

Article

# Drug Resistance Prediction Using Deep Learning Techniques on HIV-1 Sequence Data

Margaret C. Steiner <sup>1,\*</sup>, Keylie M. Gibson <sup>1</sup> and Keith A. Crandall <sup>1,2</sup>

<sup>1</sup> Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052, USA; kmgibson@gwu.edu (K.M.G.); kcrandall@gwu.edu (K.A.C.)

<sup>2</sup> Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052, USA

\* Correspondence: steinerm@gwu.edu; Tel.: +571-553-0146

Received: 22 April 2020; Accepted: 17 May 2020; Published: 19 May 2020



**Abstract:** The fast replication rate and lack of repair mechanisms of human immunodeficiency virus (HIV) contribute to its high mutation frequency, with some mutations resulting in the evolution of resistance to antiretroviral therapies (ART). As such, studying HIV drug resistance allows for real-time evaluation of evolutionary mechanisms. Characterizing the biological process of drug resistance is also critically important for sustained effectiveness of ART. Investigating the link between “black box” deep learning methods applied to this problem and evolutionary principles governing drug resistance has been overlooked to date. Here, we utilized publicly available HIV-1 sequence data and drug resistance assay results for 18 ART drugs to evaluate the performance of three architectures (multilayer perceptron, bidirectional recurrent neural network, and convolutional neural network) for drug resistance prediction, jointly with biological analysis. We identified convolutional neural networks as the best performing architecture and displayed a correspondence between the importance of biologically relevant features in the classifier and overall performance. Our results suggest that the high classification performance of deep learning models is indeed dependent on drug resistance mutations (DRMs). These models heavily weighted several features that are not known DRM locations, indicating the utility of model interpretability to address causal relationships in viral genotype-phenotype data.

**Keywords:** HIV; antiretroviral therapy; HIV drug resistance; machine learning; deep learning; neural networks

## 1. Introduction

Human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS), affects over 1.1 million people in the U.S. today [1]. While there is still no cure, HIV can be treated effectively with antiretroviral therapy (ART). Consistent treatment with ART can extend the life expectancy of an HIV-positive individual to nearly as long as that of a person without HIV by reducing the viral load to below detectable levels [2] and can reduce transmission rates [3,4]. However, the fast replication rate and lack of repair mechanisms of HIV leads to a large number of mutations, many of which result in the evolution of HIV to resist antiretroviral drugs [5,6]. Drug resistance may be conferred at the time of HIV transmission, so even treatment-naïve patients may be resistant to certain ART drugs, which can lead to rapid drug failure [7]. As a result, the analysis of drug resistance is critical to treating HIV, and thus is an important focus of HIV research.

HIV drug resistance may be directly evaluated using phenotypic assays such as the PhenoSense assay [8]. Both the patient’s isolated HIV strain and a wild type reference strain are exposed to an

antiretroviral drug at several concentrations, and the difference in effect between the two indicates the level of drug resistance. Resistance is measured in terms of fold change, which is defined as a ratio between the concentration of the drug necessary to inhibit replication of the patient virus and that of the wild type [9]. These tests are laborious, time-intensive, and costly. Additionally, phenotypic assays are reliant on prior knowledge of correlations between mutations and resistance to specific drugs, which evolve quickly and thus cannot be totally accounted for [10]. An alternative, the “virtual” or genotypic test, predicts the outcome of a phenotypic test based on the genotype using statistical methods. Several web tools exist for HIV drug resistance prediction based on known genotype profiles, including HIVdb [11] and WebPSSM [12], utilizing rules-based classification and position-specific scoring matrices, respectively, to predict drug resistance. Two additional tools utilize machine learning approaches: geno2pheno [13] and SHIVA [14].

Several machine learning architectures have been applied to predict drug resistance, including random forests such as those used in SHIVA [14–16], support vector machines as in geno2pheno [13], decision trees [17], logistic regression [16], and artificial neural networks [10,18–20]. Deep learning models (i.e., neural networks) are a major focus in current machine learning research and have been successfully applied to several classes of computational biology data [21]. Yet, one aspect of deep learning models in particular that has been overlooked to date is model interpretability. Deep learning is often criticized for its “black box” nature, as it is unclear from the model itself why a given classification was made. The resulting ambiguity in the classification model proves to be a major limitation to the utility of deep learning in translational and clinical applications [22]. As a response to this concern, recently-developed model interpretability methods seek to map model outputs back to a specific subset of the most influential inputs, or features [23]. In practice, this method allows researchers to better understand whether predictions are based on relevant patterns in the training data as opposed to bias, thus attesting to the model’s trustworthiness and, in turn, providing the potential for deep learning models to identify novel patterns in the input data.

Here, we integrate deep learning techniques with HIV genotypic and phenotypic data and analyses in order to investigate the implications of the underlying evolutionary processes of HIV-1 drug resistance for classification performance and vice versa. The objectives of this study are: (i) to compare the performance of three deep learning architectures which may be used for virtual HIV-1 drug resistance tests—multilayer perceptron (MLP), bidirectional recurrent neural network (BRNN), and convolutional neural network (CNN), (ii) to evaluate feature importance in the context of drug resistance mutations, and (iii) to explore the relationship between the molecular evolution of drug resistant strains and model performance.

## 2. Methods

Briefly, our deep learning approach included training and cross-validation of three architectures for binary classification of labeled HIV-1 sequence data: MLP, BRNN, and CNN (Figure 1). In addition to comparing performance metrics across architectures, we evaluated feature importance using a permutation-based method and interpreted these results using known DRM loci. Lastly, we reconstructed and annotated phylogenetic trees from the same data in order to assess clustering patterns of resistant sequences.

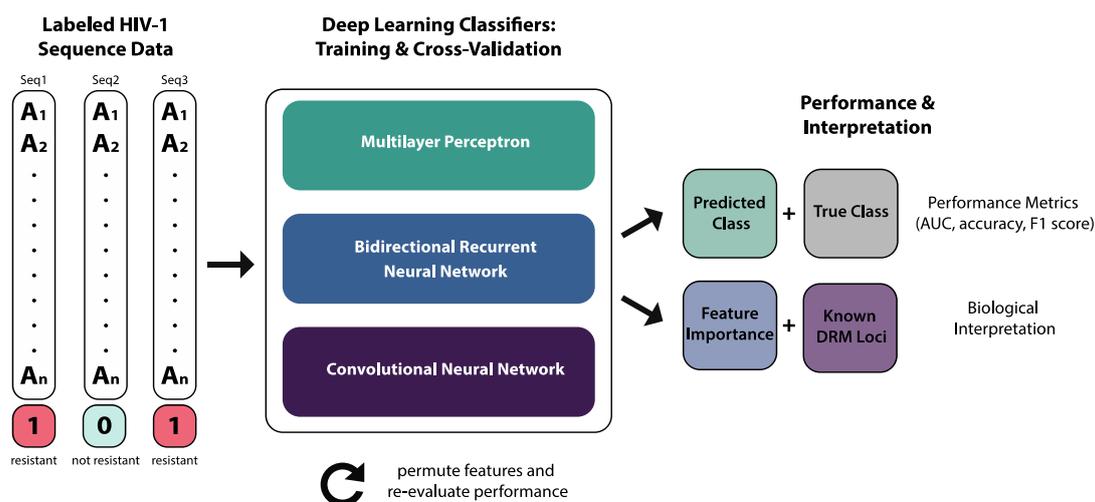


Figure 1. Overview of deep learning methods.

### 2.1. Data

Genotype-phenotype data were obtained from Stanford University’s HIV Drug Resistance database, one of the largest publicly available datasets for such data [9]. The filtered genotype-phenotype datasets for eighteen protease inhibitor (PI), nucleotide reverse transcriptase inhibitor (NRTI), and non-nucleotide reverse transcriptase inhibitor (NNRTI) drugs were used (downloaded August 2018; Table 1). While the Stanford database additionally includes data pertaining to integrase inhibitors (INIs), at the time of this study insufficient data were available for deep learning analysis, and so here we have focused on polymerase. In the filtered datasets, redundant sequences from intra-patient data and sequences with mixtures at major drug resistance mutation (DRM) positions were excluded from further analyses. In total, the data pulled from the Stanford database contained 2112 sequences associated with PI susceptibility, 1772 sequences associated with NNRTI susceptibility, and 2129 sequences associated with NRTI susceptibility (Table 1). Drug susceptibility testing results included in the Stanford dataset had been generated using a PhenoSense assay [8]. Susceptibility is expressed as fold change in comparison to wild-type HIV-1; a fold change value greater than 3.5 indicates that a sample is resistant to a given drug [10,17].

All sequences were from regions of the polymerase (*pol*) gene, which is standardly sequenced in studies of HIV-1 drug resistance. Sequences in the PI dataset include a 99 amino acid sequence from the protease (*PR*) region (HXB2 coordinates: 2253-2549), while the NRTI and NNRTI datasets include a 240 amino acid sequence from reverse transcriptase (*RT*) region (HXB2 coordinates: 2550-3269). Subtype is not explicitly listed in the filtered dataset, though the consensus sequence given is that of Subtype B. The data were separated into 18 drug-specific datasets, each of which contained all sequences for which a resistance assay value for the drug was available, as well as their respective drug resistance status. For training and evaluating deep learning models, amino acid and ambiguity codes were encoded via integer encoding.

**Table 1.** Description of datasets.

Drug	Drug Class	No. Sequences	Sequence Length	(%) Resistant
FPV	PI	1444	99	36.5
ATV	PI	987	99	43.4
IDV	PI	1491	99	43.7
LPV	PI	1267	99	44.7
NFV	PI	1532	99	52.8
SQV	PI	1483	99	37.6
TPV	PI	696	99	17.8
DRV	PI	605	99	21.5
All PI		2112	99	41
EFV	NNRTI	1471	240	41.5
NVP	NNRTI	1478	240	47.8
ETR	NNRTI	481	240	28.7
RPV	NNRTI	181	240	32
All NNRTI		1772	240	42
3TC	NRTI	1270	240	58.3
ABC	NRTI	1293	240	46.1
AZT	NRTI	1283	240	41.8
D4T	NRTI	1291	240	13
DDI	NRTI	1291	240	5.3
TDF	NRTI	1025	240	8.8
All NRTI		2129	240	40.4

Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

## 2.2. Deep Learning Classifiers

Three classes of deep learning classifiers were constructed, trained, and evaluated in each of the 18 datasets: multilayer perceptron (MLP), bidirectional recurrent neural network (BRNN), and convolutional neural network (CNN). The structure and parameters of each model are described below (Sections 2.2.1–2.2.3). All data pre-processing, model training, and model evaluation steps were completed using R v3.6.0 [24] and RStudio v1.2.1335 [25]. All classifiers were trained and evaluated using the Keras R package v2.2.4.1 [26] as a front-end to TensorFlow v1.12.0 [27], utilizing Python v3.6.8 [28].

All classifiers were evaluated using 5-fold stratified cross-validation with adjusted class weights in order to account for small dataset size and class imbalances (Table 1), implemented as follows. All data were randomly shuffled and then split evenly into five partitions such that each partition had the same proportion of resistant to non-resistant sequences. Then, a model was initiated and trained using four of the five partitions as training data and the fifth as a hold-out validation set. Next, a new, independent model was initiated and trained using a different set of four partitions for training and the fifth as the hold-out validation set. This was repeated three additional times such that each partition was used as a validation set exactly once. Thus, for each architecture and dataset, a total of five independent neural networks models were trained and evaluated using a distinct partition of the total dataset for validation and the remainder for training. This method is advantageous when working with limited available data, as it allows all data to be utilized towards evaluating the performance of a given architecture without the same data ever being used for both training and validation of any one model. Taking the average of performance metrics produced from each fold also lessens the bias that the choice of validation data may inflict on these results.

