Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 03, No. 01, Jan-June 2023
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.32.53.60

Research Paper

Machine learning in drug discovery

Deepak B*
*UG Scholar, Department of Information Technology, SNS College of Technology, Coimbatore, Tamil Nadu, India.

Article Info

Article History:
Received: 28 January 2023
Revised: 15 April 2023
Accepted: 21 April 2023
Published: 07 May 2023

Keywords:
Drug
Biological Data
Algorithms
Drug Development
Machine Learning

ABSTRACT

A drug is a substance that when put into the body can change the way the body works and a person's mental state. Discovering the accurate drug plays a vital role in saving precious lives. In traditional drug creation, scientists identify a target in the body and test a large range of chemical compounds on it until they obtain the results. The process gets quicker and more effective with the role of machine learning techniques. This is done by using the huge amount of biological data, medical data, algorithms and statistical models available today. This automation of the drug development process is a key to the current issue of low productivity rate that pharmaceutical companies currently face and helps eliminate the side effects. Machine learning might be a useful tool to further enhance the drug development process.

Corresponding Author:
Deepak B
UG Scholar, Department of Information Technology, SNS College of Technology, Coimbatore, Tamil Nadu, India.
Email: deepaksmail2004@gmail.com

Copyright © 2023 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

This era has faced large medical problems such as illness and life- killing diseases etc... These will also continue in the upcoming era, human beings should adapt to it. This can be overcome by precautions and personalized medical treatment. Discovering an accurate drug for the researched morbidity is the major part. Till today the traditional way is used for the drug discovery with the concept of bioinformatics where the biological data is organized and analyzed until the personalized drug is discovered, here the man

power rules the discovery process. The major cons of the traditional way of discovery is that it requires a long duration, but this is possible quicker and more efficiently with the concept of machine learning.

Drug development is a complex and costly process with a high degree of uncertainty that a drug is actually discovered. [1] Emphasis on human data can improve the target identification and synthesizing to obtain the final drug which can reach the market. Machine Learning is used to solve these complex problems which can be used from target validation, identification to analysis of data in clinical trials. Machine Learning also promotes decision making, potential to speed up the process which can reduce failure rates in drug discovery and use advanced smart technologies. The major important processes like molecular structure prediction and testing are done completely with machine learning by training the models by biological database with the more specialized machine learning algorithm, various outcomes are able to be deduced with more personalized ways.

1.1 Compilation on Drug Discovery

The disease discovery rate is comparatively higher than the rate of drug discovery, hence some diseases become incurable. Every drug is discovered successfully to treat lives. The problems faced by the bioinformatic team are unnoticed when a successful and personalized drug is launched to the medical market. The problems faced behind this biological success and drug management over market are most commonly discussed below.

1.1.1 Time Complicity

Drug discovery takes decades. The treatment is being held until the personalized drug is discovered in the most traditional way. This time complexity leads to loss of lives. Even when a drug is made in short duration needs to be skipped some of the tests and processes can lead to side effects that can occur in eventual time.

1.1.2 Availability

Next to the discovery of a drug, the major part is its availability. The availability depends upon source resources. A single drug for a disease creates the situation of key resource shortage as entire individual bioinformatics teams go with the same resource. The low availability of discovered drugs creates scarcity, so it might not reach all of the needed lives.

1.1.3 Price

Price plays a major part, when the availability is reduced then demand is created that leads to raise the price. The price also depends on the source resources used for the production. As the effort increases in the discovery and the production of drugs also cause the rise in price of the personalized drug.

1.1.4 Communication

Communication medium between discoveries helps the process of successful discovery of personalized drugs to be completed quickly. Now the massive ego between the countries makes the communication broken and unable to share public biological data on servers and unusable of drugs patented by a bioinformatics team to be used by other teams.

1.1.5 Discovery

The process of discovery contains numerous problems where everything has to be done more accurately. The stages where to be performed carefully are discussed below.

1.1.6 Structure

In traditional drug discovery the prediction of drug molecular structure is harder. The drug discovery process starts with the stage of predicting the structure of the proteins, this becomes a very critical stage as it depends on productivity. The process gets stuck if it gets collapsed in the wrong prediction and manages it using a simple computational process with the major human role.

1.1.7 Test

The drug effect tests are done for a long decades in the traditional way and noticed over all categories like absorption, distribution, metabolization and excretion. The variety of tests of discovery may not get successes, when the compulsion of most dreadful diseases may bring up a chance to skip or make some of the tests liberal and manually reducing the test period of drugs may cause formation of unhealthy drugs and future problematic side effects on living beings.

1.1.8 Outcome

After coming with successful results from all of the tests that had been conducted on multiple varieties of beings and satisfying the organizations like world health organization, world medical association rules the drug is introduced to the market. All of the drugs that get discovered by every bioinformatic team worldwide may not successfully reach the medical market that can be used for treatment.

