International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

AI-Driven Drug Discovery: A Comprehensive Review

Garima Sharma

M. Phil, MSc. Organic Chemistry(Education), Garima Sharma, PGT Chemistry,Presidium Indirapuram, Ghaziabad

Abstract
Artificial intelligence (AI) has become a part of modern life and continues to pique people's interests. AI enables computers and robots to learn from past behaviour and mistakes in the same way that people can. AI has the ability to make complicated drug development procedures quicker and more cost-effective, with the goal of lowering the time it takes for a new treatment to reach the patient. In small-molecule drug discovery, AI can add value in four ways: access to new biology, improved or original chemistry, higher success rates, and faster and cheaper discovery methods. Many issues and limits in traditional R&D can be addressed by the technology.The application of artificial intelligence (AI) in medicine, particularly through machine learning (ML), marked a significant progression in drug discovery. AI acts as a powerful catalyst in narrowing the gap between disease understanding and the identification of potential therapeutic agents. This review provides an inclusive summary of the latest advancements in AI and its application in drug discovery.

Keywords: Artificial intelligence, Drug discovery, AI-assisted content generation, AI-limitations, Machine learning, Disease identification, Drug design, Drug discovery.

Introduction

The traditional drug discovery process is a complex and challenging endeavor that can require up to 15 years and over $1 to 2 billion for each approved drug.

The use of artificial intelligence (AI) in medicinal chemistry has gained significant attention in recent years as a potential means of revolutionizing the pharmaceutical industry . Drug discovery, the process of identifying and developing new medications, is a complex and time-consuming endeavor that traditionally relies on labor-intensive techniques, such as trial-and-error experimentation and high-throughput screening. However, AI techniques such as machine learning (ML) and natural language processing offer the potential to accelerate and improve this process by enabling more efficient and accurate analysis of large amounts of data . The successful use of deep learning (DL) to predict the efficacy of drug compounds with high accuracy has been described recently by the authors of . AI-based methods have also been able to predict the toxicity of drug candidates . These and other research efforts have highlighted the capacity of AI to improve the efficiency and effectiveness of drug discovery processes. However, the use of AI in developing new bioactive compounds is not without challenges and limitations. Ethical considerations must be taken into account, and further research is needed to fully understand the advantages and limitations of AI in this area . Despite these challenges, AI is expected to significantly contribute to the development of new medications and therapies in the next few years.

Artificial intelligence (AI) and machine learning (ML) offer a promising path toward increased efficiency and success rates in drug development, providing the pharmaceutical industry with a solution with AI/ML implementation to correct limitations while also opening up novel opportunities using new model implementations based on AI parameters.

AI-based disease identification(Fig.1)

