ARTIFICIAL INTELLIGENCE ADVANCES IN DRUG DISCOVERY

[1]Devesh Maithani, [1]Abhishek Chauhan, [2]Himani Maheshwari, [2]Amarjeeet Rawat

[1]MCA Student, Department of Computer Applications, Uttaranchal University, Dehradun, Uttarakhand, India

[2]Assistant Professor, Department of Computer Applications, Uttaranchal University, Dehradun, Uttarakhand, India

Email: himanimaheshwari@uttaranchaluniversity.ac.in, amarjeet@uttaranchaluniversity.ac.in

ABSTRACT

In today's world, artificial intelligence has become the most important field of computer science. Machine learning (ML) is an artificial intelligence (AI) field in which we use a training dataset to forecast future outcomes. Machine learning is being used in a wide range of fields, banking, insurance, medical imaging, stock and also including healthcare. Without machine learning, drug discovery takes a long time, costs a lot of money, and has a lower success rate. In recent year's Decision tree, SVM, Naive Bayesian, ANN, and Random Forest (RF) algorithms are some of the Machine Learning applications or algorithms utilised in the field of discovery of drug.Machine learning has become a significant tool in the drug development field as a result of various applications or algorithms.

Keywords: Drug Discovery, Machine Learning, Artificial Intelligence, healthcare

1. INTRODUCTION

Healthcare is one of the applications where machine learning performs well and produces good results in a short period of time. Because high-throughput screening (HTS) procedures were widely used in the pharmaceutical sector in the 1990s, but it requires a significant amount of time and cost to screen and pick potential candidates, time is of the essence (hit compounds)[1].Machine learning may be the most effective method for discovering new pharmaceuticals, as it increases productivity and reduces the time spent searching for and implementing novel drugs. Drug discovery [2] in clinical trials is a lengthy process that takes 10-12 years to complete [3]. According to the Eastern Research [4] Group (ERG), a time window of 10-15 years is required for the development of a drug, with an average cost of US$ 2-3 billion and a success rate of 2.01% [5], so we can see why we need technology that provides the highest success rate in the shortest amount of time. Machine learning, a subset of AI (Artificial Intelligence), encompasses all capabilities, allowing us to create modules that outperform others.

To use machine learning to discover medications, one must first create a knowledge bridge between the drug and the ailment. Both traditional approaches and statistics can be used to build correlations between diseases and drugs. Identifying drug correlations using traditional methods is time-consuming, difficult, and costly.

Multidisciplinary machine learning integration approaches and procedures appear to be beneficial in drug development, thanks to advances in computer intelligence and data collecting. As a result, computational tools must be used to develop and study drug-disease relationships.

2. BACKGROUND

In the pharmaceutical industry, drug discovery, or the process of discovering novel medications, is critical. Finding novel medications is still an extremely expensive and time-consuming procedure at this point, involving Phases I, II, and III clinical trials [4]. Machine learning approaches in Artificial Intelligence (AI), particularly in-depth learning techniques, have lately become widely used and achieve high-quality performance in a variety of domains, including speech recognition, picture classification, bioinformatics, and others.The field of drug development is a most vital uses of modern AI approaches.i.e drug discovery. According to a survey, different patterns can be found in machine learning and drug discovery fields [4].

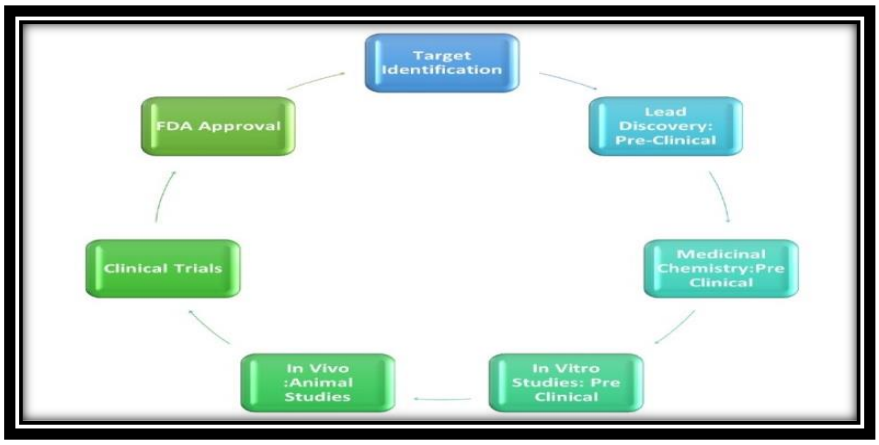

Fig 2.1: Phases in Drug Discovery

2.1. Target Identification: Identifying the target is the initial step in medication development. Understanding the objectives is more important than the medicine in this phase. Though the most likely strategy to develop a drug is to first detect the target and then concentrate on drug development, it is sometimes difficult for humans to identify every possible compound combination. This process can take up to two years for most medications.

