IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Drug Repurposing: A Comprehensive Review Of AI And ML In Accelerating Drug Discoveries

[1] G. Angel, [2]Dr. P. Sujatha
[1]Research Scholar, [2]Professor & Head
[1]Department of Computer Science, [2]Deparment of Computer Applications
[1]Vels Institute of Science, Technology and Advaned Studies (VISTAS), Chennai, India

Abstract: Drug repurposing is an important area of research for pharmaceutical companies and chemical scientists. In the recent times, there has been an increased focus on the process of finding a new therapeutic use for already-approved medications. Machine learning (ML) and artificial intelligence (AI) algorithms are transforming drug repurposing because they can minimize risks, maximize efficiency, cut down on development time, and be inexpensive. This article discusses algorithms used in drug repurposing that are based on AI and ML. It also describes the ways in which methods from AI and big data have evolved to meet the growing demands of research in terms of planning development of drug candidates more efficiently and at a lower cost. With an analysis of current trends and recommendations for best practices, this study aims to give researchers a comprehensive understanding of how AI and ML may be utilized as a resource to help them navigate the always evolving field of drug discovery.

***Index Terms* - Drug repurposing; Artificial Intelligence; Machine Learning; Deep Learning, Big data.**

Introduction

Drug repurposing or repositioning is an essential technique in the field of drug development since it allows current medications to be used to treat new conditions. When preclinical safety studies have been already been finished, repurposing existing drugs can help find novel treatments for complicated illnesses more quickly and at a lesser cost with improved success rates. The "therapeutic stratification procedure" is crucial for patients with chronic, uncommon, or complex disorders for which alternatively there are zero or very few commercially available therapies [1]. Nowadays, drug repurposing is more methodical and mostly driven by in-silico approaches, whereas in the past, successful cases were primarily fortuitous. Computational approaches include genetic connections, pathway mapping, molecular docking, and binding-site detection. By employing algorithms, massive datasets, and pattern matching to find connections between the chemical structures of medications and their targets, artificial intelligence (AI) and machine learning (ML) have the potential to address one of the main problems facing contemporary drug discovery. In order to produce potential new therapeutic target predictions, AI-based drug repurpose. identification uses Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) to extract useful information, such as gene data, patient records disposition, and drug biochemical properties [2]. In recent years, deep learning and neural networks have gained increasing popularity. They can be used to screen compounds, anticipate new drug-target combinations, and evaluate adverse drug reactions, all of which help expedite the process of choosing current targets and medications for novel uses [3]. Drug repurposing and drug development in general can benefit from new opportunities presented by AI and ML techniques, but they also provide novel and distinctive problem. In this paper, we examine various AI/ML-based approaches for drug repurposing and also draw attention to the present difficulties and problems in this area.

I. DRUG REPURPOSING APPROACHES

Drug repurposing, the possible side effects of the well-known medication, which is safer, more effective, and eliminates the need to start from scratch. Traditional de nova drug discovery is very expensive, takes longer time, and has a high attrition rate-approximately 90% of candidates fail clinical trials. In neurodegenerative disorders, where there is a failure rate of 100% and when it comes to developing possible drug candidates, the rates of failure are greater than those for illnesses such as cancer. Because of its remarkable abilities to save time and costs, drug repurposing is a strategy that has attracted growing interest from pharmaceutical companies and governments [7]. Besides discovering new uses for drugs that have already been approved, this can involve: i) reaching out to new patient groups; ii) creating new dosage forms; and iii) investigating novel administration methods or therapeutic approaches [9]. Figure 2 outlines the drug repurposing workflow. The first step in drug repurposing is the target identification (disease, gene or protein) Then collect the relevant drug and target-based datasets and preprocess the same according to the AI or ML algorithms applied. Once we get the curated datasets for drugs and target, apply the AI and ML methodology and evaluate the model. Finally, validate the model through experiments and clinical trials. Following are the most important approaches applied for drug repurposing.

