Mini Review

Open Access

CrossMark

Machine learning approaches in drug development of HIV/AIDS

Abstract

Due to the complexity of HIV/AIDS cutting edge machine learning technologies are used for drug delivery and development. In this review drug delivery methods are discussed with machine learning techniques. Combination of both these computational methods will give new hope to enhance the life of HIV infected persons. As these methods are time consuming and easy to interpret than wet lab techniques.

Keywords: drug, machine learning, computational method, wet-lab

Volume 3 Issue 1 - 2018

Anubha Dubey
Independent researcher and analyst Bioinformatics, India

Correspondence: Anubha Dubey, Independent researcher and analyst Bioinformatics, Gayatri Nagar, Katni, 483501 MP, India, Email anubhadubey@rediffmail.com

Received: May 01, 2017 | **Published:** February 01, 2018

Introduction

Globally, approximately 35million people are infected with Human Immunodeficiency Virus (HIV), the virus that causes acquired immunodeficiency syndrome (AIDS). The currently available medicines and vaccines in the development pipeline include:

A. A first class medicine intended to prevent HIV from breaking through the cell membrane.

A cell therapy that modifies a patient's own cells in an attempt to make them resistant to HIV. A therapeutic vaccine designed to induce responses from T-cells that play a role in immune protection against viral infections. Biopharmaceutical research companies are investigating new ways to treat and prevent HIV infection. Potential therapies being developed for HIV infection include:

i. Attachment inhibitor: the attachment inhibitor inhibits the attaching of virus to new cells. It means this inhibitor blocks the interaction between gp120 and the cell receptors.

ii. Gene modification: CCR5 is a co-receptor on surface of cells that allows HIV to enter and infect T-cells. These cells from patients are extracted, modified and then reinserted into the patients. This therapy provides with the population of cells that can fight HIV and other opportunistic infections of HIV patients.

iii. Involving T-cell responses: Another therapeutic vaccine in development is designed to include CD4+T cell responses in HIV infected people. CD4+T cells play a important role in immune protection against viral infections. Deficits in CD4+t cells are associated with virus reactivation, vulnerability to opportunistic infections and poor vaccine efficacy. For HIV/AIDS, the introduction of novel therapeutics and continuous research into their best use in patients have revealed the result of the development and introduction of multiple drugs (used in combinations i.e. anti retroviral drugs) are proven good in health of HIV infected patients. As there is currently no publicly available vaccine or cure for HIV or AIDS.[1] Some examples include a vaginal gel containing tenofovir, a reverse transcriptase inhibitor is developed showing good results in clinical trials.[2]

HIV infection consists of highly active antiretroviral therapy, or HAART[3] or ART, these classes are consisting two nucleoside analogue reverse transcriptase inhibitors (NARTIs or NRTIs) plus either a protease inhibitor or a non-nucleoside reverse transcriptase inhibitor (NNRTI). Abacavir–nucleoside analog reverse transcriptase inhibitors (NARTIs or NRTIs) shows good results. According to one study the average life expectancy of an HIV infected individual is 32years from the time of infection if treatment is started when the CD4 count is 350/μL.[4] If CD4counts is less than 500ART is also recommended to enhance the life expectancy of HIV infected individuals. Figure 1 describes how ARTS hindering HIV life cycle in humans and shows their responses. Anti-retroviral drugs are expensive, so there is a need to develop vaccines or drugs that enhance the life of HIV infected persons without side effects. In view of this, computational methods along with machine learning brings hope to develop such treatment which cost less and reachable to common persons.

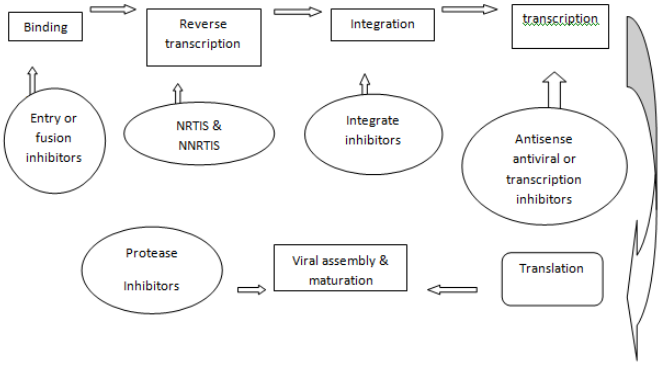

Figure 1 Schematic representation of ARTS hindering HIV life cycle.

