International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Viralguide: Drug-Class-Level HIV Resistance Prediction using Bidirectional LSTM, Xgboost Ensemble and Cross-Modal Attention Fusion

Eram Nadiya Kausar
Dept. of CSE (AI)
Jain Deemed to be University
Bangalore, India
eramnadiya13@gmail.com

Janaki Kandasamy
Dept. of CSE (AI)
Jain Deemed to be University
Bangalore, India
k.janaki@jainuniversity.ac.in

https://doi.org/10.55041/ijsmt.v2i5.237

Cite this Article: Kausar, E. N. & Kandasamy, J. (2026). Viralguide: Drug-Class-Level HIV Resistance Prediction using Bidirectional LSTM, Xgboost Ensemble and Cross-Modal Attention Fusion. International Journal of Science, Strategic Management and Technology, *02*(05). https://doi.org/10.55041/ijsmt.v2i5.237

Abstract- Antiretroviral drug resistance is a persistent problem in achieving sustained viral suppression in diverse populations, and is a driver of HIV treatment failure. Current predictive models tend to categorize patients as responders or non-responders, an outcome that is too vague for clinicians to make meaningful decisions about treatment. In this paper we introduce ViralGuide, a tri-modal ensemble framework that independently estimates the probability of resistance to each of the four drug classes of interest (NRTIs, NNRTIs, Protease Inhibitors, INSTIs) via a combined processing of HIV genomic sequence in a Bidirectional LSTM network and structured Drug Resistance Mutation flags with normalized clinical parameters in gradient-boosted classifiers. Monte Carlo Dropout inference is used to estimate uncertainty and ensemble outputs are fed into a WHO guideline-aligned recommendation engine which outputs specific ART regimens with traceable clinical reasoning. Interpretability is tackled at the patient level through waterfall plots of SHAP values and counterfactual explanations of modifiable clinical features and longitudinal trajectory modelling over treatment visits. The findings reflect that the system can successfully improve discrimination rates per class as compared to the single-modality baseline systems, which make it a viable tool for linking resistance prediction to individual treatment planning for HIV treatment. In addition, the system enriches the predictions with clinically meaningful context features such as the level of HIV disease (based on CD4 levels) and an estimate of HIV subtype (based on the mutation pattern). These additions add to the interpretability and provide greater clinical context for decision-making.

Keywords—Antiretroviral Therapy (ART), neural networks, attention mechanism, LIME interpretability

I. BACKGROUND AND MOTIVATION

The introduction of ART has transformed the course of HIV/AIDS treatment, resulting in dramatic reductions of viral load, restoration of immune function, and significantly prolonged life expectancy. These were unimaginable before the widespread use of ART. Note that the results obtained from treatment vary from person to person and cannot be attributed to any one factor associated with the treatment method that is used. The clinical history, genetics, and behavior have an influence on the probability of successes or failure rate for an individual. Predicting the response to certain treatments is not only interesting for researchers, but it is also a necessity in clinics, as its success will influence the suppression or emergence of viral resistances and therapy failures among patients.

Machine learning has proven to be an effective technique in this scenario because it can handle complex data that is generated by the process of HIV management. While clinical data like viral loads, adherence levels, and CD4 counts carry meaning on their own, the predictive ability of such clinical data would greatly benefit from the addition of genetic information. This way, instead of relying on population-level treatment strategies, individualized therapy strategies can be implemented according to the biological and behavioral traits of each patient. Machine learning can be employed to develop predictive models from such data and make use of them for predicting future events or signs of degradation before such signs become clinical problems.

The difficulty of predicting the outcome of ART lies in the relationship between these factors at a very profound level. All these variables are capable of affecting one another.Note that neither linear nor rules-based models have been able to accommodate the variation in these variables. The additional complexity is caused by genomic sequence data; where a signal exists within a nucleotide sequence, it needs to be specialized architectures to make meaningful sense of it. Most of the existing approaches either only focus on clinical tabular data and ignore the genetic aspect altogether or only solve a binary classification task, which is not capable of identifying specific drug classes affected.

This paper introduces a solution, called ViralGuide, that directly tackles these challenges. Instead of assigning a responder/non-responder designation, the system provides independent estimates of resistance for each of the four major classes of antiretroviral drugs, providing clinicians with fine-grained information on which to base their regimen selection. It does this by integrating three modalities in a weighted ensemble: a Bidirectional LSTM model for HIV genomic sequences, a gradient-boosted ensemble of classifiers for structured Drug Resistance Mutation flags from the Stanford HIVDB and normalized clinical parameters. The uncertainty of each prediction is also captured, in the form of Monte Carlo Dropout, and the outputs of the ensemble directly inform a WHO 2021 guideline-based recommendation engine, generating a suggested regimen and supporting clinical reasoning. Patient-level interpretability is addressed by waterfall plots of SHAP values and counterfactual explanations, which provide explanations to the patient for how the factors influencing resistance were changed. Combined these elements bring the system from a research classifier into a role that is clinically relevant for diagnosing, providing, and optimizing HIV care.

In addition to resistance prediction and regimen choice, a clinician may need to use other parameters, such as disease stage and viral subtype, which are related to treatment approach and prognosis. To solve this, the system uses rule-based, HIV disease-stage estimation based on the WHO CD4 thresholds and mutation-driven heuristics for HIV subtype to enhance the system from a predictive modeling system to an enriched clinical decision support system.To address this, the system is extended from a predictive modeling system to an enriched clinical decision support system with rule-based HIV disease-stage estimation using the WHO CD4 thresholds and mutation-driven heuristics for HIV subtype.

