Author/s:
Latt, PM;Soe, NN;Fairley, CK;Chow, EP;Johnson, CC;Shah, P;Maatouk, I;Zhang, L;Ong, JJ

Title:
Machine learning for personalized risk assessment of HIV, syphilis, gonorrhoea and chlamydia: A systematic review and meta-analysis

Date:
2025-08-01

Citation:
Latt, P. M., Soe, N. N., Fairley, C. K., Chow, E. P., Johnson, C. C., Shah, P., Maatouk, I., Zhang, L. & Ong, J. J. (2025). Machine learning for personalized risk assessment of HIV, syphilis, gonorrhoea and chlamydia: A systematic review and meta-analysis. International Journal of Infectious Diseases, 157, pp.107922-. https://doi.org/10.1016/j.ijid.2025.107922.

Persistent Link:
https://hdl.handle.net/11343/363151

International Journal of Infectious Diseases 157 (2025) 107922

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Machine learning for personalized risk assessment of HIV, syphilis, gonorrhoea and chlamydia: A systematic review and meta-analysis

Phyu M. Latt[1,2,*], Nyi N. Soe[1,2], Christopher K. Fairley[2,3], Eric P. F. Chow[2,3,4], Cheryl C. Johnson[5], Purvi Shah[5,6], Ismail Maatouk[5], Lei Zhang[1,2,7,#], Jason J. Ong[2,3,8,#]

[1] Artificial Intelligence and Modelling in Epidemiology Program, Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia
[2] School of Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[3] Melbourne Sexual Health Centre, Alfred Health, Melbourne, Australia
[4] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia
[5] Global HIV, Hepatitis and STIs Programmes, World Health Organization, Geneva, Switzerland
[6] Regional Support Team, Asia Pacific, UNAIDS, Bangkok, Thailand
[7] Clinical Medical Research Centre, Children's Hospital of Nanjing Medical University, Nanjing, Jiangsu, China
[8] Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK

ARTICLE INFO

Article history:
Received 12 February 2025
Revised 21 April 2025
Accepted 30 April 2025

Keywords:
Machine learning
HIV
Sexually transmitted infections
Risk assessment
Systematic review
Meta-analysis

ABSTRACT

Background: Machine learning (ML) shows promise for sexually transmitted infection (STI) risk prediction, but systematic evidence of its effectiveness remains fragmented.
Methods: We systematically searched six electronic databases, three preprint archives and conference proceedings (January 2010-April 2024). Studies reporting quantitative performance metrics for supervised ML-based STI risk prediction models were included. We used a bivariate random-effects model to estimate pooled sensitivity, specificity and area under the curve (AUC). The risk of bias was assessed using the Prediction model Risk of Bias Assessment Tool. We conducted sequential analyses of studies with complete and reconstructed confusion matrices. Subgroup analyses and meta-regression explored potential sources of heterogeneity.
Results: Among 3877 records screened, 25 studies comprising 45 unique models met inclusion criteria. For HIV, analysis of studies with complete confusion matrices (7 studies, 9 contingency tables) demonstrated summary AUC of 0.91 (95% CI: 0.88-0.93), pooled sensitivity 0.84 (0.76-0.90) and specificity 0.84 (0.70-0.93). Substantial heterogeneity persisted across subgroups ($I^2 > 98\%$). For other STIs, individual studies reported AUCs ranging from 0.75-0.87 for syphilis ($n = 5$), 0.73-1.00 for gonorrhoea ($n = 6$) and 0.67-1.00 for chlamydia ($n = 6$).
Discussion: While ML models show promising performance, particularly for HIV, significant heterogeneity complicates interpretation. Future research should prioritize external validation, standardized guidelines and multi-centred robust implementation studies to evaluate clinical impact.

Introduction

Sexually transmitted infections (STIs) represent a major challenge to global health, with the World Health Organization reporting over a million new infections occurring daily and an annual total of approximately 374 million curable STI cases worldwide as of 2020 [1]. Among these, syphilis, gonorrhoea and chlamydia are particularly prevalent and impactful [2]. In 2022, syphilis affected approximately 8 million adults, and in some settings, congenital syphilis cases increased, while gonorrhoea and chlamydia account for a substantial portion of new STI cases, with estimates of 82 million and 129 million new infections in 2020, respectively [3]. Similarly, HIV continues to be a major challenge, with approximately 39.9 people globally living with HIV in 2023 [2]. The impact of these infections extends beyond individual health, affecting communities and straining healthcare systems [4].

One of the primary challenges in effectively managing STIs is their asymptomatic nature. Many individuals infected with those

* Corresponding author: Phyu M. Latt, Artificial Intelligence and Modelling in Epidemiology Program, Melbourne Sexual Health Centre, Alfred Health, 580 Swanston St, Carlton VIC 3053, Melbourne, Australia.
E-mail address: phyu.latt@monash.edu (P.M. Latt).
# Contributed equally to this work as co-last authors.

https://doi.org/10.1016/j.ijid.2025.107922

STIs may not exhibit noticeable symptoms, leading to an inadvertent transmission to other vulnerable populations. This silent spread can result in severe health complications, including infertility and increased susceptibility to HIV infection [5]. Therefore, early detection and prompt treatment are crucial in preventing these adverse outcomes.

Since the early 2010s, machine learning (ML), a subset of artificial intelligence, has emerged as a promising approach to healthcare. Supervised ML uses labelled datasets to train algorithms to classify data or predict outcomes [6]. The application of ML in infectious disease prediction and diagnosis has garnered increasing interest, driven by its ability to process large volumes of complex data and identify subtle patterns that might elude traditional analytical methods [7]. However, the application of ML to STI risk assessment is not without challenges, including issues of data quality and relevance given evolving risk patterns over time, model interpretability and potential biases in training data [8].

While ML has shown promise in various medical fields, its application to HIV/STI risk assessment presents unique opportunities and challenges. The empirical evidence supporting these applications remains fragmented, with significant gaps in understanding implementation effectiveness across different healthcare settings and populations. Questions persist regarding model generalizability, real-world performance metrics and the impact on clinical decision-making processes. Several studies have explored the application of ML in HIV/STI risk estimation, reporting promising results [9,10].

While previous reviews have examined the role of ML in HIV testing [8,11], no comprehensive synthesis exists evaluating ML performance in risk assessment across HIV, syphilis, gonorrhoea and chlamydia collectively. Moreover, no previous reviews have included a meta-analysis of ML approaches for STIs. This gap is particularly important given the increasing adoption of ML-based tools in clinical practice and the need for evidence-based implementation guidelines.

To address these gaps, we conducted a systematic review and a meta-analysis of studies applying supervised ML techniques to the risk assessment for HIV, syphilis, gonorrhoea and chlamydia. This review aims to evaluate the performance of ML models in estimating individual risk for HIV/STIs and explore the current landscape of ML applications, identifying potential gaps in the literature, particularly regarding clinical implementation studies. The findings could inform future research directions and contribute to developing more effective, personalized STI prevention and control approaches.

Methods

Protocol registration and study design

We conducted a systematic review and meta-analysis in line with the Cochrane Handbook for Systematic Reviews of Interventions [12]. Our reporting adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Supplementary Table S1). Before starting the review, we registered our protocol in PROSPERO (CRD42024532154).

Search strategy and selection criteria

We conducted a comprehensive literature search across multiple electronic databases, including Embase, MEDLINE, Scopus, CINAHL, Web of Science and IEEE, for studies published from January 1, 2010, to April 26, 2024. The search strategy was built around three overarching concepts: ML (e.g. 'machine learning', 'artificial intelligence'), risk assessment (e.g. 'risk assessment' 'prediction model') and STIs (e.g. 'sexually transmitted diseases', 'HIV', 'syphilis', 'gonorrhoea', 'chlamydia'). Full search strings are given in the Supplementary Material. We also searched three preprint archives: medRxiv, bioRxiv and SSRN. Additionally, we manually reviewed the proceedings of key conferences, including the International AIDS Society, the International Union Against Sexually Transmitted Infections and the Conference on Retroviruses and Opportunistic Infections, for all available years.