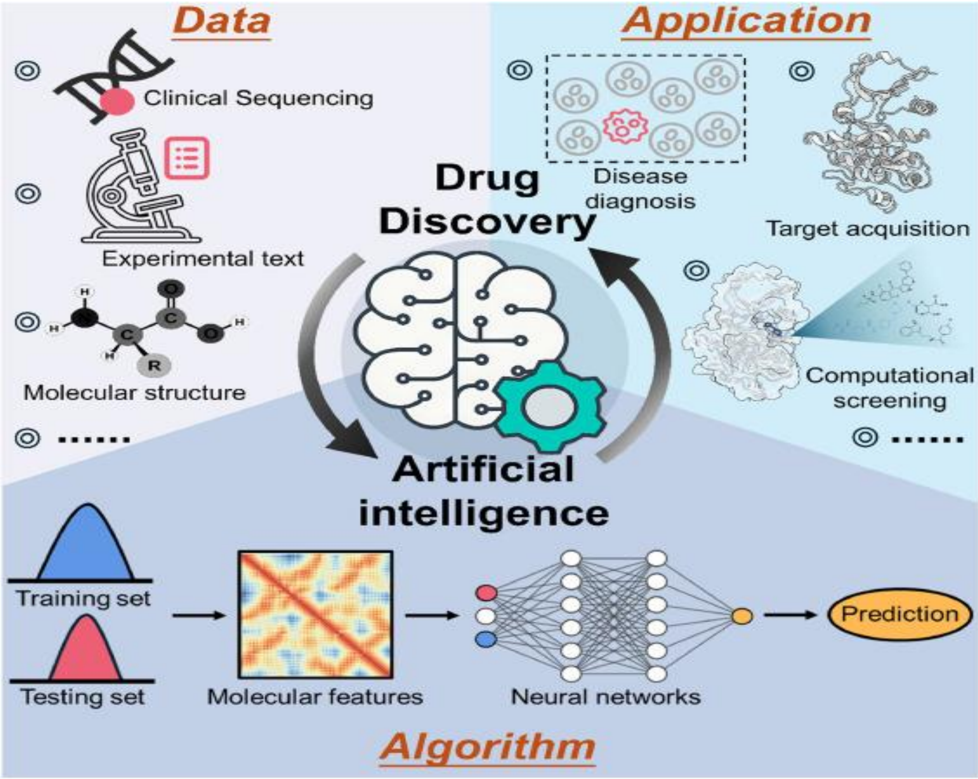

Fig.1

AI-based disease detection uses machine learning (ML) and deep learning (DL) to analyze vast medical data (images, records, labs) to find patterns, enabling faster, more accurate, and earlier diagnosis for conditions from cancer and heart disease to rare illnesses, often predicting disease risk years in advance by identifying subtle biomarkers and patterns invisible to humans, enhancing clinical decision-making and personalized treatment.

How it works

Data ingestion: AI models ingest diverse data: medical images (X-rays, MRIs), lab results, genetic data, EHRs, and even lifestyle factors.

Pattern Recognition: Algorithms, especially Convolutional Neural Networks (CNNs) for images, learn complex patterns associated with specific diseases.

Prediction & Diagnosis: Trained models identify these patterns in new patient data to flag potential illnesses, often before symptoms appear, or to help differentiate between similar conditions.

Explainability: Some models provide "treasure maps" highlighting which image areas or data points contributed most to a diagnosis, increasing trust and allowing doctors to verify.

Target identification

Key Aspects of AI in Target Identification:

- Multi-omic Data Integration: AI algorithms analyze data from genomics, proteomics, and transcriptomics, including datasets like the Cancer Cell Line Encyclopedia (CCLE), to identify disease-related molecules.
- Network-Based Approaches: AI maps interactions within cellular systems to identify critical nodes (potential targets) that, when modulated, can disrupt disease pathways.
- Generative AI and NLP: Natural Language Processing (NLP) parses millions of scientific publications and grants to extract hidden relationships, while generative AI can hypothesize novel targets.
- Precision Medicine: AI helps identify personalized targets by analyzing individual patient data, improving the specificity of treatments.
- Accelerating Discovery: AI-driven platforms like Pharma.AI have already produced novel drugs for clinical trials, such as for idiopathic pulmonary fibrosis, by automating the identification process.
- Target Assessment: Beyond identification, AI evaluates potential targets based on safety, druggability, and novelty.

AI tools are increasingly integrated into existing,, conventional drug discovery workflows to increase efficiency and reduce the,, high failure rates associated with manual target identification.

AI-enabled virtual screening in drug discovery: opportunities and challenges

The initial phase of drug discovery usually involves computational screening of numerous compounds to identify those with the desired cellular or biochemical effects. To enhance the speed, efficiency, and cost-effectiveness of this process, new methods are constantly being developed. A positive response during the first round of screening in a biochemical assay identifies primary "hit" compounds. Subsequently, additional screening is performed to assess whether the physicochemical and pharmacological properties of the hit compounds are suitable for developing a medicine. If they pass this filter, they are designated as "leads". These leads are then refined chemically and subjected to biological screening in subsequent rounds before proceeding to clinical testing.

