

**Table S5.** Characteristics of the selected studies

First Author, Year	Country	Objective	Population	Sample Size	Study type / Data Source
1. Roche et al., (2024) [55]	Kenya	To evaluate the effectiveness of AI technology in interpreting HIVST results	Male and females ≥18 years	1500 testers	Intervention study using primary data
2. Saha et al., (2024) [29]	India	To use ensemble ML for early-stage HIV diagnosis	General and KPs Age 12 to 80 years	698 total samples	Retrospective analysis of secondary data Kaggle
3. Cobre et al., (2024) [30]	-Brazil -Mozambique	To develop, validate, and evaluate ML in predicting HIV/AIDS diagnosis and other diseases	Patients' data from a general hospital Median age = 32	6418 patient data	Retrospective study / Patients' data from a general hospital collected in 2021
4. Alie, (2024) [31]	Ethiopia	To identify HIV predictors for improved HIV testing	Ethiopian adolescent DHS (EDHS) data	4502	Retrospective analysis of secondary data from EDHS
5. Latt et al., (2024) [56]	Australia	To identify individuals at the highest risk of HIV/STIs with an AI-risk-based assessment tool	Self-reported hospital data from the general population and MSM aged 25 to 35.	216 252 for HIV cases	Retrospective analysis of secondary data from hospital records
6. Nguyen et al., (2024) [46]	-Dominican Republic -Dominican Republic -India -Haiti -Guinea -Ethiopia -Cameroon -Angola	To predict HIV risk while prioritizing privacy and improved communication using lightweight federated learning	-Behavioral, clinical HIV/STIs and demographic data of 8 countries	168,459 records from 8 countries	Retrospective analysis of DHS datasets from 8 countries
7. Ni et al. (2024) [41]	China	-To evaluate the effectiveness of secondary distribution of HIVST through ML compared to the traditional approach	- MSM age ≥18 years	-1828 MSM	Primary study Quasi-experimental
8. Cheah et al. (2024) [60]	Malaysia	-To test the feasibility and acceptability of AI chatbots for promoting the uptake of HIV testing among MSM	MSM age ≥18 years	14 MSM	-Mixed-primary study
9. Majam et al. (2023) [42]	South Africa	To evaluate the accuracy of the ML-based HIV risk assessment tool	-Adult males and females from 2 trials	-Trial 1=1026 -Trial 2=593	-Primary study using data from 2 trials
10. Gupta et al., (2023) [50]	USA	To use AI for accurate detection and interpretation of HIV Transforming Rapid Diagnostic Tests (RDTs) LMIC	-HIV RTD images	888 HIV cases	-Cross-sectional study -Primary data -RDT images

11. Biri Makota et al. (2023) [35]	Zimbabwe	To determine the common risk factors of HIV positivity in Zimbabwe.	ZDHS 2005–06, 2010–11, and 2015 Youth 15-24 Adult 24-54	4457 -HIV+ 2215-HIV-  Total =6672	Retrospective analysis of ZDHS
12. Burns et al., (2023) [23]	USA	To develop an HIV risk prediction model for rapid identification of individuals vulnerable to HIV diagnosis.	EHR data from 2014 to 2016 and 2016 and 2018	Total eligible patients -998 787	Retrospective analysis of ERH data from Duke University Health System
13. Friedman et al., (2023) [38]	USA	To develop a predictive model for HIV screening among women	-Women aged 18 years and above underwent HIV testing within hospitals in Chicago from 2014 to 2020.	-48 HIV+ -192 HIV-	Retrospective analysis of EHRs from 2014 to 2020
14. Mutai et al., (2023) [58]	13 PHIA SSA countries (Tanzania, Rwanda, Uganda, Cameroon, Zimbabwe, Zambia, Ethiopia, Malawi, Cote d'Ivoire, Namibia, Lesotho, Swaziland & Kenya)	To identify a cluster of countries-based sociodemographic predictors of HIV screening using UML in SSA.	Individuals tested for HIV from 13 PHIA countries	146,733 male and 155,622 female from the combined PHIA datasets	Retrospective analysis of secondary data using survey data from PHIA countries
15. Jing et al. (2023) [39]	China	To develop ensemble ML to identify key influencers among MSM for secondary distribution of HIVST kits.	MSM	309 MSM	Retrospective analysis of dataset from RCT of HIVST secondary distribution in Zhuhai, China
16. Xu et al. (2022) [73]	Australia	To develop web-based HIV and STI risk prediction using ML	Men and women age 18 years and older	The HIV dataset consists of 88,642 participants	Retrospective analysis of HER data
17. Chikusi et al., (2022) [13]	Tanzania	To develop ML models for the Prediction and Visualization of HIV Index Testing in Northern Tanzania	Consist of males and females with or without HIV	6346 samples and 11 features were collected.	-Experimental study -Primary data
18. Chingombe et al., (2022) [47]	Zimbabwe	To apply multiple ML algorithms in predicting HIV Status among MSM	MSM from Zimbabwean Ministry of Health Data -ICAP study 2018	1538 MSM	Retrospective analysis of Zimbabwean Ministry of Health -ICAP study 2018 data
19. Chingombe et al., 2022 [36]	Zimbabwe	To predict HIV status using ML for targeting intervention.	-males and females aged 15 and above who participated in the Zimbabwean National survey	20,577 individuals	Retrospective analysis of Zimbabwean National Survey Dataset
20. Dong et al., (2022) [37]	China	To develop and validate an integrated prediction model for predicting HIV risk among MSM.	MSM- biologically male and aged 18 years and above	547 included in the statistical analysis	-Cross-sectional study -Primary study

21. Belete et al., (2022) [49]	Ethiopia	To evaluate the effectiveness of DL in predicting HIV status	EDHS	83,100 dataset instances collected	Retrospective analysis of EDHS dataset from the years 2000, 2005, 2011, and 2016
22. Bender et al., (2022) [52]	Peru	-To identify biomarkers and socio-behavioral covariates predicting HIV acquisition.	A sample drawn from the Sabes study in Peru (MSM and transgender women)	90 cases 270 controls	-Nested Case-Control -Primary study
23. Aribe et al., (2022) [51]	Philippines	Using DL to forecast HIV/AIDS in the Philippines.	COVID-19 dataset from January 2020 to February 2022 was used to forecast HIV spread in the country.	Not indicated	Retrospective analysis of Covid-19-HIV dataset from the Department of Health website and registry
24. He et al., (2022) [32]	China	To develop and validate a predictive model for HIV acquisition among MSM in China.	MSM aged 15 years and above	6,346 MSM were included	Retrospective analysis of Sentinel surveillance data
25. Laybohr Kamara et al., (2022) [53]	Mano River Union countries (Liberia, Sierra Leone, Cote d'Ivoire and Guinea)	To determine the prevalence of HIV patterns in the Mano River Union countries using ML	DHS dataset from the MRU countries from 2005 to 2020	158,408 respondents from 2005 to 2020	Retrospective analysis of DHS and AIDS Indicator Survey data at the DHS STATcompiler database
26. Xu et al. (2022) [24]	Australia	To develop several MLs in predicting timely clinic attendance and the uptake of HIV/STI testing post-clinic reminder messages	MSM aged 18 years and above	3044 consultations among 1627 MSM	Retrospective analysis of secondary data from the short message service reminder program in Melbourne Sexual Health Centre
27. Zucker et al., (2022) [57]	USA	To improve HIV testing in the era of Preexposure Prophylaxis (PrEP) and Immediate Antiretroviral Therapy (iART) using ML	Males and females 25 to 35 years	38 850 test results from 35 500 patients	Retrospective analysis of EHR data
28. Orel et al., (2022)-[40]	Southern and Eastern Africa (Angola, Burundi, Ethiopia, Lesotho, Malawi, Mozambique, Namibia, Rwanda, Zambia and Zimbabwe)	To predict HIV status using socio-behavioral characteristics in East and Southern Africa	Male and female participants from 10 DHS datasets	55,151 males and 69,626 females aged 15 -64 years	Retrospective analysis of DHS datasets
29. Mutai et al., (2021) [10]	4 PHIA SSA countries (Tanzania, Zambia, Malawi and Eswatini)	To identify HIV predictors as well as individuals at the highest risk of HIV.	Individuals tested for HIV from 4 PHIA countries	41,939 male and 45,105 female from the combined PHIA datasets	Retrospective analysis of cross-section survey data from PHIA countries
30. Haas et al., (2021) [62]	USA	-To develop a tool for automated HIV risk estimation for HIV testing	Males and females, both HIV+ and HIV- from the MIMMIC IV dataset publicly available dataset	-38765 males -46451 females -525,000 hospital stays	Retrospective analysis of the MIMMIC IV dataset publicly available dataset

