Vol. 1, Issue 11 | Nov 2026 | Open Access — Research Paper

ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY: APPLICATIONS OF MACHINE LEARNING, DEEP LEARNING, GENERATIVE AI AND COMPUTATIONAL APPROACHES

Pushp Sachin Garg[1], Deshpande Girish[1], Swati Sabne[1], Gayakwad Monali[1], More Gayatri[1], Mugale V.S[2,*]

[1]Department of Pharmacy, Shivlingeshwar College of Pharmacy, Almala, Latur, Maharashtra, India

[2]Assistant Professor, Department of Pharmacy, Shivlingeshwar College of Pharmacy, Almala, Latur, Maharashtra, India

*Corresponding author: vikasmugale@gmail.com

***Abstract*—**Artificial intelligence (AI) has been demonstrated to be a promising tool in modern drug discovery due to its ability to overcome the limitations of traditional drug discovery, including high cost, long lead time, and high attrition rate. The use of AI in drug discovery has been realized through machine learning, deep learning, and generative models, which allow efficient analysis of the huge amount of data concerning different aspects of the drug-discovery process. There are numerous applications of AI in different stages of drug discovery, such as target identification and validation, protein structure prediction, virtual screening, prediction of physicochemical properties, de novo design of molecules, optimization of drug-like molecules, drug repurposing, and prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET). The application of AI in drug discovery has been facilitated by graph neural networks, transformer-based models, and generative artificial intelligence (GAI), which have enhanced the possibility of designing therapeutic molecules. Several challenges remain, including limited data availability and quality, model interpretability, generalizability, model accuracy, computational power, and the need for clinical trials and regulatory approval. Ethical concerns regarding AI training data and regulation also remain. This review highlights the key applications, recent advances, benefits, and limitations of AI in drug discovery to illustrate its potential in accelerating drug discovery and development to generate safer and more effective therapeutics.

Keywords*—Machine learning, Deep learning, Generative artificial intelligence, Virtual screening, Molecular design*

I. INTRODUCTION

The drug-discovery process is a complex and multidisciplinary one encompassing target identification, hit discovery, lead optimization, preclinical evaluation, and clinical development. Traditional drug discovery is a time- and resource-intensive endeavor due to the exhaustive process of screening and validation that every molecule must undergo. The growing intricacy of diseases and the sheer volume of chemical and biological information have resulted in a pressing need for more sophisticated computational approaches to drug discovery [1].

AI has proven to be a valuable tool that can disrupt multiple areas of the drug-discovery pipeline. AI encompasses a wide range of data-driven methods and processes capable of learning, analyzing, and reasoning from information to mimic human cognitive abilities and assist in research and decision-making [2].

Machine learning (ML), deep learning (DL), graph neural networks (GNNs), and generative AI are some areas of AI that can be leveraged to analyze large-scale datasets of chemicals, biology, genomics, proteomics, and 3D structures. This technology can recognize patterns in data, detect associations, and build models that traditional computer algorithms and empirical methods cannot [3].

AI has been utilized in target identification and validation, protein folding, virtual screening, QSAR analysis, predicting physicochemical properties of molecules, de novo design, lead optimization, ADMET prediction, toxicity analysis, and drug repurposing. Generative AI, in particular, has been utilized for the design of novel molecular structures with desirable properties and behaviors [4].

II. EVOLUTION OF AI IN DRUG DISCOVERY

The development of artificial intelligence (AI) in drug discovery has evolved from early computer-aided drug design (CADD) to modern machine learning and generative AI. Initially, computer-aided methods such as molecular docking, pharmacophore identification, molecular dynamics simulations, and quantitative structure–activity relationship (QSAR) analysis were utilized to elucidate molecular interactions and predict biological activity. Although these approaches improved the efficiency of drug discovery, they were often limited by reliance on predefined rules, molecular descriptors, and mathematical relationships [6].