The conventional drug discovery process involves synthesizing and testing thousands of compounds, which is both time-consuming and costly, requiring large amounts of protein supplies and established laboratory methods for bioactivity testing. In contrast, VS has emerged as a cost-effective approach to scan millions of commercially available compounds and prioritize those for further testing, synthesis, or purchase. VS methods are classified into two categories: structure-based and ligand-based methods.

Prediction of drug toxicity with AI:

Proportion of drug candidates being discarded during clinical trials due to unexpected adverse effects. Predicting drug toxicity during preclinical stages is a crucial step to reduce the failure rate and improve the efficiency of drug discovery. Traditional methods of predicting drug toxicity are limited by their reliance on small datasets and simplistic models. However, AI-based approaches have emerged as promising alternatives.

AI-powered drug toxicity prediction accelerates drug discovery by using machine learning and deep learning to identify toxic compounds early, replacing traditional, costly, and time-consuming in vivo animal studies. These models, including architectures like DeepTox, analyze chemical structures and molecular data to predict hepatotoxicity, cardiotoxicity, and other hazards.

Key aspects of AI in drug toxicity prediction:

AI-powered drug toxicity prediction accelerates drug discovery by using machine learning and deep learning to identify toxic compounds early, replacing traditional, costly, and time-consuming in vivo animal studies. These models, including architectures like DeepTox, analyze chemical structures and molecular data to predict hepatotoxicity, cardiotoxicity, and other hazards.

Key aspects of AI in drug toxicity prediction:

AI-powered drug toxicity prediction accelerates drug discovery by using machine learning and deep learning to identify toxic compounds early, replacing traditional, costly, and time-consuming in vivo animal studies. These models, including architectures like DeepTox, analyze chemical structures and molecular data to predict hepatotoxicity, cardiotoxicity, and other hazards.

Key aspects of AI in drug toxicity prediction:

- Key Methodologies: AI tools analyze molecular descriptors and 2D/3D structures to predict toxicity. Major techniques include deep neural networks (DNNs), Random Forest, and ensemble models, often trained on data from initiatives like Tox21.
- Targeted Endpoints: AI models focus on critical safety endpoints, including hepatotoxicity (liver), cardiotoxicity (heart), nephrotoxicity (kidney), neurotoxicity (nerve), and genotoxicity (genetic damage).
- Tools and Data Sources: Commonly used tools for prediction include LimTox, pkCSM, admetSAR, Toxtree, Opera software, DeepTox, and TOXSCAPE.

<u>AI and gene editing technologies for developing gene therapiesAI and gene editing technologies for developing gene therapies(Fig.2)</u>

