World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

PREDICTING HIV DRUG RESISTANCE MUTATIONS: BIOINFORMATIC STRATEGIES, MACHINE-LEARNING MODELS, AND CLINICAL RELEVANCE

Aby John and Elamathi Natarajan *

Bioinformatics, Biotecnika Info Labs Pvt. Ltd., Bengaluru, Karnataka, India.
* Corresponding Author

ORCID Details
Aby John: https://orcid.org/0009-0001-0052-1901
Elamathi Natarajan: https://orcid.org/0000-0002-1944-9137

World Journal of Advanced Research and Reviews, 2026, 32(02), 1314–1324

Publication history: Received on 10 October 2026; revised on 16 November 2026; accepted on 18 November 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.32.2.2840

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

HIV-1 is characterised by a high degree of genetic diversity, and it is this diversity that leads to resistance against antiretroviral drugs. Resistance occurs when changes in the viral genome reduce the virus's sensitivity to one or more of the antiretroviral agents. By identifying and predicting the mutations associated with resistance, clinicians and researchers are able to explain why treatment fails, monitor the way in which the virus evolves, and select appropriate treatment regimens. Traditional genotypic interpretation makes use of established mutation patterns and scoring systems in order to estimate susceptibility. More recently, progress in the fields of bioinformatics and machine learning has expanded the range of methods available for deducing resistance from viral sequences, the methods extending from statistical regression and support vector machines to random forests, gradient boosting and deep neural networks. Resources that are openly available, especially the Stanford HIV Drug Resistance Database, provide paired genotype and phenotype records which enable model building and benchmarking. Yet, despite this, it is still very difficult to make reliable predictions, due to the diversity of viruses, the effect of interacting mutations, imbalanced class distributions, inconsistent definitions of resistance, and the lack of independently validated data. This review addresses the biology of HIV drug resistance, the existing interpretation tools, computational prediction methods, evaluation practices, and the open problems, and looks at how interpretable and widely applicable systems might be developed. The use of computational models together with established interpretation systems and appropriate clinical evidence could make resistance assessment more reliable.

Keywords: HIV-1; Drug Resistance Mutations; Bioinformatics; Machine Learning; Genotypic Resistance; Sequence Analysis; Antiretroviral Therapy; Computational Prediction.

1 INTRODUCTION

HIV-1 is a retrovirus that primarily infects CD4-positive T lymphocytes, progressively weakening the immune system. Combination antiretroviral therapy (ART), involving nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs),

non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), integrase strand transfer inhibitors (INSTIs), and entry inhibitors, has transformed HIV infection into a manageable chronic condition. However, drug resistance remains a major challenge, reducing treatment effectiveness and limiting therapeutic options [1,2,3].

1.1 HIV-1 Genetic Variability and Drug Resistance

HIV-1 exhibits extensive genetic diversity due to the absence of proofreading activity in reverse transcriptase, rapid replication, and frequent recombination. These processes generate genetically diverse viral populations known as quasispecies [3–5]. HIV-1 comprises four groups (M, N, O, and P), with group M responsible for the global pandemic. Group M includes multiple subtypes and recombinant forms with different geographical distributions. Naturally occurring polymorphisms may influence drug susceptibility and resistance pathways, although most resistance studies focus on subtype B [3–5].

Antiretroviral drug exposure selects resistant variants, particularly when viral replication persists because of poor adherence, drug interactions, or suboptimal treatment. Resistance mutations may reduce viral fitness, leading to compensatory mutations that restore replication capacity [3,6,7].

1.2 Transmitted and Acquired Drug Resistance

Drug resistance occurs as acquired drug resistance (ADR), developing during ART, or transmitted drug resistance (TDR), resulting from infection with an already resistant virus. Resistant variants may persist in viral populations or latent reservoirs and re-emerge during treatment. TDR can compromise first-line treatment, highlighting the importance of pretreatment resistance testing [7,8].

