**DRUG DISCOVERY**

A.Karthikeyan, P.Santhosh, K.Sathiyan, S.Yathin.

Assistant Professor, Students

Department of Computer Science Engineering,

Panimalar Engineering College, Chennai,

Tamilnadu, India

***ABSTRACT***---- **In the modern days, The advancements of information technology and related processing techniques have created a fertile base for progress in many scientific fields and industries. In the fields of drug discovery and development, machine learning techniques have been used for the development of novel drug candidates. The methods for designing drug targets and novel drug discovery now routinely combine machine learning to enhance the efficiency, efficacy, and quality of developed outputs. The generation and incorporation of big data, through technologies such as high-throughput screening and high throughput computational analysis of databases used for both lead and target discovery, has increased the reliability of the machine learning and deep learning incorporated techniques. The use of these virtual screening and encompassing online information has also been highlighted in developing lead synthesis pathways.**

1. **Introduction**

At present, there are many diseases in the world for which the vaccine or the drug to cure those diseases is not found, and also for new diseases which can come in the future it takes Time and more Capital to find a cure for these diseases. The concept of DRUG DISCOVERY benefits society, particularly the aging society in a very significant manner. The general meaning of Drug Discovery is the process that underpins the entire pharmaceutical industry, encompassing the early stages of research from target discovery and validation, right through to the identification of a drug candidate or lead compound. In the future, Technology plays a vital role in drug discovery (ex: we can find the new biologically active compounds in computer-aided drug design). As per the reports, global drug discovery is expected to grow at the rate of approximately 12.2% over the next decade to reach approximately $160 billion by 2025. Your motive is to use the Machine Learning concept in inorder to find the cure for a disease in a short time and with less usage of Money. For example, drugs can reportedly take 12 years from the initial discovery stage to licensing approval, and the Association of the British Pharmaceutical Industry estimated the amount of investment to be at £1.15 billion per drug.

1. **Related Works**

Searching over billions of molecules to find possible drug formula, drug design using machine learning. In this study, we introduce SMILES Pair Encoding (SPE), a data-driven tokenization algorithm. A dataset containing drug molecules (encoded as SMILES) and their binding affinities. The task is to use this dataset to make a regression model for binding affinity prediction. SMILES-based deep learning models are slowly emerging as an important research topic in cheminformatics.

**(EX SMILES REPRESENTATION)**:

|  |  |
| --- | --- |
| **SMILES** | **NAME** |
| CC | ethane |
| O=C=O | Carbon dioxide |
| C#N | Hydrogen cyanide |
| CCN(CC)CC | Acetic acid |
| [OH3+] | Hydronium ion |

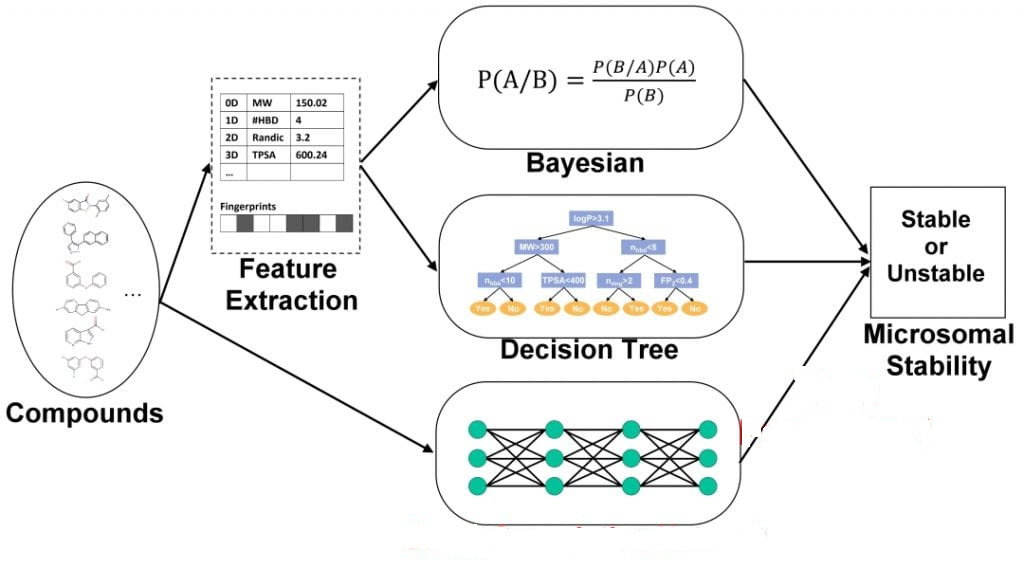
We load trained mol2vec which is trained on Morgan fingerprints with radius = 1 to yield 300-dimensional embeddings. Link to load trained mol2vec. Mol2vec is an unsupervised machine learning approach to learn vector representations of molecular substructures. It is discovered that scientists across the world are in search of new drugs. About 2.6$million is the estimated price for developing the treatment. But the system fails because it includes money spent on nine out of ten therapies that fail between phase 1 trials and regulatory approval. Here, few leading biopharmaceutical companies believe that there is a solution for this downfall. Pfizer is using 1BM Watson(i.e) a system that uses Machine Language, to power its search for immuno-oncology drugs(used for cancer treatments). There is UK start-up Exscientia's Artificial Intelligence platform used to hunt for metabolic disease therapies. A company called Sanofi has signed a deal to use it. Following this, Roche subsidiary Genentech is using an AI system for GNS healthcare in Cambridge, to help Multinational Companies search for cancer treatments.[**1**]