We included studies that developed, validated or compared ML-based risk prediction models for HIV, syphilis, gonorrhoea or chlamydia. We focused on traditional supervised ML methods. For this systematic review, we defined supervised ML as computational methods incorporating automated feature selection, regularization techniques and/or iterative optimization procedures beyond traditional statistical approaches. We recognize that methods like Least Absolute Shrinkage and Selection Operator (LASSO) exist on a continuum between classical statistics and ML, sharing characteristics of both fields. Beyond author self-identification, we classified methods as ML when they incorporated automated feature selection, complex regularization, ensemble approaches and/or nonparametric predictive algorithms. This classification aligns with current methodological frameworks, distinguishing statistical from ML approaches [13]. Regularized regression methods (such as LASSO, Ridge and Elastic Net) were included when they employed these ML characteristics, reflecting their dual positioning in both statistical and ML literature.

This review did not include studies using advanced deep learning techniques, including complex neural network architectures, image-based deep learning and generative AI models. Eligible studies reported at least one predictive performance metric (e.g. area under the receiver operating characteristic curve [AUC], sensitivity, specificity, positive predictive value or negative predictive value). All original research studies were eligible, including randomized controlled trials, cohort studies, case-control studies, cross-sectional studies and external validation studies. We placed no restrictions on publication language. We excluded studies that did not focus on risk estimation for HIV and three STIs, those using only traditional regression methods without ML enhancements, defined as those using maximum likelihood estimation without regularization, automated feature selection, or cross-validation procedures, and those not reporting our outcomes of interest. Studies estimating re-infection or non-primary/secondary syphilis infections (e.g. neurosyphilis) were also excluded. Non-original studies such as opinion pieces, editorials or narrative reviews were not considered.

Study selection and data extraction

We exported the search results from each database as RIS files and imported the data into Covidence, where duplicates were removed. Two independent reviewers (PL and NS) screened titles and abstracts. Discrepancies were resolved by consensus with a third senior researcher (JJO). Following initial screening, we conducted a full-text review of potentially eligible studies. A third reviewer (JJO) checked the process, excluding further duplicates and non-eligible articles.

Following study selection, two researchers (PL and NS) independently extracted data using a standardized form based on the CHARMS (Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies) framework [14]. The primary outcomes included performance metrics (AUC, sensitivity, specificity), model validation results (internal and external validation outcomes) and confusion matrices where available. Secondary outcomes included model development characteristics and implementation metrics, such as the feasibility of integrating the ML models into clinical workflows. For each study, we also extracted the following variables: study design characteristics

(setting, sample size), participant demographics (age, risk factors), model development features (algorithm type, type of predictors) and validation methods.

Critical appraisal of papers

To evaluate the risk of bias in the included studies, we employed the Prediction model Risk of Bias Assessment Tool (PROBAST) [15,16]. We evaluated the risk of bias for each of the four domains, categorizing it as low, high or unclear. Additionally, we assessed concerns regarding the study's applicability to our review questions for the first three domains, again using the categories of low, high or unclear. Two researchers (PL and NS) independently conducted the quality assessment for each included study. In cases of disagreement, we reached a consensus through thorough discussion and, when necessary, consulted with a third senior researcher (JJO).

Statistical analysis

We used descriptive statistics to summarize study characteristics and model performance metrics. We calculated pooled sensitivity, specificity and AUC with 95% confidence intervals using a bivariate random-effects meta-analysis model. We conducted two sequential analyses: first with studies providing complete confusion matrices (7 studies, 9 contingency tables), then expanding to include studies where matrices could be reconstructed from reported metrics (11 studies, 14 contingency tables). For studies reporting gender-stratified results, we maintained separate contingency tables to examine performance differences between male and female populations.

We conducted subgroup analyses across three domains: study population (general vs specific populations), sample size ($<$50,000 vs $\geq$50,000 participants) and income level settings (high-income vs low and middle-income). We assessed heterogeneity using the Higgins I^2 statistic, with values above 75% indicating high heterogeneity. To explore heterogeneity sources, we performed meta-regression analysis across subgroups. We visualized study-specific and pooled estimates using forest plots and assessed publication bias using Deeks' funnel plot asymmetry test. All analyses were conducted using STATA version 17 (StataCorp, College Station, TX, USA).

Results

Study selection

We identified 3877 records through database searching and other sources. We assessed 70 full-text articles for eligibility. Of these, 46 were excluded due to irrelevance (29), duplicates (6) or inaccessible full texts (11). Finally, 25 publications met the inclusion criteria and were included in the review (see Figure 1).

While this represents 25 distinct studies, our analysis captured 45 unique predictive models. The best-performing model was chosen based on the reported AUC or, if unavailable, the highest reported performance measure (e.g. sensitivity).

Study characteristics

The 25 included studies comprised data from 14 countries, with a significant representation of the United States of America (6/25, 24.0%), Australia (4/25, 16.0%) and various African countries. The sample size for the studies ranges from 547 to 4384,178 participants. While most studies focused solely on HIV estimation (17 studies), seven studies investigated multiple infections. Positive case proportions ranged from 0.1% to 40%. Details of the included studies are provided in Supplementary Table S2.

Of the 25 included studies, the majority (23/25, 92.0%) were retrospective analyses, with data sources primarily consisting of electronic health records (8/25, 32.0%), national health surveys (6/25, 24.0%) and specialized clinical databases. Laboratory confirmation was the predominant reference standard in 21 studies (84.0%). Common ML algorithms included XGBoost (6 studies), Gradient Boosting Machine (4 studies) and ensemble methods (4 studies). For model validation, most studies (22/25, 88.0%) implemented internal validation procedures, with 5-fold (8 studies) and 10-fold (10 studies) cross-validation being the most common approaches. External validation was relatively uncommon, with only 6 studies (24.0%) conducting either temporal [17–20] or geographic validation on independent datasets [21,22]. Input features typically included a combination of demographic, behavioural and clinical characteristics (see Table 1).

Model performance and reporting

Sensitivity/recall was the most commonly reported performance metric, available in 40 models (88.9%), followed by AUC in 37 models (82.2%) and specificity in 23 models (51.1%). Where reported, sensitivity ranged from 0.39 to 1.00 and specificity from 0.64 to 0.98, depending on the probability thresholds used for classification. Only 14 models (31.1%) explicitly reported their threshold cut-off values. Complete confusion matrices were available for 10 models for HIV and 2 models for syphilis, gonorrhoea and chlamydia, respectively (Supplementary Table S3). Performance varied across infection types. HIV prediction models demonstrated AUCs ranging from 0.72 to 0.98, with the Optimized Ensemble Learning model (XGBoost and Random Forest) achieving particularly high performance (AUC = 0.94-0.98). Gender-stratified analyses demonstrated consistent performance across male (AUC = 0.94) and female (AUC = 0.92) populations [28,37,41].

Other STIs showed varying results. Syphilis models (n = 5) achieved AUCs of 0.75-0.87, while gonorrhoea and chlamydia models (n = 6 each) showed AUCs of 0.73-1.00 and 0.67-1.00, respectively. One study reported perfect discrimination (AUC = 1.00) using the CatBoost Classifier, suggesting potential overfitting. Four studies evaluated all four infections comprehensively. Detailed performance metrics are presented in Supplementary Table S3.

