ANN-based Drug-isolate-fold-change model predicting the resistance profiles of HIV-1 protease inhibitors

Huseyin Tunc, Murat Sari, Serdar Durdagi, Seyfullah Enes Kotil

ABSTRACT: Drug resistance is a primary barrier to effective treatments of HIV/AIDS. Calculating quantitative relations between genotype and phenotype observations for each inhibitor with cell-based assays requires time and money consuming experiments. Machine learning models are good options for tackling these problems by generalizing the available data with suitable linear or nonlinear mappings. The main aim of this paper is to construct drug isolate fold change (DIF)-based artificial neural network (ANN) models for estimating the resistance potential of molecules inhibiting the HIV-1 protease (PR) enzyme. Throughout the study, seven of eight protease inhibitors (PIs) have been included in the training set and the remaining ones in the test set. Using the 7-in 1-out procedure, eight ANN models have been produced to measure the learning capacity of models from the descriptors of the inhibitors. The mean value of eight ANN models for unseen inhibitors is and 95% confidence interval (CI) is Predicting the fold change resistance for hundreds of isolates allowed for robust comparison of drug pairs. These eight models have predicted the drug resistance tendencies of each inhibitor pair with the mean 2D correlation coefficient 0.933 and 95% CI A classification problem has been created to predict the ordered relationship of the PIs and the mean accuracy, sensitivity and specificity values are obtained as 0.954, 0.791 and 0.791, respectively. The currently derived ANN models can accurately predict the drug resistance tendencies of PI pairs, and this observation could help test new inhibitors with various isolates.

Keywords: Machine learning; Artificial neural networks; HIV/AIDS; Drug resistance; Protease inhibitors

INTRODUCTION

Acquired immunodeficiency syndrome (AIDS) disease caused by the human immunodeficiency viruses, HIV-1 and HIV-2, began to spread in the 1970s and came into focus in the early 1980s as one of the most severe public health threats in history [1]. Detection of reverse transcription activity in cultures of lymph node cells from AIDS patients in the early 1980s revealed that AIDS was caused by a retrovirus later called human immunodeficiency virus (HIV) [2]. Zidovudine (AZT), the first nucleotide reverse transcriptase inhibitor (NRTI) that inhibits the reverse transcription enzyme of HIV, was approved in 1987, and today there are nearly thirty approved drugs [3]. HIV-1 has affected approximately 38 million people today, and just about 26 million

people are receiving "Highly Active Antiretroviral Treatment" (HAART) [4]. The HAART therapy proposed in the mid-1990s was defined as the procedure of using three or four different drugs that act on various targets in the virus's life cycle [5]. With HAART therapy, the death rate fell to 47% in 1997, just ten years after the first AIDS case was detected [6].

Drug resistance is the primary barrier to the effective treatment of HIV/AIDS [7,8]. Single drug treatments for HIV yield rapid resistance due to the high genetic diversity and error-prone replication of the virus [8,9]. Thence, the use of drug combinations through the HAART protocols increases the efficacy of the treatment [10]. However, cross-resistant isolates for available drugs encourage researchers to find novel inhibitors [11-15]. To combat with drug-resistant isolates, novel drug design methodologies have been adopted for HIV-1 protease enzyme such as phosphonate-mediated solvent anchoring [11], lysine sulfonamide-based molecular core [12], allophenylnorstatine containing inhibitors [13], nonpeptic inhibitor GRL-02031 [14], bis-tetrahydrofuranylurethane containing nonpeptidic inhibitor UIC-94017 [15]. Testing novel inhibitors with various drug-resistant isolates need experimental or computational mechanisms.

HIV protease enzyme plays a vital role in forming infectious viruses by regulating immature viruses' synthesized gag and gag-pol polyproteins [16]. Protease inhibitors are generally included in the scope of HAART therapy, and eight approved drug molecules are used effectively today [17]. Dose-response curves of protease inhibitors were shown that they have higher Hill coefficient values than the fusion (FI), integrase (II), nucleoside reverse transcriptase (NRTI) and non-nucleotide reverse transcriptase (NNRTI) inhibitors [18]. Even if a person is infected with the wild-type virion, resistant variants may emerge with dosing disruptions or the use of inappropriate combinations in the HAART therapy [19]. The success rate of HAART therapy can be increased by measuring the efficacy of existing and novel inhibitors over resistant genotypes [20-21]. The observation of drug-efficacy relations with cell-based assays is expensive and time-consuming in the presence of genotype information, and mathematical models are essential to tackle this vital problem [22-24].

Various mathematical models have been calibrated using genotype-fold change data proposed in the Stanford HIV database to predict mutational effects on viral dynamics in the literature [25-40]. The life span of patients can be considerably extended by the construction of reliable mathematical models that accurately predict suitable drugs for existing isolates. Most existing prediction models are knowledge-based and require predetermined rules on mutations and

drugs [25-28]. The most commonly used genotype interpretation algorithms have been observed to be Stanford HIVdb [25], HIV-grade [26], REGA [27] and ANRS [28]. In addition to these genotype interpretation algorithms, various machine learning models have recently been proposed to predict genotype-fold change relationships in the presence of a predetermined inhibitor [29-40]. Artificial neural network [29-34], random forest algorithm [35-41], support vector machine [37,41-42], decision trees [43], k-nearest neighbours (kNN) [36], restricted Boltzmann machine [44], support vector regression [40] and linear regression [45] are the techniques used in the literature to model the efficacy of different drugs against HIV-1 variants. All of the works mentioned above focus on predicting the fold change of a single drug. The models take the mutational genotype as an input without the need for molecular descriptors as an input. Instead, a general model that makes fold-change predictions on hundreds of isolates based on molecular fingerprints are lacking.

