# A Review of the Clinical Implication of Machine Learning Frameworks for Sexually Transmitted Diseases

Benyapa Khongwichayakupe[1], Apasara Phetsri[1], Kanokros Rotjana[1],
Ashmita Sapkota[2], Jittrapat Kongsup[3] and Pichit Boonkrong[1, *]

[1]College of Biomedical Engineering, Rangsit University, Pathum Thani 12000, Thailand
[2]Department of Microbiology, Manipal College of Medical Sciences, Pokhara 33700, Nepal
[3]Department of Surgery, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand
[*]Corresponding author, E-mail: pichit.bk@rsu.ac.th

## Abstract

Sexually transmitted diseases (STDs) continue to pose a major global health burden, with many cases remaining undetected due to asymptomatic presentation and delays in laboratory-based diagnostic procedures. Machine learning (ML) has emerged as a promising tool to enhance early detection by leveraging diverse health data sources. This study aims to address the need for faster, more accurate, and scalable diagnostic support by reviewing supervised ML models used for STD classification. The research fills a practical and scientific gap by comparing how structured clinical data, medical images, and unstructured patient-generated text contribute to diagnostic performance. Following PRISMA guidelines, this systematic review synthesized 15 supervised ML studies (2015-2025) across three primary modalities: electronic health records, diagnostic imaging, and patient-generated text. Results indicate that predictive performance for structured clinical and behavioral data typically ranged from AUC 0.68 to 0.99, while image-based models for penile lesion classification achieved diagnostic accuracies up to 0.94. The analysis reveals a clear architectural shift from classical algorithms toward deep learning to handle complex visual and textual inputs. Despite performance gains, widespread clinical deployment remains constrained by data scarcity, algorithmic bias, and a lack of interpretability. Future progress necessitates the integration of multi-modal data, explainable AI (XAI), and privacy-preserving frameworks like federated learning to ensure ethically responsible diagnostic support.

**Keywords*****:*** *machine learning, sexually transmitted diseases, clinical data classification, risk factor, data-driven diagnosis*

## 1. Introduction

The global burden of sexually transmitted diseases (STDs) remains a critical public health challenge, compounded by diagnostic delays and suboptimal risk stratification. Machine learning (ML) facilitates the transition from reactive to proactive care by enabling automated risk stratification, optimizing resource allocation through outbreak forecasting, and providing non-judgmental digital screening tools that reduce patient stigma and barriers to testing (Adlung et al., 2021; Boonkrong et al., 2025a; Zheng & Yu, 2021). By integrating ML into STD care, healthcare systems can shift from reactive management to proactive, data-driven prevention, offering a transformative approach to improving patient outcomes and reducing community transmission (Barrow et al., 2020; Fieggen et al., 2022; Marcus et al., 2020).

ML model has emerged as a powerful approach for enhancing the prediction and classification of STDs using healthcare data. Modern healthcare environments generate extensive clinical, behavioral, and laboratory information that can be leveraged to identify risk patterns and support early detection (Mazhnaya et al., 2018; Saldana et al., 2024; Xu et al., 2022b). Structured datasets, e.g., electronic health records (EHRs), offer quantifiable variables including demographic factors, symptom profiles, and diagnostic histories that are suitable for classical ML algorithms. These methods, including Binary Logistic Regression (BLR), Decision Trees (DTs), and Support Vector Machines (SVM), have been widely used to estimate infection likelihood and assess population-level risk distributions. In parallel, deep learning models are increasingly

*Published online: Copyright © 2016-2026 Rangsit University*

applied to unstructured data sources, demonstrating the potential to extract nuanced patterns from clinical narratives and medical imaging (Allan-Blitz et al., 2024). The integration of feature selection and dimensionality reduction further enhances model performance by isolating the most influential predictors while minimizing redundancy. Studies utilize diverse algorithms, e.g., Naïve Bayes (NB), Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGB), Random Forest (RF), and Categorical Boosting (CatBoost) to analyze demographic and behavioral data for diagnostic accuracy (Shija et al., 2023; Soe et al., 2024a; Soe et al., 2024b). Large-scale validations have been conducted to predict individual risks for HIV and syphilis (Latt et al., 2025; Teixeira et al., 2023). These tools increasingly support global healthcare resource allocation and early detection efforts (Kebede Kassaw et al., 2023; Simmachan et al., 2025; Yang et al., 2025). Despite these advancements, significant variability persists across studies regarding data quality, sample size, and pre-processing techniques. Differences in clinical coding systems, diagnostic criteria, and patient demographics contribute to heterogeneity in dataset composition. These disparities complicate model reproducibility and limit the generalizability of reported findings. Evaluation practices also vary, with some studies relying solely on accuracy while others adopt more comprehensive metrics, e.g., precision, recall, F1-score and AUC (Boonkrong et al., 2025b). As a result, cross-comparisons remain difficult, and standardized evaluation frameworks are urgently needed. Moreover, interpretability continues to be a central challenge, especially when models are deployed in clinical settings where practitioners require transparent justifications for diagnostic outputs.

This review bridges the gap between theoretical ML potential and practical clinical application for STDs. While various models show promise, inconsistencies in global data processing and feature selection necessitate a systematic evaluation to standardize methodologies and assess cross-population performance. This study specifically identifies which ML approaches are most clinically actionable, finding that real-world deployment remains hindered by limited external validation, poor model interpretability, and a lack of privacy-preserving frameworks. By synthesizing current evidence, this review establishes the benchmarks and standards required for the ethical integration of machine learning into public health systems.