31. Jing <i>et al.</i> , (2021) [63]	China	-To enhance the uptake of HIVST through secondary distribution for economic benefits	-MSM	-309 indexes -289 alters	-Retrospective modelling among Chinese MSM
32. Bao <i>et al.</i> , (2021) [34]	Australia	To develop and evaluate ML models in predicting HIV/SIT diagnosis among MSM	clinical records of 21,273 Australian MSM during 2011–2017	21,273 Australian MSM	Retrospective analysis of clinical records of Australian MSM during 2011–2017
33. Elkhadrawi <i>et al.</i> , (2021) [22]	USA	To classify false-positive HIV screening results using ML	-Lab samples -Both HIV-positive and negative	60,587 samples	Retrospective analysis of lab samples from 2016 to 2018
34. Ovalle <i>et al.</i> , (2021) [20]	USA	To leverage social media messages to identify health risk behavior for MSM using ML.	MSM	51 MSM	-Experimental design -Primary data  -Social media data
35. Turbé <i>et al.</i> , (2021) [48]	South Africa	To classify RDT test images using DL	HIV testers in rural South Africa	11374 samples	-Intervention study -Primary / -Captured RDTs
36. Weissman <i>et al.</i> , (2021) [68]	USA	To identify HIV predictors for the missed opportunities for early HIV diagnosis using ML	PLHIV aged 18 years and above	4725 PLHIV were included in the analysis	Retrospective analysis of EHR dataset
37. Balzer <i>et al.</i> , (2020) [33]	Kenya and Uganda	To identify HIV individuals at high risk of HIV in rural Kenya and Uganda using ML	16 rural communities from Kenya and Uganda aged 15 years and above participated in the UTT trial study in SSA	A total 75 558 persons who	Retrospective analysis of 2013 to 2017 UUT trial study data (Kenya and Uganda)
38. Scott <i>et al.</i> , (2020) [64]	USA	To develop an HIV risk assessment tool for MSM	MSM aged <35, ≥35	4067 MSM	Retrospective analysis for longitudinal data
39. Adeli <i>et al.</i> , (2019) [65]	USA	To predict HIV diagnosis using brain MRI reports	Healthy individuals, people with alcohol use disorder (AUD), PLHIV and those who have both AUD and HIV; aged 25 to 75	-199 control -222 with AUD -68 PLHIV -60 both AUD and HIV	-Comparative analysis of imaging data from the 3T SIGNA system of the
40. Oladokun <i>et al.</i> , (2019) [14]	South Africa	To predict HIV status among women in South Africa using ML	Women aged 15 to 49 years from the 2016 DHS dataset	7808 women aged 15-49 (DHS data)	Retrospective analysis of the DHS dataset
41. Xiang <i>et al.</i> , (2019) [21]	USA	To detect previous unknown HI status among MSM using ML	MSM aged 16 to 29 years	378 MSM	Retrospective analysis of multiple social network data
42. Demey <i>et al.</i> , (2019) [25]	France	To improve HIV diagnosis strategy using ML	CHIV testing episodes performed at Amiens University Medical Center	56,682 HIV serological assay results	Retrospective analysis of serological assay results data between 2012 and 2018
43. Ahlström <i>et al.</i> , (2019) [15]	Denmark	To predict HIV status from status nationwide electronic registry data using ML	Adult males and females aged 16+ (both HIV positive and negative)	A total of 4,384,178 individuals	Retrospective analysis of Nationwide Electronic Registry data
44. Rice <i>et al.</i> , (2018) [66]	USA	To improve HIV prevention among homeless youth using AI	Homeless youth aged 16 to 24	-55 in AI group -62 in the comparison group	-Quasiexperimental design among homeless youth in Los Angeles
45. Feller <i>et al.</i> , (2018) [59]	USA	To examine whether natural language processing will improve predictive models of HIV diagnosis	Clinical data notes within EHRs between 2006 and 2015 -Both HIV-positive and negative	543 HIV+ 181 HIV-	Retrospective analysis of clinical data notes within EHRs between 2006 and 2015
46. Van Heeden <i>et al.</i> , (2017) [17]	South Africa	To use conversational agents for improved rapid HIV counseling and testing.	Male and female testers (average age =30 years)	-10 testers	-Intervention study -Primary data gathered from patients' telegram conversations

47. Pan et al., (2017) [19]	USA	To identify the most relevant variables predicting HIV testing uptake using ML approach	Participants from Substance Use Disorder Treatment Programs in the US (age 18 years and above)	1281 HIV-negative or status-unknown participants	Retrospective analysis of National Institute on Drug Abuse Clinical Trials Network HIV testing and counseling study dataset
48. Wade et al., (2015) [45]	USA	To predict HIV positivity using brain 3D profiles	Elderly HIV+ and HIV- participants (mean age =65 years)	-63 HIV+ -31 Control	-Comparative study of elderly individuals from the San Francisco Bay area
49. Hoenigl et al. (2015) [43]	USA	To develop and validate HIV risk score (San Diego Early Test (SDET) score)	MSM (median age of 30)	8326 unique MSM	A retrospective analysis of MSM who tested for acute and early HIV (AEH) in San Diego from 2008 -2014.
50. Hoots et al. (2013) [44]	USA	To predict HIV among undiagnosed sexual partners	HIV-infected persons & their partners report, age 15-65 (female, MSM/W)	2100 total sample	Retrospective analysis of Disease Information System records from 2 regions
51. Menza et al., (2010) [67]	USA	To develop and validate prediction models for HIV acquisition among MSM	MSM	1903 MSM who tested for HIV	Retrospective analysis of Public Health – Seattle and King Country STD clinic EHR

AI: Artificial Intelligence; AIDS: Acquired Immunodeficiency Syndrome; CDC: Center for Disease Control and Prevention; CHIV: Centaur® HIV; Covid-19: Coronavirus Disease 2019; DHS: Demographic Health Survey; EDHS: Ethiopia Demographic and Health Survey; EHR: Electronic Health Record; HIV: Human Immunodeficiency Virus; HIVST: HIV Self-Testing; ICAP: International Center for AIDS Care and Treatment Programs; LMIC: Low and Middle-Income Countries; ML: Machine Learning; MRU: Mano River Union; MSM: Men who have Sex with Men; PHIA: Population-based HIV Impact Assessment; RCT: Randomized Control Trials; SSA: Sub-Saharan Africa; STIs: Sexually Transmitted Infections; USA: United States of America; UTT: Universal Treat and Test; ZDHS: Zimbabwe Demographic and Health Survey.