The class weights used to train each model were calculated such that a misclassification of a non-resistant sequence had a penalty of 1 and the misclassification of a resistant sequence had a penalty of the ratio of non-resistant to resistant sequences (i.e., a misclassification of a resistant sequence had a penalty higher than that of a non-resistant sequence by a factor of the observed class imbalance). Each model was compiled using the Root Mean Square Propagation (RMSprop) optimizer function and binary cross-entropy as the loss function, as is standardly used for binary classification tasks, and was trained for 500 epochs with a batch size of 64.

### 2.2.1. Multilayer Perceptron

A multilayer perceptron (MLP) is a feed-forward neural network consisting of input, hidden, and output layers that are densely connected. MLPs are the baseline, classical form of a neural network. The MLP used here included embedding and 1D global average pooling layers (input), followed by four feed-forward hidden layers each with either 33 (PI) or 99 (NNRTI or NRTI) units, Rectified Linear Unit (ReLU) activation, L2 regularization, and ending with an output layer with a sigmoid activation function.

### 2.2.2. Bidirectional Recurrent Neural Network

A bidirectional recurrent neural network (BRNN) includes a pair of hidden recurrent layers, each of which feeds information in opposite directions, thus utilizing the forward and backward directional context of the input data. BRNNs are commonly used in applications where such directionality is imperative, such as language recognition and time series data tasks. The BRNN used here included one embedding layer, one bidirectional long short-term memory (LSTM) layer with either 33 (PI) or 99 (NNRTI or NRTI) units, dropout of 0.2, and recurrent dropout of 0.2, and ended with an output layer with a sigmoid activation function.

### 2.2.3. Convolutional Neural Network

A convolutional neural network (CNN) operates in a manner inspired by the human visual cortex and consists of convolutional (feature extraction) and pooling (dimension reduction) layers. CNNs are well known for their use in computer vision and image analysis but have recently been applied to genetic sequence data with much success, especially for training on DNA sequence data directly [21]. The CNN used here included one embedding layer, two 1D convolution layers with 32 filters, kernel size of 9, a ReLU activation function, and one 1D max pooling layer in between, then ending in an output layer using a sigmoid activation function.

## 2.3. Performance Metrics

The following metrics were recorded for each evaluation step: true positive rate/sensitivity (TPR), true negative rate/specificity (TNR), false positive rate (FPR), false negative rate (FNR), accuracy, F-measure (F1), binary cross entropy (loss), and area-under-the-receiver operating characteristic curve (AUC). Formulas for the calculation of all metrics are given below. Binary cross-entropy was reported in Keras output and AUC and receiving operator characteristic (ROC) curves were generated using the ggplot2 v3.1.1 [29] and pROC v1.15.0 [30] R packages.

Accuracy is a measure of the overall correctness of the model's classification:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Sensitivity, or true positive rate (TPR), measures how often the model predicts that a sequence is resistant to a drug when it is actually resistant:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

Specificity, or true negative rate (TNR), measures how often the model predicts that the sequence is not resistant to a drug when it is actually not resistant:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

The false positive rate (FPR) measures how often the model predicts that a sequence is resistant to a drug when it is actually not resistant:

$$\text{FPR} = \frac{FP}{FP + TN} \quad (4)$$

The false negative rate (FNR) measures how often the model predicts that a sequence is not resistant to a drug when it is actually resistant:

$$\text{FNR} = \frac{FN}{FN + TP} \quad (5)$$

The F1 score measures the harmonic mean of precision of recall and is often preferred to accuracy when the data has imbalanced classes:

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

Binary cross-entropy is a measure of the error of the model for the given classification problem, which is minimized by the neural network during the training phase. Thus, this metric represents performance during training but not necessarily performance on the testing dataset, which is used for evaluation. Binary cross-entropy is calculated as follows, where  $y$  is a binary indicator of whether a class label for an observation is correct and  $p$  is the predicted probability that the observation is of that class:

$$\text{cross-entropy} = -(y \log(p) + (1 - y) \log(1 - p)) \quad (7)$$

AUC measures the area under the receiving operator characteristic (ROC) curve, which plots true positive rate against false positive rate. AUC is also commonly used in situations where the data has imbalanced classes, as the ROC measures performance over many different scenarios.

#### 2.4. Model Interpretation

Model interpretation analysis was conducted in R/RStudio using the permutation feature importance function implemented in the IML package v0.9.0 [23]. This function is an implementation of the model reliance measure [31], which is model-agnostic. Put simply, permutation feature importance is a metric of change in model performance when all data for a given feature is shuffled (permuted) and is measured in terms of 1-AUC. Feature importance plots were rendered using the ggplot2 package and annotated with known DRM positions using the Stanford database [9], both for the top 20 most important features and across the entire gene region.

#### 2.5. Phylogenetics

In addition to deep learning-based analysis, we reconstructed phylogenetic trees for all datasets in order to empirically test whether resistant and non-resistant sequences formed distinct clades and to visualize evolutionary relationships present in the data. ModelTest-NG v0.1.5 [32] was used to estimate best-fit amino acid substitution models for each dataset for use in phylogeny reconstruction. The selected models included HIVB (FPV, ATV, TPV, and all PI), FLU (IDV, LPV, SQV, and DRV)—which has been shown to be highly correlated with HIVb [33], JTT (NFV, ETR, RPV, 3TC, D4T, DDI, TDF, and all NRTI), and JTT-DCMUT (EFV, NVP, ABC, AZT, and all NNRTI). We then used RAxML v8.2.12 [34] to estimate phylogenies for each data set using the maximum likelihood optimality criterion and

included bootstrap analysis with 100 replicates to evaluate branch support. Both ModelTest-NG and RAxML were run within the CIPRES Web Interface v3.3 [35]. Trees were then annotated with drug resistance classes using iTOL v4 [36]. The approximately unbiased (AU) test for constrained trees [37] was used to test the hypothesis that all trees were perfectly clustered by drug resistance class using IQ-Tree v1.6.10 [38], with midpoint rooting used for all trees.

### 3. Results

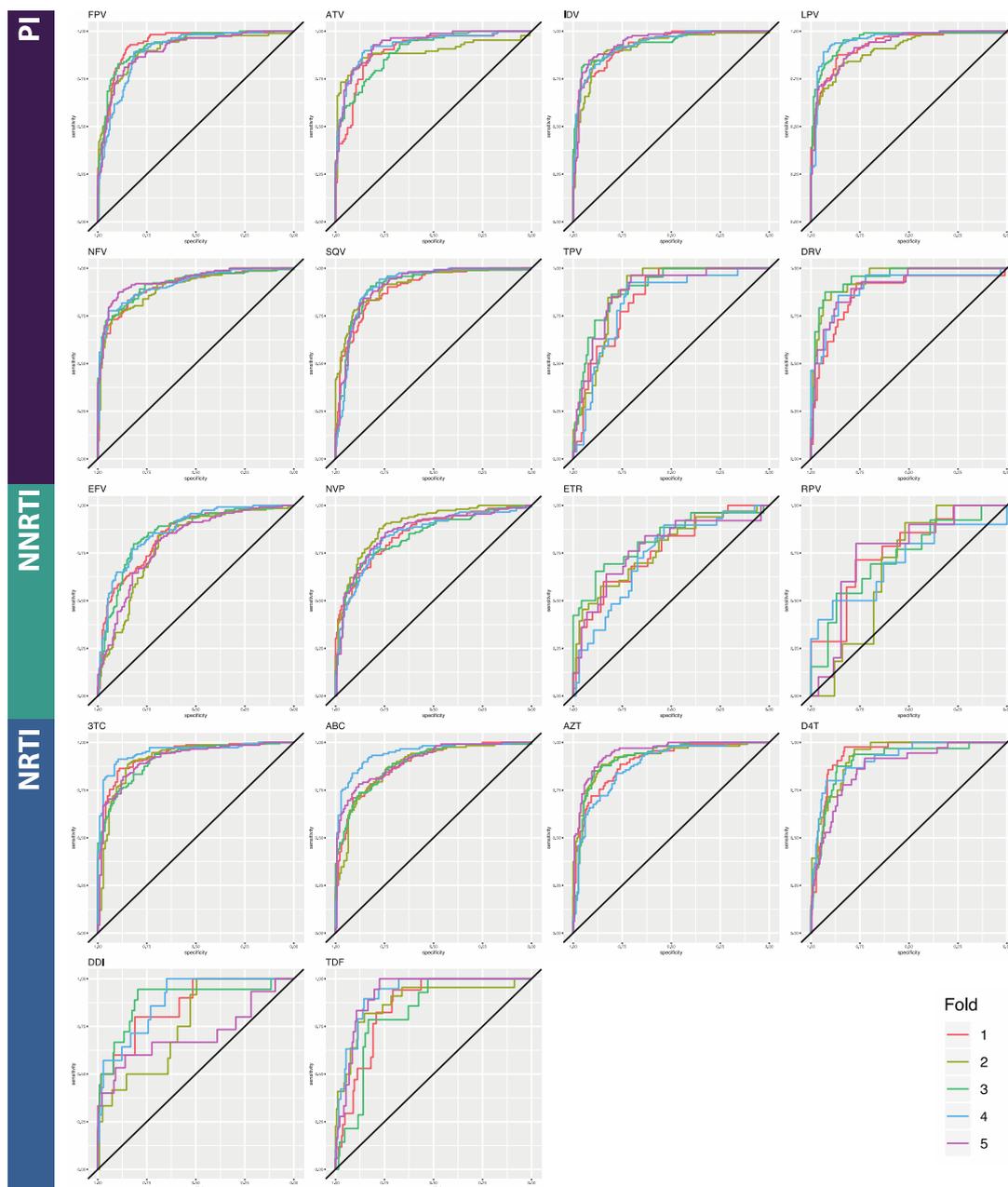
#### 3.1. Classifier Performance

Here, we compared the performance of three deep learning architectures for binary classification of HIV sequences by drug resistance: multilayer perceptron (MLP), bidirectional recurrent neural network (BRNN), and convolutional neural network (CNN) (Table 2). The reported metrics are averages taken from 5-fold cross-validation. Average accuracy across folds ranged from 65.9% to 94.6% for the MLPs, from 72.9% to 94.6% for the BRNNs, and 86.2% to 95.9% for the CNNs (Tables A1–A3). Due to the noted class imbalances in the data, accuracy is not an ideal metric to compare performance, so we additionally considered AUC and the F1 score, both of which are more appropriate in this case. Average AUC across folds ranged from 0.760 to 0.935 for the MLPs, from 0.682 to 0.988 for the BRNNs, and 0.842 to 0.987 for the CNN models (Table A4; Figures 2–4). Average F1 score across folds ranged from 0.224 to 0.861 for the MLPs, from 0.362 to 0.944 for the BRNNs, and 0.559 to 0.950 for the CNNs (Table A4). Across all models and all three overall performance metrics (accuracy, AUC, and F1), average performance was best for PI datasets, followed by NRTI and then NNRTI (Table 2). All three overall performance metrics also indicate that the CNN model showed the best performance of the three. False negative rates were similar among BRNNs and CNNs, both of which were notably lower than that of MLPs. Average false positive rate was notably lower for the BRNN and CNN models than the MLP model, while false negative rate remained within a more consistent range.