Methods and discussion

Machine learning techniques are cutting edge technologies[5] have been referred to the development of algorithms that improve their performance in pattern recognition, classification, regression and prediction based on the models derived from existing data. It is closely related to data mining as pattern recognition is one of the most important areas of research in both. Algorithms like classification have been frequently used to identify active and inactive compounds while regression approaches are applied to the training and testing the continuous data (prediction also used). Then ensemble algorithms i.e. bagging, boosting, etc make accurate and fast decisions. Although cross-validation also plays an important role in achieving the result. In drug discovery and development like target identification machine

learning methods have been widely used in quantitative structure activity relationship, ligand based virtual screening, in-silico ADMET (Adsorption, Distribution, Metabolism, Excretion and Toxicity) studies. Modern QSAR are characterized by the use of multiple descriptors of chemical structures combined with both linear and non-linear optimization techniques and a strong emphasis is done on model validation. These methods include structural, physicochemical properties of compounds in which counts of atom, bond, electrostatic and thermodynamic properties are important. Modeller, chem. sketch, DRAGON, MOE, VMD, AUTODOCK etc are the computational based cheminformatics software that is widely used in target identification, hit discovery, etc. Machine learning methods are widely used to identify best suited model obtained by these methods. QSAR models are widely used in virtual screening for hit discovery.[6] Hence it shows structure activity relationship to find the potential target for drug delivery and drug discovery. Virtual screening is major target area within the cheminformatics spectrum that has used machine learning techniques. Virtual screening (VS) is the application of computational tools to search large databases for new leads with higher probability of strong binding affinity to the target protein. VS methods can be classified into structure based (SBVS) and ligand based (LBVS) approaches depending on the amount of structural and bioactivity data available. If the 3D structure of the receptor is known, a SBVS method is used in high throughput docking,[7] but if information of receptor is scant, LBVS methods are commonly recommended. Thus taken together there is a broad spectrum of applications for machine learning methods in computer aided drug discovery. That makes it attractive to select approaches and highlight their applications. HIV has a special character of recombination and sudden & rapid mutation so it is difficult for biopharmaceuticals to formulate medicines for curing AIDS. Still many medicines are in practise and shows better results. These methods need clinical trials but drug designing with combination of computational methods with molecular basis of HIV proves better results. Here are some of the machines learning approaches that are widely used with computational methods:

Support vector machines

It is developed by Vapnik and co-workers[8,9] are supervised machine learning algorithms for facilitating compound classification, binary classification (linearly separable). Once linearly separable two classes of compound are separated by hyper plane as shown in Figure 2. There are many hyper planes developed and SVM chooses the hyper plane that maximizes the margin between the two classes as it was assumed that larger the margin, lower would be the error of the classifier when dealing with unknown data. These hyper planes are called support hyper planes (dash lines as shown in figure2) and the data points lie on these hyper planes are called support vectors (blue and red dots). In the case of non-separable classes, soft-margin hyper plane is applicable, which maximizes the margin while keeping the number of misclassified samples minimal. When high dimensionality feature space is considered SVM-kernels are used. There are four kernels which are basically used: linear, polynomial, sigmoid, and radial basis (RBF). The first three kernels are global and RBF is a local kernel. Extensive work has shown that RBF-based SVM outperforms best then other three kernels and hence used widely. Basically SVM are used for binary property or activity prediction i.e. to distinguish between drugs and non-drugs[10,11] or between compounds that have or do not have specific activity[11–13] synthetic accessibility is also an important criteria or aqueous solubility.[14]

Decision tree (DT)

A DT is commonly as a tree with the root at the top and the leaves at the bottom as shown in Figure 2. Starting from the root, tree splits from the single trunk into two or more branches. Each branch itself split into two or more branches. This process continuous until a leaf is reached which could not further split. The split of the branch is referred as an internal node of the tree (root and leaves are also called node). Each leaf node is assigned with a target property whereas a non leaf node is assigned molecular property. This method basically used in designing combinatorial libraries, predicting drug-likeliness, predicting specific biological activities. Classification of compounds into drugs and non-drugs,[15] ADME-TOX properties[16–19] and metabolic stability.[20] These models are simple to understand, interpret and validate. A moderate data size is recommended to avoid over fitting.

