www.ijarp.com ISSN 2456-9992 Page: 01-10

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 01

ROLE OF ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY

Dr. Manoj B. Jograna*[3] Suhana Shaikh[1], Bhagyashree Godbole[1], Kumudini Kasar[1], Saniya Shaikh[1], Yogeeta Girwalker[1], Manoj Jograna[2], Rashid Ansari[2]

[1]Department of Pharmacology, School of Pharmacy and Technological Management, SVKM NMIMS Global University, Dhule, 424001.

[2]Department of Pharmaceutics, School of Pharmacy and Technological Management, SVKM NMIMS Global University, Dhule, 424001.

[3]Head of the Department School of Pharmacy and Technological Management, SVKM NMIMS Global University, Dhule, 424001.

Article Received: 13 December 2025, Article Revised: 01 January 2026, Published on: 20 January 2026

***Corresponding Author: Dr. Manoj B. Jograna**

Head of the Department School of Pharmacy and Technological Management, SVKM NMIMS Global University, Dhule, 424001.

DOI: https://doi-doi.org/101555/ijarp.6726

ABSTRACT

Artificial intelligence (AI) has emerged as a transformative tool in drug discovery and development, accelerating the identification of novel therapeutic targets, lead compounds, and optimized drug candidates. Machine learning (ML), deep learning (DL), natural language processing (NLP), and generative models enable the integration of chemical, biological, and clinical data to enhance prediction of drug-target interactions, ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, and efficacy. AI has demonstrated applications across target identification, virtual screening, de novo molecule design, pharmacokinetic modeling, and drug repurposing. Despite its potential, challenges such as limited high-quality datasets, interpretability, and ethical considerations remain. Future directions involve multimodal AI, explainable AI, and integration with high-throughput experimental validation to ensure safe and effective drug development. This chapter provides a comprehensive overview of recent advancements, methodologies, challenges, and future perspectives of AI in drug discovery.

KEYWORDS: Artificial intelligence, Machine learning, Deep learning, ADMET.

1. INTRODUCTION

Drug discovery is traditionally recognized as a long, costly, and high-risk endeavor, often requiring more than a decade and substantial financial investment—frequently exceeding billions of dollars—to advance a single therapeutic from initial concept to market approval [1]. The conventional drug discovery paradigm encompasses target identification, hit discovery, lead optimization, preclinical evaluation, and multiple phases of clinical trials. Each stage is labor-intensive, time-consuming, and prone to high attrition rates. According to recent reports, the probability of success for a drug entering clinical development is less than 10%, with late-stage failures contributing substantially to the economic burden [2]. Consequently, there has been an urgent demand for strategies that can enhance efficiency, reduce costs, and improve success rates in the pharmaceutical pipeline.

Artificial intelligence (AI) has emerged as a transformative technology capable of revolutionizing drug discovery by leveraging computational power to analyze complex biological, chemical, and pharmacological datasets. AI methods, including machine learning (ML), deep learning (DL), and natural language processing (NLP), have demonstrated considerable promise in accelerating various stages of drug development. Machine learning algorithms, such as support vector machines (SVM), random forests (RF), and gradient boosting machines, have been extensively applied to predict molecular properties, drug-target interactions, and potential off-target effects [3]. Deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models, further enable the identification of intricate patterns within high-dimensional datasets, such as chemical structures, protein sequences, and pharmacokinetic parameters [4,5]. These models are particularly valuable for predicting binding affinities, identifying allosteric sites, and estimating drug toxicity profiles with higher precision than traditional in vitro and in silico methods.

Natural language processing has expanded the capacity of AI to mine unstructured biomedical literature, clinical trial databases, patents, and electronic health records. NLP algorithms can extract critical mechanistic insights, identify potential drug candidates, and detect emerging trends in therapeutic research [6]. This capability reduces manual curation efforts and facilitates rapid knowledge synthesis, enabling pharmaceutical researchers to prioritize promising targets and compounds more efficiently. For instance, AI-driven literature mining has led to the identification of novel drug-repurposing opportunities, particularly in urgent contexts such as pandemic outbreaks, where time-efficient therapeutic identification is critical [7].