So far, machine learning models for each HIV-1 inhibitor have been successfully proposed with various encoding techniques of genotypes. Here, the possibility of constructing machine learning models that simultaneously take inhibitor fingerprints and isolate descriptors as inputs and estimate the fold change values is explored. For training and testing of models, data of eight approved protease inhibitors atazanavir (AZT), darunavir (DRV), fosamprenavir (FPV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), saquinavir (SAV) and tipranavir (TPV) in the Stanford HIV drug resistance database is used. By proposing a reliable testing procedure called 7-in 1-out, our drug-isolate-fold change (DIF) based artificial neural network (ANN) models are seen to have the ability to learn from inhibitor descriptors to predict fold-change values. The model can predict the fold change of hundreds of isolates based on molecular fingerprints and the mutational genotype. To that end, the learned hundreds of predictors (fold-change of isolates) can be successfully used to assess the resistance potential of inhibitors. We used pairs of drugs to predict the more resistance prone molecule. We called these pairwise comparisons the resistance tendencies. Our DIF-based ANN models are proven to predict each PI pair's drug resistance tendencies accurately, and these quantitative results support our central arguments.

METHODS AND MATERIAL

Dataset Description

Filtered genotype-phenotype data on the Stanford HIV drug resistance database was retrieved for protease inhibitors [2]. We have regulated this data set with respect to isolates and

inhibitors, and 498 protease mutations have been observed. For the HIV-1 PI: 1218 isolates for atazanavir (ATV), 678 isolates for darunavir (DRV), 1809 isolates for fosemprenavir (FPV), 1860 isolates for indinavir (IDV), 1562 isolates for lopinavir (LPV), 1907 isolates for nelfinavir (NFV), 1861 isolates for saquinavir (SQV) and 908 isolates for tipranavir (TPV) have been analyzed for PI susceptibility. In the dataset, 436, 336, 480, 483, 472, 486, 489 and 409 different mutations have been observed for ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV, respectively.

Representation of Isolates

498 unique mutations have been observed in the complete dataset for eight protease inhibitors. To represent the isolates that occurred in the dataset, the binary barcoding technique was applied here, as also used in several studies of modelling genotype-phenotype data for various HIV-1 inhibitors [3]. Thus, 498-dimensional vector of binary entries with 0s and 1s that uniquely represent any existing isolates is considered. Assume that the 498 unique mutations produce the set $X = \{x_1, x_2, \dots, x_n\}$ where x_i is a mutation pattern that occurred in the dataset. Any isolate can be obtained from any combination of these mutations and the isolate j can be defined as $I_j = \{a_1, a_2, \dots, a_n\}$ with

$$a_k = \begin{cases} 1, & if\ x_k \in I_j \\ 0, & otherwise. \end{cases}$$

In this way, each isolate can be transformed into a unique 498-dimesional input vector used in the machine learning part.

Representation of Inhibitors

To construct a drug-isolate-fold change model for the HIV-1 protease inhibitors, the molecular representations of the inhibitors have been built with binary Morgan fingerprints. The Morgan fingerprints provide an effective way of the vector representations of molecules and are widely used in machine learning models [4]. The RDKit environment of the Python program has been used to convert the smile representations of ATV, DRV, FPV, IDV, LPV, NFV, SQV and TPV inhibitors to a binary 512-bit vector representation. 234 out of 512 bits have been seen to provide unique characteristics for 8 PI. Thus, the molecular representation of each PI needs 234-dimensional vectors.

Artificial Neural Network Model for Regression

An ANN model has been constructed with isolate-inhibitor inputs and fold change outputs with Machine Learning and Deep Learning toolbox of the MATLAB program. Since isolates and inhibitors are uniquely represented by 498- and 234- dimensional vectors, the ANN model has 732-dimensional input. The ANN architecture includes 732-dimensional input, five hidden layer neurons and one output neuron with hyperbolic tangent-sigmoid and linear activation function. Logarithms of fold-change values in the dataset are taken as output variables of the neural network models. In the training process, the scaled conjugate gradient algorithm with MATLAB built-in function "trainscg" is utilized over GPU [5].

Ensemble Processing

Since we have only eight inhibitors, measuring the molecular learning capacity of our ANN model is crucial. In this way, an ensemble learning procedure is used to improve the molecular learning performance of the model. For each PI, the 100×50 model has been trained with the data of the remaining seven inhibitors. From every 50 models, a model is chosen that yields the minimum mean square error for the interior test set of the corresponding PI data. Thus, 100 optimal models are obtained, and the final model is calculated as the average of these models.