Pooled performance of AI algorithms

In our meta-analysis of HIV prediction models, 11 studies yielded 14 contingency tables for analysis. The SROC curve analysis of studies reporting complete confusion matrices (7 studies with 9 contingency tables) demonstrated a summary AUC of 0.91 (95% CI: 0.88-0.93) (Figure 2a), with a pooled sensitivity of 0.84 (0.76-0.90) and specificity of 0.84 (0.70-0.93) (Figure 3). When we expanded the analysis to include studies where confusion matrices could be reconstructed (11 studies with 14 contingency tables), the model performance remained robust, with a summary AUC of 0.91 (0.88-0.93), pooled sensitivity of 0.84 (0.77-0.89) and specificity of 0.85 (0.76-0.91) (Figure 2b and Figure 4). Both analyses demonstrated substantial heterogeneity ($I^2 > 98\%$).

Subgroup analyses

We conducted subgroup analyses across the study population, sample size and country income level classifications. For the study population, we defined general populations as community-based samples without demographic restrictions. Specific populations comprised specific groups (MSM, male-only or female-only

Table 1
Model development and validation characteristics of 25 included studies.

No	Study	Best ML model	Data source	Study setting	Input features	Reference standard	Type of internal validation	External validation	Target infection
1	Ahlström et al. [23]	Logistic Regression with Ridge Regulariser (GLM Ridge)	Retrospective study; Danish National Hospital Registry (DNHR), Danish Civil Registration System, Employment Classification Module, Educational Classification Module	Population-based registry	Demographics and clinical features	HIV diagnosis during the index visit (ICD8: 07983 or ICD10: B20-B24.9)	10-fold cross-validation	No	HIV
2	Alehegn et al. [24]	SVM	World Data (https://data.world/datasets/hiv)	Open data repository	Demographics and clinical features	Not specified	Not specified	No	HIV
3	Balzer et al. [25]	Super Learner (Ensemble Machine Learning method)	Retrospective study; SEARCH Study (16 communities in rural Uganda and Kenya)	Population-based	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV
4	Bao et al. [7,26]	GBM	Retrospective study; Electronic health records (EHR) from the Melbourne Sexual Health Centre (MSHC), Australia	Sexual health clinic	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV Syphilis Gonorrhoea Chlamydia
5	Belete et al. [27]	GBM	Retrospective study; Ethiopian Demographic and Health Survey (EDHS)	Population-based	Not specified	Not specified	10-fold cross-validation	No	HIV
6	Birri Makota et al. [28]	XGBoost	Retrospective study; Zimbabwe Demographic Health Survey (ZDHS)	Population-based	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV
7	Burns et al. [29]	XGBoost	Retrospective study; Electronic Health Records (EHR) from Duke University Health System (DUHS), North Carolina	Hospital-based EHR	Demographics, behavioural and clinical features	Lab-confirmed	Random split-sample validation	No	HIV
8	Chingombe et al. [30]	GBM	Retrospective study; Secondary data from a 2018 prevalence study conducted by ICAP	Community-based survey (MSM population)	Demographics, behavioural and clinical features	Lab-confirmed	10-fold cross-validation	No	HIV
9	Duthe et al. [31]	Combined model (LASSO, RF and GLM)	Retrospective study; Electronic Health Records (EHR) from the eHOP Clinical Data Warehouse of Rennes University Hospital (France)	Hospital-based EHR	Demographics, behavioural and clinical features	Lab-confirmed	Not specified	No	HIV
10	Fernandez et al. [32]	k-NN	Retrospective, cross-sectional study; Screening program data from the Tropical Medicine Unit of the Hospital Universitario Central de Asturias, Spain	Specialised clinic (Tropical Medicine Unit)	Demographics and clinical features	Lab-confirmed	Leave-one-out cross-validation (LOOCV)	No	HIV Syphilis
11	Gruber et al. [17]	LASSO	Retrospective study; Electronic Health Records (EHRs) from Atrius Health, a clinical network in Massachusetts	Primary care network EHR	Demographics, behavioural and clinical features	Lab-confirmed	10-fold cross-validation	Applied to Atrius Health EHR data from 2016	HIV

(*continued on next page*)

Table 1 (*continued*)

No	Study	Best ML model	Data source	Study setting	Input features	Reference standard	Type of internal validation	External validation	Target infection
12	He et al. [18]	RF	Retrospective study; data from MSM sentinel surveillance in Zhejiang province, China	Sentinel surveillance (MSM population)	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	Prospective validation with 2020 data from Zhejiang province	HIV
13	Hu, M et al. [33]	CatBoost Classifier	Retrospective study; data from the National Health and Nutrition Examination Survey (NHANES)	Population-based	Demographics, behavioural and clinical features	Self-reported diagnoses	10-fold cross-validation	No	Gonorrhoea Chlamydia
14	Krakower et al. [21]	LASSO	Retrospective Study; Electronic Health Records (EHR) Data from Atrius Health, Massachusetts (2007-2015)	Primary care network EHR	Demographics and clinical features	Lab-confirmed	10-fold cross-validation	Prospective Validation: Atrius Health data from 2016 External Validation: Fenway Health data from 2011 to 2016	HIV
15	Latt et al. [34,35]	Ensemble (HIV) GBM (Syphilis) RF (Gonorrhoea) GBM (Chlamydia)	Retrospective study; Electronic health records (EHR) from the Melbourne Sexual Health Centre (MSHC), Australia	Sexual health clinic	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV Syphilis Gonorrhoea Chlamydia
16	Majam et al. [22]	GBM	Digital survey responses from two sequential studies (in Johannesburg, Gauteng (urban), Tshwane, Gauteng (urban); Gert Sibande, Mpumalanga (semi-urban); and Ugu, Kwa-Zulu Natal (rural))	Community-based survey	Demographics, behavioural and clinical features	Lab-confirmed	10-fold cross-validation	Trial 2 data was used as an external test set	HIV
17	Marcus et al. [19]	LASSO	Retrospective study; Electronic Health Records (EHR) from Kaiser Permanente Northern California (KPNC)	Integrated health system EHR	Demographics, behavioural and clinical features	Lab-confirmed	10-fold cross-validation	Prospective validation using data from 2015 to 2017	HIV
18	May et al. [36]	XGBoost	Retrospective study; Two Data Sources: 1. Optum's de-identified Clinformatics Data Mart Database (CDM) – national de-identified database derived from administrative health claims 2. UT Physicians (UTP) clinical data warehouse – EHR data from outpatient network in Houston, TX	Mixed (Claims database and outpatient EHR)	Demographics and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV

(*continued on next page*)

Table 1 (*continued*)