Recent AI applications in drug discovery encompass a wide spectrum of computational strategies. Virtual screening of large chemical libraries using ML models allows prioritization of compounds with desirable physicochemical and pharmacodynamic properties, thereby reducing experimental screening burdens. De novo molecule generation, powered by generative models such as variational autoencoders (VAEs) and generative adversarial networks (GANs), facilitates the design of novel chemical entities with optimized properties, including solubility, stability, and target selectivity [8,9]. Furthermore, AI-driven drug repurposing has become a prominent strategy, particularly for rare or emerging diseases, by predicting new therapeutic indications for existing drugs based on structural, biological, and phenotypic similarities [10].

Multi-omics integration represents another critical area where AI enhances drug discovery. The integration of genomics, transcriptomics, proteomics, and metabolomics data enables comprehensive understanding of disease mechanisms and target identification. AI models can identify disease-specific biomarkers, prioritize therapeutic targets, and predict patient-specific drug responses, advancing the paradigm of precision medicine [11]. In oncology, for example, AI-guided analysis of tumor genomics has enabled the identification of novel targets for small molecules and biologics, expediting the translation from bench to bedside [12]. Similarly, in neurodegenerative diseases, AI approaches have facilitated the elucidation of complex protein misfolding pathways, guiding the design of disease-modifying compounds.

Despite the transformative potential, several challenges hinder the widespread adoption of AI in drug discovery. Data quality and availability remain major bottlenecks, as AI models are highly dependent on large, curated, and accurate datasets. Inconsistent experimental data, biased training datasets, and insufficient representation of diverse chemical and biological space can compromise model reliability and reproducibility [13]. Moreover, interpretability of complex AI models, particularly deep learning networks, is a significant concern. Regulatory agencies, including the FDA and EMA, increasingly emphasize model transparency to ensure safety, efficacy, and reproducibility in clinical decision-making [14]. Ethical considerations, such as data privacy, algorithmic bias, and equitable access to AI-driven therapies, are also emerging issues that require robust frameworks for governance [15].

Integration of AI with traditional experimental approaches is likely to yield the most impactful outcomes. Hybrid strategies combining AI predictions with high-throughput screening, in vitro and in vivo validation, and cheminformatics analyses can optimize lead

selection and enhance the translational relevance of findings. Moreover, the advent of cloud computing, high-performance GPUs, and federated learning frameworks enables collaborative and decentralized model training, expanding AI's reach across academia and industry [16]. Future trends indicate increased application of reinforcement learning for multi-objective optimization of compounds, AI-driven synthesis planning, and automated laboratory robotics, which collectively have the potential to shorten development timelines, lower costs, and improve success rates across therapeutic areas.

In conclusion, AI has introduced unprecedented capabilities into the drug discovery pipeline, providing tools to analyze complex datasets, generate novel compounds, and predict pharmacological behavior with remarkable speed and accuracy. From virtual screening and de novo design to drug repurposing and multi-omics integration, AI is reshaping both the theoretical and practical foundations of pharmacological research. While challenges related to data quality, model interpretability, and regulatory compliance remain, ongoing advances in computational algorithms, high-performance computing, and ethical frameworks are steadily addressing these limitations. As AI continues to mature, its integration with conventional experimental methodologies promises to accelerate translational research, reduce attrition rates, and enable the development of safer, more effective therapeutics, ultimately transforming the landscape of modern drug discovery [17,18].

2. AI Methodologies in Drug Discovery

2.1 Machine Learning

Machine learning (ML) algorithms such as random forests, support vector machines (SVM), and gradient boosting are widely used for QSAR (quantitative structure–activity relationship) modeling, predicting biological activity and physicochemical properties of compounds [13]. ML models can rapidly screen vast chemical libraries to prioritize potential drug candidates and predict off-target effects.

2.2 Deep Learning

Deep learning (DL) approaches, including CNNs, RNNs, and transformers, have revolutionized drug discovery by learning complex patterns from molecular structures, chemical fingerprints, and biological data [14]. CNNs are particularly effective in image-based analysis of high-throughput screening data, while RNNs and transformers model sequential molecular representations and SMILES strings [15].