RESULTS

Regression performance of molecular learning models

Eight feed-forward neural network models are constructed with drug-isolate-fold- change (DIF) data by excluding one of the drugs from training in each case. The excluded results are predicted by the ANN model trained with the remaining seven DIF data. The regression performances of each model are illustrated in Figure 1 with corresponding R^2 values (square of the linear correlation coefficient). The best and worst results are obtained by predicting the outcomes of the drugs LPV and TPV with $R^2 = 0.837$ and $R^2 = 0.393$. Similarly, predicting the fold-change results of the inhibitor TPV was observed to be the worst one in literature [23]. The mean value R^2 of all predictions is 0.732 and the 95% confidence interval is $[0.613, 0.850]$. The DIF based ANN model provides accurate estimations even if the test data consists of unseen drugs. This observation implies that our ANN models learn molecular information from the Morgan fingerprints accurately. The detailed performance results of our DIF based ANN models are presented in Table 1.

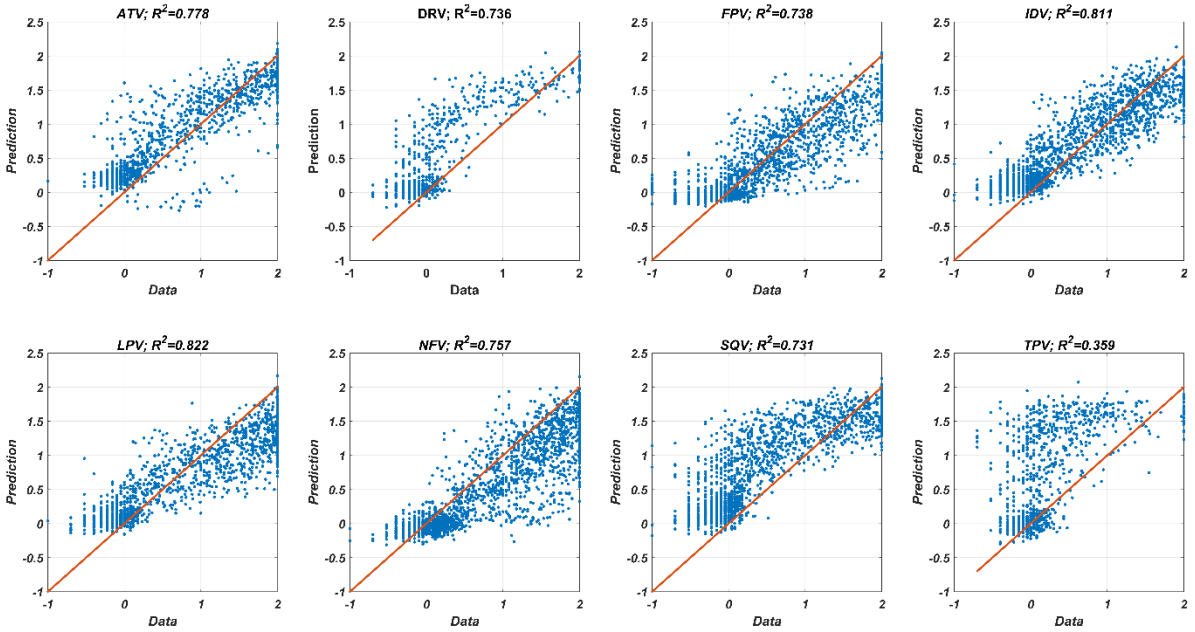

Figure 1 Data versus predicted values of fold changes obtained by DIF-based ANN models

DIF-based ANN regression models are constructed with the 7- training 1- testing methodology. For each figure, the fold-change results are estimated by an ANN model which is trained with the remaining data of the seven PI. The R^2 values correspond to the square of the linear correlation coefficient of the data and prediction.

Table 1. Mean square error (MSE) and R^2 values of the DIF-based ANN model[b].

ARVs[a]	R^2		MSE	
	Whole dataset	Test set	Whole dataset	Test set
ATV	0.865	0.778	0.087	0.166
DRV	0.857	0.736	0.092	0.227
FPV	0.849	0.738	0.097	0.160
IDV	0.861	0.811	0.090	0.131
LPV	0.852	0.822	0.096	0.188
NFV	0.845	0.757	0.101	0.215
SQV	0.833	0.731	0.109	0.283
TPV	0.821	0.359	0.116	0.560

[a] Abbreviations: ATV, atazanavir; DRV, darunavir; FPV, fosamprenavir; IDV, indinavir; LPV, lopinavir; NFV, nelfinavir; SQV, saquinavir; TPV, tipranavir.

[b] Drug-isolate-fold change models are constructed as a general neural network model taking drug fingerprints and mutation information as inputs. For each line, the corresponding drug has not been included in the training process. The test set performance of each model has been evaluated with respect to the excluded drugs. 100×50 simulations with random weights have been done, and 100 neural network models that yield minimum MSE for interior test set among 50 trials are obtained. The final neural network model is achieved by taking the mean of 100 models.

Prediction of drug resistance tendencies for each PI pair

The inhibition potential of each PI in the presence of various genotypes is known to be variable. Tendencies of the logarithmic fold change values for each PI pair provides valuable information about the resistance profiles of the inhibitors, as seen in Figure 2. Prediction of these tendencies by the DIF-based ANN models and the corresponding 2D correlation coefficients are presented in Figure 2 in a comparative way for each PI pair. For each PI, prediction is done with the ANN model trained by the data of the remaining seven inhibitors with an ensemble learning approach. This procedure shows the molecular learning capacity of our ANN models from the Morgan fingerprints. The minimum and maximum 2D correlation coefficients are 0.892 and 0.954 for TPV-DRV and LPV-DRV couples (95% CI [0.930, 0.938]). Thus, the current DIF-based ANN models can distinguish the inhibitory potentials of each PI pair.