The introduction of machine learning (ML) provided a more data-driven approach to drug discovery. Various ML algorithms including decision trees, random forests, support vector machines, and regression have been applied for predicting molecular activity, drug–target interactions, physicochemical properties, and toxicity. These algorithms have enabled the identification of patterns in large chemical and biological data sets [1].

The advent of deep learning (DL) has further advanced the capabilities of AI in pharmaceutical research. Using neural-network architectures, DL algorithms enable the automatic learning of complex molecular representations. In particular, graph neural networks have proven useful for representing molecular structures as graphs of atoms and bonds. Moreover,

deep learning has facilitated the prediction of protein structures and molecular interactions [2].

In recent years, the field has seen a shift towards generative AI, where the focus has turned from predictive modeling to generative modeling of molecular structures. This evolution highlights a transition from traditional computational approaches to a more data-driven and generative paradigm in drug discovery, although it should be noted that experimental validation is still required for the successful translation of AI-generated molecules into drug candidates [3].

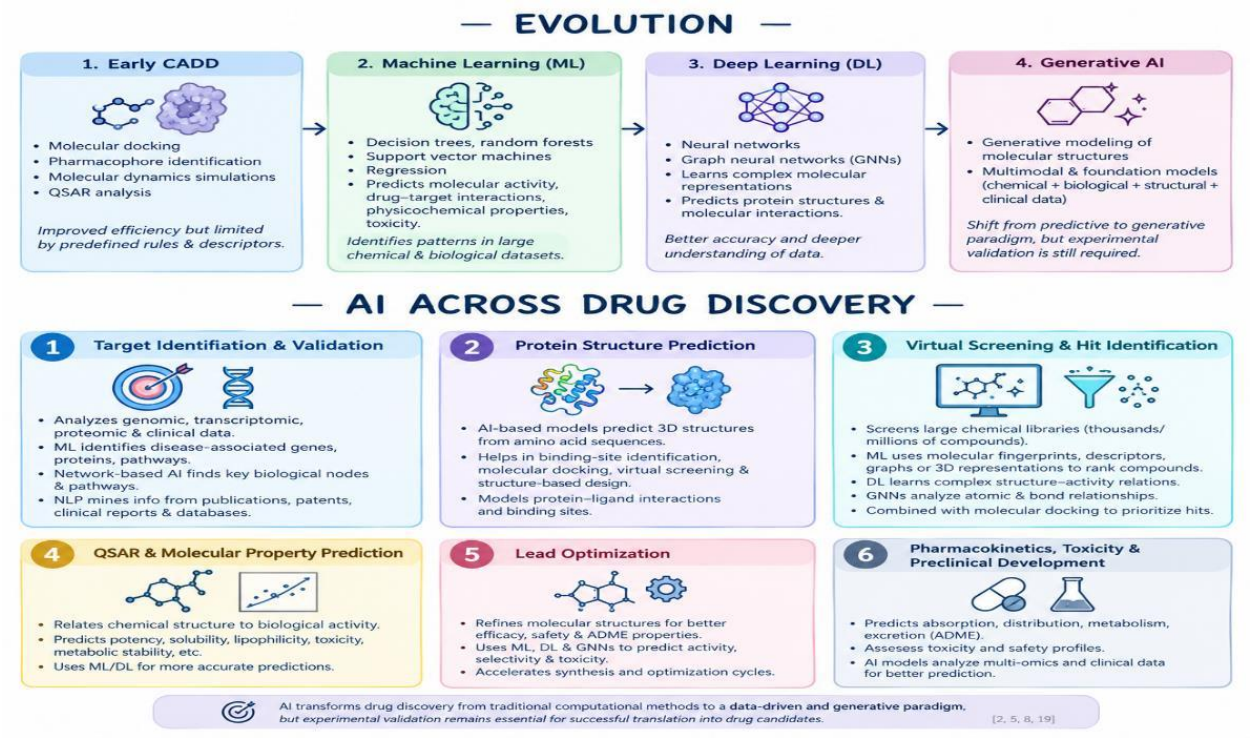


Fig. 1. Evolution of AI approaches in drug discovery and their application across the drug-discovery pipeline [1]–[3], [6], [7].

