IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY

[1]Sukhvinder Singh Sudan, [2]Urvi Marhatta, [3]Gurvinder Singh Sudan

[1]B.tech, [2]B.Tech, [3]B.Tech student

[1]Department of Biotechnology,

[1]Graphic Era deemed to be University, Dehradun, India

Abstract: The rapid advancements in the field of Artificial intelligence have opened new gateways to revolutionize the process of drug development. The use of AI tools and techniques by the pharmaceutical industry has called for its collaboration with the IT space. AI can help the pharmaceutical industry by reducing R&D costs and time thereby reducing the escalating costs of the drug development process. In this review we discuss the major possibilities and uses of artificial intelligence in the drug development process.

Index Terms **– Drug designing, artificial intelligence, machine learning, Pharmaceuticals, IT.**

I. INTRODUCTION

Artificial Intelligence is the simulation or imitations of human intelligence by machines enabling them with the ability to think and function like humans. Artificial intelligence has become an influential and crucial part of the 21st century due to its capability and potential in managing complex and large amounts of data efficiently which has proved it to be a boon for the technology industry. The basic goals of Artificial intelligence include reasoning, data sorting, learning, natural language processing and the ability to wield objects using approaches such as statistical methods, traditional symbolic AI, computational methods, economics, psychology and many other relevant methodologies [1]. Many major subjects are employed in AI research and development including computer science, information technology, mathematics, statistics, philosophy, linguistics and many other fields. Though AI has helped improving many industries it also poses many ethical issues that are a cause of concern for the scientific community. Many believe that creating machines with human intelligence could threaten our very existence if it goes unabated where as others believe that AI will encourage mass unemployment across sectors. Though these risks are a matter of concern Artificial Intelligence has proven to be a virtue for the healthcare industry specifically in expediting the process of drug discovery. In this review the possible advantages of AI in drug development process are being discussed.

II. The Process and attrition rate of Drug Development

Every year several numbers of drugs are approved for use and launched in the market for different diseases. Thousands of potential drug candidates are screened and trialed to discover one potential drug for the treatment of a particular disease which costs more than a billion dollars in funds and takes around 10 years to reach the market. The journey starts at a university or a research institution funded by various agencies and pharmaceutical groups where basic research on understanding the cellular and molecular aspects of a disease starts. It is through this understanding of the pathways and processes of a disease that leads to the identification of potential targets for the new treatments. These targets may include but not limited to genes, some proteins that are instrumental to the disease for a new treatment. After this researchers look for potential compounds with biological activity which might be chemically or artificially synthesized, plant derived, fungi or marine organisms based as well as new compounds created using computers employing knowledge of proteins and genetics. As many as 10000 compounds could be considered for a single target which is further narrowed down to 10 to 20 after they show theoretical interference in the disease process. The compound that shows activity against the selected target is called a hit [2]. Several numbers of hits are discovered while screening different compounds in the drug discovery process. The next step in the process is the identification of a lead compound. A lead compound is the one that shows the most promising activity against the selected target and can be used to develop a potential drug for the disease. Once a lead compound has been identified, its chemical structure is modified for pertaining maximum efficacy, selectivity and activity while reducing toxicity or any kind of negative effects that might prevail.

Despite the presence of an efficient system of drug likeness guidelines for drug development, pharmaceutical companies have to face rigorous challenges in improving R&D efficiency and keeping the process economically viable. R&D efficiency is simply defined as the total number of drugs approved by the US FDA per billion US$ spent on R&D alone. Since 2001 the cost of developing a new treatment has increased from US$ 800 million to approximately US$ 3 billion currently. These increasing R&D costs and increasing attrition rate are a cause of great concern for the pharmaceutical companies [3]. The major underlying reason is that out of all the chemical compounds those are in phase IIB and phase III trials, 62% of them never make it to the clinics. The major reason for these late stage failures are attributed to clinical safety and efficacy, pharmacokinetics, toxicology and bioavailability. These late stage

failures have been seen largely in therapeutic treatments for the Central Nervous System (CNS) and oncology [4]. The major reason for these failures is the encouragement of use of in vivo and in silico approaches rather than using animal models in the early stages of drug development. Other factors that have contributed to the increase in attrition rate include narrow clinical research, multicentre clinical trials and stringent and lengthy approval guidelines by the FDA. Though these factors are not universal and may vary according to area of therapeutic research.