No	Study	Best ML model	Data source	Study setting	Input features	Reference standard	Type of internal validation	External validation	Target infection
19	Mutai et al. [37]	XGBoost	Retrospective study; Population-based HIV Impact Assessment (PHIA) survey data from Tanzania (2016-2017), Zambia (2016), Malawi (2015-2016) and Eswatini (2016-2017)	Population-based	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV
20	Nisa et al. [38]	RF	Retrospective study; Electronic medical records of patients from Nai Zindagi Trust (NZ), Pakistan	Specialised clinic (PWID treatment facility)	Demographics, behavioural and biological features	Lab-confirmed	10-fold cross-validation	No	HIV
21	Saha et al. [39]	Optimised Ensemble Learning (OEL) model with XGBoost as final classifier	Retrospective study; Publicly available dataset from Kaggle	Open data repository	Demographics and behavioural features	Not specified	10-fold cross-validation	No	HIV
22	Xu et al. [20]	Boosted GLM (HIV) Ensemble (ENR + GBM + RF) (Syphilis, Gonorrhoea, Chlamydia)	Retrospective study; Electronic health records (EHR) from the Melbourne Sexual Health Centre (MSHC), Australia	Sexual health clinic	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	No	HIV Syphilis Gonorrhoea Chlamydia
23	Xu et al. [20]	Ensemble (ENR + GBM + RF)	Retrospective study; Electronic health records (EHR) from the Melbourne Sexual Health Centre (MSHC), Australia	Sexual health clinic	Demographics, behavioural and clinical features	Lab-confirmed	5-fold cross-validation	Two temporal validations: 1. Data from 2019 2. Data from 2020 to 2021	HIV Syphilis Gonorrhoea Chlamydia
24	Dong et al. [40]	LR	Cross-sectional study; Survey of 547 MSM in Shanghai, China	Community-based survey (MSM population)	Demographics, behavioural and psychosocial features	Lab-confirmed	10-fold cross-validation	No	HIV
25	Orel et al. [41]	XGBoost	Retrospective; Demographic and Health Surveys (DHS) from 10 East and Southern African countries	Population-based	Demographics and behavioural features	Lab-confirmed	Stratified 5-fold cross-validation	No	HIV

CDM, Clinformatics Data Mart Database; DNHR, Danish National Hospital Registry; EDHS, Ethiopian Demographic and Health Survey; EHR, electronic health records; ENR, elastic net regularisation; GBM, gradient boosting machine; GLM Ridge, generalised linear model with ridge regularisation; ICD: International Classification of Diseases; k-NN, k-nearest neighbours; LASSO, Least Absolute Shrinkage and Selection Operator; LR, logistic regression; MSM, men who have sex with men; NHANES, National Health and Nutrition Examination Survey; OEL, optimised ensemble learning; PHIA, Population-based HIV impact assessment; PWID, people who inject drugs; RF, random forest; SVM, support vector machine; UTP, UT physicians; XGBoost, extreme gradient boosting; ZDHS, Zimbabwe Demographic Health Survey.

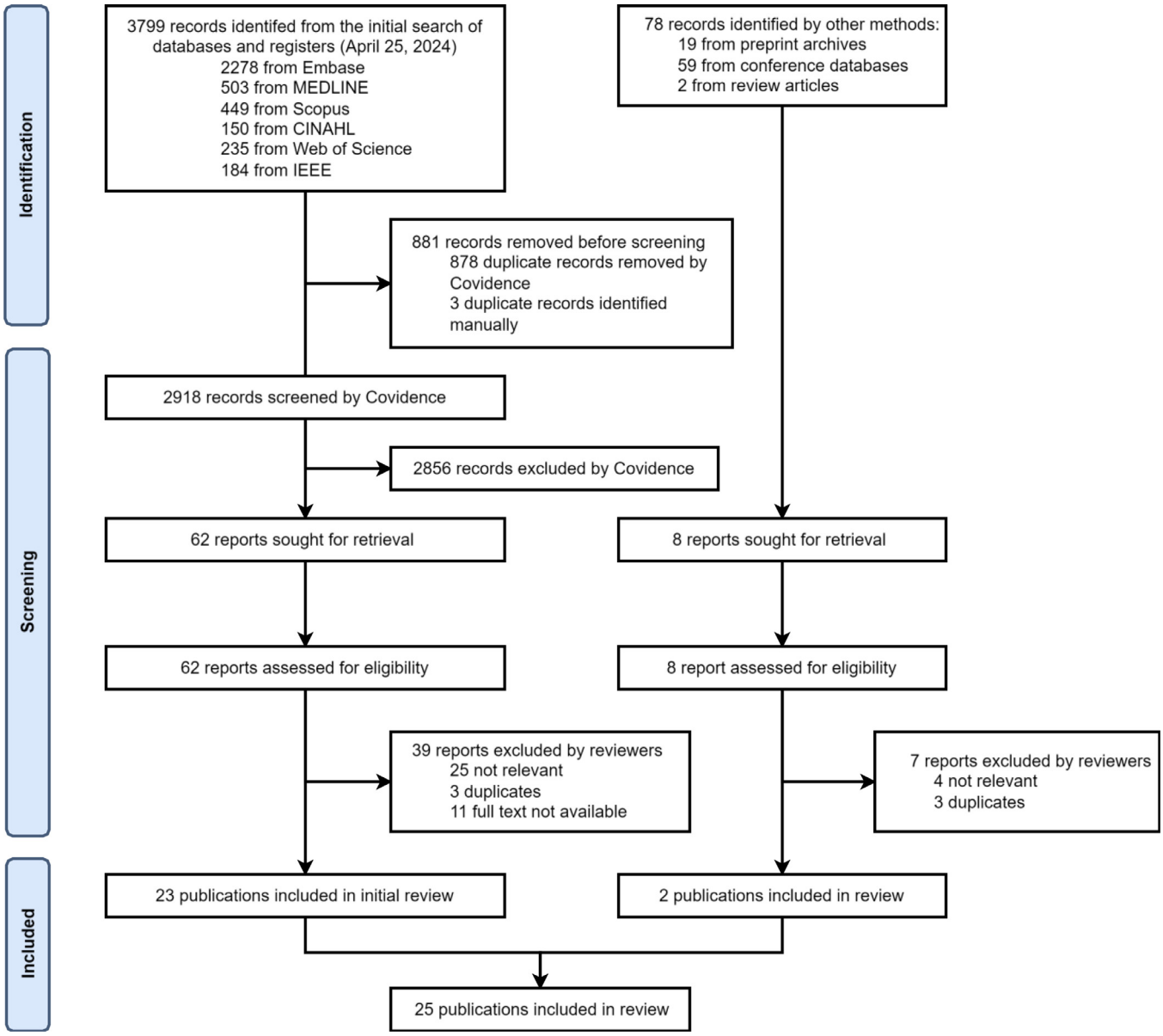

Figure 1. PRISMA flowchart for systematic review.

studies). Specific population studies ($n = 8$) showed higher performance (AUC = 0.94, 95% CI: 0.92-0.96) than general population studies ($n = 6$, AUC = 0.85, 95% CI: 0.81-0.87). Smaller studies (<50,000 participants, $n = 5$) demonstrated higher AUC (0.94, 95% CI: 0.91-0.95) compared to larger studies ($n = 9$, AUC = 0.88, 95% CI: 0.85-0.91). Studies from low and middle-income countries ($n = 9$) showed better performance (AUC = 0.94, 95% CI: 0.92-0.96) than those from high-income countries ($n = 5$, AUC = 0.79, 95% CI: 0.75-0.82). Detailed results are presented in Supplementary Figures S2-S4.

Heterogeneity

Studies with reconstructed confusion matrices showed high heterogeneity for both sensitivity ($I^2 = 99.04\%$, 95% CI: 98.86%-99.23%) and specificity ($I^2 = 100\%$). Similar patterns emerged in studies with complete matrices. Meta-regression identified significant differences in heterogeneity for sample size ($P < 0.05$), though high heterogeneity persisted across all subgroups ($I^2 > 97\%$) (Supplementary Figures S5-S7, Supplementary Table S4).

The Deeks' funnel plot suggested no publication bias for studies with complete confusion matrices ($P = 0.39$) (Supplementary Figure S8). They indicated potential publication bias for studies with reconstructed confusion matrices ($P = 0.04$) (Supplementary Figure S9).

Quality assessment

We assessed the risk of bias and applicability of included studies using the PROBAST, evaluating four domains: participants, predictors, outcome and analysis (Supplementary Figure S1). Most studies (16/25, 64.0%) demonstrated a low risk of bias across domains, particularly in participant and predictor selection. Nine studies showed a high or unclear risk of bias primarily due to insufficient documentation of participant selection criteria and unclear data sources.

