Improving case finding efficiency through use of machine learning in Kenya

Authors: J. Friedman¹, M. Ndisha², B. Otieno³, J. Mwangi², F. Miruka², B. Makone², J. Odhiambo³, P. Mwele³, K. Muthoka³, E. Munene³, D. Kimanga²

Affiliation: ¹Palladium LLC, Washington DC, United States, ²Centers for Disease Control, Nairobi, Kenya, ³Palladium Kenya, Nairobi, Kenya,

INTRODUCTION

Overview of the epidemic

- Kenya approaches epidemic control, as such HIV positivity has reduced to < 1%
- Supply disruptions for HIV test kits and resource reallocations have resulted to decreasing HIV testing volumes
- We developed and deployed a risk profiling machine learning (ML) model to maximize yield from scarce HIV testing resources.

Approach of the model

- The ML model uses probabilities to assign patients "highest," "high," "medium" and "low" into categories of risk.
- The ML model is integrated in the electronic medical records (EMR)
- During HIV testing, health providers carry HIV risk assessments to identify those eligible for testing.
- Risk scores generated by ML are used to alert providers and recommend testing for clients in "highest," "high," and "medium" risk categories



METHODS

HIV Testing Data Collection and Evolution of models

Continued Model Evaluation

Machine Learning implementation & Model-Building Process



MODEL RESULTS

The Gain chart in Figure 4 shows sensitivity for a fixed proportion target for the best performing XGBoost model. 62% of all positive tests were

The ten most important features included a mix of demographic, clinical, behavioral, temporal and locational attributes as shown in Figure 5.

Figure 6



among the 10% of patients with the highest ML risk score. Over 92% of all positives were among the 50% of highest ML risk scores.

Figure 6 shows the impact of each feature in the model. If a client tested is part of an Index testing strategy, is divorced, or presumed to have TB following screening, to suggest a client is likely to test positive.

Feature value Low

The XGBoost model performed best among all model types (AUC 87.8). The model successfully concentrated positive tests among high-risk scores: 75% of all positive tests occurred among the 18.6% highest risk scores. Table 1 below shows comprehensive results

Risk	Number of	Cumulative	Number of	Positivity	Cumulative	
category	Tests	Percent of All	Positive	Rate	Percent of	
		Tests	Tests		Positive Tests	
Highest Risk	1,876	5.6%	572	30.5%	50%	
High Risk	4,350	18.6%	286	6.6%	75%	
Medium Risk	12,681	56.4%	228	1.8%	95%	
Low Risk	14,594	100%	57	0.4%	100%	
Table 1						

Program Implementation and Evaluation

Field Results of the Model: Apr - May 2023

Figure 4

Collectively, Patients classified in the "highest," "high," and "medium" risk categories encompass 81.3% of patients with HIV positive results and 50% of the total tests conducted

Positives and Negatives by Risk Category								
	Highest Risk	High Risk	Medium Risk	Low Risk	Total			
# Tests	2,804	6,143	16,654	24,707	50,308			
# Positives	436	391	506	306	1,639			
# Negatives	2,368	5,752	16,148	24,401	48,669			
Precision (positivity rate)	16%	6%	3%	1%	3%			
Precision (positivity rate)								

Table 2: Risk categorization for patients who were tested for HIV in Kenya from April 2023 to May 2023 and uploaded to NDW.

Presented at IAS 2023, the 12th IAS Conference on HIV Science

Implementation at the site: KenyaEMR

								HTS Initial Form
	HTS date: 28-Nov-2021	00 - 21 - 16 -]			HTS Provider	User, Super	*
Population Type								
Population type : People who inject drugs Men who have sex with People in prison and oth Transgender Other	General Population men ter closed settings	Key Population	Priority Population					
Does client have any disabi Please indicated the type of Hearing impairment Visual impairment Mentally Challenged Physically Challenged Other	ility? ● Yes ○ No f disability :							
HIV Test History								
Has the client ever been tes Has the client done HIV sell Is client willing to share rest	ted for HIV by a HTS provider f test in the last 12 months? ults? O Yes No	? ● Yes ○ No ● Yes ○ No	Duration in months since	the last test: 20				
Setting								
HTS Setting: Fa	cility O Community*							
HTS Approach	Provider Initiated Test	ing(PITC) 🗸						
HTS Strategy								
Indicate HTS strategy used	VI:Integrated VCT C	enter 🗸						
HTS Entry Point								
Indicate HTS entry point:	VCT	~						
HIV Testing								
Has consent been given?	Yes O No*							
Client tested as?	Individual O Couple							
Get risk category	High Risk Clie Lot Number	nt This client has a	high probability of a HIV positi	ve test result. Testing	is strongly recommende	ed		
Final Results:	 Results given to cli 	ient? V						
Couple is discordant:	⊖yes ⊖no .n/A							
TB Screening								
Tb Screening results:	○ No TB Signs ○ Presum	ed TB ONot Done O	On TB Treatment					
Remarks								
]			

Figure 7

- Integrating the ML model into KenyaEMR, an OpenMRS based EMR (Fig 7), enables providers to generate risk predictions even in remote offline settings.
- Regular model updates are installed in the EMR through the normal upgrade process.
- The introduction of HTS-ML in Kenya improves efficiency of HTS providers by maximizing positivity while minimizing the number of tests performed.
- Adoption of ML in HIV testing catalyzed Kenya's efforts towards efficient case finding as part of epidemic control activities.

Conclusion

The model's effectiveness in segmenting patients by positivity risk demonstrates the model capability to support the HIV eligibility screening process and improving HIV case finding.

High