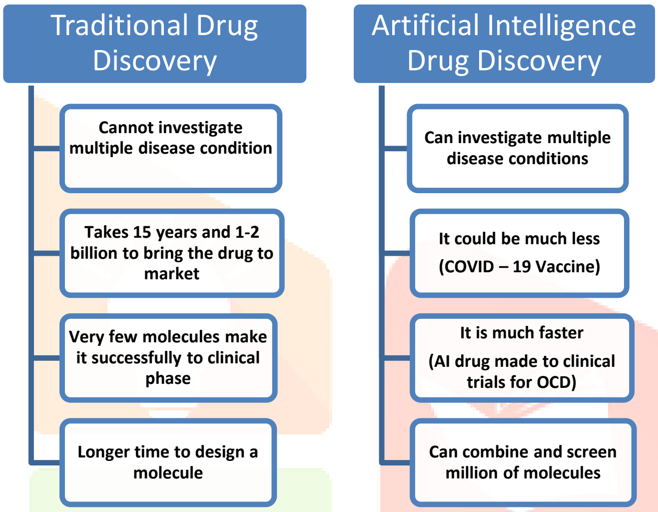

Figure 1. Traditional Drug Discovery vs. Artificial Intelligence Drug Discovery

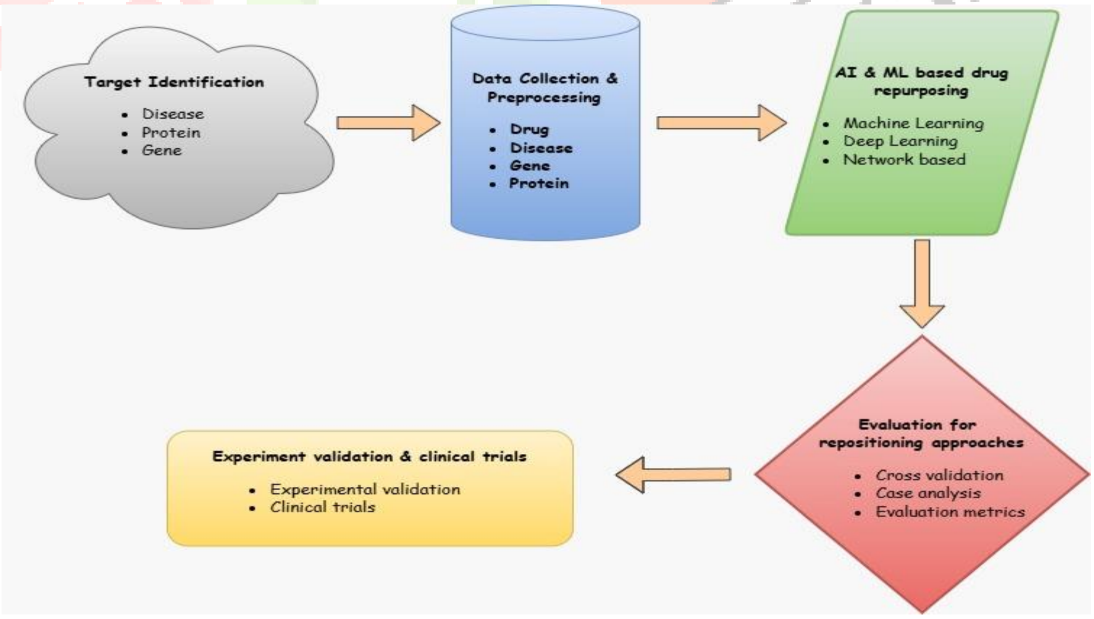

Figure 2. Drug repurposing work flow

A. Virtual Screening

In drug development, virtual screening (VS) is a computational technique that searches libraries of small compounds to find the structures that are most likely to bind to a therapeutic target, typically an enzyme or protein receptor. In other words, virtual screening (VS) is the computer counterpart of experimental HTS, in which compounds from chemical libraries are assessed for their ability to bind to a biomolecule target that may have therapeutic relevance for a particular illness. [13]. The order of various ligands is determined using a scoring function that is established to be proportional to their potential for inhibition [13]. Based on molecular characteristics, various ML algorithms such as Random Forests, Support Vector Machines, Neural Networks, and Graph Neural Networks can be employed to predict drug efficacy, toxicity, and other pertinent properties. Some crucial facets of Machine Learning (ML) in Virtual Screening include data representation, model selection, feature engineering, high-throughput screening, prediction accuracy, and the integration with other computational methods to efficiently pinpoint promising compounds exhibiting desired biological activity [ref]. Ligand-based virtual screening (LBVS) and Structure-based virtual screening (SBVS) are the two types of virtual screening. SBVS methods are advantageous for identifying actives with new scaffolds, while Finding actives with LBVS is restricted to a predetermined set of chemical descriptors linked to known ligands [14].

