A Review: Drug Discovery Methods Based on Artificial Intelligence

Jinhao Chen[1,*]

[1]Beijing Institute of Technology, Beijing, China

Corresponding author: 101010@yzpc.edu.cn

Abstract:

The discovery of drugs is recognized as a lengthy, highly costly, and extremely complex process. For example, some traditional drug discovery methods consist of millions of trials to get a druggable compound to the market. Drug discovery based on artificial intelligence can be a prompt, low-cost, and effective way to streamline drug discovery. Although some works have been proposed to use artificial intelligence tools for drug discovery, few people summarize these advances in a systematic way. In this paper, we propose an organized and comprehensive review that outlines a broad range of appliances of artificial intelligence in drug discovery including harnessing virtual screening and molecular docking techniques, utilizing pathway networks for repurposing existing drugs, lead identification, biomarker research, identification of the target, diverse variety of artificial intelligence and their comparison, etc. In addition, we shed light on predicted limitations and challenges in drug discovery based on artificial intelligence, as well as sketch the strategies to harness its potential for upcoming drug design endeavors.

Keywords: molecular docking; drug discovery; artificial intelligence.

1. Introduction

The success of drug design and optimization paves the way for certain drug discovery. Still, it is a challenging path fraught with influencing factors such as poor selectivity and affection for the specified object, off-target effects, and physicochemical properties including solubility, stability, and pharmacokinetic characteristics [16]. For this reason, this step often requires the integrated application of knowledge from sorts of disciplines. Statistically, drugs typically take 10 to 15 years to make it to market, and their anticipated costs might range from 161 million to 4.54 billion (Schlander et al., 2021) [1]. Moreover, approximately 90% of potential drug candidates still fail clinical trials (Sun et al., 2022) [1].

In 1956, John McCarthy initially coined the phrase "Artificial Intelligence" and gave a general overview of its tenets [2]. The term AI denotes systems that mimic human cognitive functions, such as learning and problem-solving, often referred to as 'thinking' and 'intelligence' in machines [3]. With artificial intelligence advancing by leaps and bounds, AI-driven drug discovery has emerged as a revolutionary technology in the field, owing to its ability to analyze and interpret vast quantities of data. For instance, AI now has an incredible potential to expedite the process of CNS drug discovery with a high success rate [4]. Innovations in AI have demonstrated valid applications in critical scenarios, such as outperforming dermatologists in skin cancer analysis [5, 6], and utilizing prostheses to predict mortality after cardiac surgery [7]. These research-

es all signify that AI has become a significant and potent component of the drug discovery industry.

Despite the rapid progress in the field of artificial intelligence (AI) and its increasingly prominent role in drug discovery, there remains a gap in the comprehensive documentation of its integration and impact within the field. The existing literature tends to ignore the broader influence of AI throughout the drug discovery and development process. The purpose of this paper is to fill this gap by providing a comprehensive review of the diverse roles of AI in drug discovery, including its role in virtual screening, molecular docking, and drug repurposing. To accomplish this goal, the paper is structured so that the background and importance of drug discovery are covered in the introduction, which also highlights the revolutionary effects of artificial intelligence in this field. After that, a thorough analysis of the conventional drug development method is provided, stressing its drawbacks and difficulties, which highlights the need for AI-driven innovation. The report additionally illustrates how AI may speed up the process of developing new drugs by discussing its application in drug repurposing. The effective use of AI in drug discovery follows through case studies and examples, offering verifiable proof of its value. These practical benefits serve to emphasize the real-world advantages and possibilities of artificial intelligence in this area. The main body of the work explores the several ways artificial intelligence (AI) is being used in drug discovery. It does this by comparing and contrasting computational methods including computer-aided drug design (CADD), machine

learning (ML), and deep learning (DL), as well as how they are integrated into drug design. This section also looks at the application of AI to molecular docking and virtual screening, two processes that are essential for successfully identifying possible medication candidates.

The remainder of the paper is arranged as follows. Firstly, we talk about the procedure for traditional drug discovery in Section 2, focusing on conventional methods and their limitations. And then we review some of the applications in drug discovery based on artificial intelligence (Sect. 3), we will explore the comparison between computer-aided drug design (CADD), machine learning (ML), deep learning (DL), and artificial intelligence (AI) in crafting novel therapeutics. Then we will delve into the power of virtual screening and molecular docking to rapidly and precisely identify potential drug candidates. Furthermore, artificial intelligence function in repurposing drugs will be highlighted. Finally, we discuss obstacles to implementing finding drugs and chart its way forward.