## 2. Objectives

This review paper aims:

1) To quantify predictive performance metrics across diverse data modalities, including structured clinical records, medical imaging, and patient text.
2) To identify high-performing algorithms for specific infections (e.g., HIV, Syphilis, Gonorrhea) and evaluate the transition from classical ML to deep learning architectures.
3) To catalog interpretability methods and assess whether the identified model risk factors are clinically plausible and transparent for healthcare providers.
4) To evaluate model generalizability by measuring the frequency of external validation and geographical dataset diversity to determine real-world deployment readiness.

## 3. Review Process and Eligibility Criteria

This review utilized a systematic approach adhering to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to evaluate ML frameworks for sexually transmitted diseases. To craft a systematic review on the clinical implications of ML for STDs, a review scheme that balances data science with medical utility are illustrated as follows:

1) *Systematic Literature Search and Selection*: The process begins with a rigorous search across medical and technical databases like *PubMed*, *Scopus*, *Google Scholar* and *IEEE Xplore* using keywords, e.g., "*Machine Learning*", "*Clinical Decision Support*", "*Healthcare Analytics*" and "*STI/STD diagnosis*". A total of 47 records were initially identified through database searching between 2015 and 2025.

2) *Data Extraction and Categorization*: Critical data points are harvested from the selected papers, including study location, sample size, and the specific ML algorithms employed. As seen in your reference table, this step involves organizing studies by their clinical objective, such as early detection, risk prediction, or visual classification of penile lesions.

*Published online: Copyright © 2016-2026 Rangsit University*

3) *Evaluation of Model Performance Metrics*: The technical efficacy of the frameworks is analyzed by comparing metrics, e.g., AUC, Sensitivity, and F1-score. During this stage, the impact of various data preprocessing techniques, e.g., synthetic minority over-sampling technique (SMOTE) for balancing datasets or one-hot encoding (OHE) on the reliability of results across different populations is examined.

4) *Assessment of Clinical Interpretability*: The interpretability of complex machine learning models is evaluated through the application of explainability methods, e.g., SHapley Additive exPlanations (SHAP) or feature importance rankings. This assessment is conducted to determine whether actionable clinical risk factors, e.g., sexual behavior or socioeconomic indicators are identified, ensuring the findings are of practical utility to clinicians during patient consultations.

5) *Synthesis of Clinical Utility and Challenges*: The final phase synthesizes the findings to identify how these ML frameworks can be integrated into real-world clinical workflows, such as web-based symptom checkers or automated triage systems. It also addresses the practical barriers to implementation, including data privacy concerns, the need for external validation, and the reduction of stigma in diagnosis.

**Table 1**. Eligibility criteria in literature selection for the systematic review.

| **Inclusion Criteria** | |
|---|---|
| 1. Study Type | Primary research articles detailing the development, training, and validation of specific machine learning or deep learning algorithms. |
| 2. Target Condition | Studies explicitly focused on the detection or risk stratification of STDs (e.g., HIV, Syphilis, Gonorrhea, Mpox, or Chlamydia). |
| 3. Data Source | Use of authentic clinical datasets, including Electronic Health Records (EHRs), patient surveys, or clinical dermatological imagery. |
| 4. Performance Metrics | Research is required to report quantitative evaluation results from ML classification, e.g., accuracy, precision, recall, F1-score, etc. |
| **Exclusion Criteria** | |
| 1. Non-Diagnostic Scope | Papers focusing on general epidemiological trends or treatment efficacy without a specific ML predictive framework. |
| 2. Incomplete Reporting | Articles that do not specify the exact algorithm architecture or provide sufficient data on model validation. |
| 3. Pre-Clinical Data | Studies relying on animal models/ purely synthetic data without clinical correlation. |
| 4. Format | Review articles, conference abstracts, or editorials that do not present empirical data. |

Evidence from high-impact literature was synthesized to evaluate the clinical efficacy and architectural diversity of machine learning frameworks in managing sexually transmitted diseases. Finally, 15 studies were selected following a critical appraisal against rigorous inclusion criteria given in Table 1. This selection process identified a transition from theoretical modeling to real-world diagnostic applications while highlighting persistent gaps in model generalizability and the need for standardized evaluation metrics.

## 4. Conceptual Framework in ML Classification for STDs Prediction

ML classification offers a structured framework for analyzing clinical and behavioral variables linked to STDs. Integrating data-driven prediction with healthcare decision support, this conceptual model enables the systematic identification of diagnostic patterns. Table 2 provides a comprehensive summary of 15 ML frameworks utilized in STD research, detailing the specific models, feature engineering techniques, and evaluation metrics across diverse global study populations. The data highlights the transition from traditional statistical methods to advanced ensemble and deep learning architectures for predicting clinical outcomes like STI incidence and risk behavior. Such precision is essential for improving early detection and optimizing public health responses. Figure 1 illustrates the methodological framework in STD classification by ML models, detailing the pipeline from multi-dimensional data collection to clinical application. This conceptual model integrates preprocessing, feature selection, and advanced ML classifiers to provide systematic decision support, essential for improving early detection and public health responses.