2.2. Lead Discovery: pre-clinical: Exploring thousands of compounds that can disrupt the disease's intentions and reducing the number of possible combinations that can work on the target in the second step. This cycle usually takes 1 to 2 years.

2.3. Medicinal chemistry: Pre-clinical: The reduced chemicals are re-evaluated at this point in order to determine the interactions that caused the sickness. The 3D configuration of the compounds, as well as the targeted interactions, were used in further investigation. The results of the analysis are added to the target. This stage takes 1 to 2 years as well.

2.4. In vitro Studies: Pre-Clinical: The cell system is used to test compounds that have been screened up to this point. The phase in which petri dish investigations take place is known as in vitro studies. The effectiveness of the medicine is evaluated at this step by looking at the combinations that interfere with the target.

2.5. In vivo Studies:Animal Studies: Composites that pass the in Vitro-phase are taken and experimented on animals such as rats in this step. The results gained from these animal experiments are fairly representative when evaluated to the 2D in vitro-cell structure-models. The failure rate in its class is significant because to changes in the structure of the animal model in animals, and the results obtained in vitro may not be comparable to those obtained in vivo.

2.6. Clinical Trails: In this stage, a combination of promising traits was identified, and clinical trials were initiated. Human volunteers are used as test subjects at this level.

2.7. FDA Approval: Compounds that fall outside of all of these categories must be approved by the FDA. It is available on the market for general usage as soon as it receives FDA approval.

The rate of drug failure has risen in later stages of clinical testing, and this has always been a major concern. The clinical studies (1998- 2008)the failure rate of phase 2 & phase 3was 54 percent [7]. Concerns about safety account for 17% of failures, while ineffectiveness is another factor that accounts for the rest. In phase 2 and phase 3 treatment failure, side-effects and the death risk are also important factors [6 & 8]. The failure of a treatment and the time-consuming process, which takes extremely lengthy periods of time and incurs significant costs, can be annoying, specifically when trials weren't sufficiently successful. ML assists in this procedure by learning from previous data and experience, removing certain unknown aspects and reducing human effort, expense, and time.

3. METHODOLOGY

In supervised learning, a labelled dataset is used to train the model. That model or algorithm separates the data and appropriately predicts the outcomes. Many disciplines and organisations benefit from supervised learning to solve real-world challenges. There are two types of supervised learning: Regression and Classification. To accurately divide and classify data into specified groups, classification algorithm is used. It detects certain entities in a database and attempts to draw judgments about how those entities should be labelled. Regression is used to understand the relationship between dependent and independent variables. SVM, decision-trees, Random Forest, K-Nearest Neighbour (KNN), Naive Bayesian classifier, polynomial regression, linear regression, and logistical regression (LR) are some of the most extensively used supervised learning algorithms [18] and we can use Artificial Neural Network (ANN) [19].

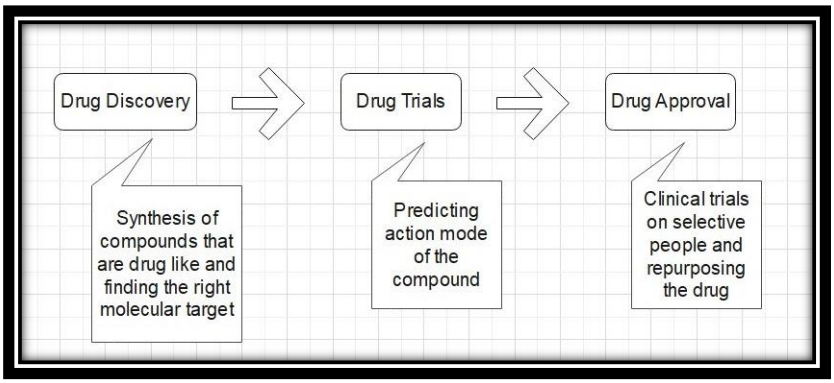

Fig 3.1: Machine learningflow in Drug Development

Unsupervised learning denotes, when there is no output variables. The basic goal of unsupervised-learning is to understand the dissemination of data in order to learn more about it (Fig 3.1). This can be divided into two categories: association and clustering. The clustering challenge requires you to collect data based on a specific pattern or behaviour. You wish to uncover rules that define a substantial percentage of your data in association. Decision tree, SVM, Naive Bayesian, ANN, and Random Forest (RF) algorithms are some of the current Machine Learning applications or algorithms utilised in the field of discovery of drug.