2. Traditional drug discovery process

2.1 Basic Methods in Traditional Drug Discovery

Identification of biological targets is the first step in drug discovery, it involves understanding the pathogenesis of the disease in question and based on pathogenesis, pinpointing the protein, enzymes, or receptors that could serve as potential therapeutic targets. Subsequently, the initial chemical molecule is identified that has therapeutic efficacy against the designated target; such a compound is termed a ‚hit' [8]. This stage often applies high-throughput screening (HTS) of large compound libraries to identify these initial leads. The next step is to isolate 'hits' compounds, they are then subjected to a series of optimization processes to improve their potency, selectivity, and properties, this process is also called lead optimization. With optimized leads, the purpose of the drug's clinical trial is to assess the compound's safety and probable impact on humans.

Fig. 1 Overall drug discovery process

2.2 Limitations and Challenges in Traditional Drug Discovery

Traditional drug discovery, while having led to numerous breakthroughs in medicine, faces several inherent limitations and challenges that can hinder the efficiency and success of the process. Apart from being time-consuming and having a high rate of failure, traditional drug discovery also faces the challenge of limited targetability. Some targets are inherently difficult to modulate due to their complex structures or their role in multiple biological pathways, which increases the risk of off-target effects and toxicity. Moreover, traditional compound screening methods are inefficient, because screening tens of thousands

of compounds may yield only a few effective candidates. Additionally, many disease models, such as animal models, may not always predict the drug's effect accurately in humans due to the complexity of human physiology and the differences in disease manifestation. These factors can significantly limit the efficiency of the drug discovery process.

3. Applications of Artificial Intelligence in Drug Discovery

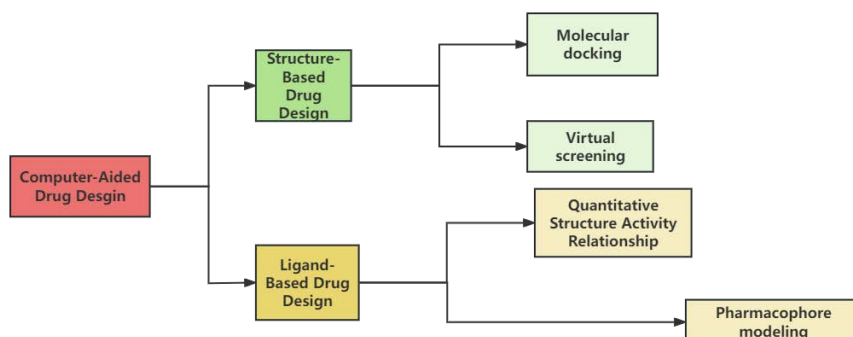### 3.1 The Comparison of CADD, ML, DL, and AI in Drug Discovery

Fig. 2 Overview of CADD

A variety of computer tools are combined in computer-aided drug design (CADD) to design and optimize lead compounds through computer simulations, calculations, and budgeting of relationships between drug and receptor biomolecules. CADD involves molecular modeling, rational drug design, computational chemistry, and molecular design (Muegge et al., 2017) [8]. CADD can be divided into two basic techniques: structure-based drug design (SBDD) and ligand-based drug design (LBDD). They are utilizing 3D structures of a protein or ligands based

on availability [8]. Machine learning (ML), a subfield of AI, can autonomously enhance its performance through experience [9]. ML is often divided into three categories: unsupervised, supervised, and reinforcement learning. It predicts computational statistics. [9]. Deep learning (DL) is a subfield of ML, comprising a class of methods focused on artificial neural networks (ANN) and spiking neural networks (SNN). It is distinguished by its capacity to learn from unstructured and unlabeled data without human intervention.
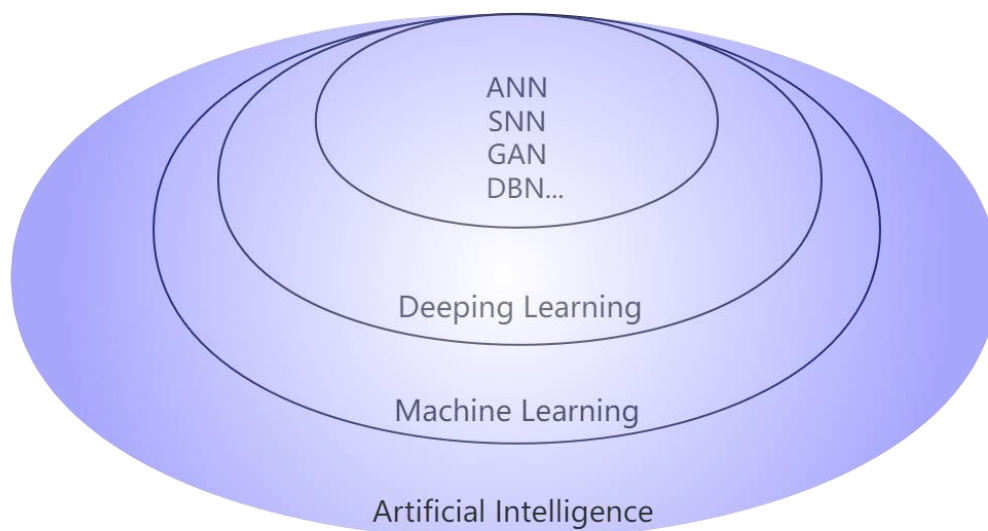

Fig. 3 Relationship of AI, ML, DL. Deep learning algorithms include artificial neural networks (ANN), spiking neural networks (SNN), generative adversarial networks (GAN), and deep belief networks (DBN).