*Published online: Copyright © 2016-2026 Rangsit University*

**Table 2.** Summary of 15 recent studies on the clinical implication of ML frameworks for STDs

| No. | Author, Year | Objectives | Country (*N*) | ML Model | Preprocessing | Disease Features (x) | Response (y) | Evaluation | Interpretation |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Mazhnaya et al., 2018 | To explore the association between injecting drugs and HIV/HCV risk | Ukraine 684 | Poisson regression | NA | Socio-demographics, Place of injection, Housing situation, Drug of choice | HIV/HCV injection risk behavior | *p*-value, Fisher's Exact | PrR, 95%CI |
| 2 | Bao et al., 2021 | To develop ML models for predicting HIV/ STI diagnoses in a cohort of men | Australia 21,273 | GBM, DL, RF, XGB, MLR | Feature selection by BLR | STI symptoms, Past syphilis infection, Length of stay in AU, Condom use, No. of | Diagnosis of HIV, Syphilis, Gonorrhea, Chlamydia | AUC, Acc, Sen., Spec., PPV, NPV | NAM, CKF |
| 3 | Xu et al., 2022a | To develop ML models to predict timely clinic attendance and the uptake of | Australia 3,044 | BLR, Penalized Regression, KNN, NB, XGB GBM, RF | NA | Clinic consultations, Responses to SMS/ email within a year | Clinic attendance, HIV/STI testing uptake | Acc., Prec., | Feature importance |
| 4 | Xu et al., 2022b | To develop ML models and a subsequent web-based risk-prediction tool for | Australia 65,043 | MLR, ENR, SVM, RF, CIRF, FDA, MARS, XGB, | OHE | Demographic, Sexual behavioral, Clinical History, Social Risk Indicators | Acquisition of an infection within the next 12 months | AUC, Rec., Sen., Spec., F1 | GBM variable importance |
| 5 | Adeboye, et al. 2023 | To analyze STD and risk factors infections reported by patients to prevent | Nigeria 400 | BLR, NB, DTs, RF, KNN, AdaBoost | NA | Age, Sex, Genitourinary symptoms, Pathological conditions | STI incidence | AUC, CA, F1, Prec., Reca. | Coefficients of BLR |
| 6 | Kebede Kassaw et al., 2023 | To identify ML–based predictors of STIs and their geographic distribution | Ethiopia 18,008 | DTs, RF, NB, LR, SVM | Boruta, SMOTE | Demographic, Geography, Sexual behavioral, Information access | STI vs. non-STI diagnosis | AUC, Acc, Sen., Spec., Prec. | Variable importance |
| 7 | Latt et al., 2053 | To validate the MySTIRisk ML tool for predicting individual HIV and STI risk | Australia 207,582 | MySTIRisk, RF, GBM, Elastic Net | Transformation, Scaling | Age, gender, Behavioral risk factors recorded during patient intake | Risk for HIV, Syphilis, Gonorrhea, Chlamydia. | AUC, Sen., Spec. | External Validation |
| 8 | Shija et al. 2023 | To develop ML model for early STI detection, reducing stigma and improving | Tanzania 13,335 | AdaBoost, RF, SVM, DTs, SGD | Imputation | Demographics, Symptoms | STI vs. non-STI diagnosis | Acc., F1, AUC | NA |
| 9 | Teixeira et al., 2023 | To develop ML models for predicting adverse congenital syphilis outcomes to | Brazil 41,762 | DTs, AF, AdaBoost, GBM, XGB, KNN, SVM | Balancing, FBS, SFA, SFS, OHE | Demographics, Family/ Housing characteristics, Lifestyles | Undesirable outcomes of congenital syphilis | Acc, Prec., Sen., Spec., F1 | Medical Experts |
| 10 | Allan-Blitz, L. T., et al., 2024 | To develop ML visual classification algorithm for penile diseases to address | USA 2,627 | U-Net architecture, Inception-ResNet V.2 | Augmented images, Pixel segmentation | Images of penile lesions: early vesicles, ulcers, and specific morphological traits of | Classification of 5 penile diseases | Sen., Prec., Spec., F1 | GradCAM++ |
| 11 | Mugurungi et al., 2024 | To apply machine learning models to predict STIs among MSM using BBS data | Zimbabwe 1,538 | MLP, DTs, XGB | SMOTE, Feature selection, OHE | Socio-demographics, Sexual Behaviors, Relationship, Medical history | STI vs. non-STI diagnosis | Acc., Reca., Prec., F1, AUC | MDA |
| 12 | Soe et al., 2024a | To develop ML model for STI prediction using demographic and symptom data | Australia 1,315 | CatBoost, RF, DTs, LR, LightGBM, GBC, XGB, | Imputation, OHE, Feature selection | Demographics, Dermatological & Genitourinary Symptoms | STI vs. non-STI diagnosis | Acc, AUC, Sen., Spec., Prec., F1 | SHAP |
| 13 | Soe et al., 2024b | To develop ML-based online symptom checker for predicting STIs/ anogenital | Australia 10,520 | Bayesian Network, CatBoost | SMOTE | Demographics, Sexual behaviors, Genitourinary symptoms | Diagnosis of STIs (Positive/Negative) | AUC, Acc, Sen., Spec., Prec., F1 | SHAP |
| 14 | Saldana et al., 2024 | To implement ML and public health datasets for notifiable STIs to predict HIV | USA 85,224 | GBM, NB, BLR, KNN, RF | Feature extraction, Selection | Age, previous STI information, provider type, race, ethnicity, SVI | HIV incidence | Acc., Prec., Reca., F1 | IFI |
| 15 | Lin, et al., 2025 | To develop ML models to predict and assess the risk of HIV/STD infection | China 1,999 | DTs, KNN, LASSO, GBM, BLR, RF, XGB | LASSO, Boruta | Socio-demographics | HIV infection, STDs | AUC, Acc, Sen., Spec., PPVb, PVc | PDP |