**Table S6.** Types of machine learning algorithms applied for enhanced HIV testing

First Author, Year	Types of ML algorithm applied	Methodology	Problem/ Type of intervention	Key findings
1. Roche et al., (2024) [55]	HealthPulse AI	Both	-Addressing misinterpretation of HIVST using ML	-Humans (PPV-100%) were more effective in reading negative test results compared to the AI interpretation (PPV-82%) interpretation - The AI algorithm (NPV-100%) was more effective in reading positive test results compared to humans (NPV-99.9%) interpretation
2. Saha et al., (2024) [29]	-RF -LR -XGBoost -KNN -AdaBoost -Extra tree -Ensemble learning	ML	-Early HIV diagnosis from hospital data using ensemble learning	-Developed a model to enhance early diagnosis of HIV -The ensemble learning outperforms individual models and existing models from previous studies. -Accuracy of 97.86
3. Cobre et al., (2024) [30]	-PLS-DA -ANN -KNN -SVM -Class analogy (SIMCA) -XGBoost -LR	ML	-Predicting HIV diagnosis	-Successfully developed and validated several ML algorithms for predicting HIV diagnosis -Mean corpuscular volume, platelets, neutrophils, and mean platelet volume were diagnostic indicators for HIV -The PLS-DA outperformed other models with 97% accuracy
4. Alie, (2024) [31]	-J48 DT -RF -KNN -SVM -LR -NB -MLP	ML	-HIV risk prediction for improved HIV testing	-The J48 outperformed other algorithms with an accuracy of 81.29% -The 5 top predictors of HIV are: - Age -Knowledge of HIV testing location -age at first sexual encounter -Recent sexual activity -Exposure to family planning
5. Latt et al., (2024) [56]	MySTIRisk ML models	ML	-HIV/STI risk prediction	-35% of participants were identified as having a high risk of STI, of which 86% were HIV cases. -The odds of HIV positivity were significantly higher among those classified as high-risk groups.
6. Nguyen et al., (2024) [46]	RF federated learning	ML	-HIV risk prediction	-A developed model that can predict HIV risk from existing data with high data privacy -The RF federated learning outperformed existing models from previous studies -The developed model can be adapted globally across different healthcare disciplines for HIV/STI risk prediction
7. Ni et al. (2024) [41]	-LR -DT -RF -SVM	Both	-Enhancing the uptake of HIVST	-The ML model motivated more alters to conduct HIVST -The difference between the ML model and the empirical scale was not significant
8. Cheah et al. (2024) [60]	AI chatbot	ML	-Enhancing HIVST and counselling	-93% of the participants perceived that the chatbot provided comprehensive information about HIV testing (performance expectancy) -100% of the participants stated the chatbot was easy to use; effective in avoiding stigma; could increase the frequency of HIV testing -79% were willing to continue using the chatbot -They suggested the inclusion of local language
9. Majam et al. (2023) [42]	LR (glm) Bayesian LR (bayesglm) Regularized LR (glmnet)	ML	-HIV risk prediction	-Trained models had high accuracy AUROCs ranging from 78.5% to 82.8% -Boosted Tree model performed best with a sensitivity of 84%

	SVM with linear weights (svmLinearWeights) RF (ranger) Gradient Boosted Tree Model (xgbTree)			-Age, duration since last HIV test, and number of male sexual partners were the most consistent predictors of HIV
10. Gupta et al., (2023) [50]	-HealthPulse AI algorithms -DL	ML	-False HIV results -HIV accuracy testing -using photos to interpret RDT results	-HealthPulse AI provides accurate and consistent results -The model can function offline in a phone with 1GB -Cost-effective and usable in LMICs -Prioritizes privacy, accountability, and accessibility -Improved accuracy of RDT results
11. Biri Makota et al. (2023) [35]	-A penalized LR (EN) -KNN -RF Classifier -DT approach -LGBM -XGBoost model	ML	-Risk of HIV transmission -Predicting HIV risk factors	-Predicted factors associated with HIV positivity -Factors associated with HIV transmission were condomless sex, multiple sex partners, lack of knowledge on HIV transmission, not working, intimate partner violence, being a male, and other socio-behavioral factors. -XGBoost algorithm achieved the best results
12. Burns et al., (2023) [23]	-XGBoost model -LASSO model	ML	-Predicting HIV risk/diagnosis	-The developed models were able to predict the incidence of HIV diagnoses, including among women, effectively. -XGBoost outperformed the LASSO model.
13. Friedman et al., (2023) [38]	LR (STI model, demographic model, and final model)	Traditional	-HIV risk among women -Predictive modeling -Predicting HIV risk among women	-Determined women newly infected with new HIV. -Pregnancy, hepatitis C diagnosis, substance use diagnosis, and STI diagnosis were used to identify women vulnerable to HIV. -The predictive model showed acceptable discrimination between those who were and were not newly diagnosed with HIV.
14. Mutai et al., (2023) [58]	-PCA	ML	-HIV risk prediction -Using unsupervised ML to characterize the distribution of HIV positivity	HIV positivity distribution was significantly associated with the following variables: -Relationship with family head -Enrolled in school -Male circumcision
15. Jing et al. (2023) [39]	-LR -SVM -DT -RF	Both	-Uptake of HIV self-testing among MSM -Identifying key influencers of secondary distribution of HIV self-testing among MSM (comparison between ML and human identification)	-ML approach outperformed human identification -The ML models achieved high accuracy -Increased efficiency 17.7% -The approach can distribute 18% of more test kits, find 13.6% more first time testers alters, and 12.0% more undiagnosed PLHIV
16. Xu et al. (2022) [73]	MySTIRisk model	ML	-HIV risk prediction	-Age, gender, having sex with MSM and condom use were associated with HIV risk -The HIV risk prediction achieved an excellent performance on testing data sets
17. Chikusi et al., (2022) [13]	-RF -XGBoost model -ANN	ML	-Unknown HIV status -HIV index testing	-Overall, the study improved our understanding of the significance of ML in the prediction and visualization of HIV index testing. -No knowledge of HIV, age, marital status, and female gender were highly helpful in the case finding. -The best performance algorithm was the RF
18. Chingombe et al., (2022) [47]	-RNN -SVM -BC -GBC	ML	-HIV testing gap among MSM -Predicting HIV status among MSM	-ML classifiers could significantly improve HIV testing capacity among MSM -RNNs performed better than the bagging classifier, gradient boosting classifier, support vector machines, and Gaussian Naïve Bayes classifier in predicting HIV status.

	-Gaussian Naïve Bayes Classifier (GNBC)			-RNNs recorded a high prediction accuracy of 0.98 as compared to other algorithms
19. Chingombe et al., 2022 [36]	-RF Classifier -SVM -LR	ML	-HIV risk -Predicting HIV status	-People with non-formal education and from Matabeleland South Province had the highest HIV prevalence -LR outperformed the random RF classifier and SVM with the prediction accuracy -RF classifier has the highest precision of 87% compared to the other models
20. Dong et al., (2022) [37]	-LASSO regression analysis -LR -Nomogram	ML	-HIV risk among MSM -HIV risk prediction	-The results indicated that the comprehensive prediction model developed had relatively good predictive performance and stability in identifying MSM at high risk for HIV infection -Involuntary Subordination, unprotected anal intercourse, multiple sexual partners, alcohol and drug use before sex were highly associated with HIV among MSM - The risk nomogram had a relatively high performance and stability in predicting HIV infection risk among MSM
21. Belete et al., (2022) [49]	-DL models -ANN -CNN -RNN -LSTM -CNNRNN -CNNLSTM.	ML	-Improving HIV accuracy testing -Predicting HIV test status	-Forecast individuals' HIV test status using the DHS dataset. -The RNN model achieved the highest accuracy, precision, and recall in the validation set compared to other DL models. -Overall, DL achieved 83% accuracy
22. Bender et al., (2022) [52]	LASSO ML	ML	-HIV Risk prediction and associated factors	-ML was used to select biomarkers and socio-behavioral covariates to predict HIV acquisition. -Dynamic changes in immune states are associated with HIV acquisition, and biomarker, demographic, and behavioral data add complementary information to HIV risk
23. Aribé et al., (2022) [51]	-SML algorithms -Multilayer Perceptron Neural Network, a form of ANN	ML	-Impact of Covid-19 on the transmission of HIV -Forecasting HIV incidence between 2022 to 2030	-Predicted monthly incidence of HIV between 2022 to 2030 -Predicted cumulative cases by 2030 will reach 145,273
24. He et al., (2022) [32]	-LR -DT -SVM -RF	Both	-Uptake of HIV testing -HIV risk prediction for improved testing (comparison between ML and conventional LR)	-RF was the best-performing algorithm -The developed ML models provide more accuracy in predicting HIV risk among MSM than conventional LR -Factors such as condom use, frequency of condom use, and diagnosis with STIs were independently associated with HIV accusation -The ML approach is feasible and highly accurate
25. Laybohr Kamara et al., (2022) [53]	LASSO model	ML	-HIV hot spots countries -Identifying HIV hot spots from MRU countries	-From 2005–2015, Cote d'Ivoire was the hot spot for HIV prevalence -From 2016 to 2020, Guinea and Sierra Leone were hot spots for HIV prevalence -HIV prevalence is higher among women than among men -HIV prevalence is higher in urban areas than in rural - The model predicted that comprehensive correct knowledge about HIV was below the UNAIDS threshold of ending HIV endemic -The HIV testing uptake and CCK are far below the benchmarks set by UNAIDS for ending the epidemic, irrespective of residence, sex, age, and socioeconomic status