III. Application of AI in Drug Discovery

Validation and identification of Drug targets

The first step in the drug development requires identifying a plausible target whose activity results in the onset of a disease or serves as a contributing factor in the disease process. The biological data that is recorded from a large variety of people is extensively rich in information and includes genetics, proteomic and transcriptomics profiling and their variation in healthy and diseased individuals as well as high quality imaging results. ML can be potentially exploited for early identification of targets that is data driven by using these large data sets that can be accessed via public databases [5]. As we confirm association between a target and the disease, we can use ML to associate them with large data sets and comparing, for example, their known properties and actions. Several ML methods that have been developed for target identification include a decision tree-based meta-classifier trained on network topology of protein-protein, metabolic and transcriptional interactions, as well as tissue expression and subcellular localization developed by Costa et al. which is helpful in identifying druggable genes associated with morbidity [6]. A Support Vector Machine classifier was built by Jeon et al. uses various genomic data sets to classify proteins into drug and non drug targets for cancers such as breast, ovarian and pancreatic [7]. The AI platform, IBM Watson has also identified 5 RNA binding proteins that are linked to pathogenesis of amyotrophic lateral sclerosis [8].

Identification of a Hit or Lead

The discovery of a small chemical molecule capable of altering the functioning of a protein or a pathway of interest is a very important step in the drug discovery process. Due to the presence of limited data in new chemistry, the use of ML is restricted in this step though elaborative work has been done to develop methods deep learning methodologies such as ligand based virtual screening. A lead compound can be used to identify similar compounds having same structure, properties and mode of action computationally using statistical methods and DNNs [9]. Since the number of compounds that go for in vivo and in vitro trials are reduced significantly due to these DL frameworks the R&D costs incurred by the pharmaceutical companies are reduced significantly thereby reducing the attrition rate. Deep Learning is specifically important for this part of drug discovery since structural data is very limited. Thus AI is an important tool that can be used for identifying these small chemical molecules that possess a potential of being used as viable therapeutic treatment [10].

Prediction of Biomarkers

To improve the success rates of clinical trials and avoiding late stage clinical failures, ML based biomarkers are employed. They not only help in better understanding the mode of action of a drug but also in identifying the right drug for the right patient. These biomarkers are predicted using the pre clinical data sets [11]. Several successful case studies have now been published in which predictive models and their corresponding biomarkers have played a critical role in drug discovery. These ML approaches were used to understand drug response in patients with rheumatoid arthritis by Tasaki et al. [12]. Genome wide polygenic scores were developed by Khera et al. for the identification of individuals at high risk of breast cancer, type 2 diabetes, atrial fibrillation coronary artery disease and inflammatory bowel disease [13]. These similar ML models are being actively employed across domains of therapeutic research specially in cancer for developing immunotherapies for cancer. Though there are still problems in using these methodologies for clinical adoptions and validation with regards to multi-institutional data sets for demonstrating the general applicability of these approaches, the scientific community is doing everything from model interpretation to parameter optimization to address this issue [14].

Prognosis of the mode of action of a compound

Prediction of activity of compounds using AI is a very exciting prospect for most researchers as it saves valuable resources such as time and money. Such AI platforms can help better understand the toxicity and pharmacokinetics of various chemical compounds long before it reaches any kind of trials. Though these kinds of platforms have been developed, their output and performance can be greatly improved by a sense of collaboration and willingness to share data among the industry that would enable the process of developing larger and mch accurate data sets. An ML platform, known as, DeepTox algorithm has proven to show significant results in identifying the toxicity of various compounds [15]. Matched Molecular Pair analysis is used for investigating changes to a drug candidate and the impact on the bioactivity of the molecule [16]. A tool known as read-across structure–activity relationships (RASAR) has also shown significant results in predicting the toxicity of unknown compounds by linking their molecular structures and toxicity done by mining large databases of chemicals [17]. Using these tools have proved to reduce the time of R&D significantly there by creating economic viability in the drug designing process.