1.3 Resistance Mutations and Computational Prediction

Important resistance-associated mutations include M184V and K65R for NRTIs, K103N and Y181C for NNRTIs, and Q148, N155, and Y143 pathways for INSTIs. Their effects depend on genetic background and interactions with other mutations [7,17].

Established interpretation systems, including Stanford HIVDB, ANRS, and Rega, predict drug susceptibility using curated mutation knowledge. Although clinically useful, they may not fully capture complex mutation interactions. Phenotypic assays directly measure drug susceptibility but are expensive and less accessible [9,11,13].

Machine-learning approaches, including regression, support vector machines (SVMs), random forests, deep learning, and protein language models (pLMs), offer alternative methods for predicting resistance from viral sequences. However, their reliability depends on representative datasets, appropriate validation, and model interpretability.

This review examines the molecular basis of HIV-1 drug resistance, bioinformatics resources, computational prediction methods, published studies, evaluation strategies, research gaps, and future developments.

2 MOLECULAR BASIS OF HIV DRUG RESISTANCE

2.1 Genetic Variation and Selection

HIV-1 genetic variability results from error-prone reverse transcription, rapid replication, and recombination. These mechanisms generate diverse viral populations containing pre-existing resistance-associated variants. ART creates selective pressure that allows resistant variants to expand when drug exposure is insufficient [3,7].

Resistance mutations may reduce viral fitness, while accessory and compensatory mutations restore replication capacity. Consequently, resistance often involves combinations of mutations rather than isolated substitutions. [3,6,7]

2.2 Viral Proteins Targeted by Antiretroviral Drugs

Antiretroviral drugs primarily target reverse transcriptase, protease, and integrase.

Reverse transcriptase (RT): Converts viral RNA into DNA. NRTIs, including tenofovir, lamivudine, emtricitabine, and zidovudine, inhibit DNA synthesis through nucleotide incorporation and chain termination. Important resistance mutations include M184V/I, K65R, Q151M, and thymidine analogue mutations (TAMs). NNRTIs, such as efavirenz, nevirapine, and rilpivirine, bind an allosteric pocket in RT. Mutations including K103N, Y181C, Y188L, and G190A/S reduce drug susceptibility [3,7].

Protease (PR): Cleaves viral polyproteins during virion maturation. Protease inhibitors, including darunavir, atazanavir, and lopinavir, block this process. Representative resistance mutations include V32I, I50V/L, I84V, V82A, and L90M. Accessory and Gag cleavage-site mutations may compensate for reduced viral fitness [3,7,17].

Integrase (IN): Inserts viral DNA into the host genome. INSTIs, including raltegravir, dolutegravir, and bictegravir, inhibit this process. Major resistance pathways include Q148H/K/R, N155H, and Y143R/C, while G118R and R263K are associated with resistance to newer INSTIs [7,18].

2.3 Genotypic and Phenotypic Resistance Testing

Genotypic testing identifies resistance-associated mutations in PR, RT, and IN genes. Interpretation systems such as HIVDB, ANRS, and Rega translate sequence variations into predicted susceptibility. However, conventional sequencing may not detect low-frequency minority variants [10,11].

Phenotypic testing directly measures drug susceptibility, generally using inhibitory concentration (IC_{50}) or fold-change values relative to a reference strain. Although it captures the combined effects of mutations, it is more expensive and time-consuming than genotypic testing [10].

2.4 Primary, Accessory, and Compensatory Mutations

Resistance-associated mutations have three major functional categories:

- Primary mutations: Directly reduce drug susceptibility, such as M184V, K65R, K103N, I50L, and N155H.
- Accessory mutations: Enhance resistance or improve viral replication alongside primary mutations, such as L10I/V, A71V, E138K, and G140S.
- Compensatory mutations: Restore viral fitness without necessarily increasing drug resistance [3,6,7].

Interactions among these mutations produce epistasis, whereby the effect of one substitution depends on other mutations in the genetic background. These interactions may be synergistic or antagonistic, highlighting the importance of modelling mutation combinations.