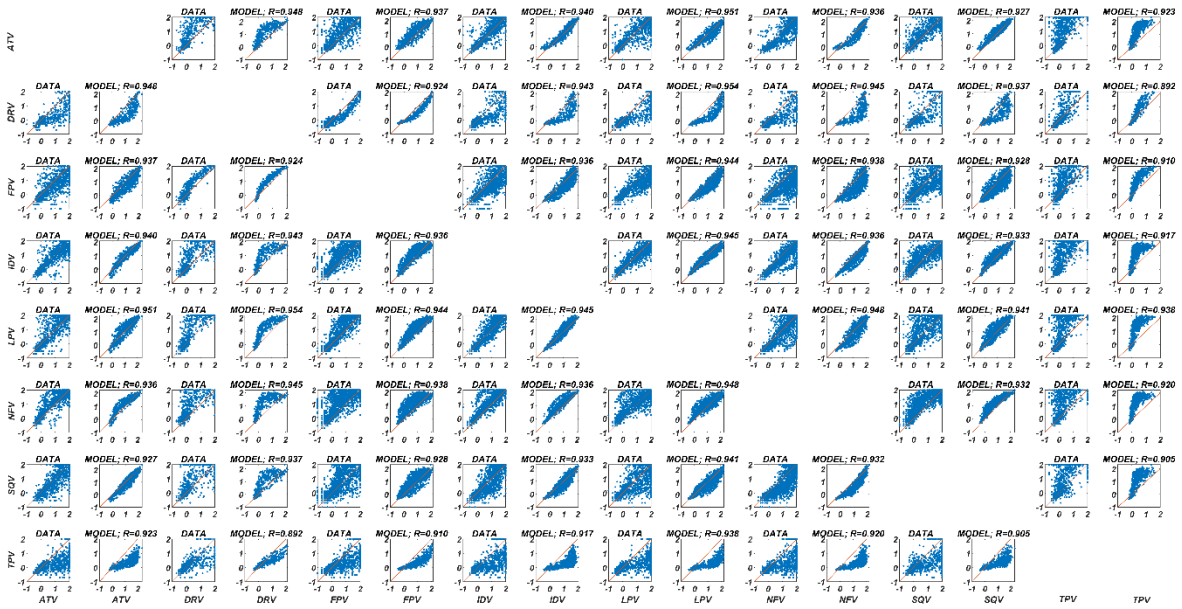

Figure 2 Prediction of the fold-change tendencies with the DIF-based ANN model for each PI pair

The common isolate data of each PI pair and the corresponding DIF-based ANN model predictions are illustrated with 2D correlation coefficients. For each PI, the prediction is constructed using the DIF-based ANN model, which is trained with the remaining seven PI data. The illustrations show the tendencies of the drug resistances for each PI pair for common genotypes.

Classification of PIs with respect to possible common isolates

Our DIF-based ANN models can distinguish the fold change values of each PI in the presence of any isolate. In this way, a classification problem measuring the relationship $log(Fold\ Change\ [A, Isolate]) > log(Fold\ Change\ [B, Isolate])$ has been constructed, where A and B are possible protease inhibitors. These relations take values 0 and 1 depending on the inhibitors and isolates. Thus, our ANN models trained with the data of seven inhibitors except that one specific inhibitor has been used to predict these binary values. The corresponding receiver operating characteristic (ROC) curves have been illustrated in Figure 3. Area under the ROC curve (AUC) values are included in the figure. The best and worst AUC values are obtained for the IDV-LPV and DRV-LPV pairs with 0.992 and 0.818 (95% CI: [0.950, 0.978]). In this context, the current DIF-based ANN models are seen to capture the binary relations between any PI pair with high approximation performance.

Performance metrics of the current ANN models for capturing binary relations of PI pairs are presented in Table 2. As indicated in the table, the DIF-based ANN models have a high rate of true prediction for each PI pair. The mean accuracy, sensitivity and specificity values are calculated as 0.954, 0.791 and 0.791 (95% CI [0.932, 0.952], [0.719, 0.863] and [0.719, 0.863]), respectively. The most conspicuous result here is that the neural network models can classify the inhibitors for resistance profiles, even if that model did not see the corresponding inhibitors in the training process.