Selection of the applicability domain, regression strategy-based biological activity prediction, feature selection for quantitative structure–activity relationship (QSAR) models, comparable hit prediction based on QSAR, and shape similarity methods are the primary objectives of LBVS [15]. In SBVS, docking calculations are a widely used and valued method for structure-based drug discovery. Databases containing small-molecule chemical structures have been expanding quickly. Currently, though, using docking for virtual screening of extensive libraries is not widespread [16]. Figure 3, shows the simple virtual screening process of SBVS and LBVS.

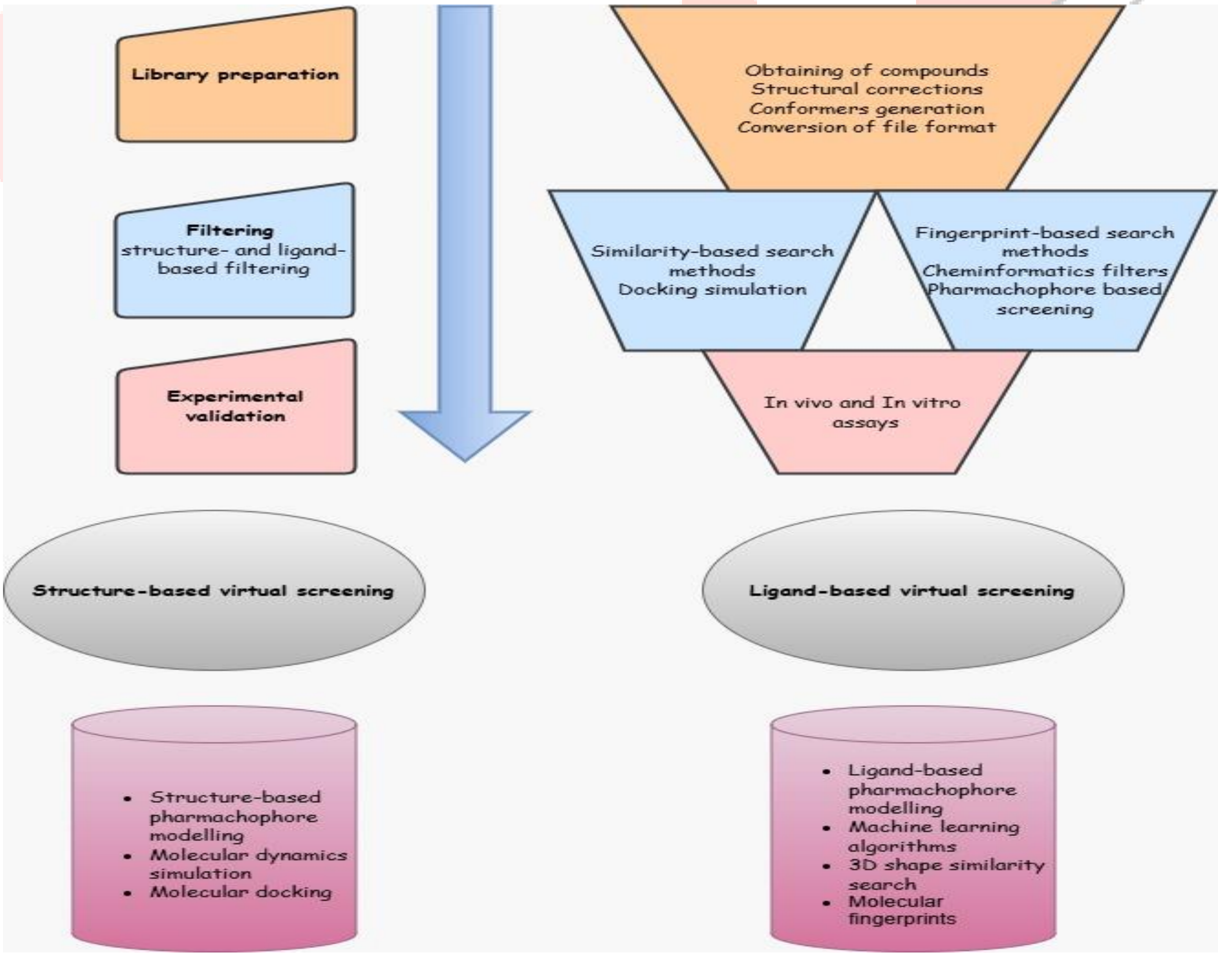

Figure 3. Virtual Screening

B. Deep Learning Approaches

Cutting-edge deep learning techniques are essential to the innovative technology of Drug-Target Interaction (DTI) predictions, which examine how medications interact with various targets to discover new uses for them. CNN, GNN, and RNN are a few of the popular DL algorithms. While GNNs depict medications and targets as graph structures that capture crucial chemical linkages that are crucial for creating an accurate interaction prediction, CNNs examine the structural features of pharmaceuticals and proteins [2]. DL primarily relies on how medications and targets are represented. Generally, drug repurposing techniques, which are classified as "target-centered" or "disease-centered" approaches, aim to forecast the unknown drug-disease or drug-target interactions [8]. Numerous techniques that encode compounds as molecular graphs have been developed recently to offer a novel end-to-end way for predicting Compound-Protein Interactions (CPIs) in the target-centered approach. This method learns real-valued vector representations of molecular graphs and protein sequences in low-dimensional space through the combination of a GNN for compounds and a CNN for proteins [8]. For drug repurposing focused on diseases, it is crucial to determine the interactions between pairs of drugs and diseases. A novel method for drug repurposing was presented by Xuan et al. [17] that combines CNN and bidirectional LSTM. The CNN component was used to determine the initial representation of drug-disease pairs based on their associations and similarities, while the BiLSTM-based module was used to learn the drug-disease path representations, balancing the contributions of different paths using an attention mechanism [8].