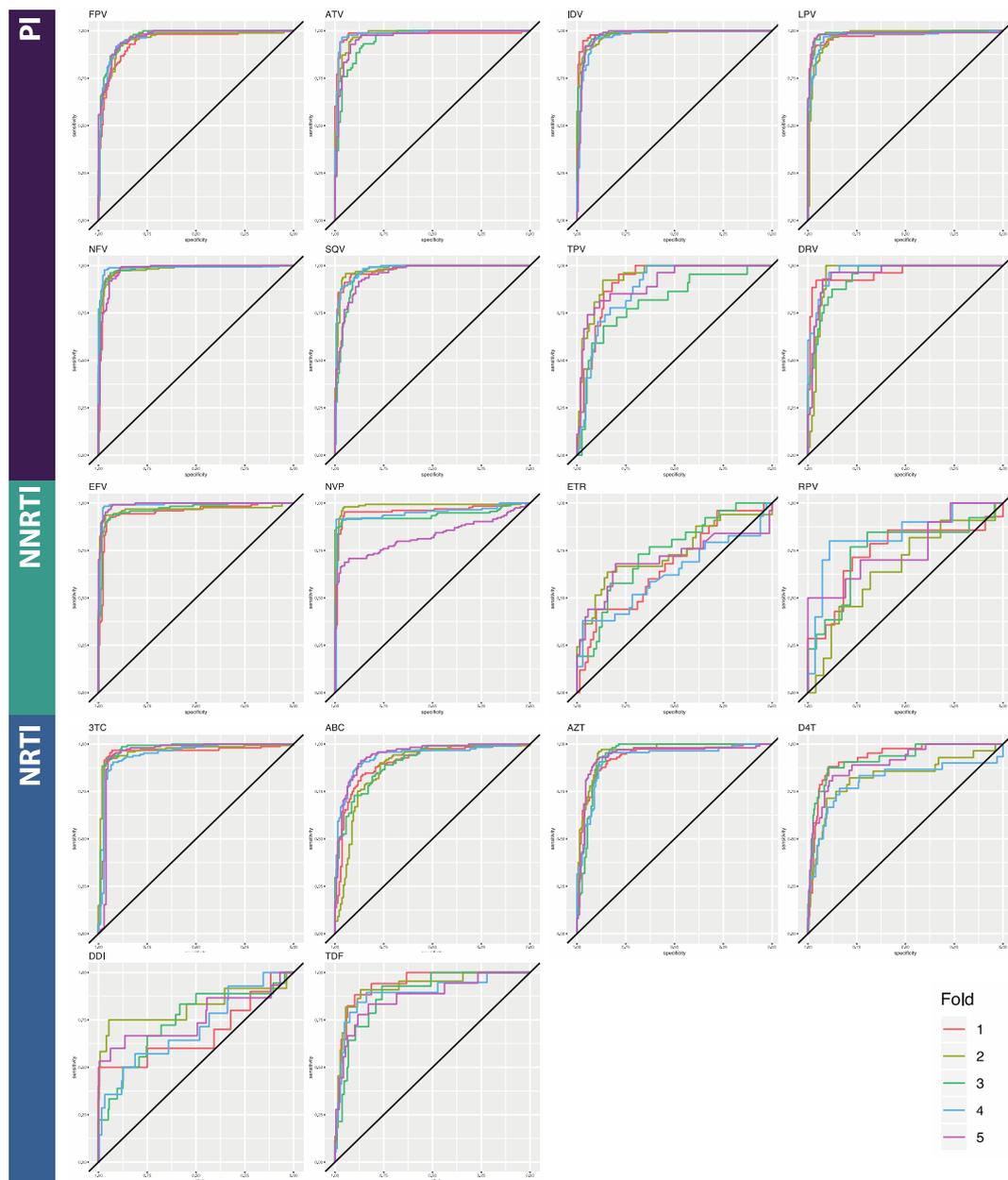
**Table 2.** Average performance metrics by drug class; standard deviation in brackets.

MLP								
Class	Accuracy	Loss	TPR	TNR	FPR	FNR	AUC	F1
PI	0.826 [0.039]	0.406 [0.058]	0.813 [0.041]	0.834 [0.061]	0.166 [0.061]	0.187 [0.041]	0.9 [0.009]	0.732 [0.135]
NNRTI	0.706 [0.095]	0.586 [0.143]	0.662 [0.140]	0.718 [0.105]	0.282 [0.105]	0.338 [0.140]	0.805 [0.006]	0.596 [0.158]
NRTI	0.795 [0.061]	0.470 [0.150]	0.725 [0.112]	0.807 [0.081]	0.193 [0.081]	0.275 [0.112]	0.895 [0.024]	0.593 [0.236]
BRNN								
Class	Accuracy	Loss	TPR	TNR	FPR	FNR	AUC	F1
PI	0.912 [0.032]	0.290 [0.087]	0.890 [0.065]	0.919 [0.028]	0.081 [0.028]	0.110 [0.065]	0.965 [0.014]	0.860 [0.106]
NNRTI	0.884 [0.104]	0.334 [0.247]	0.829 [0.165]	0.914 [0.088]	0.086 [0.088]	0.171 [0.165]	0.83 [0.021]	0.836 [0.180]
NRTI	0.896 [0.030]	0.357 [0.094]	0.751 [0.168]	0.910 [0.029]	0.090 [0.029]	0.249 [0.168]	0.89 [0.02]	0.698 [0.220]
CNN								
Class	Accuracy	Loss	TPR	TNR	FPR	FNR	AUC	F1
PI	0.919 [0.019]	0.745 [0.204]	0.865 [0.081]	0.939 [0.013]	0.061 [0.013]	0.135 [0.081]	0.968 [0.007]	0.871 [0.088]
NNRTI	0.911 [0.048]	0.614 [0.219]	0.845 [0.099]	0.945 [0.029]	0.055 [0.029]	0.155 [0.099]	0.945 [0.012]	0.861 [0.097]
NRTI	0.920 [0.035]	0.805 [0.277]	0.743 [0.182]	0.940 [0.040]	0.060 [0.040]	0.257 [0.182]	0.927 [0.016]	0.757 [0.163]

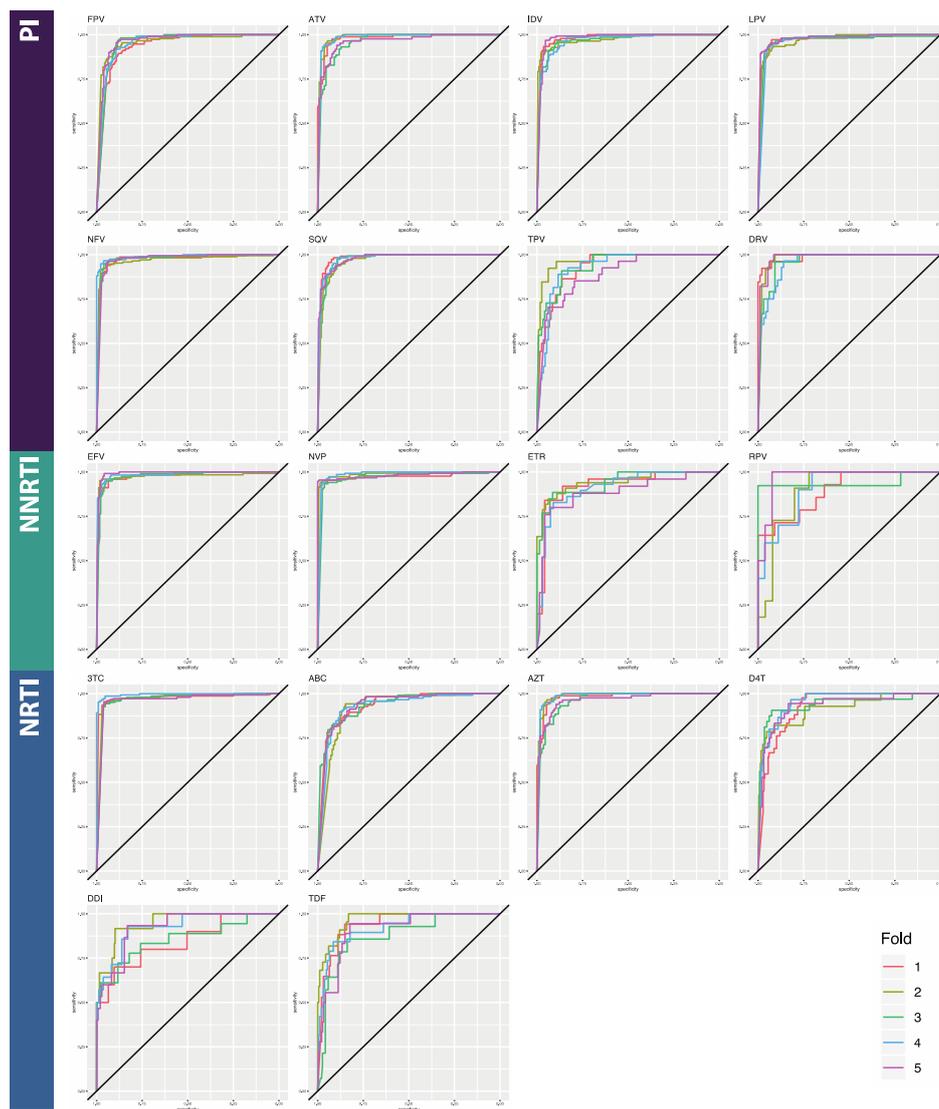
Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.



**Figure 2.** Receiving operator characteristic (ROC) curves for the Multilayer Perceptron classifier. Horizontal axes are specificity; vertical axes are sensitivity. The five curves represent the results of each cross-validation step. Curves which are closer to a ninety-degree angle in the top left corner represent better performance. Area under the curves is given in Table A4. Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor.



**Figure 3.** ROC curves for the Bidirectional Recurrent Neural Network classifier. Horizontal axes are specificity; vertical axes are sensitivity. The five curves represent the results of each cross-validation step. Curves which are closer to a ninety-degree angle in the top left corner represent better performance. Area under the curves is given in Table A4. Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

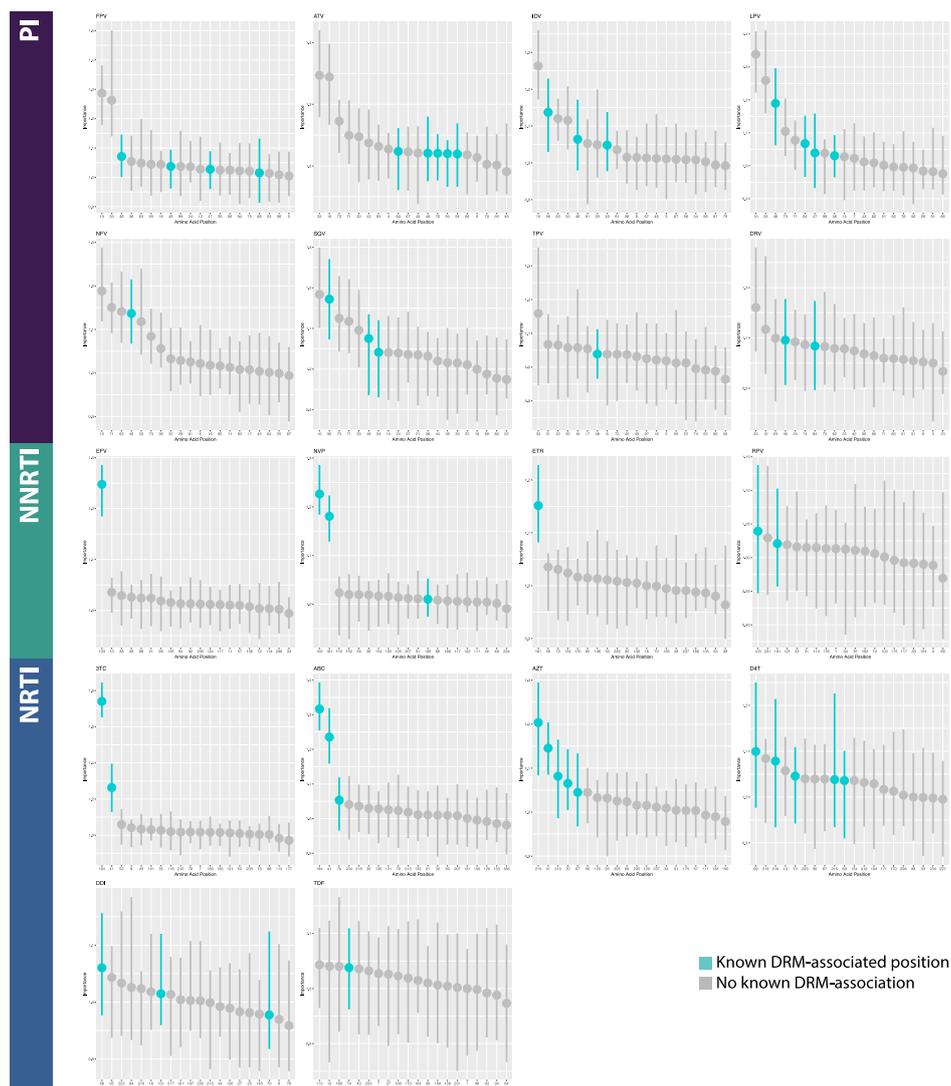


**Figure 4.** ROC curves for the Convolutional Neural Network classifier. Horizontal axes are specificity; vertical axes are sensitivity. The five curves represent the results of each cross-validation step. Curves which are closer to a ninety-degree angle in the top left corner represent better performance. Area under the curves is given in Table A4. Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