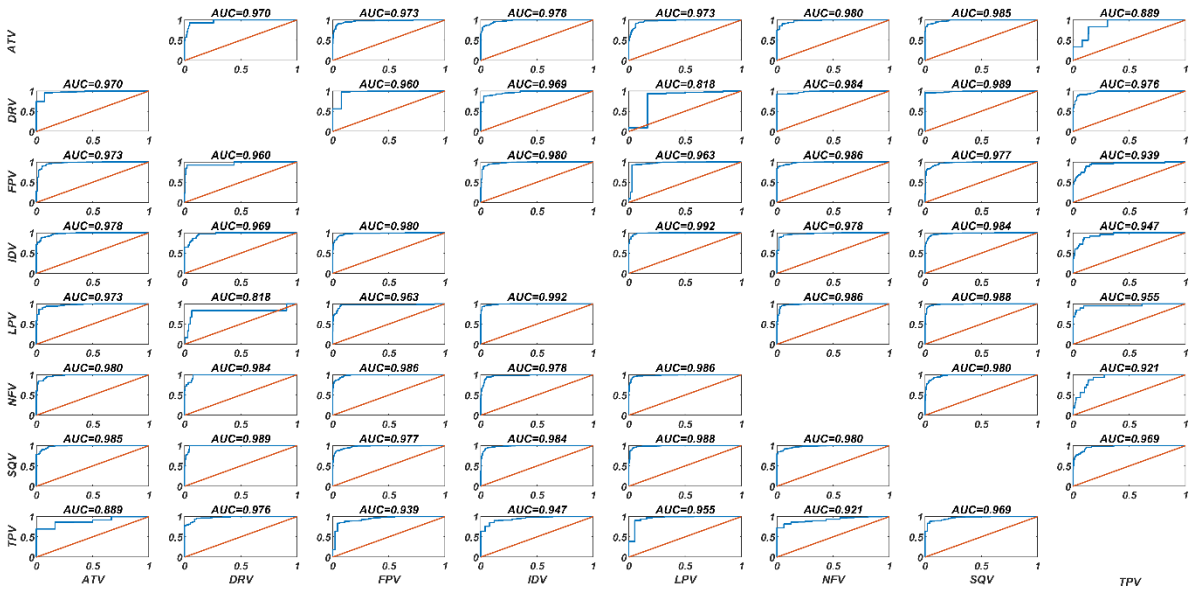

Figure 3 Classification performances of the DIF-based ANN models

DIF-based ANN classification models are constructed with the 7- training 1- testing methodology. For each PI, the classification of resistant and non-resistant isolates is estimated by an ANN model trained with the remaining data of the seven PI. The AUC values correspond to the area under the ROC curves, and the accuracy is evaluated with the true estimation rate.

Table 2. Accuracy values of the DIF-based ANN models for predicting the drug resistance tendencies for each couple of ARVs[a].

ARVs		ATV	DRV	FPV	IDV	LPV	NFV	SQV	TPV
ATV	Accuracy	-	0.970 (224/231)	0.932 (438/470)	0.923 (264/286)	0.913 (303/332)	0.948 (361/381)	0.896 (301/336)	0.983 (404/411)
	Sensitivity	-	0.500 (7/14)	0.772 (78/101)	0.850 (85/100)	0.978 (178/182)	0.986 (291/295)	0.720 (90/125)	0.000 (0/6)
	Specificity	-	1.000 (217/217)	0.976 (360/369)	0.962 (179/186)	0.833 (125/150)	0.814 (70/86)	1.000 (211/211)	0.998 (404/405)
DRV	Accuracy	0.970 (224/231)	-	0.968 (184/190)	0.917 (222/242)	0.960 (215/224)	0.976 (321/329)	0.936 (206/220)	0.897 (156/174)
	Sensitivity	1.000 (217/217)	-	0.972 (172/177)	0.966 (198/205)	0.982 (214/218)	1.000 (308/308)	0.972 (173/178)	0.714 (40/56)
	Specificity	0.500 (7/14)	-	0.923 (12/13)	0.649 (24/37)	0.167 (1/6)	0.619 (13/21)	0.786 (33/42)	0.983 (116/118)
FPV	Accuracy	0.932 (438/470)	0.968 (184/190)	-	0.936 (677/723)	0.952 (511/537)	0.964 (878/911)	0.930 (705/758)	0.932 (369/396)
	Sensitivity	0.976 (360/369)	0.923 (12/13)	-	0.993 (552/556)	0.996 (465/467)	0.996 (817/820)	0.975 (502/515)	0.511 (24/47)
	Specificity	0.772 (78/101)	0.972 (172/177)	-	0.749 (125/167)	0.657 (46/70)	0.670 (61/91)	0.835 (203/243)	0.989 (345/349)
IDV	Accuracy	0.923 (264/286)	0.917 (222/242)	0.936 (677/723)	-	0.952 (399/419)	0.952 (498/523)	0.929 (562/605)	0.957 (404/422)
	Sensitivity	0.962 (179/186)	0.649 (24/37)	0.749 (125/167)	-	0.989 (270/273)	0.994 (468/471)	0.874 (221/253)	0.280 (7/25)
	Specificity	0.850 (85/100)	0.966 (198/205)	0.993 (552/556)	-	0.884 (129/146)	0.577 (30/52)	0.969 (341/352)	1.000 (397/397)
LPV	Accuracy	0.913 (303/332)	0.960 (215/224)	0.952 (511/537)	0.952 (399/419)	-	0.944 (526/557)	0.929 (509/548)	0.982 (429/437)
	Sensitivity	0.833 (125/150)	0.167 (1/6)	0.657 (46/70)	0.884 (129/146)	-	0.979 (375/383)	0.836 (173/207)	0.632 (12/19)
	Specificity	0.978 (178/182)	0.982 (214/218)	0.996 (465/467)	0.989 (270/273)	-	0.868 (151/174)	0.985 (336/341)	0.998 (417/418)
NFV	Accuracy	0.948 (361/381)	0.976 (321/329)	0.964 (878/911)	0.952 (498/523)	0.944 (526/557)	-	0.935 (735/786)	0.966 (477/494)
	Sensitivity	0.814 (70/86)	0.619 (13/21)	0.670 (61/91)	0.577 (30/52)	0.868 (151/174)	-	0.451 (41/91)	0.188 (3/16)
	Specificity	0.986 (291/295)	1.000 (308/308)	0.996 (817/820)	0.994 (468/471)	0.979 (375/383)	-	0.999 (694/695)	0.992 (474/478)
SQV	Accuracy	0.896 (301/336)	0.936 (206/220)	0.930 (705/758)	0.929 (562/605)	0.929 (509/548)	0.935 (735/786)	-	0.898 (359/400)
	Sensitivity	1.000 (211/211)	0.786 (33/42)	0.835 (203/243)	0.969 (341/352)	0.985 (336/341)	0.999 (694/695)	-	0.146 (7/48)
	Specificity	0.720 (90/125)	0.972 (173/178)	0.975 (502/515)	0.874 (221/253)	0.836 (173/207)	0.451 (41/91)	-	1.000 (352/352)

