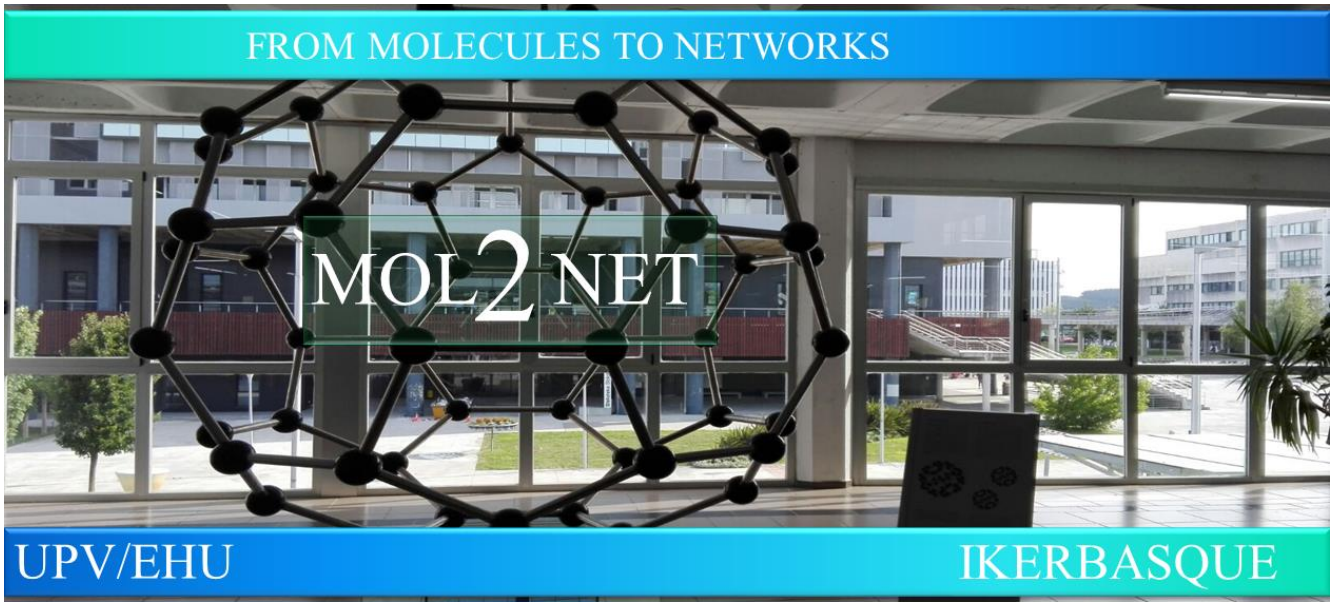
MDPI sciforum

MOL2NET'22, Conference on Molecular, Biomedical & Computational Sciences and Engineering, 8th ed.



Predicting HIV drug resistance using machine learning

Maider Baltasar Marchueta

Departamento de química orgânica e inorgânica, Universidad del País Vasco

Abstract.

Human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) is one of the major burdens of disease in developing countries, and the standard-of-care treatment includes prescribing antiretroviral drugs. Although 23 different drugs have been available, the treatment of AIDS remains challenging because the virus mutates very quickly which can lead to drug resistance. Predicting drug resistance before treatment is crucial for individual treatments. Taking that into account, different investigations undertaken with machine learning will be discussed.

Human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS) is one of the major burdens of disease in developing countries, and the standard-of-care treatment includes prescribing antiretroviral drugs. Although 23 different drugs have been available, the treatment of AIDS remains challenging because the virus mutates very quickly which can lead to drug resistance. Antiretroviral drug resistance is inevitable due to selective pressure associated with the high mutation rate of HIV. Therefore, predicting drug resistance before treatment is crucial for individual treatments. Determining antiretroviral resistance can be done by phenotypic laboratory tests or by computer-based interpretation algorithms. Besides, computer-based algorithms have been shown to have many advantages over laboratory tests. Taking that into account, different investigations undertaken with machine learning will be discussed.