III. AI ACROSS THE DRUG-DISCOVERY PIPELINE

Artificial intelligence (AI) has become an important technology across multiple stages of the drug-discovery process, ranging from identification of therapeutic targets to optimization of potential drug candidates. The conventional drug-discovery pipeline involves several interconnected stages, including target identification and validation, hit identification, virtual screening, lead optimization, pharmacokinetic and toxicity assessment, and preclinical development. Each stage can involve the analysis of large quantities of chemical and biological information by means of machine learning (ML), deep learning (DL), graph neural networks (GNNs), and generative artificial intelligence [7].

A. Target Identification and Validation

Target identification is the first step of drug discovery and involves identifying biological molecules amenable to being modulated to yield a therapeutic effect. AI can analyze large-scale genomic, transcriptomic, proteomic, and clinical datasets to prioritize disease-associated genes, proteins, pathways, and molecular interactions. Machine-learning models can identify relationships between molecular features and disease phenotypes, thereby helping to prioritize potential therapeutic targets. Network-based AI approaches can identify important biological nodes and pathways involved in disease progression [4].

B. Protein Structure Prediction

Understanding the three-dimensional structure of a target protein is important for structure-based drug discovery. Experimental techniques such as X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy yield structural information but can require considerable time and resources. AI-based protein structure prediction has offered an alternative approach using deep-learning models to predict three-dimensional protein structures from amino acid sequences. AI can assist in modeling protein–ligand interactions and identifying potential binding sites [8].

C. Virtual Screening and Hit Identification

Virtual screening is used to identify potentially active compounds from large chemical libraries before experimental testing. Traditional screening methods can require substantial time and resources because thousands or even millions of compounds may need to be evaluated. AI can improve this process by identifying which molecules are more likely to interact with a target or display a desired biological activity [1].

Machine-learning models can use molecular fingerprints, physicochemical descriptors, molecular graphs, or three-

dimensional representations to rank compounds. Deep-learning models can learn complex relationships between molecular structures and biological activity, while graph neural networks can directly analyze the relationships between atoms and chemical bonds. AI can be combined with molecular docking to prioritize candidate molecules [9].

D. QSAR and Molecular Property Prediction

Quantitative structure–activity relationship (QSAR) modelling is a well-established computational approach for relating chemical structure to biological activity. AI-based QSAR models can analyze molecular structures and predict properties such as potency, solubility, lipophilicity, permeability, and chemical stability [10].

Graph neural networks (GNNs) are particularly useful because they represent molecules as graphs, with atoms serving as nodes and chemical bonds as edges. This enables models to learn structural features directly from molecular graphs. Accurate molecular property prediction can help prioritize compounds with favorable characteristics prior to synthesis and biological testing [10].

E. De Novo Drug Design

One of the most significant developments in AI-based drug discovery is the use of generative models for de novo molecular design. Conventional medicinal chemistry often modifies known chemical scaffolds, whereas generative AI can propose entirely new molecular structures based on learned patterns in chemical space [3].

Generative approaches include variational autoencoders, generative adversarial networks, reinforcement learning, diffusion models, and transformer-based architectures. These models can generate candidate molecules according to specified objectives, such as predicted biological activity, molecular stability, solubility, or drug-likeness. AI can therefore explore chemical structures that may not be present in existing compound libraries [11].

Multi-parameter optimization is particularly important, as successful drug candidates generally need to fulfill several requirements simultaneously. A molecule with high potency may fail due to poor solubility, unfavorable pharmacokinetics, or toxicity. AI-based generative approaches can incorporate multiple objectives into molecular generation and optimization, but generated molecules must still be tested for synthetic feasibility, chemical stability, biological activity, and safety through experimental studies [12].