Published online: Copyright © 2016-2026 Rangsit University

### *4.1 Data Description*

The foundational dataset for STD classification comprises a multi-dimensional array of clinical, behavioral, and laboratory variables obtained from patients undergoing comprehensive STD screening. These datasets integrate three primary data types: structured electronic health records, medical imaging, and unstructured clinical narratives. Socio-demographic characteristics like age, gender, and geographic location provide essential context, while clinical indicators including anogenital skin symptoms, urethral discharge, and dysuria offer critical physiological indicators of infection (Bao et al., 2021; Latt et al., 2025; Saldana et al., 2024). Behavioral data elements, e.g., the number of sexual partners, condom usage frequency, and history of injecting drug use, are increasingly leveraged as influential predictors in risk-stratification models. Furthermore, laboratory-confirmed diagnostic labels for infections such as HIV, Syphilis, and Gonorrhea serve as the ground truth for training supervised ML models. The diversity of these attributes allows for a holistic view of patient health, ensuring that ML models can capture nuanced risk patterns.

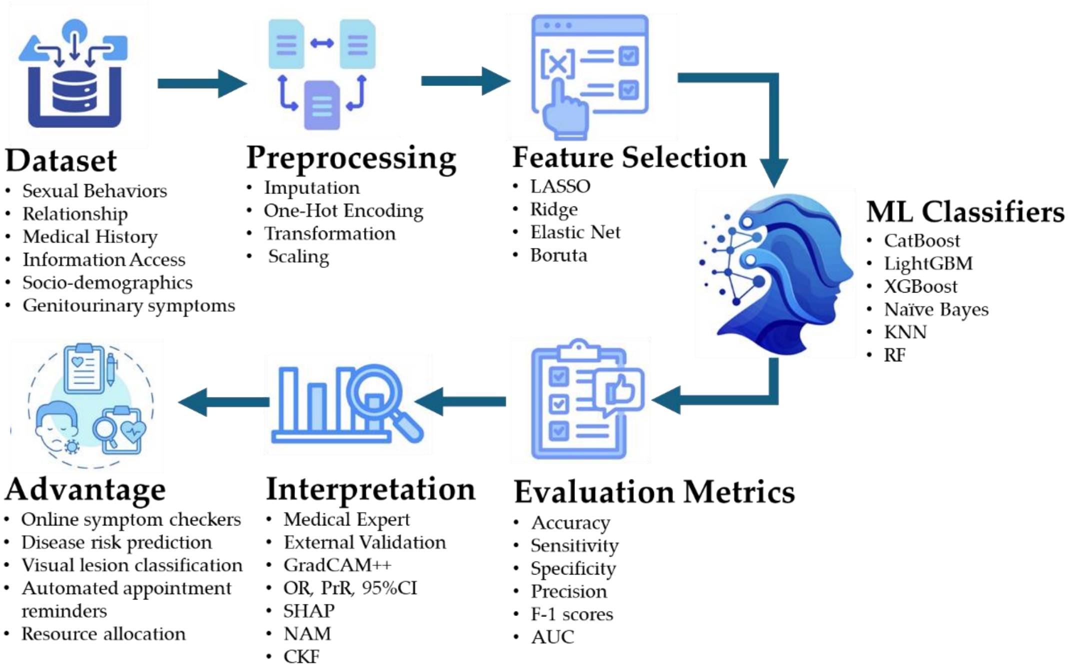

**Figure 1** Methodological framework in STD classification by ML models integrating clinical data and algorithmic analysis for diagnostic precision.

### *4.2 Pre-processing*

Pre-processing is a critical stage in the methodological framework, designed to transform raw healthcare data into a standardized format suitable for analytical consistency. This phase begins with rigorous data cleaning, addressing missing values through sophisticated imputation techniques to prevent biased results from incomplete patient records. Heterogeneous variables, ranging from numerical lab values to categorical demographic factors, are normalized and processed through techniques like OHE to ensure they are comparable during model training (Mugurungi et al., 2024; Xu et al., 2022b). Furthermore, noise reduction and outlier detection protocols are applied to filter out erroneous data points that could skew predictive performance. To address class imbalance in STI datasets, researchers frequently employ SMOTE to ensure models recognize rare infection patterns (Boonkrong et al., 2025a; Yang et al., 2025).

*Published online: Copyright © 2016-2026 Rangsit University*

### *4.3 Feature Selection*

Feature selection is a pivotal process in refining machine learning frameworks for sexually transmitted disease prediction, aimed at identifying the most informative clinical and behavioral variables. In high-dimensional healthcare datasets, many attributes can be redundant or irrelevant; therefore, researchers employ sophisticated ML-based ranking and filtering methods to eliminate noise. Algorithms such as the Boruta algorithm, LASSO, and Stepwise Forward Selection (SFS) are utilized to extract the most relevant indicators, e.g., genitourinary symptoms and sexual behaviors while reducing dimensionality (Kebede Kassaw et al., 2023; Lin et al., 2025; Mavaie et al., 2023; Simmachan et al., 2025). Additionally, Univariate analysis and Feature selection by BLR help prioritize variables that significantly impact model accuracy. For image-based diagnostics, this stage involves pixel segmentation and feature extraction from augmented images to isolate specific morphological traits of lesions. These techniques ensure that models focus on high-impact data, enhancing both predictive performance and clinical interpretability.