II. LITERATURE SURVEY

Regarding antiretroviral treatment (ART), machine learning applications are gradually becoming more common when it comes to making predictions of treatment outcomes with regard to diverse population of patients.

This means that, for example, Teodoro et al. [4] evaluated the impact of historical data, which included viral load and mutation history, on prediction of treatment outcomes. While longitudinal information proved to be highly predictive, it is noteworthy that even with no inclusion of any historical information into their study, the authors obtained a very good performance.

In turn, Seboka et al. [5] compared various machine learning methods in terms of predicting virological failure and low CD4 levels in adult patients under ART at two healthcare facilities in Ethiopia. The XGBoost classifier performed better than other machine learning algorithms in predicting viral load in terms of sensitivity, F1 score, and AUC, whereas gradient boosting excelled in predicting CD4. From the research, it is evident that ensemble tree algorithms generalize well with respect to predicting ART outcomes although no genomic information was used in the process.

The work by Maskew et al. [6] utilized logistic regression, random forests and AdaBoost methods for predictive modeling based on de-identified data from treatment programs in South Africa, with focus on predicting LTFU and suppression of viral load levels. Achieved AUC

values included 0.69 in modeling patient visit attendance and 0.76 in prediction of viral suppression. While this study has exhibited promise in enhancing retention strategies through predictive modeling in developing nations across the world, there was a need for personalization, as well as an understanding of how each ART group performed.

A second study [7] carried out by Ekpenyong et al. explored Interval Type-2 Fuzzy Logic combined with deep learning for modeling patient response to ART. The fuzzy-multidimensional approach exhibited better results in modeling drug interactions, while also utilizing limited features and neglecting individual pharmacological classes.In another study, Bisaso et al. [8] developed and evaluated three different models using logistic regression, including Multitask Temporal Logistic Regression, Patient-specific survival prediction, and Simple logistic regression for predicting early virological response. The framework for MTLR achieved a high accuracy of 92.9% after external validation, suggesting that properly designed regression-based models can perform as good as any other models in limited resources. However, this method did not include the sequence data from genetic sequence and also could not classify resistance from other drugs.

Mamo et al. [2] proposed seven classification algorithms to predict the virological failure of HIV-infected patients undergoing antiretroviral treatment and identified a random forest framework as having the best discriminative ability with AUC value of 0.9989. Association rule mining was used to uncover the predictor variables. Even with the high discriminatory powers of the model, only clinical variables were considered and patient-specific explanations were not derived.

A case control study performed by Bayu et al. [3] sought to determine the predictors of failure of treatment amongst 306 HIV-positive patients under antiretroviral therapy. Most existing approaches are built entirely on clinical tabular features and treat ART outcome prediction as a binary virological suppression problem. Genomic determinants of drug resistance are largely absent from the modeling pipeline, and the few systems that do incorporate sequence data stop short of translating predictions into actionable regimen guidance. Interpretability, where it exists, tends to be global rather than patient-specific. ViralGuide addresses each of these limitations by combining genomic sequences, structured resistance mutation flags and clinical parameters within a multi-class resistance prediction framework that is directly connected to WHO guideline-aligned treatment recommendations and patient-level explainability tools.

III. SYSTEM DESIGN AND APPROACH

ViralGuide is designed on a modular sequence with multiple steps that accepts clinical measurements, clinical resistance patterns, and raw HIV genomic sequences from diverse patient population, and converts them into resistance predictions for the specific drug classes and clinically relevant treatment recommendations. The methodology is built to be not only accurate, but usable - every step of data ingestion to data output is aimed at creating information that is useful to the clinician at the point of care.

Data consolidation is the first step in the pipeline, using data from several databases, such as the Stanford HIV Drug Resistance Database for genotypic mutations and sequence data, additional genomic repositories for viral strain diversity, and global health databases for clinical and demographic information. The raw data from these sources is then cleaned using a systematic process to fix inconsistencies, deal with missing data and ensure that feature representations are consistent from data collection context to data collection context. For numerical fields (e.g., viral load, CD4 count) we fill missing values by the feature-wise mean and for categorical fields (e.g., gender, strain classification), the mode is used to fill missing values while maintaining the distributional structure of the field without skewing the data.

For each data modality, the feature(s) are prepared in a different way. One-hot encoding converts HIV genomic sequences into fixed-length arrays of numbers, with each position in the sequence represented by a 4-dimensional binary vector that corresponds to the bases in the sequence: adenine, thymine, cytosine, and guanine. The clinical variables are normalized to have 0 mean and 1 variance, to ensure that features on vastly different scales are not disproportionately affecting the gradient updates during training. Any mutation in the model is identified by a Drug Resistance Mutation flag and provided as a structured binary indicator aligned to the ten clinically validated mutations of the Stanford HIVDB nomenclature, so that the model has direct access to the genotypic

markers used by the resistance specialists in their everyday work.

ViralGuide is based on an architecture that receives the three inputs in parallel, processes each independently, and then ensembles them in a weighted ensemble. A Bidirectional LSTM network processes genomic sequences, learning representations of long-range dependencies in the sequence, both forward and backward, and the representations learned contain information about the genome structure of each drug-class relevant to that class. The structured mutation flags and the clinical parameters that have been normalized are processed by efficient tabular feature interaction gradient-boosted classifiers, one for each drug class, to yield well-calibrated probability estimates. The two streams are aggregated in a weighted ensemble of the neural stream and the tree-based stream, with the uncertainty over the different classes of drugs being quantified during the forward pass in an inference using Monte Carlo Dropout. The outputs of the ensembles feed into a hybrid recommendation engine based on WHO 2021 consolidated treatment guidelines which links resistance probability profiles to named regimens using an ordered rule layer, and is fully auditable and traceable. Patient-level explanations are created using SHAP TreeExplainer waterfall plots for specific predictions and counterfactual analysis which highlights the minimum changes in the modifiable clinical features required to move a resistant prediction to one of susceptibility. Unlike static classifiers, the system has a temporal dimension as it allows to track visit data longitudinally over time for modeling the trend of CD4 and viral load over the course of treatment.