26. Xu et al. (2022) [24]	-LR -LASSO -Ridge -XGBoost -EN -GBM -RF -KNN -SVM -Bayesian GLM -NB -MLP	ML	-HIV/STI testing uptake -Predicting timely clinic attendance for HIV/STI testing uptake among MSM	-The model performed well in predicting timely HIV/STI testing. -XGBoost showed higher prediction accuracy than the classical multivariate LR model for the prediction of clinic attendance. -The EN regression model performed best in predicting HIV/STI testing. -The top 10 predictors for predicting timely HIV/STI testing were triage reasons as STI symptoms, triage reason as asymptomatic screen, sex overseas in the past 12 months, persons living with HIV, triage reason as contact of infection, current sex worker, screening reminder frequency, type of screening reminder method, sex with a male in the past 3 months, and condoms use in the past 3 months.
27. Zucker et al., (2022) [57]	- CART	ML	-Accuracy testing -Using SML to predict true positive HIV from multiple test results	-The results identified a few number of false positive results
28. Orel et al., (2022)-[40]	-A penalized LR (EN) -A GAM -SVM -XGBoost	ML	-High-yield HIV testing for epidemic control -Using ML to predict HIV status	-XGBoost performed best in predicting HIV -Among the ten most predictive variables, nine were identical for both sexes: longitude, latitude and altitude of place of residence, current age, age of most recent partner, total lifetime number of sexual partners, years lived in the current place of residence, condom use during last intercourse and, wealth index -Model performance based on these variables decreased minimally. For the first scenario, 7 males and 5 females would need to be tested to identify one HIV-positive person. For the second scenario, 4.2% of males and 6.2% of females would have been identified as a high-risk population
29. Mutai et al., (2021) [10]	-XGBoost -KNN -SVM -RF -EN -LGBM	ML	-HIV risk prediction & priority testing Using ML to identify people at risk of HIV from the PHIA dataset. -Enhancing screening approaches in limited resources situation	-XGBoost algorithm outperformed the remaining 8 -Amongst the eight most predictor features in both sexes were: age, relationship with family head, the highest level of education, highest grade at that school level, work for payment, avoiding pregnancy, age at the first experience of sex, and wealth quintile. -The study identified 5 males and 19 females that would require testing to find one HIV-positive individual. -It also predicted that 4.14% of males and 10.81% of females are at high risk of infection.
30. Haas et al., (2021) [62]	Association rule learning	ML	-Predicting HIV status	-Develop a novel HIV risk estimation method by building on clinical data -accurate in detecting HIV positivity -Achieved high sensitivity and AUC
31. Jing et al., 2021) [63]	Greedy algorithm	Both	-Enhancing HIVST among MSM while optimizing resources -Compared conventional and ML methods	-The ML method increased the economic benefits of HIVST kit distribution by more than 23%
32. Bao et al., (2021) [34]	-GBM - XGBoost - RF - DL -Traditional (Multivariate LR)	Both	-Comparison between ML and traditional testing -Develop and compare prediction models for the diagnosis of HIV in MSM using both ML and multivariable LR	-ML approaches consistently outperformed MLR in HIV diagnosis prediction. -GBM achieved the highest area under the receiver operator characteristic curve for HIV (76.3%)

33. Elkhadrawi et al., (2021) [22]	-SVM -PCA	ML	-HIV false positive results -HIV diagnostic accuracy	-Improved workflow with the ability to report immediately to reduce infection spread and initiate follow-up testing. -Reduced false positive -Classifiers highly accurate -The best classification was achieved by using SVM
34. Ovalle et al., (2021) [20]	SVM RF NB LR	ML	-Offline health risk behavior among MSM -Using ML to predict HIV risk among MSM from social media messages using a developed end-to-end encrypted system (a public health intervention framework)	-The study demonstrated that it is possible to build an integrative system across multiple platforms that effectively collects meaningful social media data -There was a consistently high F1 score when predicting HIV -The study found that social media data were correlated with offline sexual health and substance use, verified through biological testing.
35. Turbé et al., (2021) [48]	-DL -SVM -CNNs (ResNet50), (MobileNetV2) & (MobileNetV3)	ML	-HIV test accuracy -Using DL to classify images of rapid HIV test results	-The developed algorithms achieved higher sensitivity and specificity compared to traditional visual interpretation of HIV test results -The mHealth built has demonstrated the ability to reduce false negative HIV RDT results - MobileNet2 is the most appropriate among the other CNNs
36. Weissman et al., (2021) [68]	LASSO	ML	-HIV late diagnosis / missed opportunities. -Predicting factors associated with early HIV testing	-The findings identified old age, inpatient visits, alcohol use, and male gender as predictors of missed opportunities for HIV diagnosis - The findings highlighted the low efficiency of universal HIV testing strategy as it has been applied and underscored the need for identifying the implementation challenges
37. Balzer et al., (2020) [33]	LR Super Learner Algorithm used for validation	ML	-Priority testing in generalized settings -Predicting high risk HIV population and improved efficiency testing	-ML improved classification of individuals at risk of HIV acquisition -ML improved efficiency testing
38. Scott et al., (2020) [64]	SexPro	ML	-HIV risk prediction	-Developed and validated HIV risk assessment tool (sexpro.org) for MSM in the USA with good predictive ability for optimized HIV testing.
39. Adeli et al., (2019) [65]	Multivariate ML	ML	-Predicting HIV diagnosis	-The ML analysis revealed imaging phenotypes that predicted HIV diagnosis
40. Oladokun et al., (2019) [14]	-DT -LR	Both	-Using ML to predict HIV status (comparing DT and LR)	-Sociodemographic and socio-behavioral characteristics such as age, sex, province, sex debut, STI, race, wealth index, <u>and</u> contraceptive use were significant predictors of women's HIV status -Province was the most important predictor -LR achieved higher accuracy compared to the DT -Both models did not have high accuracy
41. Xiang et al., (2019) [21]	-GCN -LR -RF -GCN+LR -GVN+RF -LR+RF GCN+LR+RF	ML	-Unknown HIV cases -Using ML to detect unknown HIV infections among MSM	-The study produced satisfying results in identifying social network members who are at high risk of HIV seropositivity but whose status is unknown. -Modeling network information using GCN effectively increased the prediction of HIV status in the social network. - The ensemble approach achieved 96.6% on accuracy and 94.6% on F1 measure, which outperformed the baseline methods (GCN, logistic regression, and random forest: 79.0%, 90.5%, 94.4% on accuracy, respectively; and 57.7%, 80.2%, 90.4% on F1)