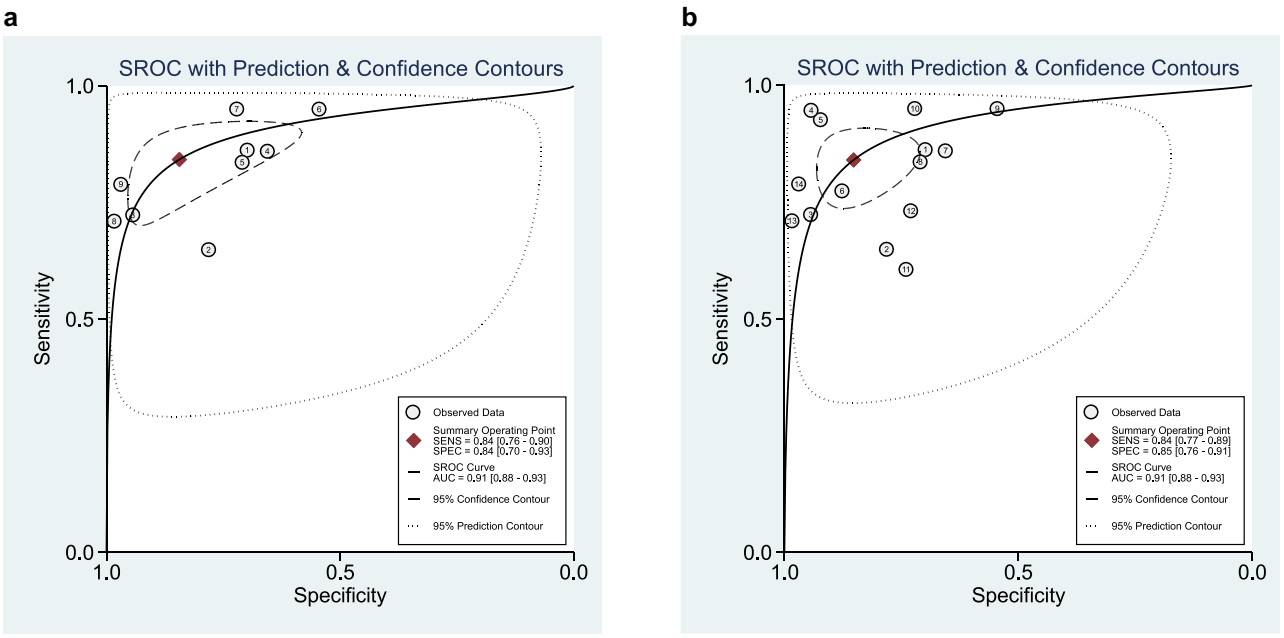

Figure 2. Pooled performance of machine learning algorithms for HIV risk prediction. (a) SROC curves of all studies with complete confusion matrices (7 studies with 9 tables), (b) SROC curves of studies with reconstructed confusion matrices (11 studies with 14 tables). ML, machine learning; SENS, sensitivity; SPEC, specificity; SROC, summary receiver operating characteristic.

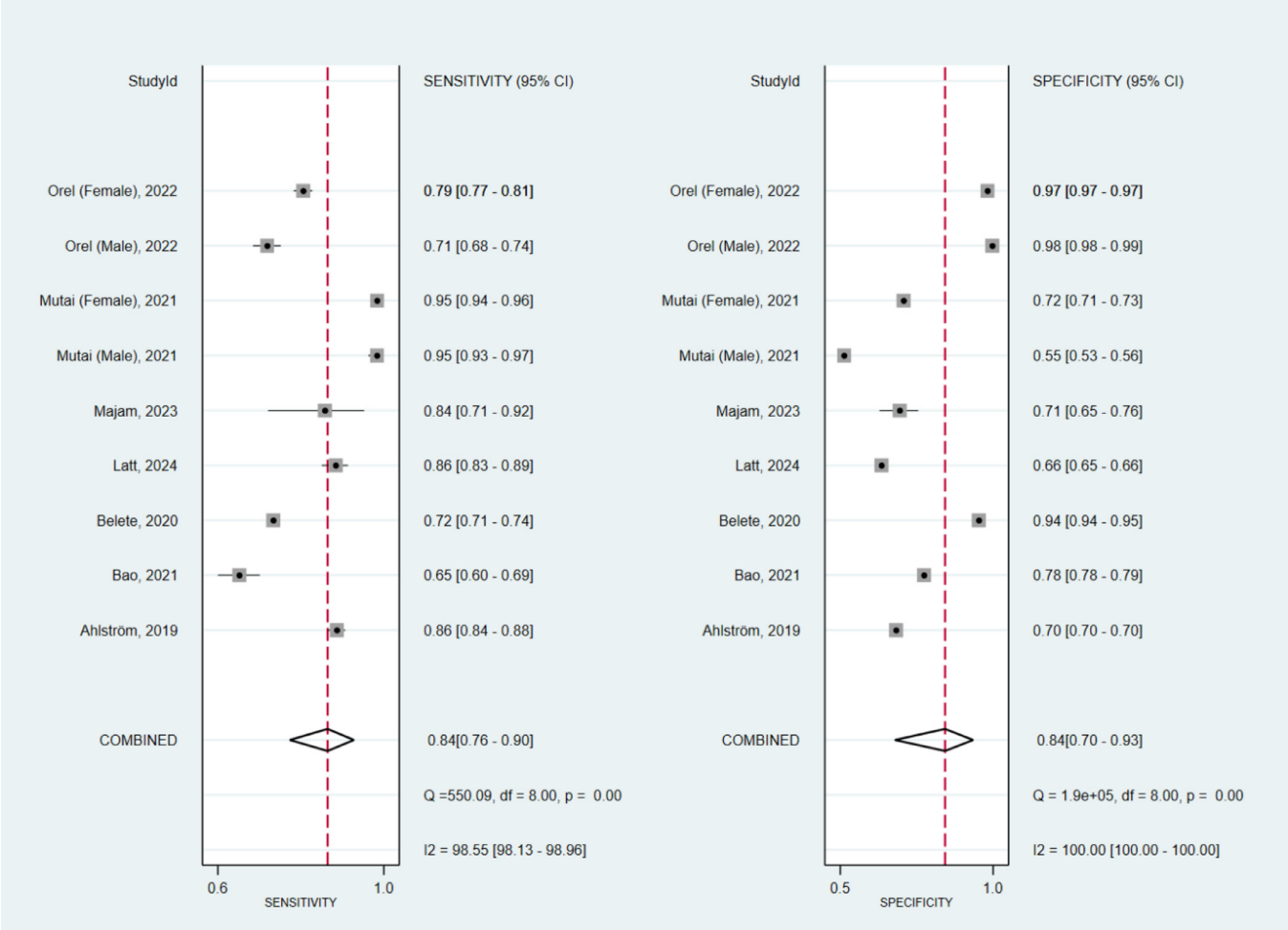

Figure 3. Forest plot of all studies with complete confusion matrices included in the meta-analysis (7 studies with 9 tables).