F. Lead Optimization

Following the identification of promising hit compounds, lead optimization seeks to improve potency, selectivity, pharmacokinetic properties, and safety. Traditional lead optimization involves repeated cycles of chemical synthesis and biological testing. AI can help reduce the number of experimental cycles by predicting how structural modifications may influence molecular properties [1].

Machine-learning models can analyze structure–activity relationships and suggest chemical modifications that may improve desired properties. Generative models can also produce new analogues of existing lead compounds. In addition, AI can assist in multi-objective optimization by balancing potency with properties such as solubility, permeability, metabolic stability, and toxicity [12].

G. ADMET and Toxicity Prediction

Absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties are critical factors determining whether a drug candidate can progress through development. Poor pharmacokinetic characteristics and toxicity are major causes of failure during drug development. AI-based prediction can help identify potential issues at an early stage [7].

Machine-learning and deep-learning models have been developed to predict intestinal absorption, blood–brain barrier permeability, plasma protein binding, metabolic stability, cytochrome P450 interactions, clearance, and various toxicity endpoints. Early prediction of these characteristics can allow researchers to eliminate unsuitable compounds or modify their structures before extensive experimental testing [10].

IV. AI TECHNIQUES

Artificial intelligence in drug discovery comprises numerous computational techniques that differ in their capacity to process chemical, biological, and structural information. The choice of a particular approach depends on the nature of the drug-discovery task, the data set used, and specific goals such as predicting or designing candidate molecules [1].

A. Machine Learning

Machine learning (ML) allows computers to train mathematical models on a particular data set and use these models to predict the output values for new, unseen examples. The algorithms most commonly used in the field of drug discovery are random forests, support vector machines, decision trees, gradient boosting, and regression [1].

B. Deep Learning

Deep learning (DL) is a part of ML that uses neural networks with multiple layers to study the input data. There are different types of neural networks, including convolutional networks that are capable of analyzing visual images and recognizing patterns, and recurrent networks that can process sequential data. Applied to drug discovery, DL allows representing molecular data by features learned in each layer and identifying patterns at various levels [2].

C. Graph Neural Networks

Graph neural networks (GNNs) are primarily used for studying properties of molecules since molecular structures can be easily represented as graphs. In a graph, molecular elements (atoms) are represented by nodes, while their interactions (bonds) are represented by edges [10].

D. Reinforcement Learning

Reinforcement learning (RL) is an ML type that is based on improving an agent's performance using rewards and punishments. Using RL, it is possible to train an agent to design molecules by rewarding promising synthetic pathways and penalizing unfavorable outcomes [11].

E. Generative AI and Transformers

Generative AI is a powerful tool that can be used for generating new molecular structures and designing molecules with specific properties. The main idea is to use variational autoencoders, generative adversarial networks, diffusion, and transformer-based algorithms to create new molecules [14].

V. APPLICATIONS

Artificial intelligence (AI) can be applied to various stages of the drug-discovery process. It can help researchers identify targets, discover hits, and predict their properties. AI can be applied to target identification and validation, wherein it can analyze genomic, proteomic, transcriptomic, and clinical data to discover disease-associated genes, proteins, and biological processes. It can be used to predict protein structures, which can help in structure-based drug design [4].

Virtual screening is another key application of AI in drug discovery. In this stage, machine-learning and deep-learning algorithms can be used to identify potentially active compounds from chemical libraries. QSAR models can also help researchers predict the biological activity of candidate molecules. These models can predict relevant properties, such as solubility, lipophilicity, permeability, and other physicochemical attributes [9].

Researchers can use AI-driven de novo design to design novel molecules with desirable properties. AI can also help in lead optimization by predicting the effect of certain modifications on the potency, selectivity, pharmacokinetics, and drug-likeness of lead compounds. Through in silico ADMET prediction, researchers can identify compounds with undesirable absorption, distribution, metabolism, excretion, and toxicity properties. Finally, AI can also be used to identify new indications or applications for existing therapeutics through drug repurposing initiatives. By facilitating these and other processes, AI can help researchers make better decisions and shorten the drug-discovery timeline [3].