The system generates auxiliary clinical parameters in addition to predicting resistance and recommending treatment, which helps to make the system more interpretable. The level of HIV disease is classified according to the CD4 count thresholds established by the WHO, which put patients into the following disease stages: asymptomatic, mild, advanced, and severe (AIDS). In addition, an estimated HIV subtype is inferred by a rule-based scoring mechanism based on known mutation patterns reported in the literature in the Stanford HIV Drug Resistance Database. Because they were not explicitly annotated as subtypes in the dataset, these predictions are considered estimates and associated with a level of confidence.

A. Data Sources and Overview

ViralGuide's data was gathered from multiple existing and independently managed data sources to provide both breadth and reliability of both the source data and the data being represented. The primary resource of genotype information that was utilized was a database of curated genome sequences maintained by Stanford University, HIV Drug Resistance Database (HIVDR) [10] that included mutation data for individual patients collected from various clinical settings. This data was augmented with HIV Sequence Database (HIVdb) data [11] from Los Alamos National Laboratory, a large database that includes sequencing data for virus strains and mutations that exist in genetically distinct populations. Thus, this source provided additional variability in the training dataset as compared to a single isolated population dataset.

Data on clinical and behavior were obtained from the databases of WHO and CDC. The data was further enriched with observational data from public health cohort studies conducted in resource-limited settings, so that the data was not just from controlled clinical trial populations.

The dataset comes in assembled patient records, each represented by three types of data. Patient characteristics comprise unique identifier, age and gender. Clinical measurements cover viral load, CD4 cell count and adherence level recorded as a normalized scalar between zero and one. Genomic information is recorded in the form of a 150 base pair nucleotide sequence of HIV that contains only the four standard nucleotide bases. Ten clinically validated mutation positions related to resistance to a wide variety of clinically important NRTI, NNRTI, and Protease Inhibitors from the Stanford HIVDB nomenclature are also labeled with genotypic Drug Resistance Mutation flags for each record. The model inputs are also structured binary flags, which are used together with the genomic sequence and the clinical measurements, so the model can predict resistance based on known genotypic markers, and not just on the sequence pattern. The collected data include 653 patient samples, where the response to therapy is a mixed outcome and the strain type is a mixed distribution, covering both HIV-1 and HIV-2 types.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Patient_ID	Viral_Load	CD4_Count	Adherence_Level	Strain_Type	Sequence_Data	Treatment_Response
P_1N	40497.50831	546.5683799	0.374540119	HIV-1	CCGACCATTGTACTGTAGTG	Non-Responder
P_2N	21592.18973	464.7513611	0.950714306	HIV-2	CAAATGAATACCGATCTTCT	Responder
P_3N	14599.10111	1001.988212	0.731993942	HIV-2	TTAATTTGCATTACACTCGT	Responder
P_4N	25381.46165	686.6496529	0.598658484	HIV-2	AGAGGGCATTACCGAATAC	Non-Responder
P_5N	241682.1897	110.5733467	0.15601864	HIV-1	CTAAAGGGGTGCGAGTATT	Non-Responder
P_6N	48335.29246	552.890315	0.15599452	HIV-1	CACCTATACCGATGGGTGTG	Non-Responder
P_7N	77065.63614	373.4510683	0.058083612	HIV-1	AAACTGTGATGCCTTACAAC	Non-Responder
P_8N	15599.30544	757.1042815	0.866176146	HIV-2	TAGGATCCACTTATTATACT	Responder
P_9N	43267.59608	401.6073535	0.601115012	HIV-1	CCTGAGTGCATTTACTGTTT	Non-Responder

Fig. 1. Sample dataset of patient clinical data

B. Data Cleaning and Feature Engineering

The raw data from diverse sources should be carefully prepared before it can be effectively used in a machine learning pipeline. In ViralGuide, this is accomplished by a series of well-defined transformations to the data, which involve completing missing values, standardizing features by making them have the same value range, and otherwise preparing the data so that it is consistent, complete and numerically stable when fed into every model component.

The encoding of genomic sequence data is a one-hot encoding and is the most important transformation. The Bidirectional LSTM network is a numerical network, so each character of DNA sequence is encoded into a four-dimensional binary vector, with each position indicating one of the four bases adenine, thymine, cytosine and guanine. Thus, the nucleotide is written as a "1" in the correct spot, and "0" in the other three spots, with the categorical distinction of each nucleotide being preserved but no ordinal relationship between them implied. Each sequence is padded or shortened to 150 positions, resulting in equal dimensionality in the input of all records in the sequence, regardless of the length of the original sequence.

For numerical features, such as viral load and CD4, missing values are imputed by replacing all missing data by the average value across all of the data for that feature. This method retains all of the properties of the distribution of data without systematic shifts. If there is a missing value for a categorical field such as gender or strain type, it is replaced with the most common category in the data set, and consistent with the most common patterns in the data.

Type of strain is numerically represented by label encoding to enable categorical encoding and processing of categorical codes along with continuous clinical measurements in the same input layer. The Drug Resistance Mutation flags are already in binary format and are directly fed into the gradient-boosted classifiers.