The history of CADD can date back to 1950 [10], with the increase in computational power, its applications also expanded. ML and DL emerged with the rise of data science in the 21st century, getting remarkable achievements in image recognition and natural language processing, and gradually applied to drug discovery to overcome some tough issues.

The chief merit of CADD is precisely simulating the possibility of the interaction between receptors and the drug, enabling efficient virtual screening. ML is accomplished by extracting features from chemical and biological data to identify the potential drug candidates. DL is capable of processing intricate data structures, such as gene expression data and protein sequences, which can unveil the intricate mechanisms behind drug action. CADD technologies have been utilized to discover and/or optimize multiple drugs at various stages. These medications, include Nolatrexed, Oseltamivir, Rupintrivir, Saquinavir, Zanamivir, Boceprevir, Captopril, Dorzolamide, and Aliskiren [11]. ML is capable of analyzing a wide range of data types, including networks of protein-protein interactions,

gene expression patterns, and proteomic and genomic data. By doing so, it can discover possible components that are probably involved in disease processes [13]. For instance, graph neural networks (GNNs) are a type of deep learning (DL) used for the creation of novel molecular structures in drug discovery. Through their ability to discern and acquire the complex connections between atoms and molecular fragments within a dataset, GNNs are capable of generating new compounds [13]. The integration of artificial intelligence (AI), machine learning (ML), deep learning (DL), and computer-aided drug discovery (CADD) will further enhance the efficiency of drug discovery. This integration allows for a comprehensive understanding of the molecular mechanisms of disease and accelerates the speed of drugs from development to market.

3.2 Virtual Screening

AI techniques can be utilized for target-specific molecular identification and virtual screening or to examine their biological activities, and to predict protein-drug interaction.

Rather than conducting physical tests on every compound in a given library, virtual screening utilizes computational methods to predict which compounds are most likely to bind effectively with the desired drug targets [20]. Identifying potential drug compounds with favorable characteristics for specific targets is an essential phase within the drug discovery process [9]. It is necessary to take physicochemical properties, bioactivity, and toxicity into consideration because they ensure the discovery of safe, cost-effective medicines.

Physicochemical properties by altering solubility, stability, and lipophilicity, affect the toxicity features of absorption, distribution, metabolism, and excretion [19] in our body. Machine learning has been applied to model the combined probability distribution of physical characteristics and molecular structures to accomplish the inverse design. In recent years, the strategy of matched molecular pair (MMP) analysis, which is AI in terms of bioactivity, has gained widespread application in the de novo design of drug compounds. This approach involves assessing the effects of bioactivity and molecular properties through the introduction of a single chemical transformation [9]. AI-based toxicity prediction contributes to assessing the safety of drugs. The DeepTox algorithm [18] is a toxicity prediction tool based on DL. This method helps to anticipate and mitigate potential side effects during the drug development phase.

3.3 AI Applications in Drug Repurposing

Drug repurposing is reusing the existing and already available drugs for novel therapeutic options. One major benefit of repurposing drugs is that it is an approach that can skip the toxicity examination and clinical trial phases using already approved drugs to cure disease, which can greatly save time and money, as well as reduce the failure rate. Alberto Paccanaro et al. proposed two methods specifically for the repurposing of drugs for COVID-19. One of these methods involves utilizing a matrix decomposition algorithm to prioritize broad-spectrum antiviral drugs. The second method is based on network medicine and introduces graph kernel methods to rank drugs[1]. This strategy is dependent on the disruptions induced by drugs in the human interactome sub-networks, specifically those that are vital for SARS-CoV-2 infection and replication. Sun et al. [12] introduced an approach named AdaDR, which intends to enhance the effectiveness of drug repositioning through the comprehensive integration of node features and topological structures. Unlike traditional graph convolutional networks, AdaDR models the relationship

between topological structures and deeply integrated node characteristics, in this way, can greatly improve the expressive ability of the model. As a new model, AdaDR suggests that the correlation between drug diseases can be utilized in drug repurposing [9].