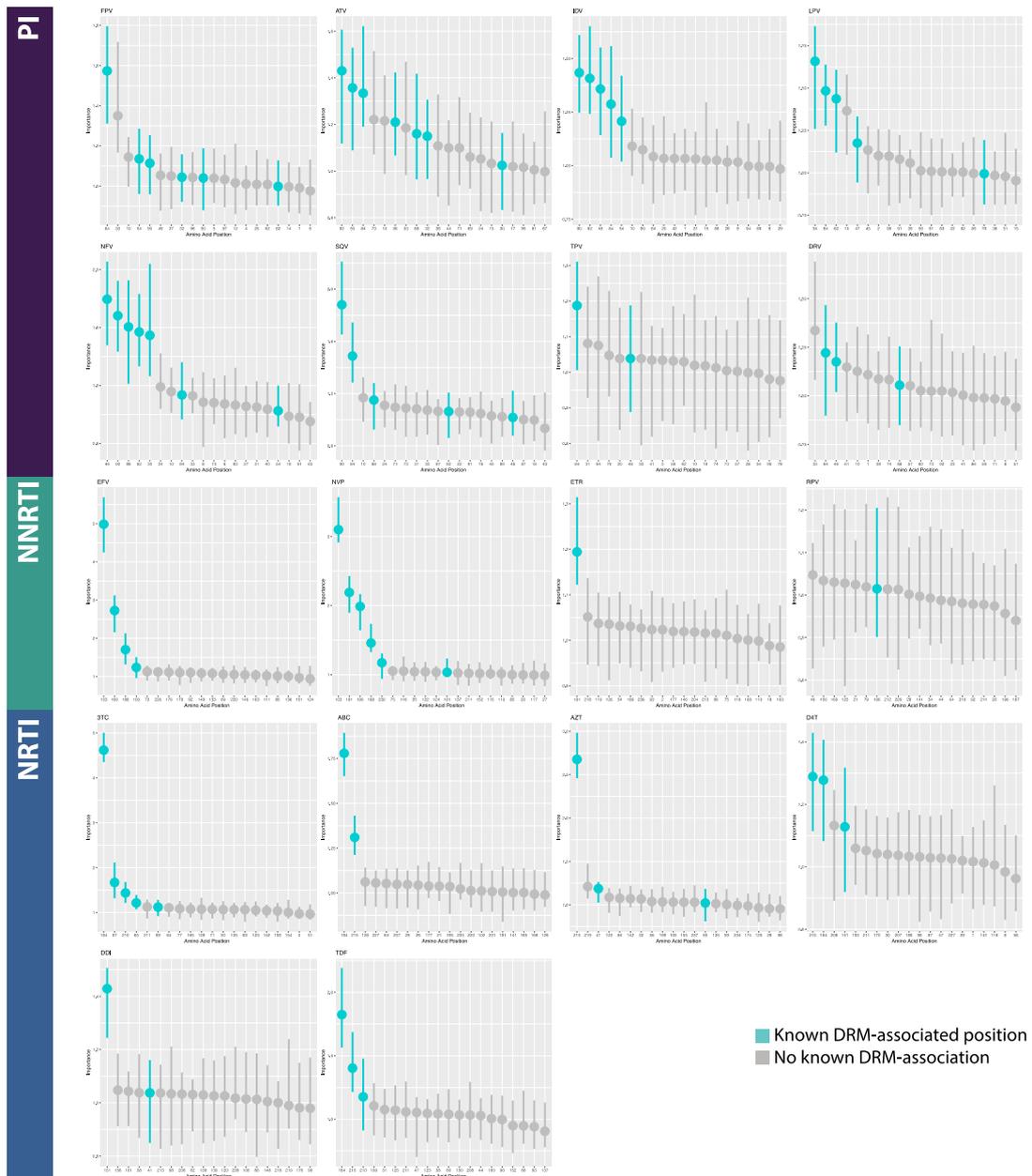
### 3.2. Model Interpretation

On average in the MLPs, of the 20 most important features returned by permutation feature importance analysis, 2.875 PI features (range 1–5 features), 3.167 NRTI features (range 1–5), and 1.75 NNRTI features (range 1–3) correlated to known DRM positions. For the BRNNs, an average of 5 PI features (range 2–7), 3 NRTI features (range 2–5), and 3 NNRTI features (range 1–6) correlated to known DRM positions. For the CNNs, an average of 5.75 PI features (range 3–7), 4.333 NRTI features (range 1–6), and 5.25 NNRTI features (range 5–6) correlated to known DRM positions. In total for MLPs, 50% (9/18) of datasets returned the most important feature being a known DRM position (PI: 0/8; NRTI: 5/6; NNRTI: 4/4), while BRNNs and CNNs fared better. A total of 88.9% (16/18) and 100% (18/18)

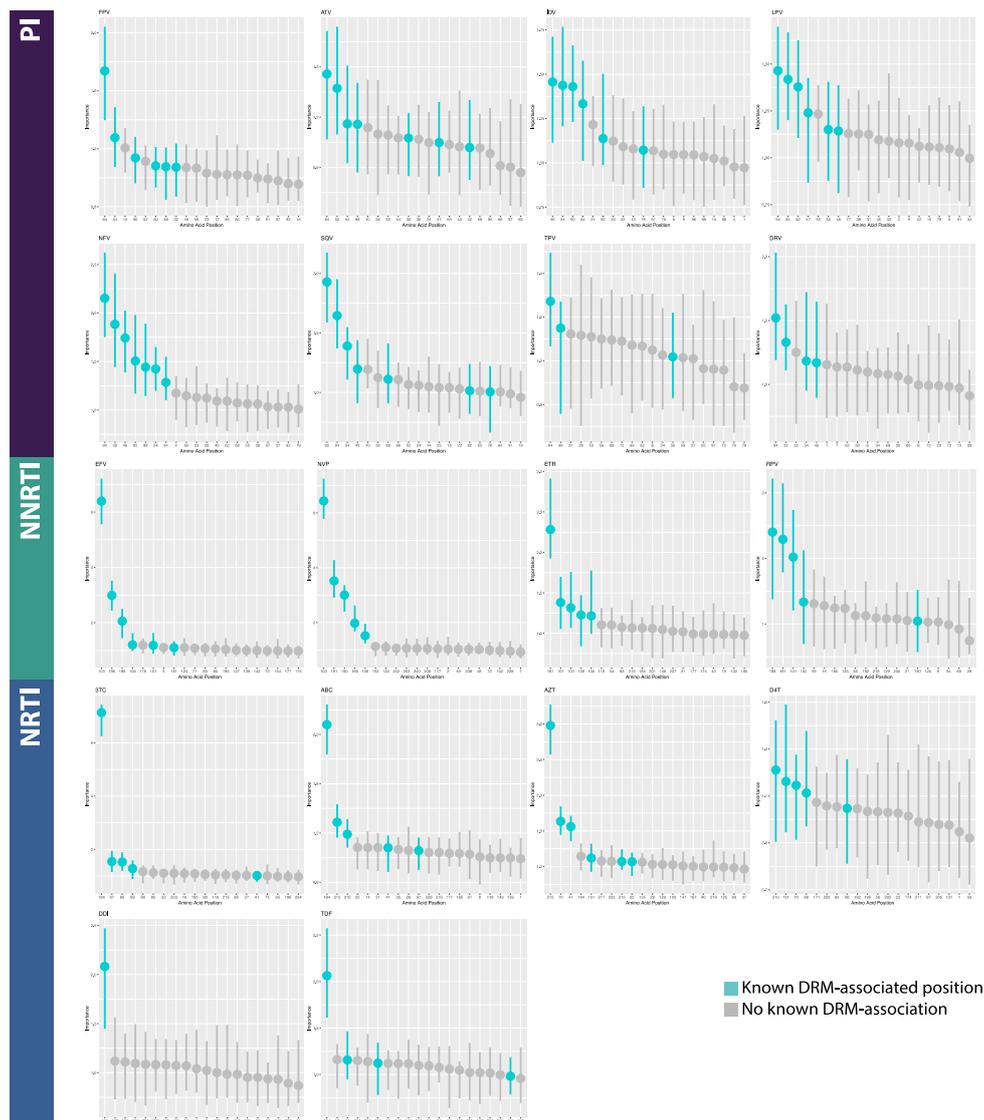
of BRNNs and CNNs, respectively, returned a known DRM position as the most important feature position (BRNN PI: 7/8; NRTI: 6/6; NNRTI: 3/4; CNN PI: 8/8; NRTI: 6/6; NNRTI: 4/4) (Figures 5–7, Supplementary Materials; Figures S1–S9). In the CNN classifiers for PI datasets, of features which were among the 20 most important features but are not known DRM-associated positions, four were identified in half (4/8) of the PI datasets. Similarly, for NNRTI datasets, five such features were identified in half (2/4) of the datasets and for NRTI datasets, three such features were identified in half (3/6) of the datasets. The majority of features with high importance but no known DRM association were identified in less than half of the datasets, across PI, NNRTI, and NRTI datasets. A majority of such features were only identified in one dataset for both NRTI and NNRTI, and in only one or two datasets in PI (Supplementary Materials; Figures S10–S12).



**Figure 5.** Annotated feature importance plots for the 20 most important features in Multilayer Perceptron classifiers. Horizontal axes are amino acid positions; vertical axes are feature importance (measured as change in 1-AUC). Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.