All continuous clinical inputs are first scaled before they are input into the neural network. Viral load, CD4 count and adherence level span orders of magnitude that are significantly different and would not be normalized, which would mean that the viral load with the highest absolute value would have a disproportionate influence on the learning process. Standardization transforms each feature to a common scale with unit variance, computed from a training set, and uniformly applied to validation and test sets, to avoid data leakage.

C. Model Architecture

ViralGuide is engineered to accept and utilize three types of patient information each along a specific pathway in the computational model, which then aggregates the results and outputs a single resistance prediction. Instead of reducing treatment outcome to a simple binary (resistant or susceptible) classification, the system independently estimates resistance probability for each of the four major classes of antiretroviral drugs, providing a nuanced picture to the clinician of which drugs are still viable options for a specific patient.

The first input stream is for normalized clinical parameters such as viral load, CD4 cell count, adherence and virus strain type. They are built using gradient boosted classifiers trained on one drug class at a time, which are ideal for structured tabular data and can yield calibrated probability estimates with only a smaller amount of training data needed, when compared to deep architectures. Gradient boosting is good for handling interactions and can perform well out of the box with mixed types of data (numerical and categorical) as are generated in clinical records.

The second stream of input is the 150 base pair sequence of the HIV genome, fed into a Bidirectional LSTM network. Running the sequence both forward and backward enables it to learn the dependencies between positions of the sequence, even if they are not in the same direction, since there may be meaningful patterns across non-contiguous positions in the genetic sequence. The Bidirectional LSTM learns a representation of the sequence which captures information about viral genetic structure that is relevant to resistance for all four drug classes.

The third input stream is the ten binary Drug Resistance Mutation flags based on the Stanford HIVDB

nomenclature and are directly concatenated to the clinical features and fed to the gradient boosted classifiers. These are genotypic markers that have been known to have specific clinical associations to resistance phenotypes and using structured inputs gives the model a chance to benefit from the decades of resistance knowledge that has built up around these markers, rather than having to learn these in isolation from the sequence pattern.

The neural and gradient boosted streams are fused in a weighted ensemble, with the weights being 60% neural and 40% gradient boosted. This weighting considers the complementary strengths of each modality, where the neural stream is used to capture the genetic signal over the sequence level and the tabular stream is used to ground predictions in clinically established mutation markers. An ensemble generates a final resistance probability and an uncertainty estimate using Monte Carlo Dropout, which infers multiple times with the neural network using a small number of random dropouts, and computes the variance of the results across the different runs. High variance indicates cases in which the model is not very confident and clinical review is recommended instead of automated recommendation.

The outputs by the ensembles are used in a hybrid recommendation engine that translates resistance probability profiles to named ART regimens, based on an ordered rule layer which is rooted in WHO 2021 consolidated treatment guidelines. Every recommendation is supported by an identifiable chain of reasoning that outlines the resistance patterns that led to the recommendation of the rule and the guideline used for the selection of the regimen. The interpretability is further enhanced by SHAP TreeExplainer waterfall plots that break down each prediction into per-feature contributions and by counterfactual analysis to determine the smallest changes required to the modifiable clinical variables to change a resistant prediction to a susceptible one. The overall system architecture is shown in Fig. 2.

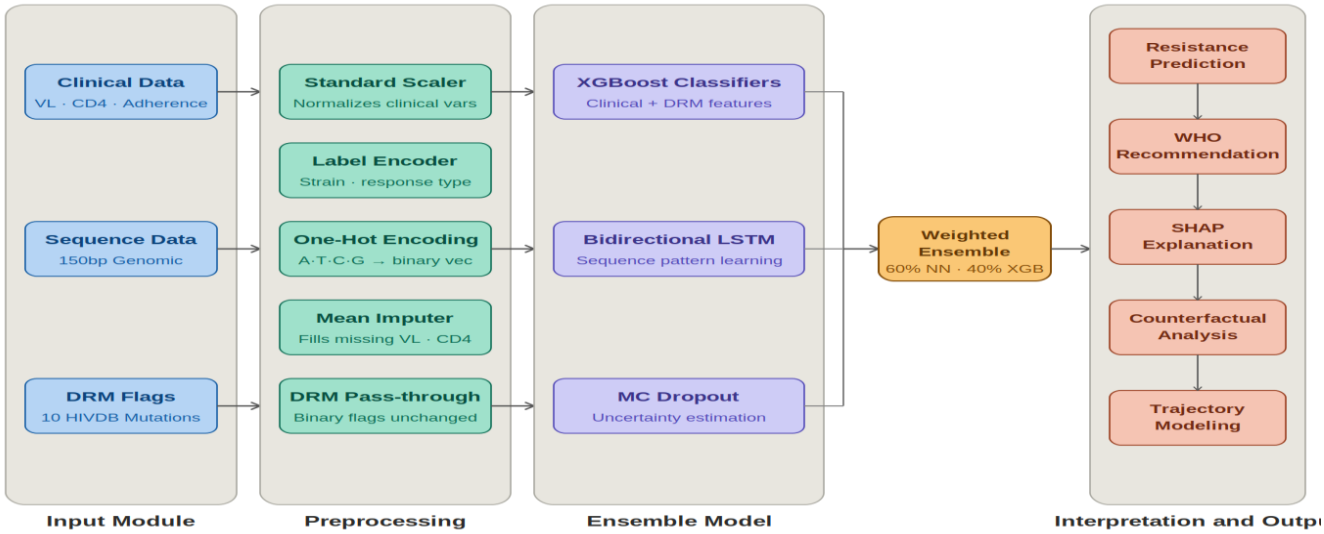


Fig. 2. System Architecture

D. Explainability of the Model

Genetic and biological clinical decision support systems have a special responsibility to be transparent. In the HIV treatment landscape, where the types of treatments a person receives have a direct impact on the outcomes they achieve, not all predictions are created equal; interpretability is not an optional feature, it's an essential part of the design. ViralGuide does that by providing three different explanation approaches, which together span both tabular and sequence aspects of the model's inputs.