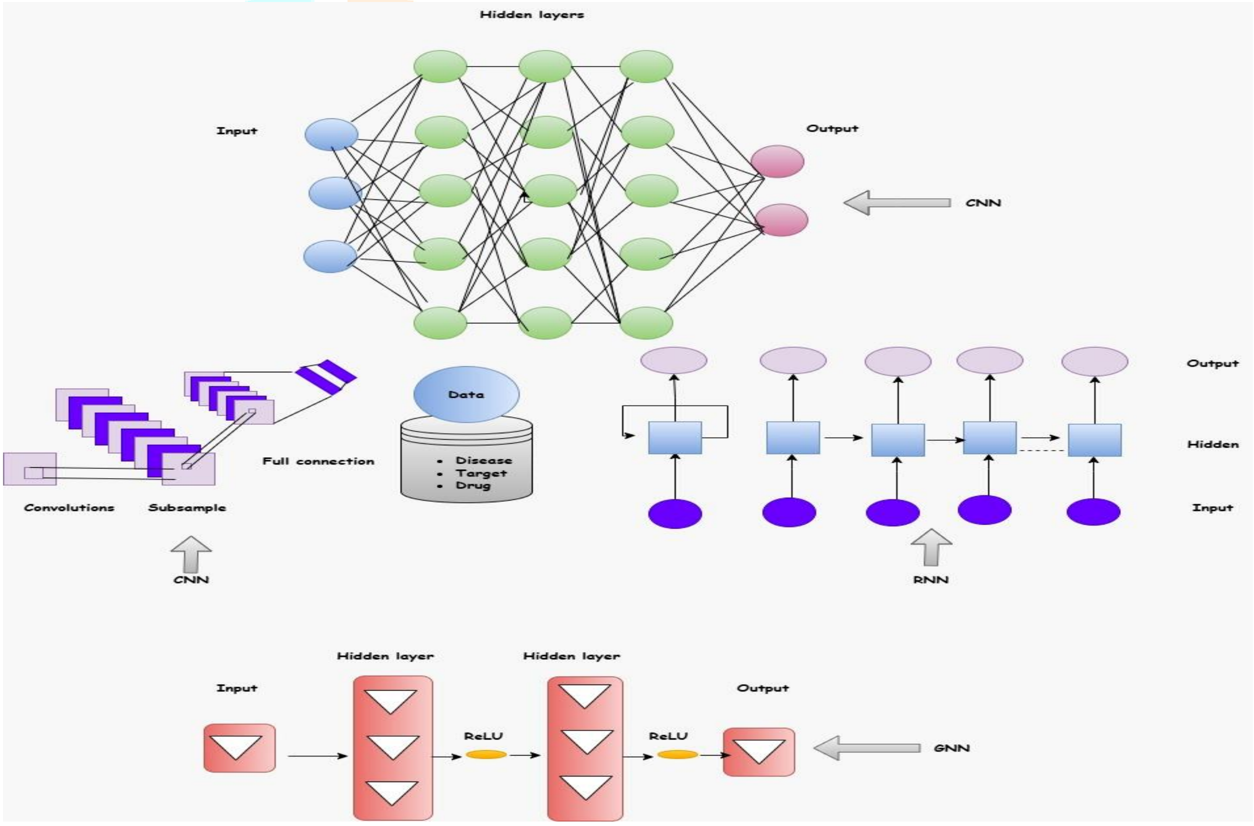

Figure 4. Deep Learning

C. Network Based Approaches

Network based approach utilize interaction networks to predict how drugs may affect different biological pathways and diseases. By focusing on repurposed drugs, these approaches not only expedite the process of finding new indications but also reduce the risks associated with clinical trials. There are two types of network-based methods which can be distinguished: matrix factorization (MF) and random walk (RW) techniques. The MF approach attains greater precision, but the RW approach is more scalable. This method can be widely applied in computational drug repurposing thanks to the developments in high-throughput technologies and the proliferation of data sources [4]. Network-based techniques have also been applied to target identification for popular drugs in effort to reduce the side effects and expedite drug repurposing [8]. There are more unidentified drug-disease interactions than the known ones in the process of drug repurposing problem. This problem could lead to the class-imbalance problem in learning-based approaches and inequalities in network-based systems. To circumvent the problems, network-based methods treat unknown interactions as unlabeled connections rather than negative associations. Additionally, this helps to prevent the issue of incorrect label assignment, which could somewhat impair model performance [4].

II. BIG DATA RESOURCES IN DRUG REPURPOSING

The concept of big data comes when it might be difficult to extract useful information from the heterogeneous data. Data from multiple fields, including structural biology, pharmacological, bio-assay, and clinical data are used in the process of drug repurposing. When combined, this data from various sources and areas creates a large, complicated dataset, and the Artificial Intelligence (AI) is essential for addressing complexity present in the data [5]. As big data continues to grow, more sophisticated computing techniques and more processing resources are needed to analyze the complicated data. Due to the need for greater computational power, the cloud computing, the high-performance computing, and the graphics processing components have replaced personal computers so as to analyze large amounts of data.

Table 1. DATA SOURCES USED IN DRUG REPURPOSING

S.no	***Database***	***Data type***	***Data size***	***Website***
1	ChEMBL	Contains bioactive and drug-like molecules	2.4 million compound and 20 million bioactivities	https://www.ebi.ac.uk/chembl/