Drug repurposing using AI

Researchers using AI have been scrutinizing existing drugs and comparing it with data of millions of anonymous patients to develop new therapies for different diseases. Using already known drugs for developing treatments reduces the time of drug development because the Food and Drug Administration permits all such drugs to go directly into stage II clinical trials [18]. Studies have been reported for drug repurposing using deep learning and transcriptomic data. A company named Pharnext used AI to go through millions of records and discover three existing drugs named baclofen (muscle relaxant), naltrexone (used for treating opioid dependence) and sorbitol (a glucose-based laxative) to develop a treatment for a very disease known as called Charcot-Marie-Tooth disease (CMT) which is neurodegenerative disorder. Deep neural networks capable of performing various data transformation are a very effective tool for drug repurposing [19]. Other techniques that are being used for thus purpose are GAN technique and reinforcement learning [20, 21].

AI for selecting potential human candidates for clinical trials

Natural Language Processing and Machine Learning have been extensively used in recruiting the right kind of patients for clinical trials. These technologies help in improving electronic phenotyping used for reducing population heterogeneity [22]. These predictive models help in increasing the success rate of clinical trials. AIcare also developed a mobile application to check and measure the adherence of a drug in patients with schizophrenia and reported to increase the adherence by 25% in comparison with the traditional therapy [23].

IV. CONCLUSION

Due to a significant rapid increase in the involvement of AI in the healthcare industry, pharmaceutical companies have started investing in joint ventures with the IT space with an objective of developing efficient tools that could help in the improvement of the drug development process. The present statistics that show a rigorous and humungous spending of US$3 billion on the development of a single treatment have been long overdue for change. The business model needs a significant refurbish and AI is the best opportunity to do so. AI will make it extremely easy to evaluate and screen millions of drugs in much reduced time as compared to what it takes. Due to this increase in demand and prevalence of AI based software systems and infrastructures large numbers of start ups are rapidly developing. Though currently there are no drugs developed through AI, looking at the advancements it would not take much time for them to hit the markets. As we witness these significant developments there are still a number of unanswered questions that need to be addressed such as predicting the accurate values of binding affinity between the target and the drug ligand and many more. Despite all this practical trials in laboratories remain of vital importance. AI shall further help in developing viable immunotherapies and gene therapies and help significantly in the field of regenerative medicine.

V. **Acknowledgment**

The authors are thankful to their parents and mentors at the Department of Biotechnology and Department of Computer Science and Engineering, Graphic Era deemed to be University and Graphic Era Hill University.

REFERENCES

[1] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chemical Reviews. 2019; 119(18):10520-10594.

[2] Hall DR, Ngan CH, Zerbe BS, Kozakov D, Vajda S. Hot spot analysis for driving the development of hits into leads in fragment-based drug discovery. Journal of Chemical Information and Modelling. 2012; 52(1):199-209.

[3] DiMasi JA, Grabowski HG, Hansen RW. The cost of drug development. New England Journal of Medicine. 2015;372(20):1972.

[4] Acharya KR, Sturrock ED, Riordan JF, Ehlers MR. Ace revisited: a new target for structure-based drug design. Nature Reviews Drug Discovery. 2003;2(11):891-902.

[5] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. Applications of machine learning in drug discovery and development. Nature reviews Drug discovery. 2019; *18*(6): 463–477.

[6] Costa PR, Acencio ML, Lemke N. A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data. *BMC Genomics*. 2010;11 Suppl 5(Suppl 5):S9.

[7] Jeon, J., Nim, S., Teyra, J., Datti, A., Wrana, J. L., Sidhu, S. S., Moffat, J., & Kim, P. M. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Medicine. 2014;6(7):57.

[8] Bakkar, N., Kovalik, T., Lorenzini, I., Spangler, S., Lacoste, A., Sponaugle, K., Ferrante, P., Argentinis, E., Sattler, R., & Bowser, R. Artificial intelligence in neurodegenerative disease research: use of IBM Watson to identify additional RNA-binding proteins altered in amyotrophic lateral sclerosis. 2018. Acta neuropathologica; *135*(2): 227–247.

[9] Ramsundar B, Liu B, Wu Z, Verras A, Tudor M, Sheridan RP, Pande V. Is Multitask Deep Learning Practical for Pharma?. Journal of Chemical Information and Modelling. 2017; 57(8):2068-2076.

[10] Okafo, G. et al. Adapting drug discovery to artificial intelligence. Drug Target Rev. 2018: 50–52.