1.1.9 Usage

Nowadays everyone depends on drugs, everyone takes the appropriate drug for the medical illness. Even if it reaches the level of addiction, people intake drugs whenever they get affected by every undangerous illness. The mindset of people mostly get addicted to drugs that may cause harmful side effects when they are used in the routine of often.

1.2 Machine Learning in Drug Discovery

All of the problems that mentioned above can be fixed with the help of machine learning technology by integrating the process of drug discovery with machine learning technology, more personalized drugs have been found. Here by the play of modern technology machine learning with a powerful physics setup can drastically reduce the time duration in the drug discovery. The process of structure prediction, testing and comparing with available drugs can be undertaken by machine learning technology.

The numerous output can be predicted on preferred duration so the more available resources can be used as key source production and due to multiple resources availability for production will cause an affordable price range for marketing. Making recording of all [2] stages and processes and creating a database will help in the communication intra medical field. Having a variety of drugs may cure all illnesses and make usable for regular endangered illnesses that can be used often by drug dependent people. Let's see the deep advantages concepts of machine learning in drug discovery.

1.2.1 Physical Setup

As every record in the medical field is managed by my computers today. Everything has been digitized and moreover become computerized, but the evaluation of the beast generation of computers that have capability 1.3 trillion floating point operations can be done per second by the powerful machine known supercomputers developed by IBM that have ultra-low latency with high performance.

Table 1. Top Supercomputers and their Regions

Sr. No	Computers	Region
1	SUMMIT	USA
2	SIERA	USA
3	TAIHULIGHT	CHINA
4	FRONTIER	USA
5	SUPERMUC-NG	GERMANY

The top most developed countries have a well-structured laboratory with multi facilities that have the capability of making every biological process easier with integration of modern machine learning technology. Having a common public biological intra server makes the discovery process quicker by creating a shared medium in between the bioinformaticians around the world.

1.2.2 Dataset

Data set is a collection of data pieces that can be managed computationally as a single set that is used for the processes like prediction, classification and regression problems. Here the data is not in the manner humans understand, it is completely created for the use of machines so it is only in machine understandable form.

1.2.3 Algorithm

Algorithms are the major source in machine learning for structural analysis and prediction. The phylogenetic and genes are predicted by the algorithms. In this stage of the process with the use of biological data with the help of machine learning algorithms the predictive models are obtained so that the final outcome is gathered by new data with the help of predicted models and also used in the testing stage. Here using the data sets available using the machine learning algorithm the predictive model is trained. Using the new data with the help of the pre-found predictive model the final outcome is obtained.
There are three types of machine learning algorithm

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

In supervised learning the models are trained with a labelled data set. The data set is to be labelled during the process of training, so that the output is identified on time of testing. The data set that to be inserted should be labelled properly for the more accurate outcome. The labelling process is to be done manually by the humans. This algorithm works by establishing the relationship between the given parameter data with the predicted trained data set. The most important stage in drug discovery like designing a new structure, property prediction and classification of compounds are made highly accurate with the appropriate algorithm.

The illness of a human is recognized by the dysfunctionality of protein, by use of algorithms the structure of the dysfunctionality protein is predicted as accurately as possible. The identified molecule is validated using the trained model. Distinguishing the blind structure by locating the very small molecules is known as visual screening. Using the same methodology like absorption, distribution, metabolization, excretion tests are done on time with more accuracy the effectiveness of personalised drug with similar drug is estimated using the supervised algorithm with the integration with the data sets. Let's see some machine learning algorithms.

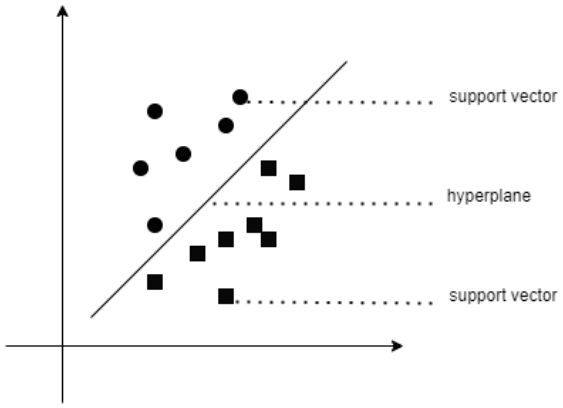


Figure1. Support Vector Machine (SVM) Classification Model

Support vector machine algorithm is a supervised learning algorithm which is used in the area of classification. The classification problems are taken with a support vector machine as it has the capability of solving it with more accuracy compared to all of the others. The support vector machine algorithm classifies the data set into different classes and creates a decision boundary or hyperplane. Support vectors are the data point that helps the support vector machine algorithm to obtain the hyperplane.

Here the first stage is to obtain the correct data set that has less toxicity and with characteristics that can be managed and produced easily in the laboratory because of the large amount of complex molecules like proteins. The sequence of compounds Like DNA, RNA, proteins and tiny molecules are gathered. The compounds are labelled to get started with the use of learning. The data set is most commonly separated into two sets as one for training and another for testing. After variables are located the compounds are trained with a particular set. Finally the test is processed and the result is obtained. Discovered from the original set here the new predictive drug is discovered successfully. This model is best economically.