### *4.4 ML Classification*

ML classification for STDs primarily functions through supervised learning, where models are trained to map input features to specific clinical outcomes. These frameworks predominantly utilize binary classification schemes, represented Positive/Negative or Presence/Absence of an infection. For instance, many studies focus on "*STI vs. non-STI*" diagnosis to provide a baseline screening result. The response variables ($Y$) in these models are highly diverse, often expanding into multi-class classification for specific disease identification. This includes distinguishing between specific pathogens, e.g., HIV, Syphilis, Gonorrhea, and Chlamydia. Advanced visual frameworks even categorize five distinct penile conditions, including herpes lesions, syphilitic chancres, and genital warts, using image-based classification (Allan-Blitz et al., 2024; Soe et al., 2024a; Teixeira et al., 2023). Beyond direct diagnosis, the response variables may also target longitudinal outcomes, such as predicting the acquisition of an infection within the next 12 months or identifying adverse congenital syphilis outcomes. By integrating these varied classification tasks, ML frameworks allow clinicians to move from generalized screening to precise, pathogen-specific diagnostic insights.

### *4.5 Evaluation Metrics*

Evaluating machine learning frameworks for STDs requires metrics that accurately reflect clinical reliability, particularly when dealing with class imbalances. The F1-score and Area Under the Curve (AUC) are the most popular metrics among others (Accuracy, Precision, Recall/Sensitivity, Specificity, Balance Accuracy, etc.), as they provide a balanced view of precision and recall (Bao et al., 2021; Boonkrong et al., 2025a). These evaluations are fundamentally rooted in the confusion matrix, which adapts based on the number of classes. In binary classification (e.g., Positive/Negative), the matrix tracks True Positives and False Negatives to ensure high sensitivity. In multi-class scenarios (e.g., differentiating between Syphilis, Gonorrhea, and HIV), the matrix expands to visualize misclassification patterns across specific diseases, allowing researchers to refine model specificity for each pathogen. In high-performing models, metrics such as AUC and F1-score ideally approach 1.0, signifying near-perfect discrimination, while a value of 0.5 suggests performance no better than random chance. Models utilizing structured clinical and behavioral data achieved a wide performance range, with predictive capabilities for HIV/STI risk reaching an AUC of 0.9995 and classification accuracies for specific infections like syphilis reported at 96% (Latt et al., 2025). Image-based deep learning frameworks for classifying penile lesions demonstrated high diagnostic precision with an accuracy of 94.4%, while models analyzing clinical text and narratives via natural language processing achieved AUC values of approximately 0.912 (Adeboye et al., 2023).

### *4.6 Interpretation*

Model interpretation is essential for bridging the gap between algorithmic complexity and clinical trust, ensuring that ML frameworks function as transparent decision-support tools. Feature importance acts as a vital interpretation mechanism by identifying and ranking the clinical and behavioral variables, e.g.,

Published online: Copyright © 2016-2026 Rangsit University

anogenital symptoms or sexual history that most significantly influence a model's prediction. By quantifying the contribution of each attribute through methods like mean decrease accuracy (MDA) or variable importance ranking, researchers can validate the model's logic against established medical knowledge to ensure clinical reliability. In STD research, interpretability tools like SHAP and Partial Dependence Plot (PDP) are frequently employed to quantify how socio-demographics and genitourinary symptoms drive individual risk scores (Boonkrong et al., 2025b; Soe et al., 2024a). Researchers also utilize variable importance ranking, MDA, and GBM variable importance analysis to prioritize high-impact predictors such as sexual behavior, medical history, and injection risk. For image-based diagnosis, salience maps using GradCAM++ provide visual explanations by highlighting specific morphological traits of lesions (Allan-Blitz et al., 2024). These methods, alongside external validation, PrR/95%CI, and medical expert review, ensure that diagnostic outputs are clinically grounded and actionable (Mazhnaya et al., 2018). By providing a clear rationale for each prediction, these techniques foster provider confidence and support targeted public health interventions across diverse global populations.

### *4.7 Clinical Implication*

ML is revolutionizing the clinical landscape for STD management by shifting care from reactive to proactive. Online symptom checkers and AI-driven virtual triage now provide accessible, anonymous entry points for patients, effectively reducing the stigma-related barriers that often delay diagnosis. By converting passive symptom concern into proactive intent to seek care, these tools improve healthcare-seeking behavior for both symptomatic and asymptomatic individuals. Beyond triage, ML models excel in disease risk prediction, leveraging electronic health records and behavioral data to identify high-risk individuals with high accuracy, e.g., achieving a pooled AUC of 0.912 for HIV risk (Adeboye et al., 2023). For visible symptoms, visual lesion classification apps enable real-time, image-based screening, helping clinicians differentiate between STIs and non-STI conditions with increased precision. Administratively, automated appointment reminders via SMS significantly reduce no-show rates (by up to 40%) and increase repeat testing frequency among key populations (Xu et al., 2022a). Combined with resource allocation planning, these models allow clinics to prioritize high-risk patients and optimize staff workflows, ultimately enhancing early detection, reducing community transmission, and improving long-term reproductive health outcomes (Barrow et al., 2020). ML models enable highly personalized prevention strategies by tailoring PrEP recommendations and clinical intervention plans based on an individual's unique behavioral data, socio-demographic factors, and specific risk profile (Calabrese, 2020).