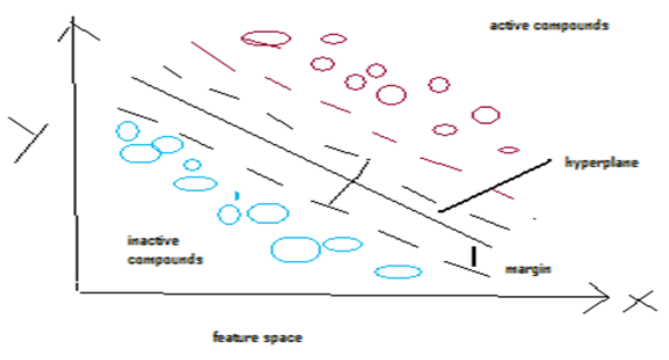

Figure 2 Hyper plane separation of two objects.

Ensemble methods

Cross-validation is necessary before choosing any of the classification method. Methods like random forest, bagging, boosting proved better results.

Naive base classifier

It is basically based on Byes theorem, which gives a mathematical framework for describing the probability of an event that might have been result of two or more causes:[21]

$$p(a/b) = \frac{p(b/a)p(a)}{p(b)} \qquad \text{Equation (1)}$$

This equation describes the probability p for state *a* existing for a given state *b*. The importance of Bayesian theorem is that probabilities of occurring new things depends upon existing knowledge. This is frequently used in chemo informatics both generally for predicting biological rather than physicochemical properties, prediction of toxicity of compound, protein target, bio active classification for drug like molecules[22,23] (v) k-Nearest neighbours: it is one of the simplest algorithms. A molecule is classified by a majority vote of its neighbours with the molecule being assigned to the class most common among its nearest neighbours. The k-NN algorithm is sensitive to the local structure of the data. Therefore it is ideal for calculating properties with strong localities as in predicting protein function.

Artificial neural networks

This method was developed to model brain structure and functioning. As neurons have certain topology they connected to each other forming neural networks. In ANN it is called feed forward network which includes multiplayer perceptrons (MLP), radial basis function (RBF) networks and Kohonen's self organizing maps (Kohonen's SOM).[24] It is mainly applied in compound classification, QSAR studies, primary VS of compounds, identification of potential

Citation: Dubey A. Machine learning approaches in drug development of HIV/AIDS. *Int J Mol Biol Open Access.* 2018;3(1):23–25.
DOI: 10.15406/ijmboa.2018.03.00044

drug target sites and localization of structural and functional features of proteins[25–27] pattern identification and others.

Concluding remarks & future directions

LVS techniques are widely used for hit identification. The methodological spectrums of these techniques are wide and time consuming, as they are simple to implement and interpret. Developing drug for HIV is difficult task as high mutation rate of virus, still biopharmaceutical companies regularly work on this area. Machine learning cutting edge technologies would provide sound effect in drug development as comparative analysis is done in seconds. Studies shows SVM prediction is good then others. These all are based on trials of algorithms according to the availability of data. In near future vaccines are also developed based on MLT. It is said that in near future more focus on the development of machine learning algorithms that reflect domain knowledge. Clearly much work needs to be done for drug delivery and development of HIV. It is truly said worlds depend on hope.

Acknowledgements

None.