TPV	Accuracy	0.983 (404/411)	0.897 (156/174)	0.932 (369/396)	0.957 (404/422)	0.982 (429/437)	0.966 (477/494)	0.898 (359/400)	-
	Sensitivity	0.998 (404/405)	0.983 (116/118)	0.989 (345/349)	1.000 (397/397)	0.998 (417/418)	0.992 (474/478)	1.000 (352/352)	-
	Specificity	0.000 (0/6)	0.714 (40/56)	0.511 (24/47)	0.280 (7/25)	0.632 (12/19)	0.188 (3/16)	0.146 (7/48)	-

[a] Accuracy, sensitivity and specificity values represent the rate of true predictions, true positive rate and true negative rate, respectively. The common genotype data is used for each PI pair by eliminating the observations satisfying $|\log(A) - \log(B)| \leq log2$ where A and B are the fold change values of drugs A and B for a specified genotype.

DISCUSSION

This study proposes a machine learning approach to predict fold-change values from the descriptors of HIV-1 protease inhibitors and isolates. The filtered PhenoSense assay datasets publicly available in the Stanford HIV drug resistance database have been utilized for training and testing machine learning models. Drug-isolate-fold change-based feed-forward artificial neural networks have been trained with seven of eight inhibitors, and the remaining one is used as test data, and this procedure has been called 7-in 1-out. In this context, the 7-in 1-out process yields an objective testing approach to identify the learning capacity of models from the descriptors of the inhibitors. Both inhibitors and isolates have been encoded through binary mappings that are observed to be computationally effective representations. The Morgan fingerprints have been used as the binary mappings of protease inhibitors due to their known advantages in molecular machine learning models [41-43]. An efficient ensemble process has been proposed and verified through various quantitative experiments to handle the overfitting trouble.

The most crucial contribution of this study is the construction of drug-isolate-fold change (DIF)-based ANN models rather than isolate-fold change (IF)-based models widely studied in the literature [29-40]. The IF models do not take the molecular fingerprints as an input, thus insensitive to the molecular structure. This study shows the possibility of achieving such a generalized model by feeding models with enough data of various PIs in the presence of isolates. With the utilization of a 7-in 1-out procedure throughout the study, the current DIF-based models have been seen to have the ability to predict the drug resistance profiles of the unseen inhibitors. Even if the number of available inhibitors in the Stanford HIV database is only eight, having many isolates for each inhibitor has contributed to the molecular learning

process, and acceptable predictions have been seen in the regression performance of remaining inhibitors.

An inevitable expectation from our DIF-based ANN models is the prediction of drug resistance tendencies for each PI pair. It has been shown here that our generalized models can predict the resistance tendencies with high 2D correlation scores. By defining classification problems from the tendency relations of each PI pair, the DIF-based models have provided satisfactory accuracy, sensitivity, and specificity values. Our all-quantitative observations have shown that the DIF-based ANN model takes valuable information from the Morgan fingerprints to predict the fold change values of hidden inhibitors.

This study provides a new perspective on this research area by including inhibitor descriptors on the input side of machine learning models, rather than creating so many individual models for each inhibitor. The most conspicuous limitation of the current study is having a limited number of protease inhibitors with enough genotype-phenotype data. Nevertheless, our positive results have proven to shed light on the construction of more general drug-isolate-fold change-based machine learning models by adding genotype-phenotype data of novel protease inhibitors. Additionally, feeding the DIF-based models with the data of various traditional and nontraditional inhibitors may lead to a unified model for predicting drug resistance tendencies for any PI pair in the presence of known genotypes.

The drug development process for evolvable diseases, such as HIV, bacterial infections, and cancer should be fundamentally different from diseases such as blood-pressure regulators. A drug needs to be effective and stay effective through the test of evolution. Predicting resistance potentials for drugs is becoming a necessity. Luckily, the experiments can measure fold-change values for many genotypes at once by sequencing. Our model aims to make sense of such data.