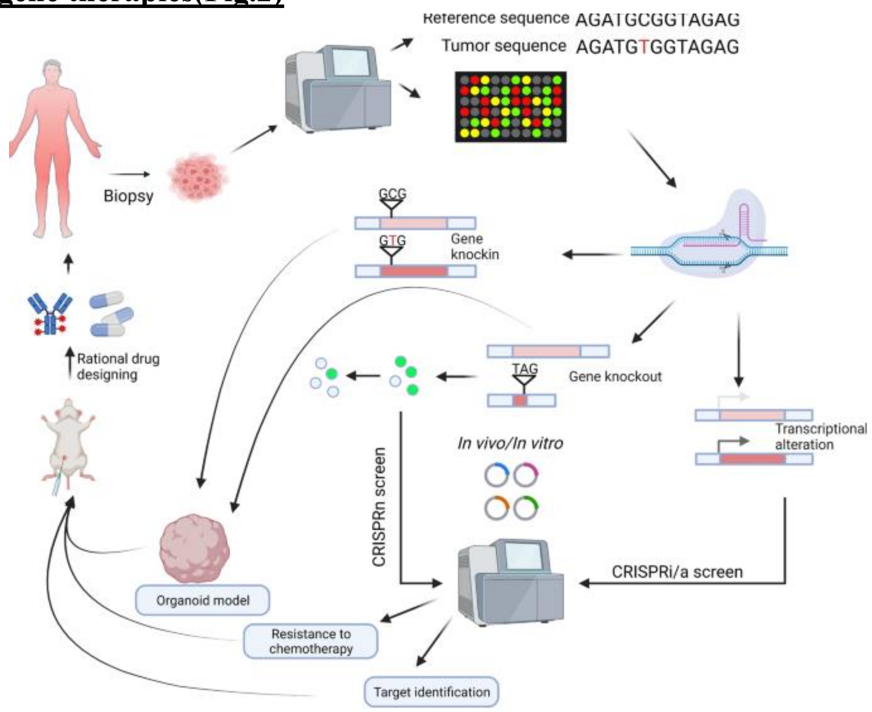

Fig.2

Various aspects of GED. CRISPR encompasses various gene-editing approaches. The most well-known is CRISPR-Cas9, utilizing Cas9 endonuclease guided by sgRNA to target and modify specific DNA sequences. Numerous other Cas nucleases have been identified in the recent past. CRISPR-Cas12 offers a different target site and has diagnostic advantages. Cas13 functions as an RNA-guided RNA endonuclease, specifically targeting and cleaving RNA. Base editing allows precise changes without double-strand breaks, while Prime editing enables versatile DNA sequence modifications. Epigenome editing controls gene expression via epigenetic marks, while CRISPRi and CRISPRa regulate gene expression without altering DNA. These techniques each have distinct applications in genetics and medical research, selected based on specific objectives and genetic contexts.

AI-based modeling for personalized drug dosing

Accurate calculation of the dose is one of the key responsibilities of pharmaceutical practice. Human or computational errors in the determination of the dose can compromise drug or formulation efficacy, cause toxicity, or reduce patient adherence to the treatment. Over the years, dosage calculations have relied on established equations, PK/PD modeling, and the experience of the clinician. With the advancement, computational systems can integrate multiple data streams, such as patient demographics, genetic factors, comorbidities, and drug–drug interactions, to optimize dosage in a patient-specific manner. Thus, AI is bridging the gap between generalized dosing guidelines and precision medicine.

By analyzing patient data, the AI algorithms, such as ANNs and BO models, are gradually applied to individualize drug dosage. For this purpose, the system will use parameters like age, body surface area, renal function (RF) and hepatic function (HF), and genetic polymorphisms (e.g., CYP450 variants) to recommend individualized doses.

An adaptive dosing system that adjusts prescriptions dynamically is another example of advancement. Gadgets such as wearable sensors, continuous glucose monitors, and smart infusion pumps provide real-time PD/PK data, helping the AI algorithms to interpret these inputs to finalize dosages automatically.

As mentioned earlier, the use of AI-based wearable sensors enables continuous monitoring of therapeutic responses, enabling dose adjustments in real-time [76]. Insulin dosing for diabetics is an example that has been transformed by closed-loop systems based on AI. Moreover, “artificial pancreas” is one of the devices that maintain near-normoglycemia while reducing episodes related to hypoglycaemia.

The role of AI in rare disease research

AI works as an accomplice in integrating and examining diversified data. Diagnostic decision support systems effectively assist the medical practitioner by providing a list of relevant differential diagnoses. These systems have previously been effectively utilized for a variety of well-known use cases. Recently, it has been harnessed for the early detection and diagnosis of coronavirus disease 2019 (COVID-19) through monitoring of demographic, clinical, and epidemiological characteristics of patients.[6] These systems are also useful for implementation for rare diseases (RDs). The RDs, also sometimes referred to as orphan diseases, can stand to benefit from quicker and more efficient diagnoses. Algorithms have been designed and are already used to compile networks and register information through patients on rare diseases to identify new cases.[6] For instance, a combination of brain function and structural imaging data can be harnessed to determine whether a person with Huntington's disease (HD) will receive a clinical diagnosis within 5 years (pre-HD) or quantifiable assessments of oculomotor function preceding HD.[6] These use cases show promising potential for future utilization of AI in rare disease diagnoses.