## 5. Data and Methodology Challenges

The successful integration of ML frameworks into clinical STD management necessitates addressing a multifaceted array of technical and ethical hurdles. While current models demonstrate significant predictive potential, their real-world utility is often constrained by data irregularities, static modeling approaches, and a lack of interpretability. The core challenges were identified in the reviewed literature, ranging from the technical complexities of data heterogeneity and temporal dependency to the broader implications of model generalizability. Furthermore, the analysis highlights the essential roles of ethical rigor, bias mitigation, and external validation in transforming these computational tools into trustworthy diagnostic supports.

### *5.1 Data Heterogeneity and Sparsity in Healthcare Records*

Healthcare datasets for STD prediction are characterized by high heterogeneity, spanning structured laboratory values, clinical narratives, and inconsistent patient-reported behaviors (Adlung et al., 2021; Mavaie et al., 2023). As shown in the synthesized data, inputs range from sociodemographic factors to complex image data for penile lesions. This diversity complicates feature integration and exacerbates data sparsity, especially for rare or stigmatized conditions. Sparse datasets hinder the identification of significant patterns, reducing predictive reliability. To mitigate these challenges, advanced imputation and multi-modal fusion techniques, e.g., combining behavioral data with clinical symptom profiles are essential. Without addressing heterogeneity, models may appear accurate during training but fail in real-world environments

*Published online: Copyright © 2016-2026 Rangsit University*

where data is often incomplete. Consequently, standardized data collection protocols are required to ensure the development of robust, scalable diagnostic systems.

### *5.2 Temporal Dependency and Causal Inference*

A critical limitation in current machine learning frameworks for sexually transmitted diseases is the inability of traditional static models to capture the temporal dynamics of infection progression and shifting patient behaviors. As reflected in the diverse methodologies in Table 2, many models provide a cross-sectional "*snapshot*" of risk, failing to account for longitudinal changes in symptom evolution or behavioral patterns over time. This omission of temporal dependency can lead to biased risk stratification and inaccurate categorization of high-risk patient groups. Furthermore, a significant methodological challenge persists in distinguishing between statistical correlations and true causal relationships. While ensemble models like Random Forest effectively identify associations, they often lack the causal inference required to predict how specific clinical interventions or behavioral modifications will alter patient outcomes. Future research must prioritize recurrent architectures or causal discovery algorithms to support more meaningful and dynamic clinical decision-making.

### *5.3 Model Interpretability and Clinical Translation*

Establishing model interpretability is a prerequisite for the clinical translation of ML frameworks in the management of sexually transmitted diseases. The clinical translation of these ML frameworks is hindered by the complexity of advanced machine learning models, which requires the integration of interpretability methods such as SHAP or GradCAM++ to ensure that algorithmic predictions are transparent, medically sound, and trustworthy for healthcare providers. As evidenced by the inclusion of SHAP and Feature Importance rankings in several studies, shifting away from opaque computational architectures is essential for clinician trust. These interpretability tools allow practitioners to identify actionable clinical risk factors, e.g., specific sexual behaviors, urogenital symptoms, or demographic indicators, ensuring that algorithmic outputs align with established medical logic. By translating complex multidimensional data into transparent insights, these models can be effectively integrated into point-of-care interventions, such as automated triage or web-based symptom checkers. Ultimately, clinical utility is determined by a framework's capacity to deliver high predictive accuracy while simultaneously providing a transparent rationale to support clinical decision-making and patient education. This transparency is regarded as essential for bridging the existing gap between sophisticated computational capabilities and practical bedside applications.

### *5.4 Generalizability and External Validation*

A recurring challenge highlighted by the diverse studies in Table 2 is the lack of model generalizability across different populations. Many frameworks, such as the GBM trained on regional cohorts like the Australian men who have sex with men (MSM) population, may exhibit reduced performance when applied to external groups with different demographics or diagnostic practices (Lin et al., 2025; Mugurungi et al., 2024). Unlike general clinical records, datasets for STDs are uniquely complex as they must integrate sensitive behavioral data, e.g., specific sexual practices and history of intravenous drug use, with biological biomarkers from blood and fluid exchange. This inherent complexity, coupled with a lack of robust external validation across diverse cohorts, currently restricts the scalability of ML-based diagnostic systems, as models often become over-optimized for specific institutional datasets. Consequently, ensuring true generalizability requires multi-site datasets that reflect global patient diversity and standardized evaluation protocols to verify that high reported accuracy is not merely an artifact of homogeneous data. Without rigorous external validation, clinical adoption will remain localized, preventing the broader public health impact these technologies could offer (Hunter et al., 2022; Kaissis et al., 2020).

### *5.5 Ethical Concerns and Bias Mitigation*

Future clinical frameworks must prioritize data security and regulatory compliance, e.g., the Personal Data Protection Act (PDPA), to protect sensitive behavioral and demographic data (Gerke et al.,

*Published online: Copyright © 2016-2026 Rangsit University*

2020; McGraw & Mandl, 2021). As ML transitions toward real-world STD diagnostics, privacy-preserving architectures like federated learning are essential to secure patient information while enabling cross-institutional model training. Beyond security, ethical rigor is required to identify and mitigate embedded biases, e.g., those related to age and sex that risk amplifying systemic healthcare inequalities. However, the implementation of ML for STDs in Thailand is governed by the PDPA, which classifies sexual health and behavior as sensitive data requiring explicit consent and rigorous de-identification to mitigate deep-seated cultural stigma. Future frameworks must integrate transparent auditing and robust encryption with explainability methods like SHAP and Grad-CAM to ensure that diagnostic accuracy remains equitable across diverse global populations without compromising patient confidentiality. By prioritizing these safeguards, developers can foster essential clinical trust and ensure that algorithmic bias does not lead to the further social or legal marginalization of vulnerable patient groups.