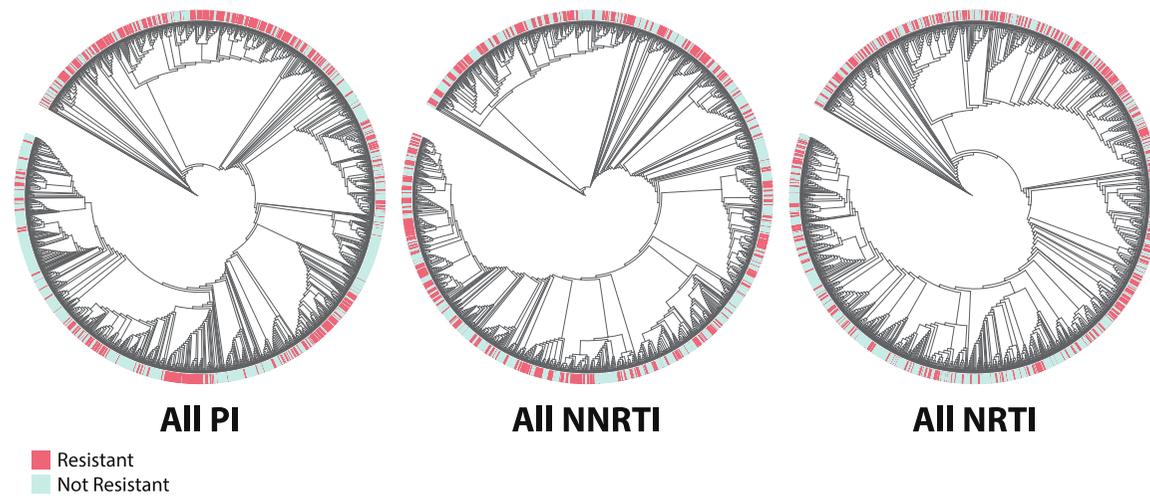
**Figure 6.** Annotated feature importance plots for the 20 most important features in Bidirectional Recurrent Neural Network classifiers. Horizontal axes are amino acid positions; vertical axes are feature importance (measured as change in 1-AUC). Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleoside reverse transcriptase inhibitor.



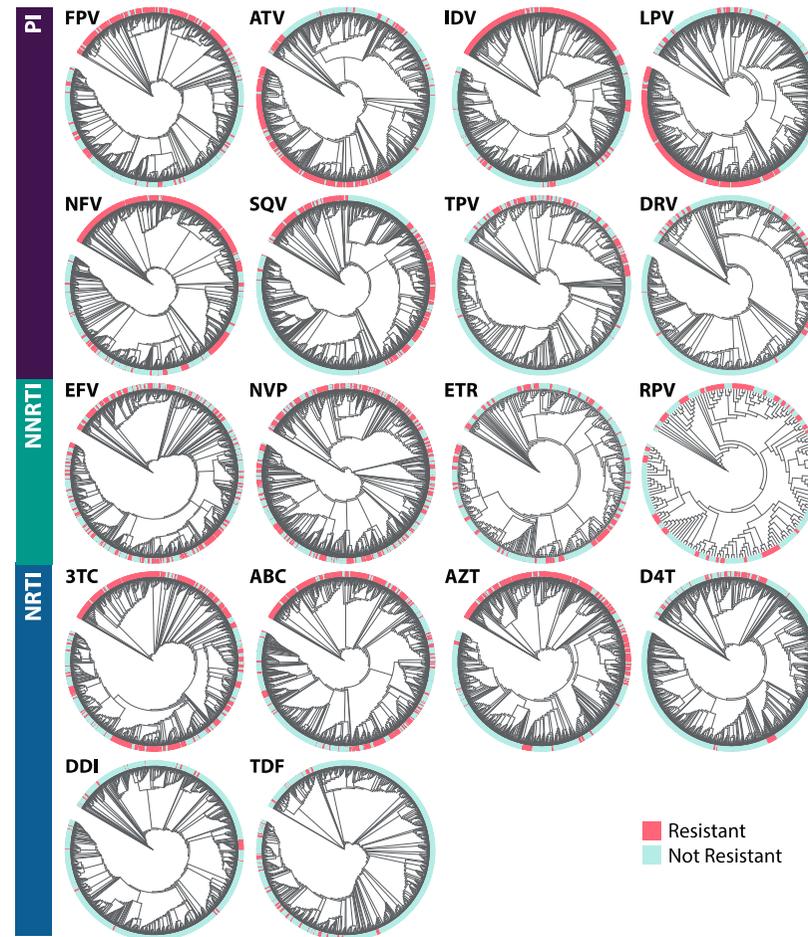
**Figure 7.** Annotated feature importance plots for the 20 most important features in Convolutional Neural Network classifiers. Horizontal axes are amino acid positions; vertical axes are feature importance (measured as change in 1-AUC). Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

### 3.3. Phylogenetics

Phylogenetic trees were constructed by drug class (Figure 8) as well as for all 18 drug-specific datasets (Figure 9). For all datasets, the AU test for constrained trees rejected the null hypothesis of perfect clustering of resistant vs. nonresistant sequences ( $p$ -value range [2.12e-85,1.57e-4]). Visual clustering of resistant sequences is seen primarily in PI datasets, with some clustering apparent in NRTI datasets and little to none in NNRTI datasets. Yet few clades were supported by bootstrap replicate values, and so no trends about drug resistance prevalence within clades could be identified with statistical support.



**Figure 8.** Annotated phylogenetic trees for each drug class. Resistant sequences are denoted by red labels; non-resistant sequences are denoted by blue labels. Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.



**Figure 9.** Annotated phylogenetic trees for each dataset. Resistant sequences are denoted by red labels; non-resistant sequences are denoted by blue labels. Abbreviations: FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

## 4. Discussion

### 4.1. Model Performance

Here we compare performance metrics from cross-validation of the described multilayer perceptron (MLP), bidirectional recurrent neural network (BRNN), and convolutional neural network (CNN) models. Overall model performance metrics (accuracy, AUC, and F1 score) indicate that the CNN models were the best performing of those included in this study, which is consistent with their known success in the direct interpretation of sequencing data [21]. Thus, our results contribute to existing literature supporting CNN models as a useful architecture for sequencing data tasks, in addition to their widespread use in biomedical imaging. A further study exploring best practices for training CNN models on genomic sequencing data would be of great benefit to the development of computational tools in the HIV sphere and for other viral and human sequencing data. BRNN models also showed high performance in many cases, particularly for PI datasets, likely due to the importance of directionality in genetic sequence data. CNN models additionally displayed the lowest false positive rates, on average, for each drug class. These results suggest that, of the three considered, the CNN architecture would be the most suitable architecture for genotypic drug resistance tests, with BRNNs also being comparable in performance.

The MLP models were included in this study as a baseline for comparison to the BRNN models and CNN models, and the notable increase in performance of the latter two confirm that more complex models are necessary for greater performance in this classification task. Further, it is notable in MLP models that none of the first most-important features from PI datasets were true DRM positions. It is possible that if variants are phylogenetically linked—as is suggested by the greater clustering of resistant sequences in these data—DRM-associated signals from these datasets are effectively non-independent, resulting in a reduction in statistical power [39]. This effect would likely be particularly troublesome for feed-forward models such as MLPs, whereas more advanced models such as BRNNs and CNNs could have greater success by identifying higher-level patterns such as those resulting from sequential ordering of features. Additional phylogenetic analysis would be necessary to address this hypothesis.

While CNN performance metrics were quite high across all datasets—with the lowest average AUC being 84% (Table A4)—it should be noted that class imbalances may have contributed to these high numbers, though AUC is more robust to class imbalance than other measures (i.e., accuracy). Overall, these results are informative for the choice of deep learning architectures in future drug resistance prediction tool-development involving deep learning models.

### 4.2. Model Interpretation

Based on permutation feature importance analysis, CNN models displayed the highest average number of known drug resistance mutation (DRM) loci within the top 20 most important features and were the only architecture for which the most important feature was a known DRM position in all 18 drug datasets (18/18). In contrast, only half of MLP models (9/18) identified a DRM position as the most important feature, while this was the case for the majority of the BRNN models (16/18). These results, combined with those of overall model performance, suggest the use of CNN models in deep learning-based drug resistance testing tools. Further, the identification of DRM positions among the most important features across datasets and architectures suggest that classification was largely based on drug resistance mutations, as opposed to artifacts or confounding factors (e.g., subtype, non-resistance mutations). Additionally, datasets for which the deep learning models had greater success in identifying known DRM loci as important features displayed higher performance, regardless of architecture, indicating a potential link between neural network functionality and evolutionary signals such as positive selection.

All models highly weighted several features which do not correspond to known DRM positions. These positions could be further explored as potential DRMs through *in vitro* studies in future work. It is possible that these sites could be linked to known DRMs, although it is thought that linkage in

HIV-1 is limited to a 100 bp range as a result of frequent recombination [40]. All but one of the features identified in at least half of the datasets for a drug class occurred within 100 bp (33 amino acid) range of a known DRM position, suggesting that linkage could be present. This observation is limited, though, by the availability of only consensus sequences—and not haplotype data—for these samples, as variants in the consensus may not appear together on any one haplotype [41,42]. It is also known that epistasis is a key factor to the evolution of HIV-1 drug resistance and that entrenchment of DRMs occurs in both protease and reverse transcriptase [43], which may also contribute to the importance of these features. The observation that a majority of highly ranked features which do not correlate to known DRMs were identified in only one or two datasets points to potential bias in the training data, which could have resulted from variation in HIV subtype or resistance to one or more additional ART drugs.

#### 4.3. Phylogenetics

Our initial, null hypothesis was that resistant and non-resistant sequences would cluster among themselves due to similarities at DRM loci. However, the AU test rejected all constrained trees with high confidence, indicating that resistant and non-resistant sequences do not entirely separate in any of the phylogenies. While some trends in the clustering of drug resistant sequences in the phylogenies are visibly apparent, low bootstrap support values precluded statistical analysis of clustering in the context of model performance. Low bootstrap support values are known to be typical in HIV-1 phylogenies due to high similarity among sequences and potential recombination events [44]. This observation demonstrates that phylogenetics alone is not sufficient to determine drug resistance status, further necessitating the development and use of computational tools designed for drug resistance prediction for clinical and public health applications.

#### 4.4. Limitations

The primary limitations of this study, and of many others which apply machine learning to biomedical data more broadly [45], are due to limited data availability. Due to sequencing costs and privacy concerns, the public availability of large quantities of HIV sequence data along with clinical outcomes such as drug resistance is inherently difficult. In addition, data that are available regarding HIV drug resistance exhibit notable class imbalance between resistant and non-resistant sequences, as shown here (Table 1), another noted challenge in deep learning for genomics [46]. In this study, these limitations were addressed using methods including stratified cross-validation and cost-sensitive learning; however, dataset size remained a limiting factor for performance in several cases.

Additionally, it is plausible that dependencies exist between resistance phenotypes for different drugs (e.g., resistance to one PI drug may confer resistance to another PI drug), which could be highly informative. This is especially relevant as combination therapy is standard in HIV care [47]. These dependencies were not addressed in this study as all categories of drug-specific phenotype data were not available for all of the sequences. Furthermore, the availability of only amino acid data—and not the corresponding nucleotide sequences—limited the extent of evolutionary analysis possible, as most established methods rely on nucleotide data. Lastly, the lack of subtype information limits the generalization of these results to the global HIV epidemic.

#### 4.5. Future Work: Further Applications of Interpretable Deep Learning in HIV

Here, we present an initial attempt to apply interpretable deep learning techniques to classify amino acid sequences by drug resistance phenotypes. We recognize that several challenges with integrating multi-omics data and machine learning are currently present [21,22,45,46,48], and so we have detailed several extensions of this work that would be beneficial to the field as more and better data become publicly available and as methods evolve to adapt to big-data difficulties. We intend for this study to serve as a proof of concept for incorporating model interpretation with biological data for novel analysis of viral genomics data.