2.3 Natural Language Processing

NLP allows extraction of knowledge from biomedical literature, clinical trial reports, and patents to identify drug-target relationships and uncover mechanistic insights [16]. Transformer-based models such as BERT and BioBERT are widely applied for text mining and knowledge graph construction [17].

2.4 Generative Models

Generative AI models, including variational autoencoders (VAE) and generative adversarial networks (GAN), are employed for de novo molecule design with optimized physicochemical and pharmacokinetic properties [18]. These approaches enable the creation of novel chemical entities with predicted high efficacy and low toxicity.

2.5 Multimodal AI

Integration of chemical, genomic, proteomic, and clinical data through multimodal AI models allows comprehensive prediction of drug efficacy, ADMET properties, and safety profiles, providing a holistic approach to drug design [19].

2.6 Explainable AI

Explainable AI (XAI) methods such as SHAP and LIME improve model transparency, enabling chemists and pharmacologists to understand predictions, reduce bias, and enhance regulatory acceptance [20].

3. Applications of AI in Drug Discovery

- **Target Identification:** AI integrates multi-omics and pathway data to identify novel therapeutic targets [21].
- **Virtual Screening:** ML and DL models predict ligand-target interactions and prioritize candidate compounds [22].
- **De Novo Drug Design:** Generative models design new molecules with desired properties [23].
- **Drug Repurposing:** AI predicts new therapeutic indications for existing drugs [24].
- **ADMET Prediction:** AI models forecast absorption, distribution, metabolism, excretion, and toxicity profiles [25].
- **Clinical Trial Optimization:** AI identifies suitable patient populations and predicts trial outcomes [26].

4. Challenges in AI-Based Drug Discovery

Challenges include limited high-quality datasets, bias in chemical and biological data, lack of interpretability, reproducibility issues, and regulatory hurdles. Data privacy, ethical concerns, and integration with experimental workflows remain important considerations [27,28].

5. Future Directions

- **Multimodal AI:** Integration of chemical, omics, and clinical datasets [29].
- **Federated Learning:** Collaborative AI development without data sharing [30].
- **Explainable AI:** Transparency for regulatory and clinical adoption [20].
- **High-throughput Integration:** Coupling AI predictions with automated experimental validation [31].
- **AI-Driven Personalized Medicine:** Designing compounds and therapies tailored to patient-specific profiles [32].

S. No	AI Application/Methodology	Description	Key References (Vancouver)
1	Virtual Screening	AI models prioritize compounds from large chemical libraries based on predicted efficacy and binding.	[1]
2	Drug Target Identification	Machine learning integrates multi-omics data to identify potential therapeutic targets.	[2]
3	Predicting Drug-Target Interactions	DL models like CNNs and RNNs predict binding affinities and interactions.	[3]
4	De Novo Molecule Generation	Generative models (VAE, GAN) design novel compounds with optimized properties.	[4]
5	Drug Repurposing	AI predicts new indications for existing drugs using structural and phenotypic similarities.	[5]
6	Pharmacokinetic Modeling	AI predicts absorption, distribution, metabolism, and excretion (ADME) properties.	[6]
7	Toxicity Prediction	ML models forecast potential adverse effects of candidate compounds.	[7]
8	Multi-Omics Integration	Combining genomics, transcriptomics, proteomics, and metabolomics for precision medicine.	[8]
9	NLP for Literature Mining	Extracts mechanistic insights and potential drug candidates from	[9]

		unstructured data.	
10	High-Throughput Screening Optimization	AI reduces experimental burden by prioritizing likely hits.	[10]
11	Graph Neural Networks (GNN)	Models complex molecular graphs for predicting interactions and properties.	[11]
12	Reinforcement Learning	Multi-objective optimization for efficacy, safety, and pharmacokinetics in drug design.	[12]
13	Automated Synthesis Planning	AI-assisted prediction of synthetic routes for novel compounds.	[13]
14	AI-Guided Clinical Trials	Predict patient stratification, dosing, and outcomes using ML.	[14]
15	Biomarker Discovery	AI identifies disease-specific biomarkers for diagnostics and targeted therapies.	[2]
16	Protein Structure Prediction	DL models predict tertiary protein structures critical for target validation.	[3]
17	Side Effect Prediction	ML predicts off-target interactions and adverse reactions.	[4]
18	Integration with Robotics	AI drives automated labs for high-throughput experimentation.	[5]
19	Explainable AI	Enhances interpretability of complex ML models for regulatory acceptance.	[6]
20	Drug Discovery Workflow Acceleration	End-to-end AI pipelines reduce timelines and costs from discovery to clinical trials.	[7]