VI. CHALLENGES

Despite the potential advantages for the pharmaceutical industry, the application of artificial intelligence (AI) in drug discovery is associated with numerous challenges that may impact its reliability and practical implementation. One of the main concerns is the quality of data used for training algorithms. First, pharmaceutical data may contain missing values or inconsistencies that negatively affect the quality of results. Second, a significant amount of data is needed to train an AI model, which may be unavailable for some pharmaceutical data sets. Third, the data sets should also include relevant examples with diverse chemical structures to cover real-life cases and avoid bias in the results [7].

The use of AI in drug discovery may also be limited by the required amount of computational power. Advanced models, such as deep learning and generative adversarial networks, are often associated with the necessity of using powerful hardware for processing large amounts of data. From a practical perspective, computational power is not the primary issue, but the possibility of obtaining accurate and experimentally verified results. The fact that AI can predict certain molecules with high efficacy and desirable properties does not ensure their successful synthesis and biological evaluation [2].

Finally, it is worth noting that the application of AI in pharmaceutical settings cannot replace the traditional multidisciplinary approach to drug discovery and development. From this perspective, challenges related to the use of AI in the pharmaceutical industry include the necessity to involve clinicians, biologists, and pharmacologists in the process and consider their needs and requirements. These issues cannot be resolved without multidisciplinary collaboration and comprehensive consideration of all aspects of drug discovery and development [4].

VII. ETHICAL AND REGULATORY ISSUES

The introduction of artificial intelligence (AI) into the drug-discovery process creates several regulatory and ethical issues. First of all, the methods of regulation and control of the decision-making processes must be established since AI uses sophisticated algorithms in its work. Moreover, problems related to the quality and quantity of data to be used for the functioning of AI systems and the security of this data are presented. The use of clinical or patient-related data is associated with the risk of medical information security breaches, which may result in professional secrecy violations and discrimination [5].

In addition, the application of AI algorithms in decision-making systems requires data validation procedures and standardized documentation. The validation step is mandatory since different data sets, algorithms, and software may lead to different conclusions and, thus, to the appearance of various drugs with the same characteristics. Finally, the problem of intellectual property rights is presented. The issue of data set ownership or the appearance of the same candidate compounds identified by different algorithms needs to be regulated at the legislative level. Additionally, AI algorithms may have biases due to the data sets used, which is critical in drug discovery and development [7].

Thus, some of the main problems that should be taken into consideration in the context of AI application in drug discovery include data governance, transparency, validation, and human factors. For this reason, the discussed issues need to be addressed by stakeholders, including researchers, pharmaceutical companies, regulators, and ethics boards, to ensure that AI is used safely and responsibly in the development of new medicines [5].

VIII. FUTURE PERSPECTIVES

The future of artificial intelligence (AI) in drug discovery will be defined by fundamentally better models, bigger data, and more tightly integrated in-silico and experimental approaches. More specifically, more powerful generative and foundation AI models capable of encoding and operating on vast amounts of multimodal data are expected to emerge. These models will be crucial for designing drug molecules with desired properties (e.g., potency, selectivity, pharmacokinetic profile, toxicity, and synthetic accessibility) [3].

Multimodal AI is also expected to play an increased role in drug discovery, where foundation models will combine information from small molecules, protein structures, genomics, transcriptomics, biomedical literature, imaging, and clinical data, enabling a more comprehensive understanding of disease biology and better-informed drug-discovery decisions [5].

AI will become further entrenched in the drug-discovery workflow, working in concert with robotic laboratory platforms to form closed loops of hypothesis generation, testing, and refinement. Generative AI tools can work in tandem with laboratory automation to suggest promising molecules for synthesis, test their biological activity, and use these results to inform future predictions [7].