#### 4.5.1. Intra-Patient and Temporal Data

This work demonstrates one application of model interpretation to deep learning-based HIV drug resistance classification. A similar framework could be quite useful in studying other aspects of HIV drug resistance via sequence data. All sequences in the Stanford database are consensus sequences, which typically do not resemble any one haplotype but, rather, an ancestral sequence of the haplotypes present in the viral population [41]. Due to the high intra-patient variability of HIV, it is plausible that novel insights about drug resistance could be made from applying a similar approach to haplotypes obtained from NGS [49,50]. A large-scale dataset including both haplotype data and clinical phenotypes would be necessary to achieve this aim. Moreover, evolutionary data is inherently dependent on time scales. As time goes on, haplotype frequencies will be variable, and so multiple sampling points would be necessary for a robust analysis of drug resistance evolution in intra-patient data.

#### 4.5.2. Multi-Omics Approaches

Here, our approach has focused largely on a population genetics framework, which has been previously applied to HIV genomics data successfully [40]. Due to limitations of working with amino acid sequences, we have been constrained in population genetic analysis. Notably, nucleotide data would facilitate selection analysis (dN/dS) and better characterized co-evolutionary interactions [51–53]. In recent years, however, it has become increasingly evident that genomics—or any other single data type—is insufficient to fully capture the etiology and progression of a disease [48]. Numerous studies have identified the role of virus-host interactions and other gene expression pathways in HIV-1 drug resistance [54–56], pointing to the need for a genomics-transcriptomics database on this subject. Furthermore, integrating proteomics data such as tertiary structure, hydrogen bonding, and amino acid polarity and composition would help to expand our knowledge of the role amino acids and the accumulation of amino acid changes—together and individually—play in the development of drug interaction and resistance pathways [57,58].

#### 4.5.3. Applications to Other Viruses

Drug resistance in treatment-naive patients coinfecting with HIV and HCV is not well studied [59]. Yet, HIV/HCV coinfection has been associated with selection patterns at DRM loci distinct from those of mono-infected HIV patients [60], and so data from both viruses may be informative for drug resistance patterns in coinfecting patients. In future work, an interpretable deep learning approach could be used to identify drug resistance patterns across both viral genomes in order to account for all possible evolutionary pathways for drug resistance. Moreover, drug resistance develops in other viruses, and thus this approach could be applied to other viruses of interest. Particularly, the recent SARS-CoV-2 pandemic has instigated the use of drug repurposing [61], highlighting the importance of characterizing antiviral drug resistance in HIV-1 patients so that this knowledge can be readily adapted to other infectious agents for which previously tested ART drugs may have clinical utility.

## 5. Conclusions

This study benchmarks the performance of deep learning methods for drug resistance classification in HIV-1. While such classification methods have been implemented previously, this study further compares the performance of several ML architectures and, most notably, addresses model interpretability and its biological implications. Identification of known DRM positions within the neural network, as revealed by feature importance analysis, showed correspondence to higher classification performance. Furthermore, this link suggests the investigation of other important features as biologically relevant loci. This work demonstrates the utility of interpretable machine learning in studying HIV drug resistance and develops a framework which has many important applications in viral genomics more broadly.

**Supplementary Materials:** The supplemental data includes additional figures, as referenced in the text. All sequencing and assay data used is publicly available through the Stanford University HIV Drug Resistance Database (<https://hivdb.stanford.edu>). Code used for deep learning analysis and additional supplementary material is available on GitHub ([https://github.com/maggiesteiner/HIV\\_DeepLearning](https://github.com/maggiesteiner/HIV_DeepLearning)). The following are available online at <http://www.mdpi.com/1999-4915/12/5/560/s1>, Figure S1: Annotated feature importance plots for the MLP classifier for all features across the whole gene-PI; Figure S2: Annotated feature importance plots for the MLP classifier for all features across the whole gene-NNRTI; Figure S3: Annotated feature importance plots for the MLP classifier for all features across the whole gene-NRTI; Figure S4: Annotated feature importance plots for the BRNN classifier for all features across the whole gene-PI; Figure S5: Annotated feature importance plots for the BRNN classifier for all features across the whole gene-NNRTI; Figure S6: Annotated feature importance plots for the BRNN classifier for all features across the whole gene-NRTI; Figure S7: Annotated feature importance plots for the CNN classifier for all features across the whole gene-PI; Figure S8: Annotated feature importance plots for the CNN classifier for all features across the whole gene-NNRTI; Figure S9: Annotated feature importance plots for the CNN classifier for all features across the whole gene-NRTI; Figure S10: Histogram denoting the frequency of amino acid positions among the 20 most important features for CNN models in PI datasets; Figure S11: Histogram denoting the frequency of amino acid positions among the 20 most important features for CNN models in NNRTI datasets; Figure S12: Histogram denoting the frequency of amino acid positions among the 20 most important features for CNN models in NRTI datasets.

**Author Contributions:** Formal analysis, M.C.S.; investigation, M.C.S.; methodology, M.C.S. and K.M.G.; software, M.C.S.; visualization, M.C.S.; writing—original draft, M.C.S.; writing—review & editing, K.M.G. and K.A.C.; funding acquisition, M.C.S. and K.A.C.; supervision, K.A.C. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** The authors acknowledge the sponsorship of the GW Data MASTER Program, supported by the National Science Foundation under grant DMS-1406984. The authors thank Drs. Kyle Crandall and Ali Rahnavard for helpful comments on earlier versions of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

FPV = fosamprenavir; ATV = atazanavir; IDV = indinavir; LPV = lopinavir; NFV = nelfinavir; SQV = saquinavir; TPV = tipranavir; DRV = darunavir; 3TC = lamivudine; ABC = abacavir; AZT = azidothymidine; D4T = stavudine; DDI = didanosine; TDF = tenofovir disoproxil fumarate; EFV = efavirenz; NVP = nevirapine; ETR = etravirine; RPV = rilpivirine; PI = protease inhibitor; NRTI = nucleotide reverse transcriptase inhibitor; NNRTI = non-nucleotide reverse transcriptase inhibitor.

## Appendix A

**Table A1.** Model performance metrics for the MLP classifier. All values are the mean over 5-fold stratified cross-validation, reported with standard deviation in brackets.

MLP							
Class	Drug	Accuracy	Loss	TPR	TNR	FPR	FNR
PI	FPV	0.839 [0.029]	0.391 [0.064]	0.738 [0.114]	0.895 [0.029]	0.105 [0.029]	0.262 [0.114]
PI	ATV	0.830 [0.026]	0.400 [0.038]	0.862 [0.061]	0.805 [0.075]	0.195 [0.075]	0.138 [0.061]
PI	IDV	0.858 [0.016]	0.360 [0.030]	0.800 [0.063]	0.908 [0.042]	0.092 [0.042]	0.200 [0.063]
PI	LPV	0.852 [0.042]	0.347 [0.055]	0.826 [0.060]	0.868 [0.075]	0.132 [0.075]	0.174 [0.060]
PI	NFV	0.819 [0.020]	0.408 [0.040]	0.839 [0.081]	0.799 [0.118]	0.201 [0.118]	0.161 [0.081]
PI	SQV	0.843 [0.015]	0.414 [0.030]	0.850 [0.101]	0.826 [0.061]	0.174 [0.061]	0.150 [0.101]
PI	TPV	0.734 [0.076]	0.539 [0.111]	0.812 [0.169]	0.720 [0.119]	0.280 [0.119]	0.188 [0.169]
PI	DRV	0.835 [0.049]	0.391 [0.073]	0.780 [0.173]	0.851 [0.093]	0.149 [0.093]	0.220 [0.173]
<b>PI AVERAGE</b>		<b>0.826 [0.039]</b>	<b>0.406 [0.058]</b>	<b>0.813 [0.041]</b>	<b>0.834 [0.061]</b>	<b>0.166 [0.061]</b>	<b>0.187 [0.041]</b>
NNRTI	EFV	0.709 [0.044]	0.584 [0.060]	0.753 [0.194]	0.683 [0.198]	0.317 [0.198]	0.247 [0.194]
NNRTI	NVP	0.767 [0.033]	0.510 [0.036]	0.816 [0.089]	0.725 [0.108]	0.275 [0.108]	0.184 [0.089]
NNRTI	ETR	0.659 [0.145]	0.622 [0.164]	0.524 [0.378]	0.721 [0.318]	0.279 [0.318]	0.476 [0.378]
NNRTI	RPV	0.690 [0.079]	0.627 [0.044]	0.554 [0.210]	0.741 [0.149]	0.259 [0.149]	0.446 [0.210]
<b>NNRTI AVERAGE</b>		<b>0.706 [0.095]</b>	<b>0.586 [0.143]</b>	<b>0.662 [0.140]</b>	<b>0.718 [0.105]</b>	<b>0.282 [0.105]</b>	<b>0.338 [0.140]</b>
NRTI	3TC	0.839 [0.030]	0.368 [0.047]	0.859 [0.072]	0.811 [0.095]	0.189 [0.095]	0.141 [0.072]
NRTI	ABC	0.811 [0.026]	0.455 [0.061]	0.742 [0.076]	0.866 [0.107]	0.134 [0.107]	0.258 [0.076]
NRTI	AZT	0.711 [0.173]	0.769 [0.605]	0.717 [0.250]	0.716 [0.415]	0.284 [0.415]	0.283 [0.250]
NRTI	D4T	0.867 [0.032]	0.308 [0.078]	0.704 [0.097]	0.891 [0.046]	0.109 [0.046]	0.296 [0.097]
NRTI	DDI	0.709 [0.256]	0.572 [0.518]	0.613 [0.301]	0.714 [0.283]	0.286 [0.283]	0.387 [0.301]
NRTI	TDF	0.831 [0.047]	0.344 [0.069]	0.717 [0.164]	0.842 [0.061]	0.158 [0.061]	0.283 [0.164]
<b>NRTI AVERAGE</b>		<b>0.795 [0.061]</b>	<b>0.470 [0.150]</b>	<b>0.725 [0.112]</b>	<b>0.807 [0.081]</b>	<b>0.193 [0.081]</b>	<b>0.275 [0.112]</b>

**Table A2.** Model performance metrics for the BRNN classifier. All values are the mean over 5-fold stratified cross-validation, reported with standard deviation in brackets.