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Robb ML. Failure of the Merck HIV vaccine: an uncertain step forward. *Lancet*. 2008;372(9653):1857–1858.
2. Karim QA, Abdool Karim SS, Frohlich JA, et al. Effectiveness and Safety of Tenofovir Gel, an Antiretroviral Microbicide, for the Prevention of HIV Infection in Women. *Science*. 2010;329(5996):1168–1174.
3. Department of Health and Human Services. A Pocket Guide to Adult HIV/AIDS Treatment. 2005.
4. Schackman BR, Gebo KA, Walensky RP, et al. The lifetime cost of current HIV care in the United States. *Med Care*. 2006;44(11):990–997.
5. Anubha Dubey. Applications of Machine Learning: Cutting Edge Technology in HIV Diagnosis, Treatment & Further research. *Computational Molecular Biology*. 2016;6(3):1–6.
6. Lavecchia A, Di Giovanni. Virtual Screening strategies in drug discovery: a critical review. *Curr med chem*. 2013;20(23):2839–2860.
7. Patel J, Chaudhari C. Introduction of artificial neural networks in QSAR studies. *ALTEX*. 2005;22:271.
8. Vapnik VN, Vladimir. The nature of statistical learning theory. New York: Springer-Verlag; 2000. p. 314.
9. Vapnik VN. Statistical learning theory. USA: Wiley; 1998. p. 768.
10. Byvatov E, Fechner U, Sadowski J, et al. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J Chem Inf Comput Sci*. 2003;43(6):1882–1889.
11. Zernov VV, Konstantin V Balakin, Andrey A Ivaschenko, et al. Drug discovery using support vector machines. The case studies of drug likeliness, agrochemical-likeness, and enzyme inhibition predictions. *J Chem Inf Comput Sci*. 2003;43(6):2048–2056.
12. Warmuth MK, Liao J, Rätsch G, et al. Active learning with support vector machines in drug discovery process. *J Chem Inf Comput Sci*. 2003;43(2):667–673.
13. Jorison RN, Gilson MK. Virtual screening of molecular databases using a support vector machines. *J Chem Inf Model*. 2005;45(3):549–561.
14. Cheng T, Li Q, Wang Y, et al. Binary classification of aqueous solubility using support vector machines with reduction and recombination feature selection. *J Chem inf Model*. 2011;51(2):229–236.
15. Schneider N, Christine Jäckels, Claudia Andres, et al. Gradual in silico filtering for drug like substances. *J Chem Inf Model*. 2008;48(3):613–628.
16. Hou T, Wang J, Li Y. ADME evaluation in drug discovery .8. The prediction of human intestinal absorption by a support vector machines. *J Chem Inf Model*. 2007;47(6):2408–2415.
17. Deconinck E, Zhang MH, Coomans D, et al. Classification tree models for the prediction blood-brain barrier passage of drugs. *J Chem Inf Model*. 2006;46(3):1410–1419.
18. Gleeson MP, Nigel J Waters, Stuart W Paine, et al. In silico human and rat Vss quantitative structure - activity relationship models. *Journal of Medical Chemistry*. 2006;(6):1953–1963.
19. Lamanna C, Bellini M, Padova A, et al. straightforward recursive partitioning model for discarding insoluble compounds in the drug discovery process. *J Med Chem*. 2008;51(10):2891–2897.
20. Sakiyama Y, Yuki H, Moriya T, et al. Predicting human liver microsomal stability with machine learning techniques. *J mol Graph Model*. 2008;26(6):907–915.
21. Jensen FV. Bayesian Networks and Decision graphs. USA: Springer; 2001. p. 448.
22. Koutsoukas A, Robert Lowe, Yasaman Kalantar Motamedi, et al. *In Silico* target predictions: defining a benchmarking dataset and comparison of performance of the multiclass Naïve Bayes and Parzen-Rosenblatt window. *Journal of Chemical Information Model*. 2013;53(8):1957–1966.
23. Nigsch F, Bender A, Jenkins JL, et al. Ligand - target prediction using Winnow and naïve Bayesian algorithms and the implications of overall performance statistics. *J Chem Inf Model*. 2008;48(12):2313–2325.
24. Antonio Lavecchia. Machine Learning approaches in drug discovery: methods and applications. *Drug Discovery today*. 2015;20(3):318–331.
25. Patel JL, Goyal RK. Artificial neural netwoks and their applications in pharmaceutical research. *Pharmabuzz*. 2007;2:8–17.
26. Patel JL, Goyal RK. Applications of artificial neural networks in medical sciences. *Curr Clin Pharmacol*. 2007;2(3):217–226.
27. Soyguder S. Intelligent control based on wavelet decomposition and neural network for predicting of human trajectories with a novel vision based robotic. *Expert Systems with Applications*. 2011;38(11):13994–14000.