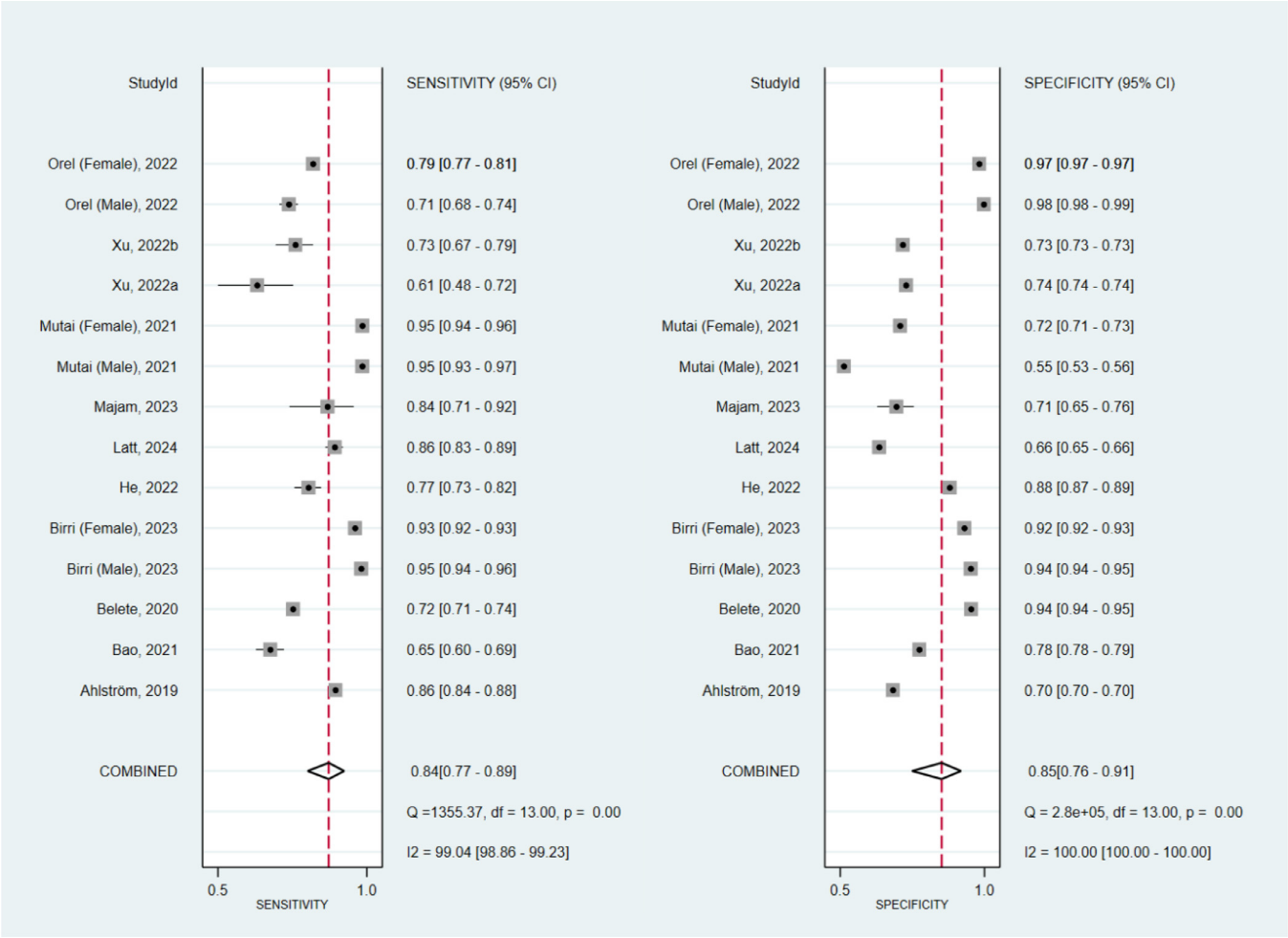

Figure 4. Forest plot of all studies with reconstructed confusion matrices (11 studies with 14 tables).

Discussion

Our systematic review and meta-analysis found substantial data on the role of ML in estimating individual risk for HIV, syphilis, gonorrhoea and chlamydia. We found that ML models demonstrated high performance in estimating HIV risk, with pooled AUC values of 0.91 (95% CI: 0.88-0.93). For the other three STIs, meta-analyses were not feasible due to insufficient studies and individual studies demonstrating lower performance compared to HIV. Tree-based algorithms, particularly XGBoost, emerged as the most frequently used and best-performing approach, followed by Gradient Boosting Machine and ensemble methods. However, we should interpret these pooled performance metrics with caution due to the high heterogeneity among the included studies. Our findings align with another systematic review [11], but extend current knowledge by examining ML applications across multiple STIs and incorporating a meta-analysis for HIV risk estimation, which previous studies did not include.

Our meta-analysis revealed substantial statistical heterogeneity ($I^2 > 97.0\%$) in ML model performance that persisted across all subgroup analyses and could not be adequately explained through meta-regression, suggesting that pooled estimates should be interpreted with caution, as the true effect might vary considerably across different settings and populations. This finding is not uncommon and frequently reported by previous systematic reviews on ML-based risk prediction models for other conditions, reporting similar unexplained heterogeneity ($I^2 > 90\%$) despite comprehensive subgroup analyses [42,43]. Several factors may contribute to this unexplained heterogeneity. First, fundamental differences in model development approaches affect performance, including variations in data quality, feature selection processes, model architectures and the number and types of parameters used for outcome prediction [42,44]. The diversity in validation strategies also impacts performance assessment. Second, the lack of standardized reporting creates significant challenges for model comparison. Notably, only one of the 25 included studies [21] followed the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) guidelines, and none explicitly mentioned adherence to the more recent TRIPOD-AI extension [45]. To address these challenges, we recommend that future studies should adhere to the TRIPOD-AI guidelines [45].

Translating ML models from research to clinical practice remains a significant challenge in STI risk prediction. Notably, only four studies transformed their ML research into a single, accessible clinical tool named '*MySTIRisk*' [20,26,34,35,46,47], though no formal implementation studies have yet evaluated its clinical impact. This limited translation to practice highlights the substantial gap between model development and practical application in healthcare settings. While other risk prediction tools for HIV exist, such as SexPro [48] and the San Diego Early Test Score [49], these rely on traditional statistical methods and face sustainability challenges, with their online platforms no longer accessible [50,51]. Implementation barriers are multifaceted: ML models often present interpretability challenges for healthcare providers, health system in-

tegration demands robust technical infrastructure, and data privacy regulations and governance frameworks can significantly impede model deployment and validation across different healthcare settings. Moreover, the lack of randomized controlled trials or implementation studies examining the clinical impact of these tools on patient outcomes and care pathways leaves uncertainty about their real-world effectiveness. These observations align with broader challenges in healthcare AI implementation, where promising research often faces practical barriers to clinical integration [52,53].

Our findings have important implications for HIV and STI risk prediction in both research and clinical practice. The pooled sensitivity and specificity of 0.84 for HIV models indicate reasonable screening performance but require careful interpretation due to high heterogeneity. These models can help prioritize resources and enable targeted interventions, especially in resource-limited settings, but false negatives must be considered in decision-making. Successful clinical integration demands robust technical infrastructure, standardized protocols and thorough evaluation of these tools' impact on decision-making and healthcare access. Healthcare systems need governance frameworks addressing technical and ethical considerations for AI implementation. Local validation, including clinician review of random cases, is essential to verify model interpretations and identify biases. These factors should be integrated early in model development, along with system interoperability, workflow integration and provider training.

Several key research priorities emerge from our systematic review. First, the field requires internationally agreed standards for evaluating ML-based tools, including frameworks for external validation, minimum performance criteria and ongoing assessment mechanisms. Second, implementation studies and randomized controlled trials are needed to evaluate the clinical impact, cost-effectiveness and real-world performance of ML-based prediction tools. Third, research is also needed to determine how best to use these tools, including what their individual and population objectives should be, how effective they are in this regard and whether they cause any harm. Multi-centre collaboration can strengthen evidence by developing larger, diverse datasets and standardized feature sets. Additionally, research should expand beyond HIV to develop robust prediction models for other STIs. Future studies should adhere to standardized reporting guidelines like TRIPOD-AI and address model interpretability and clinical integration challenges.