CONCLUSION

This study has revealed the advantages of producing DIF-based models to predict drug resistance profiles. Instead of IF-based models, the current approach has enabled us to investigate a new model that can predict the drug resistance tendencies of PI pairs. Even if the number of available PIs is only eight, the test results with a 7-in 1-out procedure show that the DIF-based model takes significant information from inhibitor descriptors and leads to satisfactory regression performance. Therefore, after completing this study, it is noted on the research agenda to train ANN models with more inhibitors by expanding the existing dataset.

In this context, it will be possible to track the drug resistance profiles of any novel protease inhibitor and it is strongly believed that these valuable predictions can be of great help to clinicians.

AUTHOR INFORMATION

Corresponding Author

Seyfullah Enes Kotil- Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey, orcid.org/0000-0002-9588-3947

Authors

Huseyin Tunc- Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey, orcid.org/0000-0001-6450-5380

Murat Sari- Department of Mathematics, Faculty of Science and Art, Yildiz Technical University, Turkey, orcid.org/0000-0003-0508-2917

Serdar Durdagi- Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey, orcid.org/0000-0002-0426-0905; Email: serdar.durdagi@med.bau.edu.tr

Data and Software Availability

All data and necessary codes are deposited to:

https://github.com/tnchsyn/hivdrugisolatefoldchange_model

ACKNOWLEDGMENTS

This work was supported by TUBITAK, 2232 - International Fellowship for Outstanding Researchers, Project number 118C244. All the results are in sole responsibility of the authors.

References

1. Sharp, P.M.; Hahn, B.H. Origins of HIV and the AIDS Pandemic. Cold Spring Harbor Perspectives in Medicine, 2011, 1, 006841.
2. Das, K.; Arnold, E. HIV-1 Reverse Transcriptase and Antiviral Drug Resistance (Part 1 of 2). Current Opinion in Virology, 2013, 3(2), 111–118.
3. Lu, D.Y.; Wu, H.Y.; Yarla, N.S. et al. HAART in HIV/AIDS Treatments: Future Trends. Infectious Disorders - Drug Targets, 2018, 18(1), 15-22.

4. Jespersen, N.A.; Axelsen, F.; Dollerup, J. et al. The burden of non-communicable diseases and mortality inpeople living with HIV (PLHIV) in the pre-, early- andlate-HAART era. HIV Medicine, 2021.
5. Palmisano, L.; Vella, S. A brief history of antiretroviral therapy of HIV infection: success and challenges. Annali dell'Istituto Superiore di Sanità 2011, 47(1), 44-48.
6. World Heath Organization. Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report. https://apps.who.int/iris/handle/10665/44787
7. Günthard, H. F.; Calvez, V.; Paredes, R. et al. Human Immunodeficiency Virus Drug Resistance: 2018 Recommendations of the International Antiviral Society–USA Panel. Clinical Infectious Diseases, 2019, 68(2), 177–187.
8. Kuritzkes, D. R. Drug resistance in HIV-1. Curr Opin Virol. 2011, 1(6), 582-589.
9. Oroz, M.; Begovac, J.; Planinić, A. et al. Analysis of HIV-1 diversity, primary drug resistance and transmission networks in Croatia. Sci Rep 2019, 9, 17307.
10. Lagnese, M.; Daar, E. S. Antiretroviral regimens for treatment-experienced patients with HIV-1 infection. Expert Opinion on Pharmacotherapy 2008, 9:5, 687-700.
11. Cihlar, T.; He, G.X.; Liu, X. et al. Suppression of HIV-1 protease inhibitor resistance by phosphonate-mediated solvent anchoring. J. Mol. Biol. 2006, 363, 635–647.
12. Stranix, B.R.; Sauve, G.; Bouzide, A.; Cote, A.; Sevigny, G.; Yelle, J. Lysine sulfonamides as novel HIV-protease inhibitors: Optimization of the Nepsilon-acyl-phenyl spacer. Bioorg. Med. Chem. Lett. 2003, 13, 4289–4292.
13. Nakatani, S.; Hidaka, K.; Ami E. et al. Combination of non-natural D-amino acid derivatives and allophenylnorstatine-dimethylthioproline scaffold in HIV protease inhibitors have high efficacy in mutant HIV, J Med Chem 2008, 51, 2992–3004.
14. Koh, Y.; Das, D.; Leschenko, S. et al. GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 in vitro. Antimicrob Agents Chemother 2009, 53(3), 997-1006.
15. Koh, Y.; Nakata, H.; Maeda K. et al. Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. Antimicrob Agents Chemother. 2003, 47(10), 3123-9.

