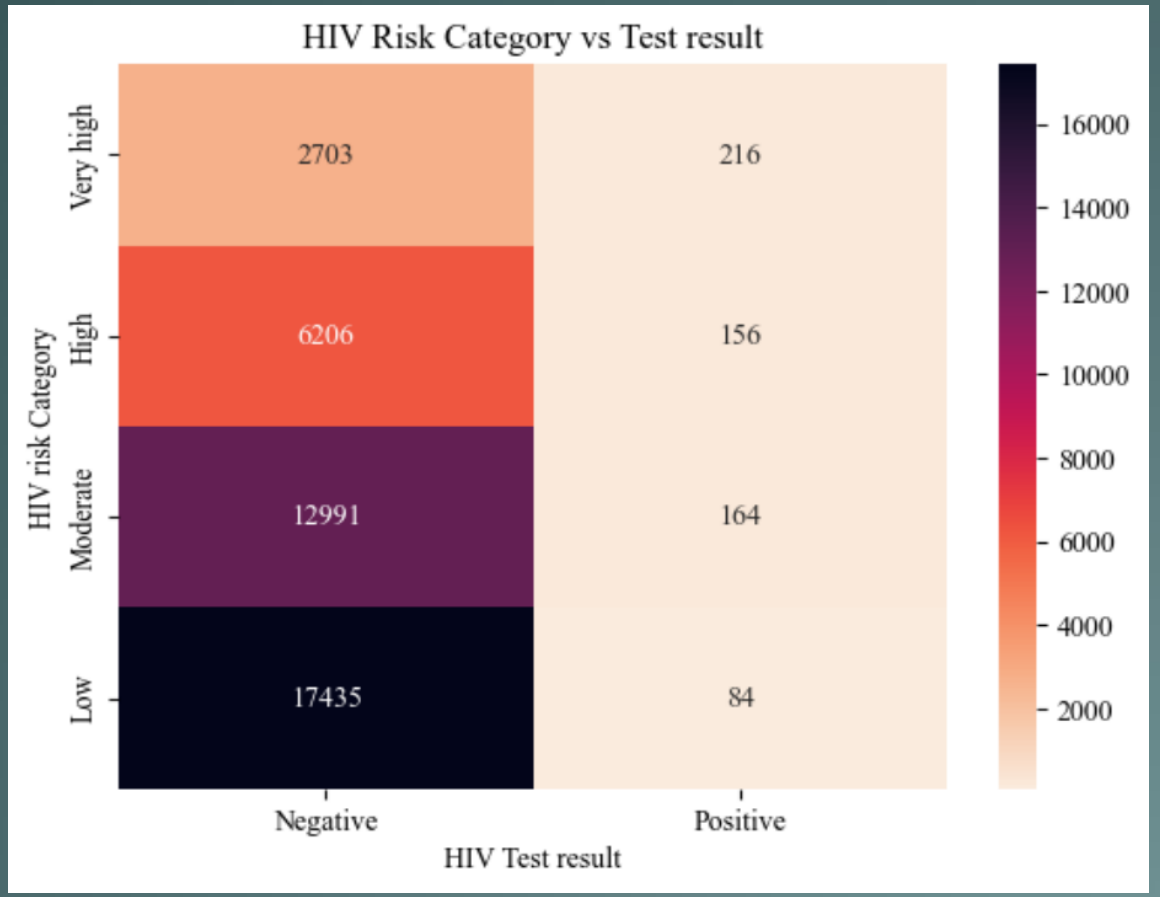


Outcome & analysis:

Primary outcome: **HIV positivity**
Main exposure: **ML-assigned risk category**
Statistical method: **Logistic regression** to identify correlates of HIV positivity

Results

- HIV testing provided to 39,995 individuals
- Overall HIV positivity: 1.6% (n = 620)
- Using machine learning, risk stratification as follows:
  - Low risk: 0.5% (84/17,519)
  - Moderate risk: 1.2% (164/13,155)
  - High risk: 2.5% (156/6,362)
  - Very high risk: 7.4% (216/2,919)
- Median age: 27 years (IQR: 22 – 35)