BRNN							
Class	Drug	Accuracy	Loss	TPR	TNR	FPR	FNR
PI	FPV	0.898 [0.012]	0.298 [0.050]	0.900 [0.024]	0.898 [0.014]	0.102 [0.014]	0.100 [0.024]
PI	ATV	0.925 [0.026]	0.259 [0.102]	0.930 [0.027]	0.921 [0.045]	0.079 [0.045]	0.070 [0.027]
PI	IDV	0.935 [0.011]	0.225 [0.050]	0.926 [0.025]	0.940 [0.029]	0.060 [0.029]	0.074 [0.025]
PI	LPV	0.936 [0.023]	0.219 [0.059]	0.919 [0.042]	0.949 [0.020]	0.051 [0.020]	0.081 [0.042]
PI	NFV	0.941 [0.014]	0.208 [0.050]	0.938 [0.017]	0.945 [0.024]	0.055 [0.024]	0.062 [0.017]
PI	SQV	0.918 [0.023]	0.292 [0.091]	0.913 [0.022]	0.922 [0.029]	0.078 [0.029]	0.087 [0.022]
PI	TPV	0.845 [0.021]	0.469 [0.130]	0.745 [0.065]	0.865 [0.021]	0.135 [0.021]	0.255 [0.065]
PI	DRV	0.896 [0.044]	0.349 [0.176]	0.851 [0.098]	0.910 [0.070]	0.090 [0.070]	0.149 [0.098]
<b>PI AVERAGE</b>		<b>0.912 [0.032]</b>	<b>0.290 [0.087]</b>	<b>0.890 [0.065]</b>	<b>0.919 [0.028]</b>	<b>0.081 [0.028]</b>	<b>0.110 [0.065]</b>
NNRTI	EFV	0.946 [0.018]	0.188 [0.055]	0.925 [0.023]	0.961 [0.022]	0.039 [0.022]	0.075 [0.023]
NNRTI	NVP	0.917 [0.063]	0.261 [0.122]	0.881 [0.102]	0.952 [0.027]	0.048 [0.027]	0.119 [0.102]
NNRTI	ETR	0.946 [0.018]	0.188 [0.055]	0.925 [0.023]	0.961 [0.022]	0.039 [0.022]	0.075 [0.023]
NNRTI	RPV	0.729 [0.033]	0.701 [0.192]	0.583 [0.189]	0.783 [0.102]	0.217 [0.102]	0.417 [0.189]
<b>NNRTI AVERAGE</b>		<b>0.884 [0.104]</b>	<b>0.334 [0.247]</b>	<b>0.829 [0.165]</b>	<b>0.914 [0.088]</b>	<b>0.086 [0.088]</b>	<b>0.171 [0.165]</b>
NRTI	3TC	0.931 [0.013]	0.247 [0.044]	0.938 [0.027]	0.923 [0.023]	0.077 [0.023]	0.062 [0.027]
NRTI	ABC	0.842 [0.029]	0.505 [0.139]	0.819 [0.025]	0.862 [0.039]	0.138 [0.039]	0.181 [0.025]
NRTI	AZT	0.890 [0.011]	0.407 [0.087]	0.891 [0.037]	0.887 [0.032]	0.113 [0.032]	0.109 [0.037]
NRTI	D4T	0.892 [0.017]	0.333 [0.041]	0.705 [0.060]	0.919 [0.018]	0.081 [0.018]	0.295 [0.060]
NRTI	DDI	0.909 [0.025]	0.278 [0.064]	0.478 [0.137]	0.935 [0.020]	0.065 [0.020]	0.522 [0.137]
NRTI	TDF	0.911 [0.023]	0.374 [0.085]	0.673 [0.178]	0.934 [0.036]	0.066 [0.036]	0.327 [0.178]
<b>NRTI AVERAGE</b>		<b>0.896 [0.030]</b>	<b>0.357 [0.094]</b>	<b>0.751 [0.168]</b>	<b>0.910 [0.029]</b>	<b>0.090 [0.029]</b>	<b>0.249 [0.168]</b>

**Table A3.** Model performance metrics for the CNN classifier. All values are the mean over 5-fold stratified cross-validation, reported with standard deviation in brackets.

CNN							
Class	Drug	Accuracy	Loss	TPR	TNR	FPR	FNR
PI	FPV	0.901 [0.017]	0.980 [0.252]	0.879 [0.053]	0.913 [0.027]	0.087 [0.027]	0.121 [0.053]
PI	ATV	0.922 [0.029]	0.610 [0.195]	0.908 [0.043]	0.932 [0.048]	0.068 [0.048]	0.092 [0.043]
PI	IDV	0.932 [0.018]	0.622 [0.134]	0.914 [0.028]	0.945 [0.021]	0.055 [0.021]	0.086 [0.028]
PI	LPV	0.946 [0.007]	0.542 [0.166]	0.942 [0.015]	0.951 [0.015]	0.049 [0.015]	0.058 [0.015]
PI	NFV	0.920 [0.008]	0.671 [0.114]	0.880 [0.012]	0.944 [0.013]	0.056 [0.013]	0.120 [0.012]
PI	SQV	0.920 [0.013]	0.671 [0.090]	0.880 [0.038]	0.944 [0.008]	0.0562 [0.008]	0.120 [0.038]
PI	TPV	0.884 [0.031]	1.133 [0.339]	0.680 [0.106]	0.929 [0.033]	0.071 [0.033]	0.320 [0.106]
PI	DRV	0.927 [0.026]	0.733 [0.353]	0.834 [0.085]	0.954 [0.029]	0.046 [0.029]	0.166 [0.085]
<b>PI AVERAGE</b>		<b>0.919 [0.019]</b>	<b>0.745 [0.204]</b>	<b>0.865 [0.081]</b>	<b>0.939 [0.013]</b>	<b>0.061 [0.013]</b>	<b>0.135 [0.081]</b>
NNRTI	EFV	0.951 [0.010]	0.398 [0.107]	0.926 [0.020]	0.969 [0.008]	0.0313 [0.008]	0.074 [0.020]
NNRTI	NVP	0.954 [0.011]	0.455 [0.076]	0.933 [0.038]	0.971 [0.024]	0.029 [0.024]	0.067 [0.038]
NNRTI	ETR	0.878 [0.022]	0.777 [0.339]	0.783 [0.056]	0.916 [0.025]	0.084 [0.025]	0.217 [0.059]
NNRTI	RPV	0.862 [0.076]	0.826 [0.409]	0.739 [0.108]	0.924 [0.091]	0.076 [0.091]	0.261 [0.108]
<b>NNRTI AVERAGE</b>		<b>0.911 [0.048]</b>	<b>0.614 [0.219]</b>	<b>0.845 [0.099]</b>	<b>0.945 [0.029]</b>	<b>0.055 [0.029]</b>	<b>0.155 [0.099]</b>
NRTI	3TC	0.946 [0.012]	0.591 [0.183]	0.945 [0.011]	0.948 [0.030]	0.052 [0.030]	0.055 [0.011]
NRTI	ABC	0.862 [0.007]	1.283 [0.143]	0.847 [0.018]	0.874 [0.012]	0.126 [0.012]	0.153 [0.018]
NRTI	AZT	0.901 [0.019]	0.874 [0.278]	0.884 [0.057]	0.913 [0.020]	0.087 [0.020]	0.116 [0.057]
NRTI	D4T	0.922 [0.025]	0.785 [0.310]	0.722 [0.071]	0.952 [0.021]	0.048 [0.021]	0.278 [0.071]
NRTI	DDI	0.959 [0.008]	0.482 [0.073]	0.497 [0.065]	0.985 [0.014]	0.015 [0.014]	0.503 [0.065]
NRTI	TDF	0.933 [0.016]	0.813 [0.168]	0.561 [0.096]	0.968 [0.013]	0.032 [0.013]	0.439 [0.096]
<b>NRTI AVERAGE</b>		<b>0.920 [0.035]</b>	<b>0.805 [0.277]</b>	<b>0.743 [0.182]</b>	<b>0.940 [0.040]</b>	<b>0.060 [0.040]</b>	<b>0.257 [0.182]</b>

**Table A4.** AUC and F1 metrics for all classifiers. All values are the mean over 5-fold stratified cross-validation, reported with standard deviation in brackets.

Model	MLP		BRNN		CNN	
	AUC	F1	AUC	F1	AUC	F1
ATV	0.899 [0.019]	0.765 [0.062]	0.980 [0.013]	0.865 [0.018]	0.981 [0.012]	0.866 [0.019]
DRV	0.872 [0.032]	0.815 [0.021]	0.960 [0.011]	0.915 [0.027]	0.987 [0.013]	0.910 [0.033]
FPV	0.934 [0.016]	0.830 [0.019]	0.945 [0.011]	0.925 [0.016]	0.937 [0.009]	0.922 [0.015]
IDV	0.928 [0.009]	0.831 [0.048]	0.988 [0.006]	0.928 [0.024]	0.979 [0.007]	0.925 [0.008]
LPV	0.926 [0.021]	0.829 [0.019]	0.975 [0.006]	0.944 [0.014]	0.973 [0.006]	0.948 [0.010]
NFV	0.906 [0.012]	0.809 [0.043]	0.979 [0.004]	0.894 [0.027]	0.976 [0.009]	0.891 [0.025]
SQV	0.897 [0.008]	0.523 [0.064]	0.976 [0.011]	0.630 [0.065]	0.981 [0.005]	0.673 [0.091]
TPV	0.842 [0.029]	0.669 [0.067]	0.916 [0.049]	0.782 [0.077]	0.93 [0.028]	0.831 [0.060]
<b>PI AVERAGE</b>	<b>0.9 [0.009]</b>	<b>0.732 [0.135]</b>	<b>0.965 [0.014]</b>	<b>0.860 [0.106]</b>	<b>0.968 [0.007]</b>	<b>0.871 [0.088]</b>
EFV	0.843 [0.033]	0.676 [0.043]	0.953 [0.017]	0.934 [0.021]	0.980 [0.007]	0.940 [0.013]
ETR	0.764 [0.032]	0.768 [0.034]	0.669 [0.049]	0.910 [0.067]	0.926 [0.023]	0.950 [0.016]
NVP	0.853 [0.022]	0.415 [0.209]	0.957 [0.062]	0.934 [0.021]	0.976 [0.008]	0.783 [0.044]
RPV	0.760 [0.037]	0.524 [0.139]	0.740 [0.062]	0.567 [0.102]	0.896 [0.032]	0.770 [0.126]
<b>NNRTI AVERAGE</b>	<b>0.805 [0.006]</b>	<b>0.596 [0.157]</b>	<b>0.830 [0.021]</b>	<b>0.836 [0.180]</b>	<b>0.945 [0.012]</b>	<b>0.861 [0.097]</b>
ABC	0.887 [0.025]	0.861 [0.031]	0.904 [0.026]	0.941 [0.103]	0.935 [0.007]	0.953 [0.010]
AZT	0.901 [0.025]	0.783 [0.017]	0.934 [0.012]	0.828 [0.028]	0.944 [0.012]	0.849 [0.009]
DDI	0.861 [0.081]	0.681 [0.100]	0.682 [0.053]	0.871 [0.009]	0.842 [0.047]	0.881 [0.027]
DFT	0.935 [0.025]	0.579 [0.050]	0.928 [0.057]	0.625 [0.083]	0.926 [0.013]	0.709 [0.051]
TDF	0.853 [0.044]	0.224 [0.107]	0.939 [0.027]	0.362 [0.116]	0.940 [0.034]	0.559 [0.095]
TTC	0.933 [0.017]	0.430 [0.096]	0.952 [0.010]	0.564 [0.100]	0.977 [0.013]	0.589 [0.119]
<b>NRTI AVERAGE</b>	<b>0.895 [0.024]</b>	<b>0.593 [0.236]</b>	<b>0.89 [0.020]</b>	<b>0.698 [0.220]</b>	<b>0.927 [0.016]</b>	<b>0.757 [0.163]</b>

## References

- Centers for Disease Control and Prevention. *HIV Surveillance Report*; Centers for Disease Control and Prevention: Atlanta, GA, USA, 2018; Volume 30.
- Wandeler, G.; Johnson, L.F.; Egger, M. Trends in life expectancy of HIV-positive adults on ART across the globe: Comparisons with general population HHS Public Access. *Curr. Opin. HIV AIDS* **2016**, *11*, 492–500. [[CrossRef](#)] [[PubMed](#)]
- Das, M.; Chu, P.L.; Santos, G.M.; Scheer, S.; Vittinghoff, E.; McFarland, W.; Colfax, G.N. Decreases in community viral load are accompanied by reductions in new HIV infections in San Francisco. *PLoS ONE* **2010**, *5*, e11068. [[CrossRef](#)]
- Quinn, T.C.; Wawer, M.J.; Sewankambo, N.; Serwadda, D.; Li, C.; Wabwire-Mangen, F.; Meehan, M.O.; Lutalo, T.; Gray, R.H. Viral load and heterosexual transmission of human immunodeficiency virus type 1. *N. Engl. J. Med.* **2000**, *342*, 921–929. [[CrossRef](#)] [[PubMed](#)]
- Rambaut, A.; Posada, D.; Crandall, K.A.; Holmes, E.C. The causes and consequences of HIV evolution. *Nat. Rev. Genet.* **2004**, *5*, 52–61. [[CrossRef](#)] [[PubMed](#)]
- Crandall, K.A.; Kelsey, C.R.; Imamichi, H.; Lane, H.C.; Salzman, N.P. Parallel evolution of drug resistance in HIV: Failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* **1999**, *16*, 372–382. [[CrossRef](#)] [[PubMed](#)]
- Hirsch, M.S.; Conway, B.; D'Aquila, R.T.; Johnson, V.A.; Brun-Vézinet, F.; Clotet, B.; Demeter, L.M.; Hammer, S.M.; Jacobsen, D.M.; Kuritzkes, D.R.; et al. Antiretroviral drug resistance testing in adults with HIV infection: Implications for clinical management. *J. Am. Med. Assoc.* **1998**, *279*, 1984–1991. [[CrossRef](#)] [[PubMed](#)]
- Zhang, J.; Rhee, S.Y.; Taylor, J.; Shafer, R.W. Comparison of the precision and sensitivity of the antiviogram and PhenoSense HIV drug susceptibility assays. *J. Acquir. Immune Defic. Syndr.* **2005**, *38*, 439–444. [[CrossRef](#)]
- Rhee, S.Y.; Gonzales, M.J.; Kantor, R.; Betts, B.J.; Ravela, J.; Shafer, R.W. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003**, *31*, 298–303. [[CrossRef](#)]
- Bonet, I.; García, M.M.; Saeys, Y.; Van De Peer, Y.; Grau, R. Predicting human immunodeficiency virus (HIV) drug resistance using recurrent neural networks. In Proceedings of the IWINAC 2007, La Manga del Mar Menor, Spain, 18–21 June 2007; Volume 4527, pp. 234–243.