3 BIOINFORMATICS RESOURCES

Bioinformatics resources provide viral sequences, genotype–phenotype measurements, clinical outcomes, and structural information required for HIV resistance prediction.

3.1 Stanford HIV Drug Resistance Database (HIVDB)

Stanford HIVDB is a major resource for HIV resistance research, providing genotype–phenotype datasets involving PR, RT, and IN sequences, measured drug susceptibility, curated resistance mutations, and interpretation algorithms [11,17].

It supports mutation annotation, supervised machine-learning development, and comparison with established interpretation systems. However, subtype B predominance, unequal drug representation, class imbalance, and sequence redundancy limit its generalizability.

3.2 NCBI and Other Sequence Resources

NCBI provides GenBank, NCBI Virus, and the Sequence Read Archive (SRA) for retrieving HIV sequences and sequencing data. These resources support sequence comparison, subtype classification, phylogenetic analysis, and minority-variant investigation. However, most sequences lack phenotypic resistance measurements and treatment histories [4,5].

The Los Alamos HIV Sequence Database provides curated sequences, subtype references, and tools for phylogenetic and recombinant analysis. UniProt and the Protein Data Bank (PDB) provide protein annotations and three-dimensional structures that support the investigation of mutation effects on drug binding [10,11].

3.3 Resistance Interpretation and Clinical Resources

Established interpretation systems include Stanford HIVDB, ANRS, Rega, and geno2pheno. HIVDB, ANRS, and Rega primarily use curated mutation rules, whereas geno2pheno employs statistical prediction approaches. IAS-USA and WHO provide updated resistance mutation information [7,8].

Clinical resources such as EuResist combine viral genotypes, treatment histories, viral loads, and clinical outcomes to investigate treatment response. However, adherence, pharmacokinetics, and differences between cohorts can influence predictions [8,22].

3.4 Dataset Selection and Sequence Processing

Different resources serve different computational purposes:

- Sequence datasets: GenBank, NCBI Virus, SRA, and Los Alamos support sequence analysis and surveillance.
- Genotype–phenotype datasets: HIVDB supports supervised resistance prediction.
- Clinical outcome datasets: EuResist supports treatment-response prediction.
- Structural datasets: PDB and UniProt support protein and drug-binding analysis.

Resistance labels must be distinguished according to whether they represent measured phenotypic susceptibility, algorithm-derived predictions, or clinical outcomes.

Sequence processing generally involves quality control, alignment against a reference such as HXB2, and identification of amino acid substitutions. Mutations may be encoded using binary indicators, categorical variables, physicochemical descriptors, or protein language model embeddings [17,20,21].

4 COMPUTATIONAL PREDICTION METHODS

Computational approaches predict HIV-1 resistance by identifying relationships between viral sequences and drug susceptibility. Prediction targets include continuous fold changes, categorical resistance levels, and clinical treatment outcomes [12–17,20,21].

4.1 Rule-Based Systems

Stanford HIVDB, ANRS, and Rega use expert-curated resistance mutations and scoring algorithms to estimate drug susceptibility. HIVDB assigns penalty scores to individual and combined mutations.

These systems are transparent and biologically interpretable but require regular updates and may not fully capture complex mutation interactions [9,11,17].

4.2 Regression Models

Linear and logistic regression predict continuous fold changes or resistance categories using mutation-based features. Ridge, LASSO, and elastic-net regularization help manage correlated and rare mutations.

Their advantages include computational efficiency and interpretability. However, additive models may fail to capture complex nonlinear mutation interactions [12,16,17].

4.3 Support Vector Machines

SVMs and support vector regression (SVR) classify resistant and susceptible isolates or predict continuous drug susceptibility. Kernel functions allow nonlinear relationships between mutations to be modelled.

They perform well with high-dimensional datasets but require careful parameter tuning and validation to prevent overfitting [13,17,20].

4.4 Random Forests and Gradient Boosting

Random forests and gradient boosting algorithms, including XGBoost and LightGBM, capture nonlinear relationships and interactions among resistance mutations.