2	PubChem	Includes substance and how they work against biological target	293 substance descriptions, 111 million chemical structures, 271 million bioactivity data points and 1.2 biological assays	https://pubchem.ncbi.nlm.nih.gov/
3	DrugBank	Includes drug and drug targets information	17,445 drugs	https://go.drugbank.com/
4	Drug central	Includes active substances, medicinal product and drug mechanism of action information.	4950 drugs	https://drugcentral.org/
5	BioGRID	Contains proteins, genetic and chemical interaction	2,826,897 proteins and genetic interaction, 31,144 chemical interactions, and 1,128,339 post translational modifications from major model organism species	https://thebiogrid.org/
6	BindingDB	Includes protein targets and information on small compounds binding affinities	3.0 million for 1.3 million compounds and 9.5k targets	https://www.bindingdb.org/
7	PDB	Contains 3D structures of protein	230,444 protein structures	https://www.rcsb.org/
8	DIP (Database of interacting proteins)	Contains protein-protein interaction	40,678 interaction information	https://pmc.ncbi.nlm.nih.gov/articles/PMC102387/

S.no	***Database***	***Data type***	***Data size***	***Website***
9	TTD (Therapeutic target database)	Contains known therapeutic protein and nucleic acid targets	3,730 targets and 39,862 drugs	https://idrblab.net/ttd/

ChEMBL, BindingDB, and PubChem are the basic libraries for compound synthesis and screening in finding the drug candidate for the target. AI algorithms like DL, RF, classification and regression, and NN are used to analyze the filtered compounds from the datasets.

Challenges faced by AI and Big Data

Improper data merging is one of the main issues in big data. This issue arises from the variety of datasets, which may include candidate data, raw data, or metadata. Currently, there is no standard technique for gathering and assembling these datasets for effective evaluation [6]. Assessing the imbalanced distribution of the available data presents is another complicated problem in big data. Resampling strategies, which are non-algorithmic specific and comprise multiple sampling techniques, and algorithmic tweaks to make them suitable for managing unbalanced datasets are the two broad types of methods used to sort the impacts of imbalanced datasets [5]. Integrating data from multiple sources can be challenging due to inconsistencies in data format and quality. Also, machine learning models can generate false positive drug-disease associations requiring careful validation.