Types of algorithms and associated benefits for disease diagnosis

1. Different AI algorithms have appreciable benefits in aiding in the diagnoses of RDs and non-RDs.[7] The ML helps in diagnosis via three types of algorithms: (a) unsupervised which works by identifying patterns, (b) supervised which classifies or predicts decisions based on former examples, and (c) reinforcement learning which uses reward and punishment processes to form a blueprint for operating in a definite obstacle. The first-generation AI works on making clinical decisions by analyzing large amounts of data for diagnosing the RDs. However, a key issue for RDs is a lack of large amounts of initial data which can be inputted. Whereas the second-generation AI works by looking for the clinical clues that are commonly overlooked in the early course of the identification of RDs. This algorithm efficiently helps in the early diagnosis for intervention and prevention of various RDs.[4] Detecting and aptly reading the diagnostic image is a necessity that is increasingly leading to the recognition of ML in AI for RDs.[1] ML, for example, has recently been utilized to identify which patients with systemic sclerosis are at a high risk of severe complications, early detection of organ involvement, and more.[3]
2. AI is an integral tool for the diagnosis of RDs as it can assist in image recognition, genetic analysis, and support clinical decision making. ML is a branch of AI that has demonstrated its effectiveness in systemic sclerosis diagnosis. ML is able to detect pulmonary involvement better than high-resolution computerized tomography (HRCT) and pulmonary function tests (PFTs), which currently serve as the standard diagnostic methods. Although HRCTs tend to be used when PFTs indicate a decline in respiratory function, ML algorithms may detect pulmonary involvement prior to the onset of deterioration resulting in improved survival rates and reduced health costs.[3] Deep learning (DL) represents a subset of ML algorithms involved in image recognition. DL algorithms break down complex mappings into clusters of simpler mappings to facilitate more efficient analysis. The algorithm can then use its visible layer to read the image and its hidden layers to extract important features from the image. By doing so, DL is able to provide superior recognition with extensive and high-dimensional data. In some cases, these images may contain important clinical features that can aid in rapid and accurate diagnosis. Compared with more traditional biometric methods, AI has shown greater flexibility and scalability which allows it to contribute to understanding complex relationships, improving early detection, and making routine tasks more efficient.

Conclusion

The use of AI technology in drug design has grown rapidly due to its predictive ability and accuracy. This review highlights the numerous applications of AI in all phases of drug development, from disease diagnosis to post-marketing analysis. AI helps in the early prediction of diseases, the development of personalized medicine, optimization of drug doses, and the prediction of treatment outcomes.

AI-driven approaches improve patient stratification, recruitment, monitoring, and follow-ups in clinical trials, and can even assist in FDA approvals and pharmacovigilance. The integration of AI in drug design has resulted in faster drug discovery, cost savings, reduced resource and manpower usage, and decreased attrition rates in clinical trials. Additionally, AI helps to minimize the use of *in vivo* bioassays, reducing animal sacrifice. AI has far-reaching applications beyond medicine, including healthcare management, surgeries, mRNA vaccination, preventive treatments, and nutrigenomics. However, it is important to note that AI models are meant to complement human intelligence, not replace it. AI models may have comparable or better predictive ability than human researchers, but they still lack human intuition. Predictions made by AI machines must be verified by humans, as AI models can provide false positive and false negative results, compromising the sensitivity and specificity of the model. Additionally, resource sustainability needs holistic solutions like cost-aware cross-layer co-design, integrating hardware, algorithms, and models for efficient exploration of resource-sustainable configurations.

AI has the potential to revolutionize the drug discovery process, offering improved efficiency and accuracy, accelerated drug development, and the capacity for the development of more effective and personalized treatments (Fig.3). However, the successful application of AI in drug discovery is dependent on the availability of high-quality data, the addressing of ethical concerns, and the recognition of the limitations of AI-based approaches.