42. Demey et al., (2019) [25]	DT	ML	-HIV false positive results - HIV diagnostic accuracy -Improve the efficiency of diagnostic tools	-Detected 38% false positives and 62% confirmed true positives -Being able to improve the efficiency of the Centaur® CHIV assay by adjusting its cut-off positivity -Reduced the number of false positive CHIV assay results from 171 to 12.
43. Ahlström et al., (2019) [15]	-Simple LR -LR with lasso regularizer ( $GL_{Lasso}$ ) -Ridge regularizer ( $GLM_{Ridge}$ ) -LR with EN regularizer -RF	ML	-Late HIV diagnosis -Predict HIV status using ML algorithms.	-Increased predictive performance. -384 individuals would have to be tested to find one undiagnosed person with HIV.
44. Rice et al. (2018) [66]	AI algorithm	Both	-Predicting HIV testing	-HIV testing was increased by 18.8% in the AI group compared to the comparison group (8.1%)
45. Feller et al., (2018) [59]	-Baseline model -Unigram model -LDA model	ML	-Cost associated with universal HIV screening. -Automated HIV risk assessment -HIV risk prediction using clinical notes	-NLP improved the predictive performance of automated HIV risk assessment by extracting terms in clinical text indicative of high-risk behavior. -Drug use and risky sexual behaviors provided valuable information on HIV risk factors -The clinical keyword model achieved the highest performance by identifying terms in clinical notes indicative of high-risk behavior
46. Van Heeden et al., (2017) [17]	AI conversational agents	ML	-Addressing HIV self-testing barriers (mental health risk with unautomated tests) through the development of an ML conversational agent.  -Automated self-HIV counseling and testing	Male and female AI conversational agents/models were developed on Android phones, which enabled participants to conduct self-HIV counseling, risk determination, rapid testing, and linkage to their care for positive cases. The models had the following features: -Introduction and orient the client to a session -Conduct brief risk screening. -Prepare for and conduct rapid HIV tests. -User Feedback -Referral of positive case to nearest health facility -Six testers found that talking to the agent felt natural and equivalent to chatting with a human. -Seven said they would feel comfortable taking a real HIV test with the agent
47. Pan et al., (2017) [19]	-RF -LR	ML	-HIV testing -Predicting HIV testing uptake using ML	-RF variable importance revealed that several types of condomless sex behaviors, condom use self-efficacy and attitudes towards condom use, and level of depression are the most important predictors of receipt of HIV testing results.
48. Wade et al., (2015) [45]	-RF	ML	-Predicting HIV status	-HIV status was classifiable using subcortical morphometry
49. Hoenigl et al. (2015) [43]	-LR	Traditional	-HIV risk prediction	-Receptive condomless sex with HIV+ MSM, having multiple partners and being diagnosed with bacterial STI were significantly associated with HIV diagnosis. -With a sensitivity of 60%, SDET score $\geq 5$ was associated with AEH
50. Hoots et al. (2013) [44]	-LR	Traditional	-HIV risk prediction	-Developed risk scores for parents risk of HIV -identified 5 factors predicting partnership between newly diagnosed HIV individuals and undiagnosed partners -With a sensitivity of 90% or more, the model suggests that the number of sexual partners who need to be located and interviewed by healthcare specialists can be reduced to 25%

51. Menza et al., (2010) [67]	Cox Hazard regression models	Proportional regression	Traditional	-HIV risk prediction	-Developed risk scores predicting HIV infection among MSM -With a risk score of 0 – 12+, the risk of acquiring HIV over 4 year period ranges from 3.9% to 14.3%
-------------------------------	------------------------------	-------------------------	-------------	----------------------	--

AdaBoost: Adaptive Boosting; AI: Artificial Intelligence; AIDS: Acquired Immunodeficiency Syndrome; ANN: Artificial Neural Network; BC: Boosting Classifiers; CCK: Comprehensive Correct Knowledge; CHIV: Centaur® HIV; CNN: Convolutional Neural Network; Covid-19: Coronavirus Disease 2019; DHS: Demographic Health Survey; DL: Deep Learning; DT: Decision Tree; EDHS: Ethiopia Demographic and Health Survey; EHR: Electronic Health Record; EN: Elastic Net; GAM: Generalized Additive Models; GBM: Gradient Boosting Machine; HIV: Human Immunodeficiency Virus; KNN: K-Nearest Neighbours; LASSO: Least Absolute Shrinkage Selection Operator; LDA: Linear Discriminant Analysis; LMIC: Low and Middle-Income Countries; LR: Logistic Regression; LSTM: Long-Short-Term Memory; ML: Machine Learning; MRU: Mano River Union; MSM: Men who have Sex with Men; PCA: Principal Component Analysis; PHIA: Population-based HIV Impact Assessment; PLS-DA: Partial Least Squares Discriminant Analysis; RDT: Rapid Diagnostic Test; RF: Random Forest; RNN: Recurrent Neural Network; SIMCA: Soft Independence Modelling of Class Analogies; SML: Supervised Machine Learning; SSA: Sub-Saharan Africa; STI: Sexually Transmitted Infection; SVM: Support Vector Machine; UNAIDS: The Jointed United Nations Programme; USA: United States of America; UTT: Universal Treat and Test; ZDHS: Zimbabwe Demographic and Health Survey; ZIMPHIA: Zimbabwe Population-based HIV Impact Assessment

**Table S7.** Strengths and Limitations

First Author, Year	Types of ML algorithm applied	Model validation/ Train-test split	Max Accuracy/ Metrics	Successes/ Strengths	Limitations/ Gaps	Future directions
1. Roche et al., (2024) [55]	HealthPulse AI	-Hold-out set -82% training set -18% test set	- NPV-100% -Sensitivity =100%	-Enhanced self-testing -Outperformed human interpretation for positive HIV test results	-Human interpretation of negative results was more accurate -Human interpreters received extra training -Camera quality -One algorithm -Not generalizable -Cost intensive	-Measure the effect size of AI algorithm -Assess accessibility, compatibility and ethical issues -Conduct cost-effectiveness studies -Assess and address biases -Incorporate into HIV care services within health facilities
2. Saha et al., (2024) [29]	-RF -LR -XGBoost -KNN -Adaptive Boosting (AdaBoost) -Extra tree -Ensemble learning	-10-fold cross-validation  -70% training set -30% test set	Accuracy =97.86%	-The use of several SML algorithms -The ensemble learning method helped overcome bias and overfitting -Outperformed existing models	-Recall bias -Limited variables (only 10) used -Small sample	-Future studies should use a generalized model with a larger dataset consisting of various socio-behavioral characteristics
3. Cobre et al., (2024) [30]	-Partial least squares discriminant analysis (PLS-DA). -ANN -KNN -SVM -class analogy (SIMCA) -XGBoost -LR	-Cross-validation  -70% training set -30% test set	Accuracy = 97%	-Proposed a novel method of diagnosing HIV -The use of several ML algorithms helped ascertain the generalizability of the models -Large sample size -Contributes to the ongoing efforts to improve HIV diagnostic accuracy	-Imbalanced data	- The use of primary data capturing relevant information
4. Alie, (2024) [31]	-J48 DT -RF -KNN -SVM -LR -NB -MLP	-10-fold cross-validation -80% training set -20% test set	ROC = 86.3%	-Highly effective in predicting HIV predictors -Large dataset with several ML algorithms -Representative and generalizable	-Irregularities and imbalances of the DHS dataset -Self-reported data is subject to bias	-Future studies should consider behavioral futures
5. Latt et al., (2024) [56]	MySTIRisk ML models	Yuden's index	For HIV Sensitivity -86% Specificity - 65.6%	-Large dataset and sample -Used Youden's index to identify risk threshold that can effectively	-Self-reported information is subject to recall, nonresponse, and desirability biases.	-Future research should evaluate the impact of MySTIRisk assessment on

				differentiate high-risk- and average-risk groups	-Self-reported information led to misclassification of sexual orientation	HIV testing uptake -Further external validation
6. Nguyen et al., (2024) [46]	Random forest federated learning	Pearson correlation was used for feature selection	Accuracy = 93%	-Large sample & representative data from 8 different countries -Model achieved high accuracy compared to existing models -Designed to minimize data sharing & ensure data privacy -Adaptable model for global usage -	-only RF applied -Self-reported data -Focused on the general population only	-Future research should apply different algorithms with different populations including KPs
7. Ni et al. (2024) [41]	-LR -DT -RF -SVM	-5-fold cross-validation	Mean difference =0.88	-Primary study -Potentially a viable option for improving the uptake of HIVST	-subject to selection bias -Ethical issues -ML did not outperform the empirical scale	-Future research using a similar approach should consider additional variable selection and the weight given during modelling
8. Cheah et al. (2024) [60]	AI chatbot	-Participants' feedback using a 10-point rating scale questionnaire	-Most participants gave positive feedback	-Novel approach to enhance HIV testing among MSM -Privacy convenience	-Small sample size -Generalizability -Only suitable for MSM who can read in English	-Future studies should apply similar approach incorporating multiple languages
9. Majam et al. (2023) [42]	LR (glm) Bayesian LR (bayesglm) Regularized LR (glmnet) SVM with linear weights (svmLinearWeights) RF (ranger) Gradient Boosted Tree Model (xgbTree)	-10-fold cross-validation  -70% training set -30% training set	-AUCROC =82.8%	-Build ML model using digital survey data with high sensitivity -The use of primary data -First study to apply the method at the time of publication	-Data discrepancies from the two trials -Non-random sampling techniques introduce bias -Generalizability issue	-Tool could be integrated into clinical practice -Future studies to apply larger from multiple regions
10. Gupta et al., (2023) [50]	-HealthPulse AI algorithms -DL	-Trained with lab-generated data -Fine-tune with sets of field images	F1 Score = 99%	-Substantially improve access to timely quality care across healthcare -Detection and interpretation of RDTs for HIV that function offline in phones with 1GB -Enhances privacy, accountability, and accessibility	-Data source not explicitly defined -Could be costly for LMICs	-To move towards a general-purpose IQA using deep learning and making use of synthetic imagery for model training, especially for rare and difficult-to-obtain training examples.