## 6. Conclusion

The findings of this systematic review demonstrate that ML techniques provide significant improvements in the classification and prediction of STDs across diverse healthcare data sources. By leveraging complex algorithms, e.g., Penalized Regression, BLR, NB, KNN, SVM, XGB, CatBoost, etc., diagnostic accuracy for both structured clinical records and unstructured medical imaging has been successfully enhanced through the application of these frameworks. These advancements facilitate a transition from reactive care to proactive management by enabling automated risk stratification and non-judgmental digital screening tools. However, persistent limitations continue to restrict widespread clinical deployment, notably the opaque nature of advanced computational architectures which can hinder clinician trust. Issues such as data scarcity, inherent algorithmic bias, and a lack of model generalizability across diverse global populations remain critical challenges. Many current models provide only a cross-sectional snapshot of risk, failing to account for the temporal dynamics of infection progression and longitudinal behavioral changes. Future studies should prioritize the integration of multi-modal data and the development of explainable ML techniques to ensure diagnostic outputs are transparent and medically sound. Adopting privacy-preserving frameworks, such as federated learning, will be essential for secure cross-institutional training while maintaining regulatory compliance. Ultimately, shifting toward causal discovery algorithms and longitudinal validation will be vital for ensuring ethically responsible and clinically reliable diagnostic support systems.

## 7. Acknowledgements

The authors would like to express sincere thanks to the anonymous reviewers for their perceptive feedback and constructive comments, which greatly improved the caliber of this research. Furthermore, we gratefully acknowledge the Research Institute of Rangsit University for the financial support that enabled our participation in the RSU International Research Conference 2026.

## 8. References

Adeboye, N. O., Bashiru, K. A., Afolabi, H. A., & Ojurongbe, T. (2023). Diagnosing Sexually Transmitted Disease from Some Symptoms Using Machine Learning Models. *Journal of Statistical Modeling and Analytics (JOSMA)*, *5*(1), 65–80. https://doi.org/10.22452/JOSMA.VOL5NO1.5

Adlung, L., Cohen, Y., Mor, U., & Elinav, E. (2021). Machine learning in clinical decision making. *Med*, *2*(6), 642–665. https://doi.org/10.1016/j.medj.2021.04.006

Allan-Blitz, L. T., Ambepitiya, S., Tirupathi, R., & Klausner, J. D. (2024). The Development and Performance of a Machine-Learning Based Mobile Platform for Visually Determining the Etiology of 5 Penile Diseases. *Mayo Clinic Proceedings: Digital Health*, *2*(2), 280–288. https://doi.org/10.1016/j.mcpdig.2024.04.006

Bao, Y., Medland, N. A., Fairley, C. K., Wu, J., Shang, X., Chow, E. P., ... & Zhang, L. (2021). Predicting the diagnosis of HIV and sexually transmitted infections among men who have sex with men using

*Published online: Copyright © 2016-2026 Rangsit University*

machine learning approaches. *Journal of Infection*, *82*(1), 48–59. https://doi.org/10.1016/j.jinf.2020.11.007

Barrow, R. Y., Ahmed, F., Bolan, G. A., & Workowski, K. A. (2020). Recommendations for providing quality sexually transmitted diseases clinical services, 2020. *MMWR. Recommendations and Reports*, *68*(5), 1–20. https://doi.org/10.15585/mmwr.rr6805a1

Boonkrong, P., Shakya, S., Kraunamkam, W., & Simmachan, T. (2025a). Dietary patterns and psoriasis severity in Thai patients: a machine learning approach for small sample data. *Scientific Reports*, *15*(1), Article 33088. https://doi.org/10.1038/s41598-025-17657-z

Boonkrong, P., Shakya, S., Yang, J., & Simmachan, T. (2025b). Risk factors in males and females for disease classification based on International Classification of Diseases, 10th Revision codes. *Engineering Proceedings*, *108*(1), Article 26. https://doi.org/10.3390/engproc2025108026

Calabrese, S. K. (2020). Understanding, contextualizing, and addressing PrEP stigma to enhance PrEP implementation. *Current Hiv/Aids Reports*, *17*(6), 579–588. https://doi.org/10.1007/s11904-020-00533-y

Fieggen, J., Smith, E., Arora, L., & Segal, B. (2022). The role of machine learning in HIV risk prediction. *Frontiers in Reproductive Health*, *4*, Article 1062387. https://doi.org/10.3389/frph.2022.1062387

Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare* (pp. 295–336). Elsevier.

Hunter, C. M., Salandy, S. W., Smith, J. C., Edens, C., & Hubbard, B. (2022). Racial Disparities in Incidence of Legionnaires' Disease and Social Determinants of Health: A Narrative Review. *Public Health Reports®*, *137*(4), 660–671. https://doi.org/10.1177/00333549211026781

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, *2*(6), 305-311. https://doi.org/10.1038/s42256-020-0186-1

Kebede Kassaw, A. A., Melese Yilma, T., Sebastian, Y., Yeneneh Birhanu, A., Sharew Melaku, M., & Surur Jemal, S. (2023). Spatial distribution and machine learning prediction of sexually transmitted infections and associated factors among sexually active men and women in Ethiopia, evidence from EDHS 2016. *BMC Infectious Diseases*, *23*(1), Article 49. https://doi.org/10.1186/S12879-023-07987-6