Feature importance and SHAP analysis help identify influential sequence positions. However, feature importance does not establish biological causality, and performance may be affected by small or imbalanced datasets [3,7].

4.5 Deep Learning and Protein Language Models

Deep-learning approaches, including multilayer perceptrons (MLPs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, identify complex sequence patterns and mutation interactions. Multitask learning can improve prediction by sharing information across related drugs [20,21].

Protein language models such as ESM and ProtTrans generate sequence embeddings that capture evolutionary and contextual information. These representations can be combined with downstream machine-learning algorithms to predict drug susceptibility, potentially improving the representation of uncommon mutations.

However, deep-learning models require careful validation because of limited labelled data, overfitting, computational requirements, and restricted interpretability. Methods such as SHAP, integrated gradients, and saliency analysis can identify influential sequence positions but require biological validation [13,17,20].

4.6 Model Evaluation

Model evaluation depends on the prediction target. Continuous susceptibility prediction requires metrics such as RMSE and R^2, whereas resistance classification requires AUROC, sensitivity, specificity, precision, PR-AUC, and MCC.

Random data splitting may produce information leakage when closely related sequences occur in training and testing datasets. Patient-level grouping, sequence-similarity-based splitting, and independent external validation provide more reliable estimates of model generalizability. [20,21,24]

5 COMPARISON OF COMPUTATIONAL APPROACHES AND PUBLISHED STUDIES

Computational approaches differ in interpretability, data requirements, and ability to capture mutation interactions.

Table 1: Comparative Overview of Machine Learning Approaches for HIV Drug Resistance Prediction

Approach	Main advantage	Main limitation
Rule-based systems	Transparent interpretation	Depend on curated evidence
Regression	High interpretability	Limited nonlinear modelling
SVM	Captures nonlinear patterns	Requires parameter tuning
Random forest	Models interactions	Limited interpretability
Gradient boosting	Strong structured-data performance	Sensitive to data quality
Deep learning	Learns complex sequence patterns	Requires careful validation
Protein language models	Captures evolutionary information	Requires further external validation

5.1 Published Prediction Studies

Selected studies demonstrate the progression of HIV resistance prediction from traditional statistical approaches to deep learning and protein language models.

Table 2: Summary of Published Studies on Computational Approaches for HIV Drug Resistance Prediction

Study	Method and dataset	Main findings
Beerenwinkel et al. (2002)	Decision trees; 471 clinical isolates	Prediction errors of 9.6–15.5% for most drugs.
Beerenwinkel et al. (2003)	Linear SVR; 652 genotype–phenotype pairs	Cross-validated R^2 ranged from 0.30–0.79; non-B representation was limited.
Wang & Larder (2003)	Neural networks; lopinavir data	A 28-mutation model outperformed an 11-position model.
Rhee et al. (2006)	Regression, decision trees, neural networks, and SVR	LARS achieved 80.3% mean accuracy.
Rosen-Zvi et al. (2008); Altmann et al. (2008)	EuResist combined models	Clinical and viral features improved AUC from 0.75 to 0.77.
Heider et al. (2013); Riemenschneider et al. (2016)	Multi-label classification	Incorporated cross-resistance information across drugs.

Masso (2015)	Integrase sequence n-gram analysis	Balanced accuracy reached 89% for raltegravir and 85% for elvitegravir.
Sheik Amamuddy et al. (2017)	Regression neural networks	Mean R^2 exceeded 0.95 for most drugs.
Tarasova et al. (2018)	Random forests	Balanced accuracy was approximately 0.84–0.86.
Steiner et al. (2020)	MLP, LSTM, and CNN	CNN achieved mean AUCs of 0.968 (PI), 0.945 (NNRTI), and 0.927 (NRTI).
Farquhar (2026)	ESM-2 embeddings and XGBoost; 6,308 HIVDB sequences	ESM-2 achieved mean AUC of 0.968 compared with 0.955 for XGBoost.