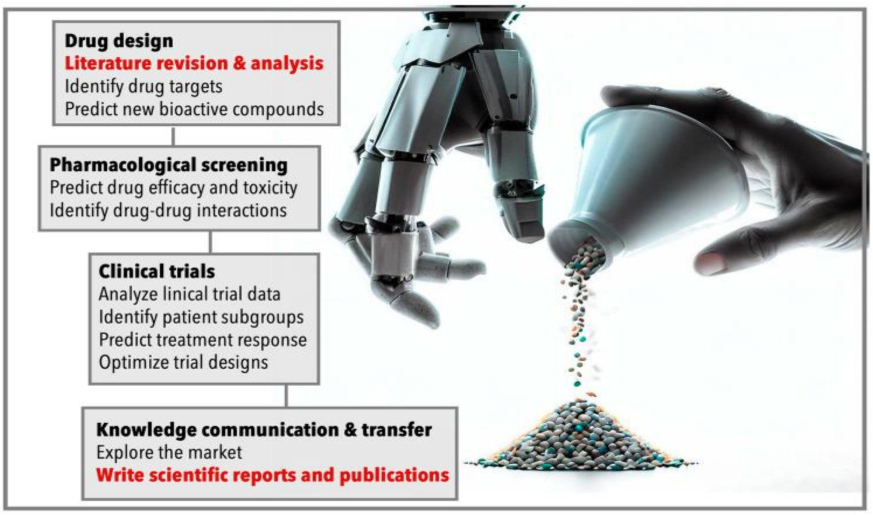

Fig.3

References

1. R. Barker A flexible blueprint for the future of drug development Lancet, 375 (2010), pp. 357-359
2. D. Sun, W. Gao, H. Hu, *et al.* Why 90% of clinical drug development fails and how to improve it? Acta. Pharm. Sin. B, 12 (2022), pp. 3049-3062
3. H. Dowden, J. Munro Trends in clinical success rates and therapeutic focus Nat. Rev. Drug Discov., 18 (2019), pp. 495-496
4. J.D. Bolcer, R.B. Hermann The development of computational chemistry in the United States Rev. Comput. Chem. (1994), pp. 1-63
5. Y. Ding, J.H. Sohn, M.G. Kawczynski, *et al.* A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain Radiology, 290 (2019), pp. 456-464
6. A. Narin, C. Kaya, Z. Pamuk Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks
7. F. Gentile, J.C. Yaacoub, J. Gleave, *et al.* Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking Nat. Protoc., 17 (2022), pp. 672-697
8. K.A. Carpenter, D.S. Cohen, J.T. Jarrell, *et al.* Deep learning and virtual drug screening Future Med. Chem., 10 (2018), pp. 2557-2567

9. M. Bule, N. Jalalimanesh, Z. Bayrami, *et al.* The rise of deep learning and transformations in bioactivity prediction power of molecular modeling tools Chem. Biol. Drug Des., 98 (2021), pp. 954-967
10. C. Berzuini, R. Bellazzi, S. Quaglini, *et al.* Bayesian networks for patient monitoring Artif. Intell. Med., 4 (1992), pp. 243-260
11. S. Kapsiani, B.J. Howlin Random forest classification for predicting lifespan-extending chemical compounds Sci. Rep., 11 (2021), pp. 1-13
12. K. Heikamp, J. Bajorath Support vector machines for drug discovery Expert Opin. Drug Discov., 9 (2014), pp. 93-104
13. F. Zhong, J. Xing, X. Li, *et al.* Artificial intelligence in drug design Sci. China Life. Sci., 61 (2018), pp. 1191-1204
14. N. Brown, P. Ertl, R. Lewis, *et al.* Artificial intelligence in chemistry and drug design J. Comput. Aided Mol. Des., 34 (2020), pp. 709-715
15. Y. Shen, T. Liu, J. Chen, *et al.* Harnessing artificial intelligence to optimize long-term maintenance dosing for antiretroviral-naive adults with HIV-1 infection Adv. Ther. (Weinh), 3 (2020), Article 1900114
16. M. Zhang, Q. Su, Y. Lu, M. Zhao Application of machine learning approaches for protein-protein Interactions prediction Med. Chem., 13 (2017), pp. 506-514