16. Zhang, S.; Kaplan, A.H.; Tropsha, A. HIV-1 protease function and structure studies with the simplicial neighborhood analysis of protein packing method. Proteins 2008, 73(3), 742–753.
17. World Heath Organization. Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV, https://www.who.int/publications/i/item/WHO-CDS-HIV-18.51.
18. Rosenbloom, D.I.S.; Hill, A.L.; Rabi, S.A. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. Nature Medicine 2012, 18(9), 1378-1386.
19. Jilek, B.L.; Zarr, M.; Sampah, M.E. A quantitative basis for antiretroviral therapy for HIV-1 infection. Nature Medicine 2011, 18(3), 446-452.
20. Xing, H.; Ruan, Y.; Li, J. et al. HIV Drug Resistance and Its Impact on Antiretroviral Therapy in Chinese HIV-Infected Patients. PLoS ONE 2013, 8(2), e54917.
21. Lima, V. D.; Gill, V. S.; Yip, B. et al. Increased Resilience to the Development of Drug Resistance with Modern Boosted Protease Inhibitor-Based Highly Active Antiretroviral Therapy. The Journal of Infectious Diseases 2008, 198(1), 51–58.
22. Wei, Y.; Li, J.; Chen, Z. et al. Multistage virtual screening and identification of novel HIV-1 protease inhibitors by integrating SVM, shape, pharmacophore and docking methods. Eur. J. Med. Chem. 2015, 101, 409–418.
23. Yu, X.; Weber, I.T.; Harrison, R.W. Prediction of HIV drug resistance from genotype with encoded three-dimensional protein structure. BMC Genomics 2014, 15, 1-13.
24. Hosseini, A.; Alibés, A.; Noguera-Julian, M. Computational Prediction of HIV-1 Resistance to Protease Inhibitors. Journal of Chemical Information and Modeling 2016, 56(5), 915–923.
25. Talbot, A.; Grant, P.; Taylor, J. et al. Predicting tipranavir and darunavir resistance using genotypic, phenotypic, and virtual phenotypic resistance patterns: an independent cohort analysis of clinical isolates highly resistant to all other protease inhibitors. Antimicrobial Agents Chemotherapy 2010, 54, 2473-2479.
26. Obermeier, M.; Pironti, A.; Berg, T. et al. HIVGRADE: a publicly available, rules-based drug resistance interpretation algorithm integrating bioinformatic knowledge. Intervirology 2012, 55, 102-107.

27. Van Laethem, K.; De Luca, A.; Antinori, A. et al. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. Antiviral Therapy 2002, 7, 123–129.
28. Meynard, J.L.; Vray, M.; Morand-Joubert L. et al. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. AIDS 2002, 16, 727–736.
29. Amamuddy, O.S.; Bishop, N.T.; Bishop, Ö.T. Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks. BMC Bioinformatics 2017, 18, 369-376.
30. Amamuddy, O.S.; Bishop, N.T.; Bishop, Ö.T. Characterizing early drug resistance-related events using geometric ensembles from HIV protease Dynamics. Scientific Reports 2018, 8, 17938.
31. Wang, D.; Larder, B. Enhanced Prediction of Lopinavir Resistance from Genotype by Use of Artificial Neural Networks. The Journal of Infectious Diseases 2003, 88(5), 653–660.
32. Drăghici, S.; Potter, R.R. Predicting HIV drug resistance with neural networks. Bioinformatics 2002, 19(1), 98-107.
33. Kjaer, J.; Høj, L.; Fox, Z.; Lundgren, J. Prediction of phenotypic susceptibility to antiretroviral drugs using physiochemical properties of the primary enzymatic structure combined with artificial neural networks. HIV Medicine 2008, 9, 642-652.
34. Steiner, M.C.; Gibson, K.M.; Crandall, K.A. Drug Resistance Prediction Using Deep Learning Techniques on HIV-1 Sequence Data. Viruses 2020, 12, 560.
35. Wang, D.; Larder, B.; Revell, A. et al. A Comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy. Artificial Intelligence in Medicine 2009, 47, 63-74.
36. Shen, C.H.; Yu, X.; Harrison, R.W.; Weber, I.T. Automated prediction of HIV drug resistance from genotype data. BMC Bioinformatics 2016, 17, 278-285.
37. Shah, D.; Freas, C.; Weber, I.T.; Harrison, R.W. Evolution of drug resistance in HIV protease. BMC Bioinformatics 2020, 21, 497-512.
38. Tarasova, O.; Biziukova, N.; Kireev et al. A Computational Approach for the Prediction of Treatment History and the Effectiveness or Failure of Antiretroviral Therapy. International Journal of Molecular Sciences 2020, 21(3), 748.

39. Tarasova, O; Biziukova, N; Filimonov, D.; Poroikov, V. A Computational Approach for the Prediction of HIV Resistance Based on Amino Acid and Nucleotide Descriptors. Molecules 2018, 23(11), 2751.
40. Ota, R.; So, K.; Tsuda, M.; Higuchi, Y; Yamashita, F. Prediction of HIV drug resistance based on the 3D protein structure: Proposal of molecular field mapping. PLoS ONE 2021, 16(8), e0255693.
41. Cai, Q.; Yuan, R.; He, J. et al. Predicting HIV drug resistance using weighted machine learning method at target protein sequence-level. Molecular Diversity 2021, 25, 1541–1551.
42. Beerenwinkel, N.; Däumer, M.; Oette, M. et al. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. Nucleic Acids Research 2003, 200331, 3850-3855.
43. Beerenwinkel, N.; Schmidt, B.; Walter, H. et al. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. Proceedings of the National Academy of Sciences 2002, 99, 8271-8276.
44. Pawar, S.D.; Freas, C.; Weber, I.T.; Harrison, R.W. Analysis of drug resistance in HIV protease. BMC Bioinformatics 2018, 19, 362-368.
45. Rhee, S.Y.; Taylor, J.; Fessel, W.J. HIV-1 Protease Mutations and Protease Inhibitor Cross-Resistance. Antimicrobial Agents and Chemotherapy 2010, 54(10), 4253–4261.