1. **Literature survey**

Many professionals have investigated this concept, For the past several decades, research in the molecular basis of human muscle aging has progressed. The major challenge is the development of accessible tissue-specific biomarkers of human muscle aging, which will evaluate the effectiveness of therapeutic interventions. Biomarkers could predict the functional capacity at some later age better than chronological age. Now, this is the method for tracking-related changes of human skeletal muscle. Here, they analyze the differential gene expression and pathway analysis of young and old tissue from healthy donors to preprocess the resulting data for a set of machine learning algorithms by Mamoshina, P. et al. [**2**]. Drug discovery and development is a long-term process, complex, and depend on several factors. Machine learning can improve discovery and decision making for well-specified questions with abundant, data,  target validation, prognostic biomarkers, and analysis data in clinical trials. The lack of interpretability and repeatability of Machine learning-generated results, which may limit their application. To tackle this situation, the application of machine learning can promote data-driven decision-making by itself, expeditious in the development process, and minimize the failure rate in drug discovery and development by Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham [**3**]. The process of discovery and development of drug-using manpower takes several years and much complicated. Finding the target element, clinical validation, completing the trial phases, and getting approval for the drug is the process. But novel target identification and validation can be done by various methods, tools, and data related to the target which will pre-validate a drug target in multiple views that will result in the new drug. This requires various aspects of knowledge and technologies. Based on the detailed specification and various comparisons the compatible elements are found by Byung-Cheol Kim; Sunghoon Kim [**4**].The process implementing using the Intel Open VINO toolkit for the identification of drugs. Using this technology one can identify the reactants which are added to the drugs and it automates the entire flow of the cycle. Using a custom object detection technique with the Intel OpenVINO toolkit the model can be trained using the R-CNN (Region-based Convolutional Neural Network) with the help of labeled drugs that also act as reactants. The total clinical trial process can be reduced by nearly 4-5 months, Originally it takes nearly 10-11 months in general to complete this process. Thus we can make a stimulating drug to see the behavior of the process and the implementation becomes much faster compared to the general process, implemented by [Risab Biswas](https://ieeexplore.ieee.org/author/37086930333), [Avirup Basu](https://ieeexplore.ieee.org/author/37088492312), [Abhishek Nandy](https://ieeexplore.ieee.org/author/37086930623), [Arkaprova Deb](https://ieeexplore.ieee.org/author/37088490832)[**5**].

1. **Proposed System**

According to the report based on the Application of Machine learning in drug discovery by Jessica Vamathevan, Dominic Clerk, Drug Discovery is a long and complex process. Machine Learning helps in this process by improving discovery and decision making. Machine Learning is also applied in all stages like validation, identification of prognostic biomakers[**6**]. Due to the lack of interpretability and repeatability of machine learning generated results may challenge the limit of their application. Systematic and comprehensive high-dimensional data still need to be generated. In this task, we are provided with the dataset of drug molecules in the form of SMILES and their binding affinity towards the disease. The molecules contain protein which is capable of replication and the transfer of the disease. One of the proteins is to be targeted and the drug is created which is capable of blocking the protein. The data has been generated using Protein-Ligand docking. Protein-ligand docking is a molecular modeling technique. The goal of protein-ligand docking is to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein receptor or enzyme.