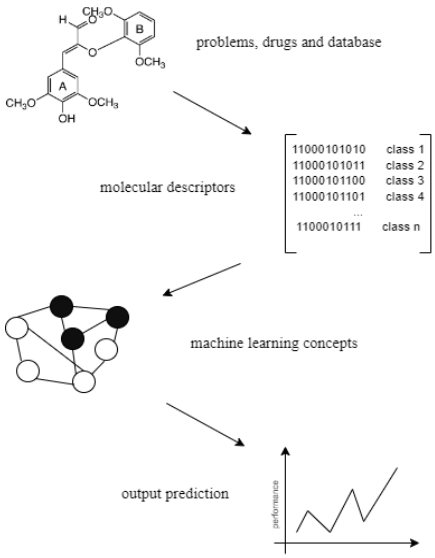


Figure 2. Machine Learning Workflow in Drug Discovery

1.2.4 Problem, Drugs and Database

The problem is documented with various tests being registered. The current drugs which are used in the treatment are also taken into account and it is studied. Medical databases are collected that are present world-wide. After the collection of all biological data form various medium than it is description

These biological data are to be digitized in superfast computers with labelling for the respective use of machine learning algorithms. The current medicine is studied and its formation data are also collected so it makes the discovery as a simple and faster way to completion. These biological data are used to train the model. The prebuilt dataset is also used as it is trained with predefined labelling so the dataset is created in a large storage manner. Only the supervised machine learning can make this huge dataset in the prediction of drug.

1.2.5 Molecular Descriptors

The data collected in the previous stage is here trained by converting it into the machine understandable form that can only be processed by the machine. Each and every data that is needed to feed in the process of discovery is to be converted in to machine understandable manner.

The molecules conversion plays a vital role here as it contains a huge chemical representation, though it is huge it can be converted with the help of high bit representation. These are stored in the matrix format as it becomes easier to label for the next stage.

1.2.6 Machine Learning Concepts

The biological data which is collected and computerized are taken out to the next step, the data are to be labelled or can be trained directly. This is determined by the use of the respective machine learning algorithm, if the supervised machine algorithm is used then the data are to be trained with labelling, on the other hand if the use of unsupervised machine learning algorithm, then the data can be directly trained without the process of labelling. But this algorithms world with the large dataset as the data were not

labelled, so the time complicity obtained here is larger than the time complexity of the supervised algorithm.

There is a special algorithm that does not need any of the dataset or data as it learns of itself. The reinforcement machine learning algorithm contains three elements namely action, agent, environment. The agent produces the action and implements the environment and the feedback is gathered. If the gathered feedback is a positive one then it is pretended to reoccur, on the other hand if the gathered feedback is negative then it is blocked from reoccurring.

1.2.7 Output Prediction

The trained dataset are registered in the fast recursive manner so that they can be utilized by the tool or algorithm. Here the testable digitized drug data are tested with the dataset with an appropriate algorithm. Here the prediction is done in all of the possible order and the prediction data is stored. After the completion of gathering the prediction output data, the collected prediction data is represented in the graphical representation.

The graphical representation the predicted outcome is given to an analytics and the best of the predicted data is obtained, thus the prediction is over the dataset which is created to this process stores in any of the medium and it might be used as the database for the other drug discovery process.

2. CONCLUSION

The drug discovery process that contains various critical stages are done very precisely by the bioinformatics team successfully until today, this becoming a deed of living beings. In supporting the traditional way of drug discovery the machine learning technology strengthens the medical field. The future discovery process is completely handed over to modern machine learning technology which can obtain a variety of outcome results faster and with higher accuracy, by support vector machines algorithm this is possible as it is highly reliable compared to all of other machine learning algorithms.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Deepak B	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not Applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCE

[1] Archana. 'Machine Learning in Modern World'. Journal of Artificial Intelligence, Machine Learning and Neural Network, no. 31, HM Publishers, Jan. 2023, pp. 39-46, doi.org/10.55529/jaimlnn.31.39.46

[2] Dara, Suresh, et al. 'Machine Learning in Drug Discovery: A Review'. Artificial Intelligence Review, vol. 55, no. 3, Springer Science and Business Media LLC, 2022, pp. 1947-1999, doi.org/10.1007/s10462-021-10058-4

How to Cite: Deepak B. (2023). Machine learning in drug discovery. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 3(1), 104–110. https://doi.org/10.55529/jaimlnn.32.53.60

BIOGRAPHY OF AUTHOR



Deepak B, is an enthusiastic undergraduate scholar in the Department of Information Technology at SNS College of Technology, Coimbatore, Tamil Nadu, India. He is passionate about emerging technologies, software development, and innovative research in the field of information technology. His academic interests include artificial intelligence, data analytics, cloud computing, and cyber security. Deepak actively participates in technical projects, research activities, and knowledge-sharing initiatives to enhance his practical and analytical skills in modern computing technologies. Email: deepaksmail2004@gmail.com