5.2 Major Findings Across Studies

Five important findings emerge:

- Performance differences: Studies use different datasets, resistance definitions, cutoffs, and evaluation metrics, preventing direct comparison.
- Dataset limitations: Farquhar (2026) reported reduced AUCs of 0.884–0.930 across subtype and temporal holdouts compared with 0.968 in the main evaluation.
- Algorithm selection: More complex models do not always outperform traditional approaches. Rhee et al. (2006) demonstrated strong performance using LARS.
- Interpretability: Feature importance and sequence-attribution methods can identify known resistance positions and potential novel associations.
- Uneven drug coverage: Most studies focus on PI, NRTI, and NNRTI resistance, with limited representation of newer integrase inhibitors and non-B subtypes.

Overall, prediction performance depends on dataset quality, drug class, sample size, resistance definitions, and validation strategies rather than algorithm choice alone.

6 EVALUATION OF HIV DRUG RESISTANCE PREDICTION MODELS

Reliable evaluation requires appropriate metrics and validation strategies because HIV resistance datasets frequently contain imbalanced classes and genetically related sequences.

6.1 Classification and Regression Metrics

For binary resistance prediction, performance is evaluated using true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Table 3: Commonly used performance metrics for evaluating HIV-1 drug resistance prediction models: purpose and main limitations

Metric	Purpose	Main limitation
Accuracy	Overall prediction correctness	Misleading with class imbalance
Sensitivity	Detects resistant isolates	Must be assessed with specificity
Specificity	Identifies susceptible isolates	Does not measure resistance detection
Precision	Reliability of resistant predictions	Depends on prevalence
F1-score	Balances precision and sensitivity	Ignores true negatives
AUROC	Measures discrimination across thresholds	Does not assess calibration
PR-AUC	Evaluates minority-class detection	Depends on class prevalence
MCC	Evaluates all confusion-matrix outcomes	Less intuitive to interpret

For continuous fold-change prediction, RMSE, MSE, and R^2 are commonly used. Researchers should report per-drug performance, resistance cutoffs, class distributions, and confidence intervals [20,21,24].

6.2 Validation Strategies

Cross-validation estimates model performance by repeatedly dividing data into training and testing folds. Stratified and nested cross-validation help preserve class proportions and reduce optimistic estimates during feature selection and hyperparameter tuning.

Independent testing uses data excluded from model development, whereas external validation assesses generalizability across laboratories, populations, subtypes, and time periods.

Patient-level and sequence-similarity-based splitting are particularly important because closely related HIV isolates may otherwise appear in both training and testing datasets [20,21,24].

6.3 Data Leakage

Data leakage occurs when information from testing samples influences model development, resulting in artificially high performance. Major sources include repeated patient samples, closely related sequences, duplicate isolates, preprocessing performed before data splitting, and resistance labels derived from the same interpretation system used for evaluation.

Recommended safeguards include patient-level grouping, sequence-similarity clustering, independent temporal and subtype holdouts, and performing feature selection and normalization exclusively within training folds.

6.4 Model Interpretability

Interpretability identifies sequence positions and mutations contributing to resistance predictions. Regression coefficients, feature importance, SHAP, and sequence-attribution methods can highlight biologically relevant mutations.

However, these explanations do not establish causality. Candidate mutations should be compared with established resistance knowledge and validated through independent datasets or experimental studies [7,17,20,21].

7 CHALLENGES AND RESEARCH GAPS

Despite promising prediction performance, several limitations restrict the generalizability and clinical applicability of HIV resistance models.

7.1 HIV-1 Subtype Diversity

Most genotype–phenotype datasets are dominated by subtype B, limiting representation of non-B subtypes and recombinant forms. Natural polymorphisms may be incorrectly interpreted as resistance mutations, while resistance pathways may differ across subtypes [4,5,7].

Beerenwinkel et al. (2003) included only 7.4% non-B sequences. Farquhar (2026) reported AUC values of 0.924 for subtype B, 0.900 for B-divergent, and 0.884 for non-B sequences.