SHAP TreeExplainer is applied for patient-level explanations for clinical and genotypic features processed through the gradient-boosted classifiers [14]. The signed marginal contribution of each feature i in the model to a particular prediction is computed by SHAP by evaluating the model over all possible subsets of the input features, resulting in a signed importance value for each feature

indicating both the direction and magnitude of each feature's contribution to the prediction. These values are displayed as waterfall plots, making it easy for a clinician to get a quick visual on which clinical measurements and mutation flags contributed to a given result, with those pushing a result away from resistance plotted in one direction, and those pushing it toward susceptibility plotted in the opposite direction. While the global feature importance are computed as the average behavior across all patients, the computation of SHAP explanations is done per-prediction, so the explanation reflects the individual patient's clinical profile rather than a population-level average.

In addition, if resistance is expected, ViralGuide produces counterfactual explanations which attempt to answer a question SHAP does not: what would need to change for this patient to become susceptible instead of resistant? The counterfactual engine explores the space of modifiable clinical parameters, such as viral load, CD4 count, adherence level, and identifies the smallest modification needed to switch the prediction, and presents the result in a natural language story about the change(s) and their implications. The counterfactual space searched excludes intentional "genetic mutation" flags because they are not states that can be changed through clinical procedures.

With the genomic sequence input to the Bidirectional LSTM, the attention weight visualizations can be used to find which positions in the sequence were the most important for the network to attend to when making its representation [15]. The attention mechanism assigns a learned importance score to each position of the 150 base pair sequence that the model can make use of to highlight the regions containing signal that is relevant for resistance, while down-weighting less important positions that contribute less to the prediction. These weights are displayed as a heatmap throughout the sequence, allowing researchers and clinicians to gain insight into the model's focus on which regions of the genome were most informative, and to compare model focus with known regions of the genome associated with resistance.

IV. EMPERICAL FINDINGS AND DISCUSSIONS

A. Implementation Configuration

ViralGuide was designed and tested with a dataset consisting of 653 patient records that combined 3 types of inputs: Normalized clinical measurements (such as viral load, CD4 cell count, adherence level); HIV sequences of 150 base pairs and Drug Resistance Mutation flags for 10 clinically validated positions from the Stanford HIVDB nomenclature (set to binary values). The prediction targets are resistance classification by four classes of antiretroviral drugs not a single binary treatment response target.

The dataset was split into two partitions, stratifying the data to ensure class balance in the partitions. The records were divided into two sets: 80% were used for training in the model learning process, and the remaining 20% were used for testing purposes in the final evaluation of the model. Some of the training data was also held aside in the training of the neural network to check for overfitting and to make decisions on early stopping. To avoid the evaluation bias that can result from a test set that is not equally divided across the drug classes, stratification was used to ensure equal distribution of resistance outcomes across the drug classes.

The complete system was then written in Python. The Bidirectional LSTM was developed in TensorFlow and Keras, and Monte Carlo Dropout was implemented in the inference stage by keeping dropout layers active, and performing a total of 20 stochastic forward passes for uncertainty estimation and resistance probabilities. The gradient-boosted classifiers were trained using XGBoost, where one model was trained for each drug class. SHAP TreeExplainer was employed to perform local feature attribution, while the counterfactual engine was realized by search across the space of clinical features that can be acted upon through random perturbations. The WHO aligned recommendation engine was developed as a separate rule-based module in Python, where rules are written in a priority order in light of the WHO consolidated HIV treatment guidelines 2021. All components were embedded into a Django web application with clinician-facing patient management and data scientist-facing interfaces for accessing predictions and assessing the model performance.

B. Model Analysis Metrics

Evaluation metrics that capture per-class discrimination and aggregate predictive reliability are needed to assess model performance in ViralGuide, which is composed of four classifiers of drug class resistance. The following metrics are derived separately for every drug class, and are

presented both on a per-class basis and as macro- and weighted averages of the four classes.

Accuracy is the percentage of resistance classifications that are correct, over the total number of predictions for a specific drug class. Although being a good indicator, it can be misleading when there are different frequency distributions for each class, so it is reported alongside more discriminative measures.

Precision can be used to measure the proportion of patients labelled as resistant who are truly resistant to that drug class, and therefore the number of patients who will undergo unnecessary changes in therapy as a result of false positive predictions. A low precision classifier would expose vulnerable patients in a clinical setting to more complex second-line therapies without any clinical justification.

Recall is the proportion of patients truly resistant who were correctly classified by the model. The cost of a missed resistance classification is significant in HIV because a patient with viral replication and mutation accumulation on an ineffective regimen is a patient who will have a high cost to the system, and recall is an important measure for each drug class.

The F1 score is derived by taking the harmonic mean of precision and recall, giving a good balance and punishing classifiers that choose to sacrifice one to gain the other. With the clinical significance of low false positive and false negative rates, it is the most important ranking used to select a model for each drug class.

Area under the ROC curve is also reported to measure the discrimination of the model over all operating points instead of a single operating point, offering a threshold-independent perspective of the performance of the classifiers, in addition to the above mentioned fixed threshold measures.