[11] Li B, Shin H, Gulbekyan G, Pustovalova O, Nikolsky Y, Hope A, Bessarabova M, Schu M, Kolpakova-Hart E, Merberg D, Dorner A, Trepicchio WL. Development of a Drug-Response Modeling Framework to Identify Cell Line Derived Translational Biomarkers That Can Predict Treatment Outcome to Erlotinib or Sorafenib. PLoS One. 2015; 10(6):e0130700.

[12] Tasaki S, Suzuki K, Kassai Y, Takeshita M, Murota A, Kondo Y, Ando T, Nakayama Y, Okuzono Y, Takiguchi M, Kurisu R, Miyazaki T, Yoshimoto K, Yasuoka H, Yamaoka K, Morita R, Yoshimura A, Toyoshiba H, Takeuchi T. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. Nature Communications. 2018; 9(1):2755.

[13] Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature Genetics. 2018; 50(9):1219-1224.

[14] Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Molecular Systems Biology. 2016;12(7):878.

[15] A Mayr, G Klambauer, T Unterthine. DeepTox: toxicity prediction using deep learning Frontiers in Environmental Science. 2016; 3:80.

[16] Tyrchan, C. and Evertsson, E. Matched molecular pair analysis in short: algorithms, applications and limitations. Computational and Structural Biotechnology Journal. 2017; 15: 86–90.

[17] Luechtefeld T, Marsh D, Rowlands C, Hartung T. Machine Learning of Toxicological Big Data Enables Read-Across Structure Activity Relationships (RASAR) Outperforming Animal Test Reproducibility. Toxicological Sciences. 2018;165(1):198-212.

[18] Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, Johnston SE, Vrcic A, Wong B, Khan M, Asiedu J, Narayan R, Mader CC, Subramanian A, Golub TR. The Drug Repurposing Hub: a next-generation drug library and information resource. Nature Medicine. 2017;23(4):405-408..

[19] Lozano-Diez A, Zazo R, Toledano DT, Gonzalez-Rodriguez J. An analysis of the influence of deep neural network (DNN) topology in bottleneck feature based language recognition. PLoS One. 2017;12(8):e0182580.

[20] Galbusera F, Niemeyer F, Seyfried M, Bassani T, Casaroli G, Kienle A, Wilke HJ. Exploring the Potential of Generative Adversarial Networks for Synthesizing Radiological Images of the Spine to be Used in *In Silico* Trials. Frontiers of Bioengineering and Biotechnology. 2018;6:53.

[21] Ozerov IV, Lezhnina KV, Izumchenko E, Artemov AV, Medintsev S, Vanhaelen Q, Aliper A, Vijg J, Osipov AN, Labat I, West MD, Buzdin A, Cantor CR, Nikolsky Y, Borisov N, Irincheeva I, Khokhlovich E, Sidransky D, Camargo ML, Zhavoronkov A. In silico Pathway Activation Network Decomposition Analysis (iPANDA) as a method for biomarker development. Nature Communications. 2016;7:13427.

[22] Perez-Gracia JL, Sanmamed MF, Bosch A, Patiño-Garcia A, Schalper KA, Segura V, Bellmunt J, Tabernero J, Sweeney CJ, Choueiri TK, Martín M, Fusco JP, Rodriguez-Ruiz ME, Calvo A, Prior C, Paz-Ares L, Pio R, Gonzalez-Billalabeitia E, Gonzalez Hernandez A, Páez D, Piulats JM, Gurpide A, Andueza M, de Velasco G, Pazo R, Grande E, Nicolas P, Abad-Santos F, Garcia-Donas J, Castellano D, Pajares MJ, Suarez C, Colomer R, Montuenga LM, Melero I. Strategies to design clinical studies to identify predictive biomarkers in cancer research. Cancer Treatment Reviews. 2017;53:79-97..

[23] Bain EE, Shafner L, Walling DP, Othman AA, Chuang-Stein C, Hinkle J, Hanina A. Use of a Novel Artificial Intelligence Platform on Mobile Devices to Assess Dosing Compliance in a Phase 2 Clinical Trial in Subjects With Schizophrenia. JMIR Mhealth Uhealth. 2017 ;5(2):e18.