Research direction: Expand non-B genotype–phenotype datasets and apply subtype-specific and leave-one-subtype-out validation. Protein language models should also be evaluated for their ability to capture subtype diversity.

7.2 Mutation Interactions

Resistance involves interactions among primary, accessory, and compensatory mutations. Their effects may be additive, synergistic, or antagonistic. For example, G140S and E138K can modify the effects of Q148 integrase substitutions [7,18].

Rhee et al. (2006) found that adding two-way interactions did not improve prediction, suggesting that interaction modelling may overfit limited datasets.

Research direction: Develop interaction-focused models and validate candidate interactions using structural analysis, site-directed mutagenesis, and phenotypic assays.

7.3 Class Imbalance and Data Scarcity

Resistance datasets frequently contain unequal class distributions. Rhee et al. (2006) reported high-level resistance in only 7% of didanosine and 10% of amprenavir isolates. Steiner et al. (2020) reported sensitivities of 0.50 for didanosine and 0.56 for tenofovir despite accuracies of 0.96 and 0.93, respectively.

These findings demonstrate that high accuracy does not necessarily indicate reliable detection of resistant isolates.

Research direction: Use class-aware metrics, appropriate weighting, multitask learning, transfer learning, and additional phenotype-labelled datasets. Synthetic oversampling should be used cautiously to avoid biologically unrealistic mutation combinations.

7.4 Inconsistent Resistance Labels

Studies use different phenotypic assays, fold-change thresholds, and resistance classifications. Some predict binary resistance, while others use susceptible, intermediate, and resistant categories or continuous fold changes. [7,10,20]

These differences limit direct comparison and may introduce label noise. Models trained on HIVDB, ANRS, or Rega interpretations may also reproduce existing rules rather than independently predict biological resistance [9,11,13].

Research direction: Report assay platforms, drug-specific cutoffs, and label sources. Evaluate alternative classification thresholds and distinguish phenotypic susceptibility from clinical treatment outcomes.

7.5 Limited External Validation

Many studies use random cross-validation within a single database, particularly Stanford HIVDB. Such evaluations may overestimate performance because related sequences can occur in both training and testing sets. [20,21,24]

Farquhar (2026) reported reduced performance on subtype and temporal holdouts compared with the main evaluation.

Research direction: Introduce patient-grouped, phylogeny-aware, temporal, and cross-database validation. Shared benchmark datasets and prospective testing are needed to assess generalizability.

7.6 Clinical Applicability

Laboratory drug susceptibility does not necessarily predict treatment success. Clinical outcomes are influenced by adherence, pharmacokinetics, treatment history, and drug combinations. [8,22]

EuResist studies demonstrated that combining clinical and viral features improved prediction compared with genotype-only models. Furthermore, conventional population sequencing may fail to detect low-frequency minority variants [3,7].

Research direction: Integrate clinical variables, evaluate complete treatment regimens, assess prediction uncertainty, and compare computational models with established systems. Models should complement clinical judgment rather than replace established resistance interpretation.

7.7 Summary of Major Research Gaps

Table 4: Key challenges in HIV-1 drug resistance prediction, their main limitations, and corresponding research priorities

Challenge	Main limitation	Research priority
Subtype diversity	Subtype B predominance	Expand non-B datasets
Mutation interactions	Complex epistatic effects	Interaction-focused models
Class imbalance	Poor detection of rare resistance	Class-aware evaluation
Inconsistent labels	Different assays and cutoffs	Standardized resistance definitions
External validation	Dependence on limited datasets	Independent validation
Clinical applicability	Phenotype differs from treatment outcome	Clinical outcome-based evaluation

8 FUTURE PERSPECTIVES

Future developments should prioritize representative datasets, biologically interpretable models, reproducible research, and independent clinical validation.

8.1 Protein Language Models

Protein language models provide sequence representations that capture evolutionary and contextual information. Farquhar (2026) reported a mean AUC of 0.968 for ESM-2 compared with 0.955 for XGBoost across 18 PI, NRTI, and NNRTI drugs.