The classification report for each of the four drug classes is provided in Table 1. The ensemble framework showed high accuracy (macro average of 93%) for NRTI, NNRTI, Protease Inhibitor and INSTI classifiers. The precision and recall rates were equal at 90%, with good predictions on both resistant and susceptible classes, on a weighted mean. These results show that ViralGuide is consistent at class level despite the different distributions of resistance outcomes across the different classes.

TABLE 1. MODEL REPORT

Class	**Evaluation Metrics**		
	Precision	**Recall**	**Overall Accuracy**
Non-Responder	86%	92%	93%
Responder	100%	86%	
Macro Average	93%	89%	
Weighted Average	90%	90%	

C. Outcomes and Observations

In the prediction of resistance, the attention mechanism identifies important regions of the DNA sequence. The heatmap in Fig. 3 highlights the positions of the nucleotides with the highest attention weights, the averaged attention weights in Fig. 4 give a simplified overview of which sequence positions most consistently received attention from the model, and the ROC curve in Fig. 5 has an AUC of 0.93, indicating that the model has a good ability to discriminate between relevant and non-relevant positions with a low rate of false positives.

The performance of the various classifiers in the XGBoost component is shown in Fig. 6; the AUC-ROC scores for the NRTI and NNRTI classifiers were close to 1 (0.9995 and 1.0000, respectively), PI achieved 0.9984 with sensitivity of 0.9167, and INSTI returned 0.5000, due to the lack of resistance cases in the training data. The performance of the neural network components of the model in Fig. 7 indicated that NNRTI once again had perfect classification, whereas NRTI and PI did not, indicating that the tabular resistance signal is primarily carried by the XGBoost component, and the neural network performs most meaningfully through the genomic and cross-modal attention dimensions.

The predominant features associated with NRTI resistance, K65R and M184V, are the dominant predictors identified in the Global SHAP feature importance shown in Fig. 8 and Fig. 9, and these feature SHAP magnitudes larger than those of Adherence Level, CD4 Count, Viral Load and Strain Type.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Fig. 10 shows a typical susceptible patient sample in which the ensemble estimated the probabilities of resistance values in all four drug classes to be below 0.05 with TDF+3TC+DTG recommended by the WHO recommendation engine with a HIGH confident level. For active suppression, the 6-month trajectory forecast (Fig. 11) predicted that the CD4 count would be 957 cells/mm³ and the viral load would be 10^1.7 copies/mL, while the visit timeline (Fig. 12) showed the time series of the visits to be monitored longitudinally for treatment.

The system not only provides resistance predictions, it also provides an estimation of HIV disease stage and the likely HIV subtype for each patient. The disease stage matched closely with the immune status derived from CD4 counts, and the estimation of subtype was based on known patterns of mutations, yielding additional clinical information as well as prediction.

The system also provides the predicted resistance pattern and predicts the HIV disease stage and estimated HIV subtype for each patient. Correlation between disease staging and immune status was found to be high, and subtype prediction showed consistency with mutation patterns, thus providing more clinical information in addition to predictive information.

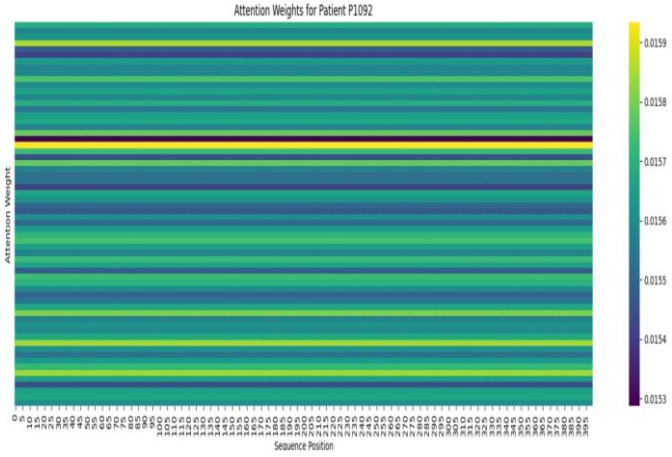


Fig 3. Heatmap of Attention Weights

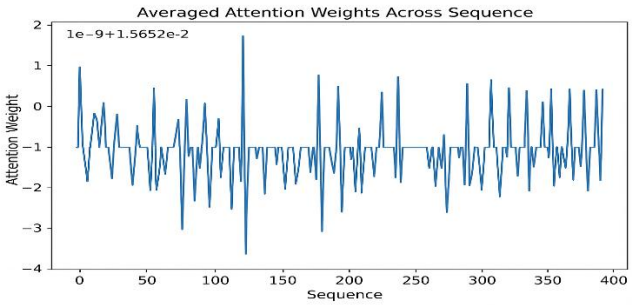


Fig 4. Average of Attention Weights across each position

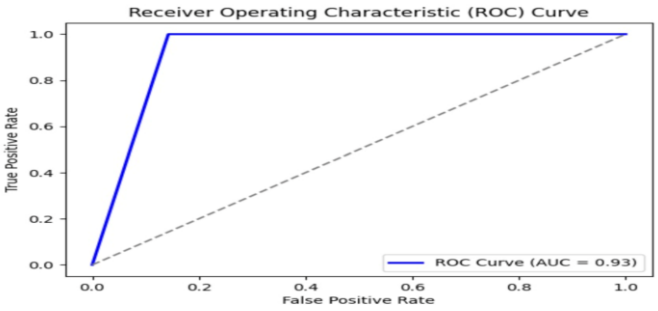
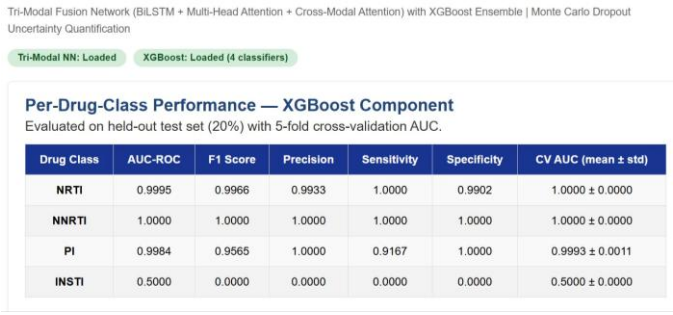