III. APPLICATIONS OF DRUG REPURPOSING

Drug repurposing has become a viable approach for drug discovery and development across a range of human disorders, including cancer, neurological diseases, uncommon diseases, and infectious diseases, as a result of numerous changes in traditional de novo drug discovery [8]. The first AI model created especially to identify possible medication options for uncommon illnesses and ailments without cures is called TxGNN. It identified potential drugs from existing pharmaceuticals for more than 17,000 illnesses, many of which have no known cure. Numerous common disease categories have seen the application of AI, which has been found to be a promising avenue for drug repurposing. These include COVID-19 (21.63%), cancer (18.18%), neurodegenerative diseases such as Parkinson's and Alzheimer's (5.73%), and infectious diseases, excluding COVID-19 (4.19%) [18]. Rare diseases make up 2.63% of the literature references, making them the fifth most discussed category [18]. The loss of dopaminergic neurons in the substantia nigra, which causes movement symptoms in Parkinson's Disease (PD), the second most prevalent neurodegenerative illness, is another instance of drug repurposing [10]. It impacts over 6-7 million people globally and remains without a cure, despite the use of levodopa since the 1960s, which only postpones the advancement of the disease. A recent study focused on selecting the best possible treatment candidates from FDA-approved medications based on their anticipated PINK1 expression-enhancing activity, despite the fact that AI and ML applications in PD are not well researched [10]. To evaluate the model's capacity in identifying molecular patterns, and extrapolate those to novel, unobserved molecular representations, and their study utilized a supervised machine learning framework using in vitro data and the scaffold splitting methodology [10]. To find novel pharmacotherapies on the expected expression of the PINK1 target, the predictions of models are merged in a meta-ensemble configuration [10]. The most promising pharmacological candidates for Parkinson’s Disease (PD) with PINK1 expression were identified by their study as nitazoxanide and representatives of the imidazolidines, trifluoromethylbenzenes, anilides, nitriles, stilbenes, and steroid esters [10].

Table 2. Repurposed drugs [9]

S.no	***Compounds***	***Original Use***	***Repurposed Use***
1	Dapoxetine	Antidepressant	Premature ejaculation
2	Clofazimine	Leprosy	Tuberculosis
3	Colchicine	Gout	Cardiovascular disease, pericarditis (clinical trials)
4	Propranolol	Hypertension	Anxiety (clinical trials)
5	Tamoxifen	Breast cancer	Bipolar disorder (clinical studies), infertility
6	Berberine	Antimicrobial digestive aid	Ulcerative colitis, cancer

IV. CONCLUSION AND FUTURE WORK

Pharmaceutical repurposing is crucial for cutting costs and speeding up the release of new drugs onto the market. As the toxicity profiles of approved experimental drugs are already well established, the safety profile of repurposed drugs becomes the most critical aspect. By offering the therapies for unmet targets like neurodegenerative diseases or medical requirements, drug repurposing also aids in drug discovery. AI-based technologies are used at several phases the drug reprocessing process which includes protein structure prediction, toxicity prediction, bioactivity prediction, and physical property prediction. Apart from various applications in AI-based technology, there are still some limitation and challenges faced when it comes to data. Big data's many "V" characteristics – volume, velocity, variety, and volatility – call for better data management and curation as well as user-friendly online web portals. The use of AI and ML in neurodegenerative diseases has gotten very minimal attention, despite the fact that AI-based drug repurposing is revolutionizing the drug designing and development. The main reason for this is the complicated CNS conditions, the absence of highly effective pharmacological targets, the difficulty in identifying prospective medications in the early stages, and the blood-brain barrier (BBB). However, this may be overcome by AI and ML algorithms like GNN and CNN as they can efficiently integrate information from drugs, targets, and diseases, capturing the complex relationships between them. Regardless of the drawbacks and difficulties, Artificial Intelligence and Machine Learning have transformed the drug development and discovery process and will eventually play a very crucial role in finding new medications and their targets.