IX. CONCLUSION

The development of therapeutics is a long, complicated, and expensive process covering six stages, namely, target identification, virtual screening, molecular design, lead optimization, ADMET prediction, and drug repurposing. Recent advancements in artificial intelligence (AI), including machine learning, deep learning, graph neural networks, and generative models, have the potential to revolutionize therapeutic discovery by enabling the mining of vast amounts of information and the identification of patterns undetectable to humans. Nevertheless, the application of AI in drug discovery involves substantial challenges, such as data quality control, interpretability, generalizability, reproducibility, and experimental validation. To fully benefit from the advantages of AI applications while minimizing associated risks and costs, it is essential to combine these emerging technologies with traditional approaches and disciplines. The combination of cutting-edge technologies, improved data quality, and regulatory strategies will allow AI to fundamentally change the way therapeutics are discovered and enable the development of safer and more effective drugs, while reducing the time and cost of this lengthy process [7].

REFERENCES

[1] J. Vamathevan, D. Clark, P. Czodrowski, et al., "Applications of machine learning in drug discovery and development," Nat. Rev. Drug Discov., vol. 18, pp. 463–477, 2019.

[2] H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, and T. Blaschke, "The rise of deep learning in drug discovery," Drug Discov. Today, vol. 23, no. 6, pp. 1241–1250, 2018.

[3] Y. Jing, Y. Bian, Z. Hu, L. Wang, and X.-Q. Xie, "Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era," AAPS J., vol. 20, p. 58, 2018.

[4] H. Kim, E. Kim, I. Lee, B. Bae, M. Park, and H. Nam, "Artificial intelligence in drug discovery: a comprehensive review of data-driven and machine learning approaches," Biotechnol. Bioprocess Eng., vol. 25, no. 6, pp. 895–930, 2020.

[5] A. Bender, M. C. Thomas, J. W. Scannell, et al., "Artificial intelligence in drug discovery—what it is, where we stand and the path forward," Nat. Rev. Drug Discov., 2026.

[6] P. Schneider, W. P. Walters, A. T. Plowright, et al., "Rethinking drug design in the artificial intelligence era," Nat. Rev. Drug Discov., vol. 19, pp. 353–364, 2020.

[7] Y. Zhang, M. Mastouri, and Y. Zhang, "Accelerating drug discovery, development, and clinical trials by artificial intelligence," Med, vol. 5, no. 9, pp. 1050–1070, 2024.

[8] J. Jumper, R. Evans, A. Pritzel, et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, pp. 583–589, 2021.

[9] L. Xue and J. Bajorath, "Molecular descriptors in virtual screening," Methods Mol. Biol., vol. 156, pp. 255–266, 2000.

[10] Z. Wu, B. Ramsundar, E. N. Feinberg, et al., "MoleculeNet: a benchmark for molecular machine learning," Chem. Sci., vol. 9, no. 2, pp. 513–530, 2018.

[11] M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, "Molecular de-novo design through deep reinforcement learning," J. Cheminform., vol. 9, p. 48, 2017.

[12] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, et al., "Deep learning enables rapid identification of potent DDR1 kinase inhibitors," Nat. Biotechnol., vol. 37, pp. 1038–1040, 2019.

[13] J. M. Levin, T. I. Oprea, S. Davidovich, et al., "Artificial intelligence, drug repurposing and peer review," Nat. Biotechnol., vol. 38, pp. 1127–1131, 2020.

[14] P. Schwaller, T. Laino, T. Gaudin, et al., "Molecular Transformer: a model for uncertainty-calibrated chemical reaction prediction," ACS Cent. Sci., vol. 5, no. 9, pp. 1572–1583, 2019.

[15] K. Tunyasuvunakool, J. Adler, Z. Wu, et al., "Highly accurate protein structure prediction for the human proteome," Nature, vol. 596, pp. 590–596, 2021.