Fig 5. ROC Curve



Drug Class	AUC-ROC	F1 Score	Precision	Sensitivity	Specificity	CV AUC (mean ± std)
NRTI	0.9995	0.9966	0.9933	1.0000	0.9902	1.0000 ± 0.0000
NNRTI	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000 ± 0.0000
PI	0.9984	0.9565	1.0000	0.9167	1.0000	0.9993 ± 0.0011
INSTI	0.5000	0.0000	0.0000	0.0000	0.0000	0.5000 ± 0.0000

Fig.6. XGBoost Component: Model Evaluation Dashboard

Per-Drug-Class Performance — Neural Network Component

Drug Class	AUC-ROC	F1 Score	Precision	Sensitivity	Specificity
NRTI	0.2221	0.0000	0.0000	0.0000	1.0000
NNRTI	1.0000	1.0000	1.0000	1.0000	1.0000
PI	0.4616	0.2391	0.1486	0.6111	0.4112
INSTI	0.5000	0.0000	0.0000	0.0000	0.8680

Fig. 7. Neural Network Component Performance

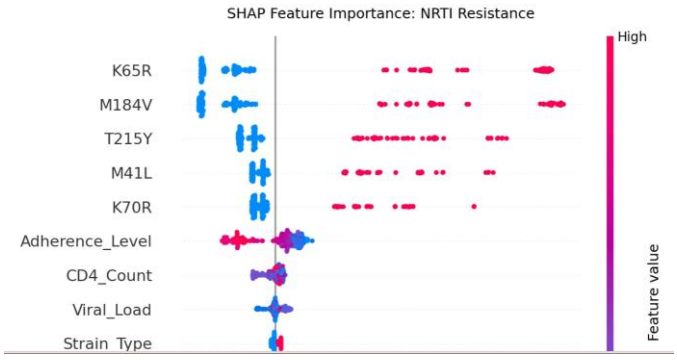


Fig. 8. Global SHAP Beeswarm Plot: NRTI Resistance

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **05** May-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