Despite this improvement, reduced performance on non-B subtypes and temporal holdouts highlights the need for further validation.

Future research should investigate domain-specific pretraining, rare mutation prediction, subtype generalization, and structure-informed representations [10,21].

8.2 Explainable Artificial Intelligence

Explainable AI methods, including SHAP, integrated gradients, and sequence-attribution analysis, can identify mutations influencing model predictions. [20,21,23]

Steiner et al. (2020) identified known resistance positions among important CNN features, while Farquhar (2026) reported enrichment of attention weights at recognized resistance positions. However, attribution scores do not establish biological causality.

Future work should combine explainability with structural analysis and experimental validation to identify potential resistance-associated mutations [3,7].

8.3 Integration of Curated Knowledge and Machine Learning

Combining established resistance knowledge with machine learning may improve prediction reliability and interpretability. [9,11,17,21]

Potential approaches include incorporating HIVDB penalty scores and curated mutation lists as model features, predicting differences between rule-based scores and measured phenotypes, and investigating disagreements between HIVDB, ANRS, Rega, and machine-learning models.

Integrating structural and viral fitness information may further improve the identification of mutation interactions [3,7].

8.4 Reproducibility and Open Science

Reproducible research requires transparent reporting of dataset versions, sequence-processing procedures, resistance labels, model parameters, and validation strategies. [24]

Studies should provide accessible code, independent test datasets, and standardized performance metrics. Relevant reporting guidelines, including TRIPOD and its AI extension, may support transparent model development [24].

8.5 Independent and Clinical Validation

Translation into clinical applications requires progressive validation:

- Internal validation: Patient-grouped or sequence-similarity-based cross-validation.
- External validation: Testing across independent datasets, subtypes, and time periods.
- Retrospective clinical validation: Evaluating associations between predictions and treatment outcomes.
- Prospective evaluation: Assessing whether model-guided decisions improve treatment outcomes.
- Implementation and monitoring: Monitoring performance, uncertainty, and changes in resistance patterns.

Until clinical benefit is demonstrated, computational predictions should complement established interpretation systems and clinical judgment.

9 CONCLUSION

HIV-1 drug resistance develops through selection within genetically diverse viral populations and involves primary, accessory, and compensatory mutations with context-dependent effects. Computational prediction has progressed from expert rule-based systems to statistical learning, ensemble methods, deep learning, and protein language models. Stanford HIVDB remains a major source of genotype–phenotype data for developing these approaches.

Published studies demonstrate promising predictive performance, including a mean AUC of 0.968 for ESM-2 compared with 0.955 for XGBoost. However, improvements in benchmark performance do not necessarily indicate better generalizability. Subtype B predominance, class imbalance, inconsistent resistance definitions, sequence leakage, and limited external validation remain major challenges.

Future progress requires representative genotype–phenotype datasets, standardized evaluation, interpretable machine learning, and independent validation across diverse subtypes and populations. Integrating curated resistance knowledge, protein language models, structural information, and clinical variables may improve prediction reliability. Ultimately, computational models must demonstrate clinical benefit before being adopted for treatment-guiding applications.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors acknowledge the researchers and publicly available literature and datasets that contributed to this review.

Disclosure of conflict of interest

The authors declare no conflicts of interest.

REFERENCES

[1] Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nat Rev Dis Primers. 2015;1:15035. doi:10.1038/nrdp.2015.35.

[2] Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. Cold Spring Harb Perspect Med. 2012;2(4):a007161. doi:10.1101/cshperspect.a007161.

[3] Clavel F, Hance AJ. HIV drug resistance. N Engl J Med. 2004;350(10):1023–1035. doi:10.1056/NEJMra025195.

[4] Hemelaar J. The origin and diversity of the HIV-1 pandemic. Trends Mol Med. 2012;18(3):182–192. doi:10.1016/j.molmed.2011.12.001.