3.1. Support Vector Machine (SVM)

In drug discovery, SVM is used to divide classes of composites mostly based on function selector by creating a hyperplane. It generates an endless number of hyperplanes by using commonalities between lessons. It divides lessons, such as chemicals, into chemical function spaces based on established functions to convey linear knowledge.A first hyperplane is one that is obtained by maximising margin among lessons in N-dimensional space; it is represented by a hyperplane, which is used to categorise information factors using choice limitations. Regression models are important for establishing the relationship between ligand and medication by using queries to forecast.

3.2. Naive Bayesian (NB)

The NB algorithm has evolved into a valuable tool for categorization in predictive modelling. To categorise the capabilities of datasets, it is employed, and depending on the input features, aspect correlation, and dimensionality of the data, it may be one of the most successful solutions for the task. It is still unknown how well NB works with decision tree algorithms for textual content mining. These methods increase the precision of recovered data sets, which are often acquired from large, jumbled sources. In the drug development process, biomedical data classification is critical, especially in the goal discovery subgroup. This method has shown tremendous promise as a classification- tool for biomedical data, which is often congested with unrelated records and data, which is referred to as "noise." It can also be used to predict ligand-goal interactions, which is a significant advancement in lead discovery. This approach lately been included into a range of medication development procedures by researchers. In a study with possible hobby as estrogenic receptor antagonists in breast cancer, Pang et al. used NB designs and other approaches as classifiers for active and inactive medicines.

3.3. Random Forest (RF)

It is a broadly used set of rules for big datasets with several capabilities, since it simplifies the process of eliminating outliers, as well as classifying and designating datasets based solely on relative capabilities defined for the specific set of rules. It's commonly utilised for big inputs and variables, as well as accessibility based solely on statistics series from several databases. It can be used for a range of tasks, such as attributing absent statistics, dealing with outliers, and calculating classification attributes. The basic mathematical structure of RF is made up of an ensemble of uncorrelated choice trees, each of which is in charge of making a single prediction. The candidate who receives the most votes is deemed the most suitable. Although false- positives can occur in any statistical study, when compared to other methods, RF, along with NB and SVM has been recommended to make the fewest number of errors. Character mistakes are reduced when more than one choice timber is used because it assembles a variety of predictions rather than focusing solely on one. RFs are commonly used in drug discovery for character selection, classifiers, and regression.

3.4. Decision Tree: Decision trees are used to classify data and provide suggestions using a set of rules. In the pharmaceutical industry, decision trees are used to solve problems such as drug similarity prediction, developing combinatorial libraries, and generating chemical profiling data. Models based on decision trees are simple to validate and comprehend Drug absorption, permeability, and solubility, as well as metabolic stability, and penetration, are all predicted using decision trees. Because of the hierarchical nature of decision trees, large or substantial datasets are required. Small changes in the dataset can induce splits in the outcomes. The decision tree's effectiveness is also determined by the order in which contentious attributes are selected. Separation traits should be considerate of value or worth [7].

4. DISCUSSION:

In recent years, machine learning has been applied in a variety of fields, including healthcare, where a machine learning model predicts the optimum outcome. Prior to machine learning, the drug development process took far too long (6-10 years), and the success rate of new drugs was extremely low. It takes a lot less time now that machine learning algorithms have been developed, and the success rate is very high. Machine learning to identify novel drugs has three primary steps or phases (Drug Discovery, Drug Trials, and Drug Approval). In the subject of healthcare, machine learning has numerous algorithms, including decision trees, support vector machines, Naive Bayesian, and Random Forest (RF) algorithms. Machine learning has become a significant tool in the drug development field as a result of various applications or algorithms.

5. REFERENCES:

[1] M. R, “Critical review of the role of HTS in drug discovery,” pp. 11(7-8):277–9, 2006.

[2] A. B. A. B. a. J. E. A. Sertkaya, “Examination of clinical trial costs and barriers for drug development,” US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, pp. 1-92, 2014.