Firstly, Cai *et al.*[1] analyzed 21-drug resistance caused by mutated residues using machine learning (ML) methods. They used seven physicochemical properties were used to transform target sequences into numeric vectors. Then, principal component analysis (PCA) method was adopted to reduce the feature dimensionality. Radial basis function-based (RBF) support vector machine (SVM) method gave a comparative performance with random forest model. Moreover, they added the weight information to RBF-based SVM method by four different weight evaluation methods of RF, eXtreme Gradient Boosting (XGB), CfsSubsetEval and ReliefFAttributeEval, respectively. Their results showed that the RF-weighted RBF-based SVM yield the superior performance. Indeed, 13 out of 21 drug models provide the correlation coefficients over 0.8 and 3 of them are higher than 0.9. After all, position-specific importance analysis indicated that most of the mutation residues with high RF weight scores were proved to be closely related with drug resistance.

In other study, Singh *et al.* [2] performed a study to improve the prediction of the ANRS (Agence Nationale de Recherches sur le SIDA) gold standard in predicting HIV drug resistance. To accomplish their goal, they obtained genome sequence and HIV drug resistance measures from the Stanford HIV database. Feature selection was used to determine the most important mutations associated with resistance prediction. Then, these mutations were added to the ANRS rules, and the difference in the prediction ability was measured. This study uncovered important mutations that were not associated with the original ANRS rules. On average, the ANRS algorithm was improved by 79% ± 6.6%. The positive predictive value improved by 28%, and the negative predicative value improved by 10%.

Last but not least, Blassel *et al.* [3] did a research on high-quality sequence data and machine learning methods to study HIV potential drug resistance mutations (DRM). They trained classifiers to discriminate between Reverse Transcriptase Inhibitor (RTI)-experienced and RTI-naive samples on a large HIV-1 reverse transcriptase (RT) sequence dataset from the UK, using all observed mutations as binary representation features. Important representation features for each classifier were then extracted as potential DRMs. When removing features corresponding to known DRMs, their classifiers retained some prediction accuracy, and six new mutations significantly associated with resistance were identified. These results likely indicate that all mutations directly conferring resistance have been found, and that our newly discovered DRMs are accessory or compensatory mutations. Moreover, apart from the accessory nature of the relationships found, any significant signal of further were not found, more subtle epistasis combining several mutations which individually do not seem to confer any resistance.

To conclude, it is visible that different methods are helpful to improve the prediction of drug resistance. The methos that Cai *et al.* used can be a supplementary tool for predicting HIV drug resistance for newly discovered mutations. Besides, Sigh *et al.* improved the prediction capacity of ANRS gold standars. And finally, Bessel *et al.* tried to study potential HIV DRMs, however, the newly discovered DRMs ought to be accessory or compensatory mutations.

References

1. Cai Q, Yuan R, He J, Li M, Guo Y. Predicting HIV drug resistance using weighted machine learning method at target protein sequence-level. Mol Divers. 2021 Aug;25(3):1541-1551. doi: 10.1007/s11030-021-10262-y. Epub 2021 Jul 9. PMID: 34241771.
2. Singh Y. Machine Learning to Improve the Effectiveness of ANRS in Predicting HIV Drug Resistance. Healthc Inform Res. 2017 Oct;23(4):271-276. doi: 10.4258/hir.2017.23.4.271. Epub 2017 Oct 31. PMID: 29181236; PMCID: PMC5688026.
3. Blassel L, Tostevin A, Villabona-Arenas CJ, Peeters M, Hué S, Gascuel O; UK HIV Drug Resistance Database. Using machine learning and big data to explore the drug resistance landscape in HIV. PLoS Comput Biol. 2021 Aug 26;17(8):e1008873. doi: 10.1371/journal.pcbi.1008873. PMID: 34437532; PMCID: PMC8425536.