6. CONCLUSION

AI is reshaping drug discovery by accelerating target identification, compound design, and predictive modeling of efficacy and safety. Machine learning, deep learning, NLP, and generative models have enabled high-throughput analysis, virtual screening, de novo molecule generation, and drug repurposing. Multimodal AI and explainable models promise more reliable and interpretable predictions. While challenges in data quality, interpretability, and regulatory compliance remain, integration of AI with experimental and clinical pipelines holds immense potential to reduce costs, shorten timelines, and improve the precision of drug development. Continued advancements in AI-driven approaches are poised to transform pharmaceutical research and the development of next-generation therapeutics.

AUTHORS' CONTRIBUTIONS

Conceptualization, Methodology: Rashid Ansari, Manoj Jogarana; Formal analysis, Writing-original draft preparation, Writing-review and editing: Suhana Shaikh, Bhagyashree Godbole, Kumudini Kasar, Saniya Shaikh, Yogeeta Girwalke.

All authors have read and agreed to the published version of the manuscript.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENT

Not Applicable

REFERENCES:

1. Paul SM, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov*. 2010;9:203–14.
2. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23:1241–50.
3. Vamathevan J, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18:463–77.
4. Zhavoronkov A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. 2019;37:1038–40.
5. Gao K, et al. Generative AI in drug discovery: recent advances. *Drug Discov Today*. 2022;27:2350–60.
6. Lee J, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020;36:1234–40.
7. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. *Nat Rev Drug Discov*. 2005;4:649–63.
8. Venkatesh P, et al. Artificial intelligence in drug repurposing. *Front Pharmacol*. 2021;12:664412.

9. Gómez-Bombarelli R, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 2018;4:268–76.
10. Zitnik M, et al. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34:i457–66.
11. Reker D, Schneider P. Ethical and regulatory considerations in AI-assisted drug discovery. *Drug Discov Today*. 2020;25:1313–9.
12. Li X, et al. Regulatory perspectives on AI in pharmaceutical R&D. *J Pharm Sci*. 2022;111:1725–33.
13. Cherkasov A, et al. QSAR modeling: where have you been? Where are you going? *J Med Chem*. 2014;57:4977–5010.
14. Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. *ACS Cent Sci*. 2017;3:283–93.
15. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. *J Cheminform*. 2017;9:48.
16. Wang Q, et al. Text mining for drug discovery. *Brief Bioinform*. 2021;22:1345–60.
17. Lee J, et al. BioBERT: pre-trained biomedical language representation for biomedical text mining. *Bioinformatics*. 2020;36:1234–40.
18. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv*. 2018;4:eaap7885.
19. Huang S, et al. Multimodal learning for healthcare: methods and applications. *IEEE Trans Biomed Eng*. 2020;67:2829–40.
20. Lundberg SM, Lee S. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–74.
21. Chen H, et al. AI-assisted target identification for novel therapeutics. *Nat Biotechnol*. 2020;38:1086–94.
22. Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. *J Comput Chem*. 2017;38:1291–307.
23. Gómez-Bombarelli R, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 2018;4:268–76.
24. Vamathevan J, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18:463–77.
25. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2018;9:513–30.

26. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics*. 2019;20:273–86.
27. Reker D, Schneider P. Ethical and regulatory considerations in AI-assisted drug discovery. *Drug Discov Today*. 2020;25:1313–9.
28. Li X, et al. Regulatory perspectives on AI in pharmaceutical R&D. *J Pharm Sci*. 2022;111:1725–33.
29. Huang S, et al. Multimodal learning for healthcare: methods and applications