11. Biri Makota et al. (2023) [35]	-A penalised LR (EN) -KNN -RF Classifier -DT approach -LGBM -XGBoost model	-5-fold cross-validation  -80% training set -20% test set	-Accuracy = 95%	-being able to develop a model capable of using socio-behavioral factors to predict HIV screen in low-resource settings	-Self-reported variables subject to recall bias	-To improve the ML approach, further research is required to integrate and implement them in a real-world primary care context.
12. Burns et al., (2023) [23]	-XGBoost model -LASSO model	-Hold-out validation -70% training set -30% test set	AUROC = 89% Specificity = 99% Sensitivity = 93%	-The refined sample of the incident HIV diagnoses used as outcomes, carefully adjudicated by chart review by 2 experienced physicians to obtain the most optimal surrogate for PrEP candidacy	-It is possible that patients were missed. - Low rates of HIV incidents limit the number of patients for the models -The short timeframe for the study limits the models' performance for individuals with a shorter duration in the health system.	-Further pre-implementation studies are needed to identify strategies to integrate such EHR-based models into clinical workflows.
13. Friedman et al., (2023) [38]	LR (STI model, demographic model, and final model)	N/A	AUC = 0.74	-Developed model for targeted HIV testing	-Risk factors identified by the model are biased towards males and MSM. -small sample -Incomplete data -Results may not be generalizable to different patient populations	-Additional studies to identify persons at high risk for HIV should consider the use of stratified ML models. -Larger cohort of women
14. Mutai et al., (2023) [58]	-PCA	-UML	-PCA1 = 62.3% -PCA2 = 70.1%	-Large sample size from multiple countries -generalizability	-Entire dependency on the generated data from the previous study, which suffered from a high degree of missingness and inconclusiveness from self-reported data that potentially impacted the training data	
15. Jing et al. (2023) [39]	-LR -SVM -DT -RF	-5-fold cross-validation -80% training set -20% test set	Accuracy = 93%	- Comprehensively combined the predictions of all four models to mitigate overfitting and model biases by adopting an ensemble learning approach -ML model superior to human identification	-Retrospective modelling research - only compared our ML approach with one kind of human identification	- A prospective trial to compare machine learning and conventional methods is needed. - Future studies should compare the ML approach with the other methods for human identification.

16. Xu et al. (2022) [73]	MySTIRisk model	5 x 10 nested cross-validation	AUC = 78%	-Used 35 ML algorithms to build an HIV prediction model -Model could be used for HIV screening on clinic websites or digital health platforms	-Predictors are self-reported information -Imbalance data could lead to overfitting	-Further validation is required to be used in other countries
17. Chikusi et al., (2022) [13]	-RF -XGBoost model -ANN	-8-% training set -20% test set	Mean Absolute Errors (MAE) =1.23	client partner notification is vital and can help to yield the target of 95 95 95	-Missing information in health care data -Lack of enough information in health care, such as socio-economic and socio-behavioral information.	-More research and development are required to capture all the required data for better results. -HIV knowledge awareness should be continuously given to the community
18. Chingombe et al., (2022) [47]	-RNN -SVM - BC -GBC -Gaussian Naïve Bayes Classifier (GNBC)	-10-fold validation	-Accuracy = 98%	-Models were highly accurate in predicting the HIV status of MSM	-A cross-sectional study lacks temporality -Self-reported information (condom use, history of HIV testing, and awareness status) may lead to social desirability and recall bias. -Stigma attached to MSM could limit data -study limited to two areas	-Bigger sample with higher statistical power in future studies - Prediction model will assist HIV programs in Zimbabwe and southern Africa to better address the needs of MSM -it is essential to include this model and its attributes in health ministry policies and guidelines to ensure that the model is fully utilized
19. Chingombe et al., 2022 [36]	-RF Classifier -SVM -LR	-Heatmap plot for Pearson's correlation coefficient -70% training set -30% test set	Accuracy = 85%	-Ability to handle datasets that contain both continuous and categorical predictors -Cope with datasets in which there are more variables than observations -Has high efficiency and accuracy	-The study only used socio-demographic and behavioral predictors to predict HIV status	There is a need to include other HIV clinical predictors to optimize HIV status prediction models and further integrate them into some real-world healthcare settings to assist policymakers and healthcare

						professionals in HIV prevention, screening, and diagnosis.
20. Dong et al., (2022) [37]	-LASSO regression analysis -LR -Nomogram	-10-fold validation -75% training set -25% test set	AUC = 83%	-Developed and externally validated an HIV infection risk assessment model, especially incorporating psychosocial and behavioral factors for MSM in China.	-Cross-sectional study: the ability to infer causality from these findings was insufficient. -Face-to-face interviews: participants' responses might not be completely honest because sexual-related questions are relatively private in China	-Large prospective cohort studies containing more comprehensive variables are needed in the future to expand further the applicability of this HIV infection risk assessment tool
21. Belete et al., (2022) [49]	-DL models -ANN -CNN -RNN -LSTM -CNNRNN -CNNLSTM.	-10-fold cross-validation -80% training set -20% test set	Accuracy = 88%	The first study applied DL in the EDHS dataset to predict HIV test status.  The models achieved high accuracy in predicting HIV test status.	-Some of the models did not achieve the best accuracy -Used only six DL algorithms	-Health domain should investigate the use of DL models that analyze individual HIV test status to enhance and re-evaluate health policies and intervention mechanisms
22. Bender et al., (2022) [52]	LASSO ML	-10-fold cross-validation	-Sensitivity analysis	-The use of longitudinal sampling of plasma as well as detailed data on sexual behavior and substance use.	- Limited sociodemographic information on MSN for HIV risk identification. - Did not also have detailed data on anogenital infections or mucosal immune samples over time -Unable to assess cellular markers because of the high burden and complexity entailed in collecting monthly PBMCs in more than 2,000 participants over 20,200 longitudinal visits.	-Identification of interventions for conditions that contribute to high-risk systemic immune profiles remains an important consideration to augment HIV prevention strategies and is an important consideration for vaccine development.
23. Aribe et al., (2022) [51]	-SML algorithms -Multilayer Perceptron Neural Network, a form of ANN	-70% training set -30% validation and test set	RMSE = 36.9200 MAE = 180.7600 MAPE = 28.5400 R2 = 0.5872	-Being able to predict the future of HIV in the Philippines -The model produces credible and correct outcomes based on evaluation metrics	Did not predict which population is at the highest risk of HIV for optimize testing	-Optimizers such as particle swarm optimization should be used to improve the performance of the existing model.