[3] R. Biswas and K. Haque, “Drug Discovery and Drug Identification using AI,” pp. 49-51, 2020.

[4] N. Stephenson, E. Shane and J. Rowland, “Survey of Machine Learning Techniques in Drug Discovery,” pp. 185-193, 2019.

[5] A. V. Singh and A. Negi, “Towards Better Drug Repositioning Using Joint Learning,” pp. 1-4, 2019.
[6] G. M. Schuhmacher O, “Hinder Changing R&D models in research based pharmaceutical companies,” J Transl Med, p. 14 (1):105, 2016.
[7] R. M. Manne, “Machine Learning Techniques in Drug Discovery and Development,” International Journal of Applied Research, vol. 7(4), pp. 21-28, 2021.
[8] C. D. L. J. W. B. F. J. K. A. Hwang TJ, “Failure of investigational drugs in late-stage clinical development and publication of trial results,” JAMA Internal Med, vol. 176(12), pp. 1826- 1833., 2016.
[9] B. G. Rajagopal and M. Arock, “Application of Machine Learning Techniques for study of drug interactions using clinical parameters for Creutzfeldt-Jakob disease,” 2020 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 415-420, 2020.
[10] G. Shobana, “Drug Administration Route Classification using Machine Learning Models,” Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020] IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3, pp. 654-659, 2020.
[11] A. Akay, “Deep Learning: Current and Emerging Applications in Medicine and Technology,” pp. 1-15, 2019.
[12] K. Babaria and S. Das, “Algorithms for Ligand based Virtual Screening in Drug Discovery,” in 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015.
[13] A. N. Lima and E. A. Philot, “Use of machine learning approaches for novel drug discovery,” pp. 2-42, 2016.
[14] M. Alam, J. Li and J. Wang, “Drug Discovery Approaches using Quantum Machine Learning,” pp. 1-4, 2021.
[15] R. Gupta, D. Srivastava and S. Tiwari, “Artifcial intelligence to deep learning: machine intelligence approach for drug discovery,” pp. 1315-1360, 2021.
[16] S. S. Kadiyala, “APPLICATION OF MACHINE LEARNING in Drug Discovery,” pp. 1-61, 2018.
[17] S. Barrett and W. L. B., “Advances in the Application of Machine Learning Techniques in Drug Discovery, Design and Development,” pp. 99-100.
[18]H. Maheshwari, P. Goswami and R. Isha, “*A Comparative Study of Different Machine Learning Tools”,* International Journal of Computer Sciences and Engineering, Vol.7, Issue.4, pp.184-190, 2019.
[19] Rana, A., Rawat, A. S., Bijalwan, A., & Bahuguna, H. (2018). Application of Multi Layer (Perceptron) Artificial Neural Network in the Diagnosis System: A Systematic Review. Proceedings of the 2018 3rd IEEE International Conference on Research in Intelligent and Computing in Engineering, RICE 2018.

BIOGRAPHIES

Devesh Maithani pursuingMaster of Computer Application from Uttaranchal University Dehradun. He completed his bachelor's degree in Information Technology from Uttaranchal University in 2020. Currently he is doing internship from Cognizant as a Programmer Analyst Trainee.

Abhishek Chauhanpursuing Master of Computer Application from Uttaranchal University Dehradun. He completed his graduation in Bachelors in Computer Application from Uttaranchal University in 2020. Currently he is doing internship from de-facto InfoTech as a Software Developer Trainee.

Dr Himani Maheshwari is an Assistant Professor in the Department of Computer Applications, Uttaranchal University, Uttarakhand, India. Artificial Intelligence, Machine Learning, Data Science, Web GIS, DSS, and Computer Vision are among her key research interests. Dr. Himani Maheshwari completed her MTech from Uttarakhand Technical University and Ph.D. from IIT Roorkee. She has published more than 25 research papers (9 are in Scopus), 3 book chapters and also registered 4 software copyrights. She received Young Scientist Award in 2020 from STEM Research Society. She also qualified GATE and UGC NET exams.

Mr. Amarjeet Rawat is an Assistant Professor in the Department of Computer Applications, Uttaranchal University, Uttarakhand, India. Recommender System, Machine Learning, Data Science, Big Data, and Web Technologies are among his key research interests. Mr. Amarjeet Rawat completed her MTech from Graphic Era University and pursuing Ph.D. from Uttaranchal University, Dehradun. She has published various research papers and book chapters in reputed journals and conferences.