Latt, P. M., Ray, A., Lu, H., Soe, N. N., Xu, X., Bao, Y., ... & Fairley, C. K. (2025). External validation of a web- and artificial intelligence-based HIV/STI risk assessment tool: performance evaluation using data from Sydney sexual health centre. *BMC Infectious Diseases 2025 25:1*, *25*(1), 1647-. https://doi.org/10.1186/S12879-025-12087-8

Lin, B., Liu, J., Li, K., & Zhong, X. (2025). Predicting the risk of HIV infection and sexually transmitted diseases among men who have sex with men: cross-sectional study using multiple machine learning approaches. *Journal of medical Internet research*, *27*, Article e59101. https://doi.org/10.2196/59101

Marcus, J. L., Sewell, W. C., Balzer, L. B., & Krakower, D. S. (2020). Artificial intelligence and machine learning for HIV prevention: emerging approaches to ending the epidemic. *Current HIV/AIDS Reports*, *17*(3), 171–179. https://doi.org/10.1007/s11904-020-00490-6

Mavaie, P., Holder, L., & Skinner, M. K. (2023). Hybrid deep learning approach to improve classification of low-volume high-dimensional data. *BMC Bioinformatics*, *24*(1), Article 419. https://doi.org/10.1186/s12859-023-05557-w

Mazhnaya, A., Tobin, K. E., & Owczarzak, J. (2018). Association between injection in public places and HIV/HCV risk behavior among people who use drugs in Ukraine. *Drug and Alcohol Dependence*, *189*, 125–130. https://doi.org/10.1016/j.drugalcdep.2018.04.036

McGraw, D., & Mandl, K. D. (2021). Privacy protections to encourage use of health-relevant digital data in a learning health system. *NPJ Digital Medicine*, *4*(1), Article 2. https://doi.org/10.1038/s41746-020-00362-8

*Published online: Copyright © 2016-2026 Rangsit University*

Mugurungi, O., Mbunge, E., Birri-Makota, R., Chingombe, I., Mapingure, M., Moyo, B., ... & Musuka, G. (2024). Predicting sexually transmitted infections among men who have sex with men in Zimbabwe using deep learning and ensemble machine learning models. *PLOS Digital Health*, *3*(7), Article e0000541. https://doi.org/10.1371/JOURNAL.PDIG.0000541

Saldana, C. S., Burkhardt, E., Pennisi, A., Oliver, K., Olmstead, J., Holland, D. P., ... & Ochoa, K. V. S. (2024). Development of a Machine Learning Modeling Tool for Predicting HIV Incidence Using Public Health Data From a County in the Southern United States. *Clinical Infectious Diseases*, *79*(3), 717–726. https://doi.org/10.1093/CID/CIAE100

Shija, J. (2025). A machine learning model for early detection of sexually transmitted infections. *Computational Ecology and Software*, *15*(2), 30.

Simmachan, T., Lerdpraserdpakorn, N., Deesrisu, J., Sriwipat, C., Shakya, S., & Boonkrong, P. (2025). A penalized regression and machine learning approach for quality-of-life prediction in psoriasis patients. *Healthcare Analytics*, *8*, Article 100417. https://doi.org/10.1016/j.health.2025.100417

Soe, N. N., Latt, P. M., Yu, Z., Lee, D., Kim, C. M., Tran, D., ... & Zhang, L. (2024a). Clinical features-based machine learning models to separate sexually transmitted infections from other skin diagnoses. *Journal of Infection*, *88*(4), Article 106128. https://doi.org/10.1016/J.JINF.2024.106128

Soe, N. N., Towns, J. M., Latt, P. M., Woodberry, O., Chung, M., Lee, D., ... & Fairley, C. K. (2024b). Accuracy of symptom checker for the diagnosis of sexually transmitted infections using machine learning and Bayesian network algorithms. *BMC Infectious Diseases*, *24*(1), Article 1408. https://doi.org/10.1186/S12879-024-10285-4

Teixeira, I. V., da Silva Leite, M. T., de Morais Melo, F. L., da Silva Rocha, É., Sadok, S., Pessoa da Costa Carrarine, A. S., ... & Endo, P. T. (2023). Predicting congenital syphilis cases: A performance evaluation of different machine learning models. *PloS one*, *18*(6), Article e0276150. https://doi.org/10.1371/journal.pone.0276150

Xu, X., Fairley, C. K., Chow, E. P., Lee, D., Aung, E. T., Zhang, L., & Ong, J. J. (2022a). Using machine learning approaches to predict timely clinic attendance and the uptake of HIV/STI testing post clinic reminder messages. *Scientific Reports*, *12*(1), Article 8757. https://doi.org/10.1038/s41598-022-12033-7

Xu, X., Ge, Z., Chow, E. P., Yu, Z., Lee, D., Wu, J., ... & Zhang, L. (2022b). A machine-learning-based risk-prediction tool for HIV and sexually transmitted infections acquisition over the next 12 months. *Journal of clinical medicine*, *11*(7), Article 1818. https://doi.org/10.3390/JCM11071818

Yang, J., Simmachan, T., Shakya, S., & Boonkrong, P. (2025). Classification of infectious and parasitic diseases by smart healthcare system. *Engineering Proceedings*, *108*(1), Article 14. https://doi.org/10.3390/engproc2025108014

Zheng, J., & Yu, Z. (2021). A novel machine learning-based systolic blood pressure predicting model. *Journal of Nanomaterials*, *2021*(1), Article 9934998. https://doi.org/10.1155/2021/9934998

*Published online: Copyright © 2016-2026 Rangsit University*