11. Liu, T.; Shafer, R. Web Resources for HIV type 1 Genotypic-Resistance Test Interpretation. *Clin. Infect. Dis.* **2006**, *42*, 1608–1618. [[CrossRef](#)]
12. Jensen, M.A.; Coetzer, M.; van 't Wout, A.B.; Morris, L.; Mullins, J.I. A Reliable Phenotype Predictor for Human Immunodeficiency Virus Type 1 Subtype C Based on Envelope V3 Sequences. *J. Virol.* **2006**, *80*, 4698–4704. [[CrossRef](#)]
13. Beerenwinkel, N.; Däumer, M.; Oette, M.; Korn, K.; Hoffmann, D.; Kaiser, R.; Lengauer, T.; Selbig, J.; Walter, H. Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res.* **2003**, *31*, 3850–3855. [[CrossRef](#)]
14. Riemenschneider, M.; Hummel, T.; Heider, D. SHIVA—A web application for drug resistance and tropism testing in HIV. *BMC Bioinform.* **2016**, *17*, 314. [[CrossRef](#)] [[PubMed](#)]
15. Riemenschneider, M.; Cashin, K.Y.; Budeus, B.; Sierra, S.; Shirvani-Dastgerdi, E.; Bayanolhagh, S.; Kaiser, R.; Gorry, P.R.; Heider, D. Genotypic Prediction of Co-receptor Tropism of HIV-1 Subtypes A and C. *Sci. Rep.* **2016**, *6*, 1–9. [[CrossRef](#)] [[PubMed](#)]
16. Heider, D.; Senge, R.; Cheng, W.; Hüllermeier, E. Multilabel classification for exploiting cross-resistance information in HIV-1 drug resistance prediction. *Bioinformatics* **2013**, *29*, 1946–1952. [[CrossRef](#)] [[PubMed](#)]
17. Beerenwinkel, N.; Schmidt, B.; Walter, H.; Kaiser, R.; Lengauer, T.; Hoffmann, D.; Korn, K.; Selbig, J. Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8271–8276. [[CrossRef](#)]
18. Wang, D.; Larder, B. Networks Enhanced Prediction of Lopinavir Resistance from Genotype by Use of Artificial Neural Networks. *J. Infect. Dis.* **2003**, *188*, 653–660. [[CrossRef](#)]
19. Sheik Amamuddy, O.; Bishop, N.T.; Tasthan Bishop, Ö. Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks. *BMC Bioinform.* **2017**, *18*, 369. [[CrossRef](#)]
20. Ekpenyong, M.E.; Etebong, P.I.; Jackson, T.C. Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy. *Heliyon* **2019**, *5*, e02080. [[CrossRef](#)]
21. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [[CrossRef](#)]
22. Min, S.; Lee, B.; Yoon, S. Deep learning in bioinformatics. *Brief. Bioinform.* **2017**, *18*, 851–869. [[CrossRef](#)]
23. Molnar, C. *iml: An R package for Interpretable Machine Learning*. *J. Open Source Softw.* **2018**, *3*, 786. [[CrossRef](#)]
24. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
25. RStudio Team. *RStudio: Integrated Development Environment for R*; R Studio, Inc.: Boston, MA, USA, 2015.
26. Chollet, F.; Allaire, J.J. Keras: R Interface to Keras; Keras Team. 2017. Available online: <https://keras.rstudio.com/index.html> (accessed on 21 April 2020).
27. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [White Paper]*; Google Research: Palo Alto, CA, USA, 2015.
28. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009; ISBN 1441412697.
29. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
30. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)] [[PubMed](#)]
31. Fisher, A.; Rudin, C.; Dominici, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.
32. Darriba, D.; Posada, D.; Kozlov, A.M.; Stamatakis, A.; Morel, B.; Flouri, T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol. Biol. Evol.* **2019**, *37*, 291–294. [[CrossRef](#)] [[PubMed](#)]
33. Dang, C.C.; Le, Q.S.; Gascuel, O.; Le, V.S. FLU, an amino acid substitution model for influenza proteins. *BMC Evol. Biol.* **2010**, *10*, 99. [[CrossRef](#)] [[PubMed](#)]

34. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
35. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010.
36. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]
37. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **2002**, *51*, 492–508. [[CrossRef](#)]
38. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
39. Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **1985**, *125*, 1–15. [[CrossRef](#)]
40. Zanini, F.; Brodin, J.; Thebo, L.; Lanz, C.; Bratt, G.; Albert, J.; Neher, R.A. Population genomics of intrapatient HIV-1 evolution. *Elife* **2015**, *4*, e11282. [[CrossRef](#)] [[PubMed](#)]
41. Posada-Céspedes, S.; Seifert, D.; Beerenwinkel, N. Recent advances in inferring viral diversity from high-throughput sequencing data. *Virus Res.* **2017**, *239*, 17–32. [[CrossRef](#)] [[PubMed](#)]
42. Wirden, M.; Malet, I.; Derache, A.; Marcelin, A.G.; Roquebert, B.; Simon, A.; Kirstetter, M.; Joubert, L.M.; Katlama, C.; Calvez, V. Clonal analyses of HIV quasiespecies in patients harbouring plasma genotype with K65R mutation associated with thymidine analogue mutations or L74V substitution. *AIDS* **2005**, *19*, 630–632. [[CrossRef](#)]
43. Biswas, A.; Haldane, A.; Arnold, E.; Levy, R.M. Epistasis and entrenchment of drug resistance in HIV-1 subtype B. *Elife* **2019**, *8*, e50524. [[CrossRef](#)]
44. Castro-Nallar, E.; Pérez-Losada, M.; Burton, G.F.; Crandall, K.A. The evolution of HIV: Inferences using phylogenetics. *Mol. Phylogenet. Evol.* **2012**, *62*, 777–792. [[CrossRef](#)] [[PubMed](#)]
45. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.M.; Zietz, M.; Hoffman, M.M.; et al. Opportunities and Obstacles for Deep Learning in Biology and Medicine. *J. R. Soc. Interface* **2018**, *15*, 20170387. [[CrossRef](#)] [[PubMed](#)]
46. Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamani, A.; Telenti, A. A primer on deep learning in genomics. *Nat. Genet.* **2019**, *51*, 12–18. [[CrossRef](#)] [[PubMed](#)]
47. Margolis, D.M.; Koup, R.A.; Ferrari, G. HIV antibodies for treatment of HIV infection. *Immunol. Rev.* **2017**, *275*, 313–323. [[CrossRef](#)]
48. Zitnik, M.; Nguyen, F.; Wang, B.; Leskovec, J.; Goldenberg, A.; Hoffman, M.M. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Inf. Fusion* **2019**, *50*, 71–91. [[CrossRef](#)]
49. Gibson, K.M.; Jair, K.; Castel, A.D.; Bendall, M.L.; Wilbourn, B.; Jordan, J.A.; Crandall, K.A.; Pérez-Losada, M.; Subramanian, T.; Binkley, J.; et al. A cross-sectional study to characterize local HIV-1 dynamics in Washington, DC using next-generation sequencing. *Sci. Rep.* **2020**, *10*, 1–18. [[CrossRef](#)] [[PubMed](#)]
50. Eliseev, A.; Gibson, K.M.; Avdeyev, P.; Novik, D.; Bendall, M.L.; Pérez-Losada, M.; Alexeev, N.; Crandall, K.A. Evaluation of haplotype callers for next-generation sequencing of viruses. *Infect. Genet. Evol.* **2020**, *82*, 104277. [[CrossRef](#)] [[PubMed](#)]
51. Shindyalov, I.N.; Kolchanov, N.A.; Sander, C. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Eng. Des. Sel.* **1994**, *7*, 349–358. [[CrossRef](#)] [[PubMed](#)]
52. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. [[CrossRef](#)] [[PubMed](#)]
53. Kosakovsky Pond, S.L.; Frost, S.D.W.; Muse, S.V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **2005**, *21*, 676–679. [[CrossRef](#)]
54. Arenas, M.; Araujo, N.M.; Branco, C.; Castelhana, N.; Castro-Nallar, E.; Pérez-Losada, M. Mutation and recombination in pathogen evolution: Relevance, methods and controversies. *Infect. Genet. Evol.* **2018**, *63*, 295–306. [[CrossRef](#)]
55. Kis, O.; Sankaran-Walters, S.; Hoque, M.T.; Walmsley, S.L.; Dandekar, S.; Bendayan, R. HIV-1 infection alters intestinal expression of antiretroviral drug transporters and enzymes. *Top. Antivir. Med.* **2014**, *22*, 53.

56. Balasubramaniam, M.; Pandhare, J.; Dash, C. Are microRNAs important players in HIV-1 infection? An update. *Viruses* **2018**, *10*, 110. [[CrossRef](#)]
57. Whitfield, T.W.; Ragland, D.A.; Zeldovich, K.B.; Schiffer, C.A. Characterizing Protein-Ligand Binding Using Atomistic Simulation and Machine Learning: Application to Drug Resistance in HIV-1 Protease. *J. Chem. Theory Comput.* **2020**, *16*, 1284–1299. [[CrossRef](#)]
58. Conti, S.; Karplus, M. Estimation of the breadth of CD4bs targeting HIV antibodies by molecular modeling and machine learning. *PLoS Comput. Biol.* **2019**, *15*, e1006954. [[CrossRef](#)]
59. Deng, H.; Deng, X.; Liu, Y.; Xu, Y.; Lan, Y.; Gao, M.; Xu, M.; Gao, H.; Wu, X.; Liao, B.; et al. Naturally occurring antiviral drug resistance in HIV patients who are mono-infected or co-infected with HBV or HCV in China. *J. Med. Virol.* **2018**, *90*, 1246–1256. [[CrossRef](#)]
60. Domínguez-Rodríguez, S.; Rojas, P.; McPhee, C.F.; Pagán, I.; Navarro, M.L.; Ramos, J.T.; Holguín, Á. Effect of HIV / HCV Co-Infection on the Protease Evolution of HIV-1B: A Pilot Study in a Pediatric Population. *Sci. Rep.* **2018**, *8*, 4–10. [[CrossRef](#)] [[PubMed](#)]
61. Li, G.; De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **2020**, *19*, 149–150. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).