This is the first comprehensive systematic review and meta-analysis examining ML-based risk estimation models for multiple STIs. While a recent systematic review focused solely on the estimation of HIV risk, our study provides broader insights into ML applications across HIV and three other STIs. Our review has several key strengths. First, we included meta-analysis, offering quantitative evidence of ML model performance through rigorous statistical analysis, including subgroup analyses and meta-regression. Second, our review included a substantial combined sample size of over 13 million participants and consultations across diverse healthcare settings, enhancing the generalizability of our findings. Third, by examining technical performance and implementation challenges, we provide practical insights for translating these models into clinical practice. Fourth, our review addresses the issue of model interpretability and the need for standardized reporting in ML studies, which is often overlooked in individual studies. Finally, by synthesizing evidence from various ML techniques and comparing their performance, our study provides valuable guidance for researchers and clinicians in selecting appropriate ML approaches for STI risk estimation.

Our study has several limitations. First, substantial statistical heterogeneity persisted despite subgroup analyses and meta-regression, suggesting pooled estimates should be interpreted with caution. The perfect discrimination (AUC = 1.00) reported for gonorrhoea and chlamydia in one small study suggests potential overfitting and raises concerns about generalizability. Second, significant variations in analytical approaches and reporting quality were observed, with some studies using rigorous cross-validation and others basic train-test splits, which we accounted for using PROBAST. Third, reporting standards varied, with many studies lacking complete confusion matrices, thresholds or confidence intervals, complicating direct comparisons. We conducted separate analyses for studies with complete and reconstructed matrices, but future studies should adopt standardized reporting. Fourth, we extracted only the best-performing model based on AUC values, suggesting future reviews might benefit from analysing all reported models. Fifth, this review focused on traditional supervised ML methods to ensure a comprehensive and methodologically consistent synthesis of the existing literature. However, we acknowledge the growing importance of advanced methods, such as deep learning and generative AI. Future research should explore deep learning in STI risk estimation and could expand the scope to explore the performance and applicability of these emerging techniques. Finally, the scarcity of external validation studies limits our ability to assess the generalizability and robustness of these ML models across different populations and settings. Given the rapidly evolving nature of ML applications in STI risk prediction, future research could consider adopting a living systematic review approach to continuously capture new developments and methodological advances in this field.

In conclusion, ML models for HIV/STI risk assessment show promise but face challenges due to methodological heterogeneity and inconsistent reporting. Our analysis identifies three critical gaps: the need for standardized reporting, external validation and real-world implementation frameworks. Future progress requires robust data standards, multi-centre collaboration and rigorous studies to evaluate clinical impact, which is essential for translating ML's potential into effective STI prevention strategies.

Author Contributions

PL, JJO and LZ conceived the study idea. PL and NS did the screening and data extraction. PL conducted the statistical analysis and wrote the first draft of the manuscript. JJO accessed the analysed data and verified the analyses. All authors contributed to interpreting the results and subsequent edits of the manuscript and had final responsibility for the decision to submit for publication. JJO supervised the project and is the guarantor of the review.

Declarations of competing interest

PL, NS, CKF and LZ have licenced their models to Helfie (https://www.helfie.ai/), and these models are included in this systematic review.

Acknowledgements

The authors sincerely thank Monash University for providing a PhD scholarship for PL. We also extend our thanks to all contributors who played a role in this study, with a special mention to Lorena Romero, Research and Training Librarian, for her invaluable assistance with the literature search strategy. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the World Health Organization.

Funding

CKF is supported by a National Health and Medical Research Council (NHMRC) Leadership Investigator Grant (GNT1172900).

EPFC is supported by an NHMRC Emerging Leadership Investigator Grant (GNT1172873) and an NHMRC Leadership Investigator Grant (GNT2033299). JO is supported by the NHMRC Emerging Leadership Investigator Grant (GNT1193955).

Data availability

Data will be made available upon request made to the corresponding author.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2025.107922.