				and performance results.		
24. He et al., (2022) [32]	-LR -DT -SVM -RF	-5-fold-cross validation -70% training set -30% validation and test set	Accuracy =94%	-Believe to be the first study to apply ML on AIDS sentinel surveillance data in China for HIV risk prediction among MSM -Achieved high accuracy compared to conventional LR	-Subjected to the recall bias. -Absence of some behavioral characteristics from the questionnaire (e.g., the number of sexual partners, sex role of accessor/recipient) -3 models used for classification. -Research subjects selected for models building came from only Zhejiang province -Models may be of overfitting	-Other useful models and methods can be explored in future research, including natural language processing in unstructured data (39). -Further exploration is needed in generalizing the optimal model to the whole country and making it universally applicable - Regularization and penalization of model complexity to address overfitting
25. Laybohr Kamara et al., (2022) [53]	LASSO model	-10-fold cross-validation	-Cross-validation – glmnet repeated 1000 times	-This is the first study of HIV prevalence in the MRU with geospatial techniques to describe clusters and hot spots of HIV prevalence and leverage ML to predict the combinations of CCKs that will be accompanied by an increase in future HIV testing uptake using DHS data	-HIV testing uptakes in surveys rely on self-reporting data, which is subject to recall bias. -The analysis was limited to aged 15–49 years for both sexes, the possibility of excluding some HIV-positive males 50-59 years -The LASSO model might over-predict CCK combinations	-Early and continuous testing of all social classes with a stamp-out strategy could significantly disrupt the transmission chain in the MRU
26. Xu et al. (2022) [24]	-LR -LASSO -Ridge -XGBoost -EN -GBM -RF -KNN -SVM -Bayesian GLM -NB -MLP		-AUC = 82.7 -F1 score =85.3%	- Believe to be the first study using machine learning algorithms to predict timely clinic visits and timely HIV/STI testing among MSM post clinic reminder messages. -Combination several ML features	-Self-reported information subject to bias -Some important might factors might be excluded -Generalizability to the general MSM population may be limited -included data up to April 30, 2020, which might introduce selection bias due to the COVID-19 lockdown	Further validation is recommended for the prediction model using data from other clinics.

27. Zucker et al., (2022) [57]	- CART -LR	-Cross-validation -Train-test split	-Sensitivity =100% -Specificity = 100 -PPV = 100% -NPV =100%	-Identified patients with false positive results -The model could allow for more informed choices to allow patients to wait for confirmatory test results before starting ART	-Only sociodemographic factors were included -The data are limited as they come from a single source	-Other risk factors should be captured in future studies -The preliminary results suggest prospective investigations
28. Orel et al., (2022)-[40]	-A penalized LR (EN) -A GAM -SVM -XGBoost	-5-fold cross-validation -80% training set -20% testing set	-F1 score = 78.8% -Sensitivity = 75.6% -PPV =82.5%	-Being able to accurately predict the HIV status of individuals using demographic and socio-behavioral characteristics -probably the first study to use ML to predict HIV in generalized HIV epidemic East and Southern African countries using routinely collected survey data.	-Generalizability -The distribution of risk factors varies between countries, and the accuracy of the prediction decreased for countries not used to train the algorithm -Some HIV results were inconclusive	-The algorithm performed well with only a limited number of variables, which do not require extensive interviews or questionnaires. This approach may be implemented by clinicians and community health care workers or utilized through additional HIV case-finding modalities such as call centers, social media and self-testing initiatives.
29. Mutai et al., (2021) [10]	-XGBoost -KNN -SVM -RF -EN -LGBM	-5-fold cross-validation -80% training set -20% test set	-F1 Score =92%	-Large sample size-generalizability -Established predictive HIV status	-Validity of models: There was a high degree of missingness and inconclusiveness from self-reported data that potentially impacted the training data	-A need to adapt better screening strategies
30. Haas et al., (2021) [62]	Association rule learning	-80% training set -20% testing set	- AUC =88%	-High sensitivity and AUC -	-Generalizability and complexity -Data from only ICU and emergency units	-To apply similar techniques using multiple databases
31. Jing et al., (2021) [63]	Greedy algorithm	-Compared the results of the two models	- Model 1 and Model 2 increased economic benefits by 25 and 45%, respectively	-A developed model that optimizes resources in the secondary distribution of HIVST	-The accuracy of models not reported -Utilized secondary data, which is subject to biased findings	-Apply a similar approach in low-income countries using a quasi-experimental design
32. Bao et al., (2021) [34]	-GBM - XGBoost - RF	-5-fold cross-validation	Accuracy= 78%	-Being able to develop models that outperform	-Self-reported information, subjected to recall	- Additional ML approaches may be

	- DL -Traditional (Multivariate LR)	-80% training set -20% test set		traditional HIV testing prediction adds to that knowledge gap that ML improves HIV testing	and non-response bias -ML models tend to overfit -Imputation for missing data may introduce uncertainty in model calibration and prediction inaccuracy - The generalizability of the models to a non-clinical environment may be limited	explored as a part of future investigation. -Prospective predictions should be a part of future investigation
33. Elkhadrawi et al., (2021) [22]	-SVM -PCA	-10-fold cross-validation	Accuracy = 95.1%	-Demonstrated the usefulness of ML to aid in triaging positive HIV screens for further assessment -Robust prediction of false-positive test results for HIV5G testing -Minimize the risk of post-test psychological distress caused by positive results -Enhanced personalized patient care	-Low prevalence population	-An independent assessment of this algorithm in a similar or larger data set is required. -Future work should also assess the incorporation of other available laboratory tests to improve the robustness of this classifier further
34. Ovalle et al., (2021) [20]	SVM RF NB LR	-5-fold cross-validation	-F1 score = 70.5% for HIV	-This is believed to be the first empirical evaluation of a social media-based public health intervention framework for men who have sex with men. -A mixed study qualitative informed risk assessment tool -Built an automated social media data collector for MSM	-Self-report information is subject to recall bias and social desirability bias	-These findings suggest that using social media data for future public health interventions is feasible.
35. Turbé et al., (2021) [48]	-DL -SVM -CNNs (ResNet50), (MobileNetV2) & (MobileNetV3)	-10-fold cross-validation -80% training set -20% test set	-Accuracy = 98.9% -Specificity = 100% -Sensitivity = 97.8%	-Built mHealth software that can be used to reduce false-positive HIV results -Focused on the original population	-Small sample in the pilot study	-In future, larger evaluation studies and clinical trials will be needed to assess the performance of the system -The real-time alerting

						capability of connected RDTs could also support public health outbreak management by mapping 'hotspots' for epidemics, including COVID-19, to protect populations.
36. Weissman et al., (2021) [68]	LASSO	-10-fold cross-validation -80% training set -20% test set	AUC =82%	-This study is one of the few attempts to use population-based EHR data and ML approaches to address the issue of missed opportunities for earlier HIV diagnosis in the context of a universal HIV testing strategy	-Recall bias -Missed opportunities could have been underestimated due to potential data bias. -Data on the history of HIV testing prior to HIV diagnosis was not documented in the dataset, which could be an important predictor -The healthcare visit classification by facility types and by physician specialty may not be mutually exclusive -The methodological framework might not estimate the "real" missed diagnosis	- To apply real-time data capturing relevant information for predicting HIV testing.
37. Balzer et al., (2020) [33]	LR Super Learner Algorithm used for validation	-Cross-validation -5-fold sample splitting	Sensitivity =78%	-ML effectively predicted HIV acquisition in a generalized setting, thus enhancing the reallocation of prevention resources across all regions. ML method outperformed	-Individuals not tested may have distinct risk profiles. -Imputations changed comparative performance. -Missing data	-Further studies on risk performance both in and outside the selected regions. -Generalized epidemic settings to use Super Learner for improved risk prediction
38. Scott et al., (2020) [64]	SexPro	-10-fold cross-validation	C-statistic=79.8%	-Developed web-based self-HIV risk assessment with good sensitivity -Can improve self-risk assessment by MSM for enhanced HIV testing	-Retrospective analysis with potential bias -Generalizability -Potential ethical issues do not address	Future studies should use primary data