4. Challenges and Future Outlook

In spite of the possible merits that AI offers in the realm of drug development, it is essential to consider several challenges and limitations that are inherent to its application. Among the challenging problems is the lack of adequate data [13]. AI-driven methods often require an extensive amount of data for training. Accordingly, the quality, completeness, and consistency of data can determine the reliability as well as accuracy of the outputs [14]. Concerns about ethics provide a further challenge [15], as early AI-based techniques lack explainability; they are frequently perceived as 'black boxes', which do not explain the reason behind their predictions [1]. That's why it is difficult for people to have confidence in the decisions generated by AI. To tackle this issue, an explainable AI was developed, such as E-AI [17], to provide an explanation of the decisions or predictions for humans in a straightforward way. Notwithstanding AI has its demerits, and there is always room for further improvement. AI holds the promise of transforming the pharmaceutical sector by accelerating the discovery and development of drugs. Looking ahead, further use of AI in drug discovery can invent drugs for the treatment of incurable diseases with a greater possibility, so that people can no longer suffer from diseases.

5. Conclusion

It is commonly recognized that the drug discovery process is a tedious, costly, and complex one. Using conventional techniques, it frequently takes several trials to get a molecule from the lab to the market as a useful medication. Artificial intelligence (AI), on the other hand, presents a potentially quicker, more economical, and more successful way to expedite the drug discovery process. Although an amount of research has proposed using AI for drug discovery, comprehensive assessments of these advancements are few. This review presents a comprehensive and systematic analysis of the diverse applications of artificial intelligence in the field of drug discovery. It addresses the application of AI to molecular docking and virtual screening, as well as the repositioning of current medications via pathway analysis, lead compound identification, biomarker research, target identification, and a comparison between multiple AI techniques. The study also explains how to take advantage of AI for upcoming drug design schemes and addresses the expected obstacles and restric-

---

1 The experiment was conducted by Alberto Paccanaro came from Network Medicine - Paccanaro Lab.

tions of AI in drug research.

References

[1]Pushkaran A C, Arabi A A. From understanding diseases to drug design: can artificial intelligence bridge the gap? Artificial Intelligence Review, 2024, 57(4): 1

[2]Patel U K, Anwar A, Saleem S, et al. Artificial intelligence is an emerging technology in the current care of neurological disorders. Journal of Neurology, 2019, 268(5): 1623–1642.

[3]Nilsson N J. Artificial intelligence: A modern approach. Artificial Intelligence, 1996, 82(1–2): 369–380.

[4]Vatansever S, Schlessinger A, Wacker D, et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions. Medicinal Research Reviews, 2020, 41(3): 1427–1473.

[5]Haenssle H A, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 2018, 29(8): 1836–1842.

[6]Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017, 542(7639): 115–118.

[7]Brock, G. N., Mukhopadhyay, P., Pihur, V., Webb, C., Greene, R. M., & Pisano, M. M. MmPalateMiRNA, an R package compendium illustrating the analysis of miRNA microarray data. Source code for biology and medicine, 2013, 8(1): 1.

[8]Vemula D, Jayasurya P, Sushmitha V, et al. CADD, AI, and ML in drug discovery: A comprehensive review. European Journal of Pharmaceutical Sciences, 2023, 181(2): 106324.

[9]Bijral R K, Singh I, Manhas J, et al. Exploring Artificial Intelligence in Drug Discovery: A Comprehensive Review. Archives of Computational Methods in Engineering, 2021, 29(4): 2513–2529.

[10]Turing A. Computing machinery and intelligence. Mind, 1950, LIX (236): 433–460.

[11]Talele T, Khedkar S, Rigby A. Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. Current Topics in Medicinal Chemistry, 2010, 10(1): 127–141.

[12]Sun X, Jia X, Lu Z, et al. Drug repositioning with adaptive graph convolutional networks. Bioinformatics, 2023, 40(1): 1-10.

[13]Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 2019, 18(6): 463–477.

[14]Gómez-Bombarelli R, Wei J N, Duvenaud D, et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Central Science, 2018, 4(2): 268–276.

[15]Basu T, Engel-Wolf S, Menzer O. The Ethics of Machine Learning in Medical Sciences: Where Do We Stand Today? Indian J Dermatol. 2020, 65(5): 358-364.

[16]Minh D, Wang H X, Li Y F, et al. Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review, 2021, 55(5): 3503–3568.

[17]Kannappan B, R M N J, N S, et al. Recent Advances of Artificial Intelligence in Drug Discovery Process. 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), 2022, 245-249

[18]Murugan N A, Priya G R, Sastry G N, et al. Artificial intelligence in virtual screening: Models versus experiments. Drug Discovery Today, 2022, 27(7): 1913–1923.

[19]Carpenter K A, Cohen D S, Jarrell J T, et al. Deep learning and virtual drug screening. Future Medicinal Chemistry, 2018, 10(21): 2557–2567.

[20]Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science, 2018, 361(6400): 360–365.