Results

- Odds of HIV positivity (low-risk group)
  - Moderate risk: 2.52 (aOR = 2.52, 95 CI: 1.95 – 3.32) higher odds vs low risk
  - High risk: 6.61 (aOR = 6.61, 95 CI: 5.04 – 8.67) higher odds vs low risk
  - Very high risk: 22.23 (aOR = 22.23, 95 CI: 17.03 – 29.01) higher odds vs low risk
- Protective factors:
  - Population at Elevated Risk (PER): 87% lower odds (aOR = 0.13 (95% CI: 0.09 – 0.18)) vs General population.
  - PMTCT service delivery point: 52% lower odds (aOR = 0.48(95% CI: 0.38 – 0.60)) vs OPD
  - Repeat testers: 23% lower odds (aOR = 0.77 (95% CI: 0.65 – 0.93)) vs 1st time testers

Results: Factors associated with HTS Positivity (Entire Population)

Variable	Overall	Number Positive	% Positive	P value	Unadjusted Odds Ratios	Adjusted Odds ratios
N	39995	620	2%			
Age at last visit in Years Mean (SD))	29.9 (12.7)	34.7 (13.0)			**1.02 (1.01 - 1.03)**	**1.01 (1.01 - 1.02)**
HTS Risk Category						
Low	17519	84	0.5%	<0.001	ref	ref
Moderate	13155	164	1.2%		**2.62 (2.01 - 3.41)**	**2.52 (1.95 - 3.32)**
High	6362	156	2.5%		**5.21 (4.0 - 6.81)**	**6.61 (5.04 - 8.67)**
Very High	2919	216	7.4%		**16.59 (12.85 - 21.41)**	**22.23 (17.03 - 29.01)**
Sex						
Female	30423	425	1%	<0.001	ref	ref
Male	9532	195	2%		**1.47 (1.24 - 1.75)**	0.85 (0.71 - 1.03)
Population type						
GP	34597	575	2%	<0.001	ref	ref
PER	5358	45	1%		**0.50 (0.37 - 0.68)**	**0.13 (0.09 - 0.18)**
Department						
OPD	19272	416	2%	<0.001	ref	ref
Others	4565	91	2%		0.92 (0.73 - 1.16)	1.36 (1.07 - 1.74)
PMTCT	15498	113	1%		**0.34 (0.27 - 0.42)**	**0.48 (0.38 - 0.60)**
Tested Before						
No	9222	222	2%	<0.001	ref	ref
Yes	30733	398	1%		**0.53 (0.45 - 0.63)**	**0.77 (0.65 - 0.93)**

Results: Factors associated with HTS Positivity (15+)

Variable	Overall	Number Positive		P value	Adjusted Odds ratios
N	38372	602	2%		
Age at last visit in Years Mean (SD))	30.8 (12.2)				**1.01 (1.00 - 1.02)**
HTS Risk Category					
Low	16801	79	0.5%	<0.001	ref
Moderate	12518	156	1.2%		**2.60 (1.98 - 3.42)**
High	6200	153	2.5%		**6.90 (5.22 - 9.09)**
Very High	2853	214	7.5%		**23.38 (17.80 - 30.71)**
Sex					
Female	29457	416	1%	<0.001	ref
Male	8915	186	2%		0.84 (0.70 - 1.02)
Population type					
GP	33016	557	2%	<0.001	ref
PER	5356	45	1%		**0.12 (0.09 - 0.17)**
Department					
OPD	18527	399	2%	<0.001	ref
Others	4366	90	2%		**1.52 (1.19 - 1.94)**
VTP	15479	113	1%		**0.47 (0.37 - 0.60)**
Tested Before					
No	7943	204	3%	<0.001	ref
Yes	30429	398	1%		**0.76 (0.63 - 0.91)**

Adjusted results are similar for the population above 15 years only

Conclusion

- ML enhances the precision of HIV screening, allowing for more accurate risk categorization of patients.
- Improved risk categorization supports efficient resource allocation and enables targeted public health interventions

Recommendations

- Further integration of ML in HTS across more regions.
- Continued evaluation of ML models to refine accuracy and effectiveness.
- Policy support for ML use in HIV prevention programs to optimize intervention strategies.

Acknowledgements

- County Department of Health – Machakos, Kitui and Makueni
- Ministry of Health
- Palladium – technical partner supporting Kenya EMR
- PEPFAR - CDC Kenya for the financial and technical support
- CIHEB –Kenya colleagues

This work has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under Grant Number NU2GGH002333