References

[1] World Health Organization. Sexually transmitted infections (STIs): fact sheet. Geneva: World Health Organization; 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis). Accessed 22 April 2024.
[2] UNAIDS. Global HIV & AIDS statistics – fact sheet. 2024. Available from: https://www.unaids.org/en/resources/fact-sheet. Accessed 22 April 2024.
[3] World Health Organization. Implementing the global health sector strategies on HIV, viral hepatitis and sexually transmitted infections, 2022–2030: report on progress and gaps 2024. Geneva: 2024. Licence: CC BY-NC-SA 3.0 IGO.
[4] Bovungana Q, Arumugam T, Ramsuran V. The association of host genes with specific sexually transmitted infections. *Front Reprod Health* 2023;**5**:1124074.
[5] Ong JJ, Bourne C, Dean JA, Ryder N, Cornelisse VJ, Murray S, et al. Australian sexually transmitted infection (STI) management guidelines for use in primary care 2022 update. *Sex Health* 2023;**20**(1):1–8.
[6] An Q, Rahman S, Zhou J, Kang JJ. A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. *Sensors (Basel)* 2023;**23**(9):4178. doi:10.3390/s23094178.
[7] Bao Y, Medland NA, Fairley CK, Wu J, Shang X, Chow EPF, et al. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. *J Infect* 2021;**82**(1):48–59.
[8] Fieggen J, Smith E, Arora L, Segal B. The role of machine learning in HIV risk prediction. *Front Reprod Health* 2022;**4**:1062387.
[9] Song C, Chen X, Tang C, Xue P, Jiang Y, Qiao Y. Artificial intelligence for HPV status prediction based on disease-specific images in head and neck cancer: a systematic review and meta-analysis. *J Med Virol* 2023;**95**(9):e29080.
[10] Marcus JL, Sewell WC, Balzer LB, Krakower DS. Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. *Curr HIV/AIDS Rep* 2020;**17**(3):171–9.
[11] Jaiteh M, Phalane E, Shiferaw YA, Voet KA, Phaswana-Mafuya RN. Utilization of machine learning algorithms for the strengthening of HIV testing: a systematic review. *Algorithms* 2024;**17**(8):362.
[12] Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane handbook for systematic reviews of interventions version 6.3 [updated February 2022]*. London: Cochrane; 2022. [Available from training.cochrane.org/handbook].
[13] Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;**319**(13):1317–18.
[14] Fernandez-Felix BM, López-Alcalde J, Roqué M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. *BMC Med Res Methodol* 2023;**23**(1):44.
[15] Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;**11**(10):e1001744.
[16] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;**170**(1):51–8.
[17] Gruber S, Krakower D, Menchaca JT, Hsu K, Hawrusik R, Maro JC, et al. Using electronic health records to identify candidates for human immunodeficiency virus pre-exposure prophylaxis: an application of super learning to risk prediction when the outcome is rare. *Stat Med* 2020;**39**(23):3059–73.
[18] He J, Li J, Jiang S, Cheng W, Jiang J, Xu Y, et al. Application of machine learning algorithms in predicting HIV infection among men who have sex with men: model development and validation. *Front Public Health* 2022;**10**:967681. doi:10.3389/fpubh.2022.967681.
[19] Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. *Lancet HIV* 2019;**6**(10):e688–95.
[20] Xu X, Yu Z, Ge Z, Chow EPF, Bao Y, Ong JJ, et al. Web-based risk prediction tool for an individual's risk of HIV and sexually transmitted infections using machine learning algorithms: development and external validation study. *J Med Internet Res* 2022;**24**(8):e37850. doi:10.2196/37850.
[21] Krakower DS, Gruber S, Hsu K, Menchaca JT, Maro JC, Kruskal BA, et al. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. *Lancet HIV* 2019;**6**(10):e696–704.
[22] Majam M, Segal B, Fieggen J, Smith E, Hermans L, Singh L, et al. Utility of a machine-guided tool for assessing risk behaviour associated with contracting HIV in three sites in South Africa. *Inform Med Unlocked* 2023;**37**:101192. doi:10.1016/j.imu.2023.101192.
[23] Ahlström MG, Ronit A, Omland LH, Vedel S, Obel N. Algorithmic prediction of HIV status using nation-wide electronic registry data. *EClinicalMedicine* 2019;**17**:100203. doi:10.1016/j.eclinm.2019.10.016.
[24] Alehegn M. Application of machine learning and deep learning for the prediction of HIV/AIDS. *HIV & AIDS Rev Int J HIV-Relat Probl* 2022;**21**(1):17–23.
[25] Balzer LB, Havlir DV, Kamya MR, Chamie G, Charlebois ED, Clark TD, et al. Machine Learning to Identify Persons at High-Risk of Human Immunodeficiency Virus Acquisition in Rural Kenya and Uganda. *Clinical Infectious Diseases* 2020;**71**(9):2326–33.
[26] Bao YN, Medland NA, Fairley CK, Wu JR, Shang XW, Chow EPF, et al. Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using machine learning approaches. *J Infect* 2021;**82**(1):48–59.
[27] Belete DM, DH M, editors *Classification of Binary Class HIV/AIDS Test Results Using Ensemble Learning Models, 2020 International Conference on Decision Aid Sciences and Application (DASA)*. Bahrain: Sakheer; 2020. p. 1240–7. doi:10.1109/DASA51403.2020.9317082.
[28] Birri Makota RB, Musenge E. Predicting HIV infection in the decade (2005–2015) pre-COVID-19 in Zimbabwe: a supervised classification-based machine learning approach. *PLoS Digit Health* 2023;**1**(6):1–20.
[29] Burns CM, Pung L, Witt D, Gao M, Sendak M, Balu S, et al. Development of a human immunodeficiency virus (HIV) risk prediction model using electronic health record data from an academic health system in the southern United States. *Clin Infect Dis* 2023;**76**(2):299–306.
[30] Chingombe I, Dzinamarira T, Cuadros D, Mapingure MP, Mbunge E, Chaputsira S, et al. Predicting HIV status among men who have sex with men in Bulawayo & Harare, Zimbabwe using bio-behavioural data, recurrent neural networks, and machine learning techniques. *Trop Med Infect Dis* 2022;**7**(9):231. doi:10.3390/tropicalmed7090231.
[31] Duthe JC, Bouzille G, Sylvestre E, Chazard E, Arvieux C, Cuggia M. How to identify potential candidates for HIV pre-exposure prophylaxis: an AI algorithm reusing real-world hospital data. *Stud Health Technol Inform* 2021;**281**:714–18. doi:10.3233/SHTI210265.
[32] Fernandez-Martınez JL, Boga JA, de Andres-Galiana E, Casado L, Fernandez J, Menendez C, et al. A machine learning model for evaluating imported disease screening strategies in immigrant populations. *Am J Trop Med Hyg* 2021;**105**(5):1413–19. doi:10.4269/ajtmh.20-1443.
[33] Hu M, Peng H, Zhang X, Wang L, Ren J. Building gender-specific sexually transmitted infection risk prediction models using CatBoost algorithm and NHANES data. *BMC Med Inform Decis Mak* 2024;**24**(1):24. doi:10.1186/s12911-024-02426-1.
[34] Latt PM, Soe NN, Xu X, Ong JJ, Chow EPF, Fairley CK, et al. Identifying individuals at high risk for HIV and sexually transmitted infections with an artificial intelligence-based risk assessment tool. *Open Forum Infect Dis* 2024;**11**(3).
[35] Latt PM, Soe NN, King AJ, Lee D, Phillips TR, Xu X, et al. Identifying individuals at high risk for HIV and sexually transmitted infections with an artificial intelligence-based risk assessment tool. *Open Forum Infect Dis* 2024;**11**(3):ofae011. doi:10.1093/ofid/ofae011.
[36] May SB, Giordano TP, Gottlieb A. Generalizable pipeline for constructing HIV risk prediction models across electronic health record systems. *J Am Med Inform Assoc* 2024;**31**(3):666–73. doi:10.1093/jamia/ocad217.
[37] Mutai CK, McSharry PE, Ngaruye I, Musabanganji E. Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa. *BMC Med Res Methodol* 2021;**21**(1):159. doi:10.1186/s12874-021-01346-2.
[38] Nisa SU, Mahmood A, Ujager FS, Malik M. HIV/AIDS predictive model using random forest based on socio-demographical, biological and behavioral data. *Egypt Inform J* 2023;**24**(1):107–15.
[39] Saha R, Malviya L, Jadhav A, Dangi R. Early stage HIV diagnosis using optimized ensemble learning technique. *Biomed Signal Process Control* 2024;**89**:105787. doi:10.1016/j.bspc.2023.105787.
[40] Dong Y, Liu S, Xia D, Xu C, Yu X, Chen H, et al. Prediction model for the risk of HIV infection among MSM in China: Validation and stability. *Int J Environ Res Public Health* 2022;**19**(2):1010. doi:10.3390/ijerph19021010.
[41] Orel E, Esra R, Estill J, Thiabaud A, Marchand-Maillet S, Merzouki A, et al. Prediction of HIV status based on socio-behavioural characteristics in East and Southern Africa. *PLoS One* 2022;**17**(3):e0264429.
[42] Liu T, Krentz A, Lu L, Curcin V. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. *Eur Heart J Digit Health* 2024;**6**(1):7–22. doi:10.1093/ehjdh/ztae080.
[43] Sazzad F, Ler AAL, Furqan MS, Tan LKZ, Leo HL, Kuntjoro I, et al. Harnessing the power of artificial intelligence in predicting all-cause mortality in transcatheter aortic valve replacement: a systematic review and meta-analysis. *Front Cardiovasc Med* 2024;**11**:1343210.
[44] Morán-Fernández L, Bólon-Canedo V, Alonso-Betanzos A. How important is data quality? Best classifiers vs best features. *Neurocomputing* 2022;**470**:365–75.
[45] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical predic-

tion models that use regression or machine learning methods. *BMJ* 2024;**385**: e078378.
[46] King AJ, Latt PM, Soe NN, Temple-Smith M, Fairley CK, Chow EP, et al. User experiences of an AI application for predicting risk of sexually transmitted infections. *Digit Health* 2024;**10**:20552076241289646.
[47] Mon Latt P, Nyi Soe N, Christopher F, Xianglong X, Alicia K, Rashidur R, et al. Assessing the effectiveness of HIV/STI risk communication displays among Melbourne Sexual Health Centre attendees: a cross-sectional, observational and vignette-based study. *Sex Transm Infectc* 2024;**100**:158–65 sextrans-2023-055978.
[48] Scott H, Vittinghoff E, Irvin R, Liu A, Nelson L, Del Rio C, et al. Development and validation of the personalized sexual health promotion (SexPro) HIV risk prediction model for men who have sex with men in the United States. *AIDS Behav* 2020;**24**(1):274–83.
[49] Hoenigl M, Weibel N, Mehta SR, Anderson CM, Jenks J, Green N, et al. Development and validation of the San Diego early test score to predict acute and early HIV infection risk in men who have sex with men. *Clin Infect Dis* 2015;**61**(3):468–75.
[50] mysexpro.org, Available from: https://mysexpro.org/en/home/. Accessed 22 April 2024.
[51] SDET. San Diego early test score. Available from: https://sdet.ucsd.edu/. Accessed 22 April 2024.
[52] Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innov* 2020;**6**(2):45–7.
[53] Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Serv Res* 2022;**22**(1):850.