[5] Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. N Engl J Med. 2008;358(15):1590–1602. doi:10.1056/NEJMra0706737.

[6] Coman RM, Robbins RM, Fernandez S, et al. (Reference on HIV-1 mutation rate and fitness costs). Virus Evol. 2017;3(1):vex003. doi:10.1093/ve/vex003.

[7] Wensing AM, Calvez V, Ceccherini-Silberstein F, Charpentier C, Günthard HF, Paredes R, et al. 2022 update of the drug resistance mutations in HIV-1. Top Antivir Med. 2022;30(4):559–574.

[8] Gupta RK, Gregson J, Parkin N, Haile-Selassie H, Tanuri A, Andrade Forero L, et al. HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. Lancet Infect Dis. 2018;18(3):346–355. doi:10.1016/S1473-3099(17)30702-8.

[9] Eberle J, Gürtler L. The evolution of drug resistance interpretation algorithms: ANRS, REGA and extension of resistance analysis to HIV-1 group O and HIV-2. Intervirology. 2012;55(2):128–133. doi:10.1159/000332009.

[10] Petropoulos CJ, Parkin NT, Limoli KL, Lie YS, Wrin T, Huang W, et al. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. Antimicrob Agents Chemother. 2000;44(4):920–928. doi:10.1128/AAC.44.4.920-928.2000.

[11] Tang MW, Liu TF, Shafer RW. The HIVdb system for HIV-1 genotypic resistance interpretation. Intervirology. 2012;55(2):98–101. doi:10.1159/000331998.

[12] Beerenwinkel N, Schmidt B, Walter H, Kaiser R, Lengauer T, Hoffmann D, et al. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. Proc Natl Acad Sci U S A. 2002;99(12):8271–8276. doi:10.1073/pnas.112177799.

[13] Beerenwinkel N, Däumer M, Oette M, Korn K, Hoffmann D, Kaiser R, et al. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Res. 2003;31(13):3850–3855. doi:10.1093/nar/gkg575.

[14] Drăghici S, Potter RB. Predicting HIV drug resistance with neural networks. Bioinformatics. 2003;19(1):98–107. doi:10.1093/bioinformatics/19.1.98.

[15] Wang D, Larder B. Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. J Infect Dis. 2003;188(5):653–660. doi:10.1086/377453.

[16] Wang K, Jenwitheesuk E, Samudrala R, Mittler JE. Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. Antivir Ther. 2004;9(3):343–352.

[17] Rhee SY, Taylor J, Wadhera G, Ben-Hur A, Brutlag DL, Shafer RW. Genotypic predictors of human immunodeficiency virus type 1 drug resistance. Proc Natl Acad Sci U S A. 2006;103(46):17355–17360. doi:10.1073/pnas.0607274103.

[18] Masso M. Sequence-based predictive models of resistance to HIV-1 integrase inhibitors: an n-grams approach to phenotype assessment. Curr HIV Res. 2015;13(6):497–502. doi:10.2174/1570162X13666150624100535.

[19] Sheik Amamuddy O, Bishop NT, Tastan Bishop Ö. Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks. BMC Bioinformatics. 2017;18:369. doi:10.1186/s12859-017-1782-x.

[20] Steiner MC, Gibson KM, Crandall KA. Drug resistance prediction using deep learning techniques on HIV-1 sequence data. Viruses. 2020;12(5):560. doi:10.3390/v12050560.

[21] Farquhar H. Protein language model embeddings improve HIV drug resistance prediction: a comprehensive benchmark with attention-based interpretability. Bioinformatics. 2026;42(5):btag260. doi:10.1093/bioinformatics/btag260.

[22] Rosen-Zvi M, Altmann A, Prosperi M, Aharoni E, Neuvirth H, Sönnerborg A, et al. Selecting anti-HIV therapies based on a variety of genomic and clinical factors. Bioinformatics. 2008;24(13):i399–i406. doi:10.1093/bioinformatics/btn141.

[23] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems. 2017;30:4765–4774.

[24] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.