References

[1] Dotolo, S., Marabotti, A., Facchiano, A., & Tagliaferri, R. (2020). A review on drug repurposing applicable to COVID-19. *Briefings in Bioinformatics*, 22(2), 726–741. https://doi.org/10.1093/bib/bbaa288

[2] *View of AI-Driven Drug Repurposing: Uncovering hidden potentials of established medications for rare disease treatment*. (n.d.). https://bpasjournals.com/libraryscience/index.php/journal/article/view/2837/2636

[3] Anokian, E., Bernett, J., Freeman, A., List, M., Santamaría, L. P., Tanoli, Z., & Bonnin, S. (2024). Machine learning and Artificial intelligence in drug Repurposing – Challenges and Perspectives. *Drug Repurposing*. https://doi.org/10.58647/drugarxiv.pr000007.v1

[4] Sadeghi, S., Lu, J., & Ngom, A. (2021), "A network-based drug repurposing method via non-negative matrix factorization," *Bioinformatics*, *38*(5), 1369–1377. https://doi.org/10.1093/bioinformatics/btab826.

[5] Tripathi, M. K., Nath, A., Singh, T. P., Ethayathulla, A. S., & Kaur, P. (2021). Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. *Molecular Diversity*, *25*(3), 1439–1460. https://doi.org/10.1007/s11030-021-10256-w

[6] Parashar, K. (2023). Use of artificial intelligence and big data for drug discovery model. *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 1–7. https://doi.org/10.1109/icses60034.2023.10465293

[7] Yang, F., Zhang, Q., Ji, X. *et al.* Machine Learning Applications in Drug Repurposing. *Interdiscip Sci Comput Life Sci* **14**, 15–21 (2022). https://doi.org/10.1007/s12539-021-00487-8

[8] Pan, X., Lin, X., Cao, D., Zeng, X., Yu, P. S., He, L., Nussinov, R., & Cheng, F. (2022). Deep learning for drug repurposing: Methods, databases, and applications. *Wiley Interdisciplinary Reviews Computational Molecular Science*, *12*(4). https://doi.org/10.1002/wcms.1597

[9] Pinzi, L., Bisi, N., & Rastelli, G. (2024). How drug repurposing can advance drug discovery: challenges and opportunities. *Frontiers in Drug Discovery*, *4*. https://doi.org/10.3389/fddsv.2024.1460100

[10] Haneczok, J., Delijewski, M., & Moldzio, R. (2023). AI molecular property prediction for Parkinson's Disease reveals potential repurposing drug candidates based on the increase of the expression of PINK1. *Computer Methods and Programs in Biomedicine*, *241*, 107731. https://doi.org/10.1016/j.cmpb.2023.107731

[11] *The quintessence of basic and clinical research and scientific publishing*. (2023). https://doi.org/10.1007/978-981-99-1284-1 .

[12] Shukla, A. K., & Pradhan, J. (2024). Applications of molecular docking techniques in repurposing of drug. In *IntechOpen eBooks*. https://doi.org/10.5772/intechopen.1004703N.

[13] Murugan, N. A., Priya, G. R., Sastry, G. N., & Markidis, S. (2022). Artificial intelligence in virtual screening: Models versus experiments. *Drug Discovery Today*, *27*(7), 1913–1923. https://doi.org/10.1016/j.drudis.2022.05.013

[14] Gautam, V., Gaurav, A., Masand, N., Lee, V. S., & Patil, V. M. (2022). Artificial intelligence and machine-learning approaches in structure and ligand-based discovery of drugs affecting central nervous system. *Molecular Diversity*, *27*(2), 959–985. https://doi.org/10.1007/s11030-022-10489-3

[15] Kumar, N., & Acharya, V. (2023). Advances in machine intelligence-driven virtual screening approaches for big-data. *Medicinal Research Reviews*, *44*(3), 939–974. https://doi.org/10.1002/med.21995

[16] Bande, A. Y., & Baday, S. (2024). Accelerating Molecular Docking using Machine Learning Methods. *Molecular Informatics*, *43*(6). https://doi.org/10.1002/minf.202300167.

[17] Xuan, P., Ye, Y., Zhang, T., Zhao, L., & Sun, C. (2019). Convolutional neural network and bidirectional long Short-Term Memory-Based method for predicting Drug–Disease associations. *Cells*, *8*(7), 705. https://doi.org/10.3390/cells8070705

[18] Cortial, L., Montero, V., Tourlet, S., Del Bano, J., & Blin, O. (2024). Artificial intelligence in drug repurposing for rare diseases: a mini-review. *Frontiers in Medicine*, *11*. https://doi.org/10.3389/fmed.2024.1404338.