39. Adeli <i>et al.</i> , (2019) [65]	Multivariate ML	Acc= 78.1%	10-fold cross-validation	-Accurately predicted HIV diagnosis with unseen data	-Generalizability -small sample	-To be replicated using larger samples with a diverse population
40. Oladokun <i>et al.</i> , (2019) [14]	-DT -LR	-Cross-validation -75% training set -25% training set	Accuracy =68%	-Predictive performance -SML	-The DT is subject to overfitting and underfitting -Generalizability -secondary data -only 20% were HIV positive -the data was highly skewed, and this influenced the logistic regression results.	-Future studies should explore a wider range of social science and epidemiological applications of machine learning
41. Xiang <i>et al.</i> , (2019) [21]	-GCN -LR -RF -GCN+LR -GVN+RF -LR+RF GCN+LR+RF		-Accuracy =63.9%	-Effectively utilized a combination of rich ML features	-Small sample size insufficient for training model	-Future studies should collect additional data and design ways to include other types of edges, such as venue affiliation ties between pairs of nodes. -To further investigate how to improve the feature extraction part of GCN.
42. Demey <i>et al.</i> , (2019) [25]	DT	-10-fold cross-validation	Accuracy = 99%	-Being able to reduce 171 false positives to 12 with 100% sensitivity	- Only one algorithm was used	-To explore other models such as RF and KNN -The results could help the manufacturers of the assays to improve diagnostic quality
43. Ahlström <i>et al.</i> , (2019) [15]	-Simple LR -LR with lasso regularizer (GL <sub>Lasso</sub> ) -Ridge regularizer (GLM <sub>Ridge</sub> ) -LG with EN regularizer -RF	-10-fold cross-validation -70% training set -30% test set	Sensitivity = 90%	A nationwide population-based design including PLWH and uninfected individuals with identical electronic registry data from well-validated and comprehensive registries.	-Algorithms assume that PLWH and uninfected individuals are classified correctly. - Unknown medical history of undiagnosed PLWHIV - Algorithms do not have universal applicability -generalizability due to variations by country -Data quality and completeness	- Evaluate the usefulness and effects of the algorithms applied, which should be aimed toward improving the algorithms.

					-The study could not determine if the algorithms should replace existing HIV identification strategies	
44. Rice et al. (2018) [66]	AI algorithm	-	-	-AI-enhanced HIV prevention methods for homeless youth	-Small sample	-Future studies should use large samples from diverse groups
45. Feller et al., (2018) [59]	-Baseline model -Unigram model -LDA model	-10-fold cross-validation	Precision =66.9%	-Enhanced, cost-effective HIV testing using automated HIV risk assessment	-Small sample	-Future studies should explore whether comprehensive data from Health Information Exchanges and more advanced NLP techniques can improve model performance and drive innovative HIV prevention interventions.
46. Van Heeden et al., (2017) [17]	AI conversational agents	The conversational agent was trained with different kinds of data, including text, images, videos	User feedback was used to evaluate the models	The agents had the following strengths: -Automated self-HIV counselling and testing in a non-clinical setting -Anonymity & privacy. -Speed and easy access -Reduced post-counselling mental health risk -Linkage to care	-A fairly linear yes / no conversation emerges from the current design and prevents the level of depth and understanding that a realistic counselor would display. -Very small number of testers and limited feedback. -Limited theoretical, technical, and methodological rigor. - Only one experiment was conducted using the agent. -Many of the technical details underlying the agent were taken for granted rather than fully explored	- To develop a more robust agent cable of responding to broader questions -A more realistic agent -Language improvement -Alternative linkage to care approach
47. Pan et al., (2017) [19]	-RF -LR	-10-fold cross-validation	-Accuracy = 83.5%	-RF can handle arbitrarily large numbers of predictors and predictors with relatively high levels of multicollinearity, whereas	-LR was only compared with RF -RF cannot produce hypothesis testing results, i.e. relative risks, odds ratios, or p-values like in classical regression methods and is thus	-RF seems promising in discovering important factors related to HIV testing uptake among large numbers of predictors

				<ul style="list-style-type: none"> <li>-The RF approach will include any important (potentially complex) interactions in the construction of the forest.</li> <li>-The data are from a randomized clinical trial with a relatively large sample size and complementary information for participants' risk behaviors.</li> </ul>	<ul style="list-style-type: none"> <li>better for model exploration</li> <li>-RF has been considered by some to be a black-box prediction method. The complexity of the forest obscures which variables or features drive the prediction from simple inspection.</li> </ul>	<ul style="list-style-type: none"> <li>and should be encouraged in future HIV prevention and treatment research and intervention program evaluations.</li> <li>-LR should be compared with different ML models in future studies</li> </ul>
48. Wade et al., (2015) [45]	RF	-2-fold cross-validation	AUC = 72	<ul style="list-style-type: none"> <li>-Being able to classify HIV status using brain MRI results</li> </ul>	<ul style="list-style-type: none"> <li>-Could not establish if HIV causes brain ageing</li> <li>-Small sample</li> <li>-Generalizability</li> <li>-Only elderly</li> <li>-Inconclusive</li> </ul>	<ul style="list-style-type: none"> <li>-Future studies to use longitudinal designs with larger samples, including younger adults, to determine if brain profiles can really be used for HIV diagnosis.</li> </ul>
49. Hoenigl et al. (2015) [43]	LR	N/A	AUC=0.741 Acc =76.19%	<ul style="list-style-type: none"> <li>-Developed a risk score that could optimize resource allocations towards HIV prevention</li> </ul>	<ul style="list-style-type: none"> <li>-Single centre and retrospective design with a small sample</li> <li>-Only MSM</li> <li>-Generalizability issue</li> </ul>	<ul style="list-style-type: none"> <li>-To test the model among other populations</li> </ul>
50. Hoots et al. (2013) [44]	LR	N/A	AUC =0.662 Acc =85.25%	<ul style="list-style-type: none"> <li>-Being able to identify undiagnosed HIV infection in sexual partners</li> </ul>	<ul style="list-style-type: none"> <li>-Results may not be generalized</li> </ul>	<ul style="list-style-type: none"> <li>-Future studies to capture data on partner notification cost</li> </ul>
51. Menza et al., (2010) [67]	Cox Proportional Harzard regression models	N/A	AUC =0.66	<ul style="list-style-type: none"> <li>-Used large multi-center behavioural intervention trial data</li> <li>-Model's simplicity and excellent callibration</li> </ul>	<ul style="list-style-type: none"> <li>-Discriminatory accuracy was only modest</li> <li>-The model was only validated among MSM and thus cannot be generalized to other populations.</li> </ul>	<ul style="list-style-type: none"> <li>To undergo additional validation among other populations</li> </ul>

AdaBoost: Adaptive Boosting; AI: Artificial Intelligence; AIDS: Acquired Immunodeficiency Syndrome; ANN: Artificial Neural Network; AUC: Area Under Curve; BC: Boosting Classifiers ; CCK: Comprehensive Correct Knowledge; CNN: Convolutional Neural Network; Covid-19: Coronavirus Disease 2019; DHS: Demographic Health Survey; DL: Deep Learning; DT: Decision Tree; EDHS: Ethiopia Demographic and Health Survey; EHR: Electronic Health Record; EN: Elastic Net; GAM: Generalized Addictive Models; GBM: Gradient Boosting Machine; HIV: Human Immunodeficiency Virus; KNN: K-Nearest Neighbours; LASSO: Least Absolute Shrinkage Selection Operator; LDA: Linear Discriminant Analysis; LMIC: Low and Middle-Income Countries; LR: Logistic Regression; LSTM: Long-Short-Term Memory; ML: Machine Learning; MRU: Mano River Union; MSM: Men who have Sex with Men; NPV: Negative Predictive Value; PCA: Principal Component Analysis; PHIA: Population-based HIV Impact Assessment; PLS-DA: Partial Least Squares Discriminant Analysis; PPV: Positive Predictive value; RDT: Rapid Diagnostic Test; RF: Random Forest; RNN: Recurrent Neural Network; ROC: Receiver Operating Characteristics; SIMCA: Soft Independence Modelling of Class Analogies; SML: Supervised Machine Learning; SSA: Sub-Saharan Africa; STI: Sexually Transmitted Infection; SVM: Support Vector Machine; USA: United States of America; UTT: Universal Treat and Test; ZDHS: Zimbabwe Demographic and Health Survey; ZIMPPIA: Zimbabwe Population-based HIV Impact Assessment