Bayesian networks are a type of probabilistic graphical model that uses Bayesian inference for probability computations. Bayesian networks aim to model conditional dependence, and therefore causation, by representing conditional dependence by edges in a directed graph. Microsomal stability or Metabolic stability is defined as the percentage of parent compound lost over time, is assessed in the presence of liver microsomes. The microsomal stability assay is primarily used to investigate phase I metabolism using NADPH as an enzyme co-factor. An algorithm like Ridge, SVM, LASSO, etc is used to get the desired output.

1. **Conclusion**

We found that Drug Discovery is a long-term process by clinical experts. It will have it's pre-clinical &amp;

clinical phases, and drug approvals from IND(Investigational New Drug) and NDA(New drug Approval). Nevertheless, this long-term process can be diminished using Machine learning. The time consuming will be reduced, the limited success rate will raise, investment can be shortened and we can avoid ethical challenges. Despite the advantage in technology and understanding of biological systems, drug discovery is still a long process with a low rate of new therapeutic discovery. Data indicates the new targets, as opposed to established targets, are prone. Although combinatorial approaches have provided a new and effective way to discover drug leads, there still exists an abundance of natural organisms that have not been screened for potential new leads. Therefore, we feel the search for new drugs from natural sources (bioprospecting) should be continued even with the advent of combinatorial methods to drug discovery.

1. **Reference**
2. N. Fleming, "How artificial intelligence is changing drug discovery", Nature News, vol. 557, no. 7706, May 2018.
3. Mamoshina, P. et al. Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. *Front. Genet.* 9, 242 (2018).

# Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee. Applications of machine learning in drug discovery and development. *PMID: 30976107*

# [Byung-Cheol Kim](https://ieeexplore.ieee.org/author/38552255200); [Sunghoon Kim](https://ieeexplore.ieee.org/author/37857273000). Assemblable project formulation for drug target discovery.[IEEE International Conference on Bioinformatics and Biomedicine Workshops](https://ieeexplore.ieee.org/xpl/conhome/6449399/proceeding).*IEEE954-955 (2012).*

1. [Risab Biswas](https://ieeexplore.ieee.org/author/37086930333), [Avirup Basu](https://ieeexplore.ieee.org/author/37088492312), [Abhishek Nandy](https://ieeexplore.ieee.org/author/37086930623), [Arkaprova Deb](https://ieeexplore.ieee.org/author/37088490832). Drug Discovery and Drug Identification using AI.[*10.1109/Indo-TaiwanICAN48429.2020.9181309*](https://doi.org/10.1109/Indo-TaiwanICAN48429.2020.9181309)*(2020)*
2. Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee. Application of Machine learning in Drug Discovery. *PMID: 30976107(2019)*
3. M. Benhenda, "ChemGAN challenge for drug discovery: Can AI reproduce natural chemical diversity?", arXiv preprint, 2017.
4. P. Szymanski, M. Markowicz and E. Mikiciuk-Olasik, "Adaptation of high-throughput screening in drug discovery-toxicological screening tests", Int J Mol Sci., vol. 13, no. 1, pp. 427-452, 2012.
5. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436 (2015).
6. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. & Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* 23, 1241–1250 (2018).
7. Hinton, G. Deep learning — a technology with the potential to transform health care. *JAMA* 320, 1101–1102 (2018).
8. Wong, C. H., Siah, K. W. & Lo, A. W. Estimation of clinical trial success rates and related parameters.  *Biostatistics* “<https://doi.org/10.1093/biostatistics/kxx069>” (2018).
9. eon, J. et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design, and high-throughput screening. *Genome Med.* 6, 57 (2014).
10. J.P. Hughes et al., "Principles of early drug discovery", British Journal of Pharmacology, vol. 162, no. 6, pp. 1239-1249, 2011.