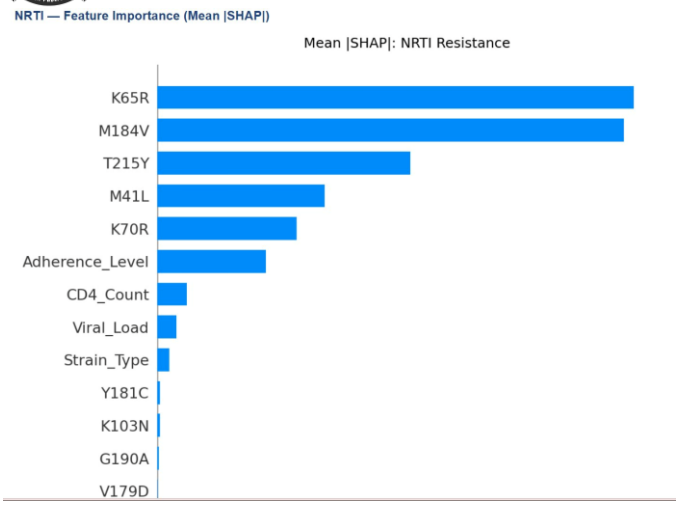


Fig. 9. Mean Absolute SHAP Values : NRTI Resistance Feature Ranking

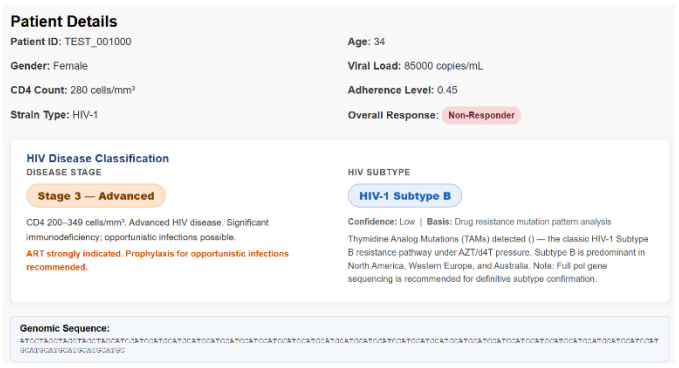


Fig. 10. Multi-Drug Resistance Prediction and WHO Recommendation

Fig. 7. Neural Network Component Performance

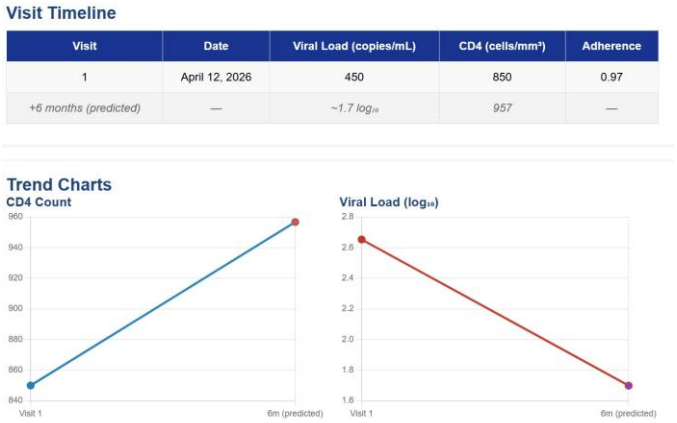


Fig. 12. Visit Timeline and Longitudinal Trend Charts

V. FINAL REMARKS

ViralGuide presents a clinically grounded ensemble framework for HIV drug resistance prediction that moves beyond binary treatment response classification toward drug-class-level resistance estimation across NRTI, NNRTI, Protease Inhibitor and INSTI categories, combining a Bidirectional LSTM processing genomic sequences with gradient-boosted classifiers operating on structured Stanford HIVDB mutation flags within a weighted ensemble that feeds directly into WHO 2021 guideline-aligned treatment recommendations. However, experimental results showed outstanding predictive accuracy for NRTIs, NNRTIs, and PIs, while the classifier for INSTIs performed poorly since there were not enough instances of resistance in the dataset, which highlights the importance of having a balanced dataset when constructing resistance classifiers. SHAP and counterfactual explanations guarantee the interpretability of predictions at the individual patient level. In future work, we would consider scaling the dataset with multi-center, real-world clinical data, including multi-ethnic genomic data sets with transfer learning specific to each region, as well as other modalities such as pharmacogenomic and socioeconomic variables to bring ViralGuide closer to a clinical decision support tool. The incorporation of HIV staging and estimated HIV subtype classification makes our system more clinically relevant by adding context beyond just resistance predictions. Even though subtype estimation was heuristically done because of the nature of the available data, it shows the possibilities of genomic knowledge in the domain of decision support systems.

REFERENCES

[1] G. A. McComsey, M. Lingohr-Smith, R. Rogers, J. Lin, and P. Donga, "Adherence patterns to antiretroviral therapy among HIV-1 infected individuals across United States clinical settings," Advances in Therapy, vol. 38, no. 9, pp. 4961–4974, 2021.

[2] D. N. Mamo, T. M. Yilma, M. Fekadie, Y. Sebastian, T. Bizuayehu, M. S. Melaku, and A. D. Walle, "Applying machine learning classifiers to predict virological failure in HIV patients receiving antiretroviral treatment at a comprehensive hospital in Amhara Region, Ethiopia," BMC Medical Informatics and Decision Making, vol. 23, no. 1, p. 75, 2023.

[3] B. Bayu, A. Tariku, A. B. Bulti, Y. A. Habitu, T. Derso, and D. F. Teshome, "Clinical and behavioral determinants of virological failure in patients on highly active antiretroviral therapy: a case-control analysis," HIV/AIDS — Research and Palliative Care, pp. 153–159, 2017.

[4] G. Di Teodoro, M. Pirkl, F. Incardona, I. Vicenti, A. Sonnerborg, R. Kaiser, L. Palagi, M. Zazzi, and T. Lengauer, "Leveraging temporal mutation dynamics to improve antiretroviral therapy outcome prediction in HIV-1 infection," Bioinformatics, vol. 40, no. 6, 2024.

[5] B. T. Seboka, D. E. Yehualashet, and G. A. Tesfa, "Machine learning approaches for predicting viral load and CD4 cell count in people living with HIV on antiretroviral treatment in Gedeo Zone public hospitals," International Journal of General Medicine, pp. 435–451, 2023.

[6] M. Maskew, K. Sharpey-Schafer, L. De Voux, T. Crompton, J. Bor, M. Rennick, A. Chirowodza et al., "Predictive modeling for retention and viral load suppression in South African HIV treatment programs," Scientific Reports, vol. 12, no. 1, p. 12715, 2022.

[7] M. E. Ekpenyong, P. I. Etebong, and T. C. Jackson, "A fuzzy-multidimensional deep learning framework for predicting antiretroviral therapy response," Heliyon, vol. 5, no. 7, 2019.

[8] K. R. Bisaso, S. A. Karungi, A. Kiragga, J. K. Mukonzo, and B. Castelnuovo, "Comparative analysis of logistic regression-based machine learning models for predicting early virological suppression in antiretroviral-initiating HIV patients," BMC Medical Informatics and Decision Making, vol. 18, pp. 1–10, 2018.

[9] D. Bukenya, B. N. Mayanja, S. Nakamanya, R. Muhumuza, and J. Seeley, "Barriers to long-term antiretroviral therapy adherence among individuals with poor viral suppression in Uganda," AIDS Research and Therapy, vol. 16, pp. 1–9, 2019.

[10] Stanford University HIV Drug Resistance Database.

[11] HIV Sequence Databases at Los Alamos National Laboratory.

[12] S. R. Choi and M. Lee, "A comprehensive review of transformer architecture and attention mechanisms in genomic data analysis," Biology, vol. 12, no. 7, p. 1033, Jul. 2023, doi: 10.3390/biology12071033.

[13] X. Li, M. Li, P. Yan, G. Li, Y. Jiang, H. Luo, and S. Yin, "Attention mechanisms in deep learning for medical image analysis: foundations and emerging directions," International Journal of Network Dynamics and Intelligence, vol. 2, no. 1, pp. 93–116, 2023, doi: 10.53941/ijndi0201006

.

[14] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, and G. Menegaz, "Comparative perspectives on explainable artificial intelligence: SHAP and LIME methodologies," Advanced Intelligent Systems, vol. 7, no. 1, p. 2400304, 2025.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[16] S. M. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017.