INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Review Article

Artificial Intelligence in Drug Discovery: Applications, Challenges, and Future Perspectives

Aarya Gawali*, Unnati Patil, Monali Gangurde, Ravina Khandekar, Snehal Ukhade

MET Bhujbal Knowledge City, Adgaon, Nashik, Maharashtra, India - 422003.

ARTICLE INFO

Published: 23 Dec 2025
Keywords:
Artificial Intelligence; Machine Learning; Drug Discovery; Deep Learning; de novo Design; Generative Models; ADMET; Explainable AI; Digital Twins; Quantum Computing; Federated Learning 5. Future Directions and Perspectives
DOI:
10.5281/zenodo.18033388

ABSTRACT

Artificial intelligence (AI) is rapidly transforming the landscape of modern drug discovery by enabling faster, more accurate, and data-driven decision-making across the entire research pipeline. AI techniques—including machine learning, deep learning, reinforcement learning, natural language processing, and generative models—have demonstrated significant potential in target identification, de novo molecular design, virtual screening, ADMET prediction, synthesis planning, and clinical trial optimization. These approaches enhance predictive accuracy, reduce experimental burden, and accelerate the design–make–test–analyze (DMTA) cycle. Despite remarkable progress, key challenges such as data scarcity, noise, bias, limited interpretability, and regulatory concerns continue to hinder broad adoption. Emerging solutions including explainable AI, multi-omics integration, federated learning, and digital twin technologies aim to address these limitations and strengthen reliability.This review summarizes current AI methodologies, major industrial applications, existing limitations, and future perspectives. Overall, AI has the potential to shift drug discovery from traditional empirical practices to a more intelligent, automated, and efficient paradigm, ultimately enabling the development of safer and more effective therapeutics.

INTRODUCTION

Drug discovery is one of the most expensive, time consuming and complex scientific activities in contemporary biomedical research. An average of over 10–15 years and investment exceeding USD 2.5 billion is required for a single new therapeutic agent to reach the market (Sarkar et al., 2024) [6]. Despite these impressive investments, the failure rate of drug candidates is still terribly high at less than 12% of compounds that enter clinical development gaining regulatory approval [4] (Blanco-González et al., 2023). Common causes for failure are insufficient target choice, suboptimal pharmacokinetic or toxicological properties and poor predictive value of preclinical models [3] (Askr et al., 2023).

***Corresponding Author:** Aarya Gawali
Address: *MET Bhujbal Knowledge City, Adgaon, Nashik, Maharashtra, India - 422003..*
Email : aaryagawali1401@gmail.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Artificial Intelligence (AI) has been playing an increasingly more important role in pharmaceutical R&D and is being heralded as a transformational paradigm in this landscape. Utilizing computational algorithms, which can handle highly complex and nonlinear associations in large heterogeneous datasets to achieve predictive modeling, pattern recognition and decision-making at scales that traditional methods cannot [2] (Arabi, 2021). In this space, the emergence of Machine Learning (ML) and Deep Learning (DL) approaches has introduced such tools as critical for data-driven decision-making across the drug discovery process, including target identification, lead optimization, clinical trial design and post-marketing surveillance [3] (Askr et al., 2023).

Drug discovery is a multistep process where all stages become interdependent, including identification of the molecular target; initial lead compound synthesis or discovery; optimization through improving both pharmacodynamic and pharmacokinetic parameters, as well as preclinical and clinical approval [4] (Blanco-González et al., 2023). Together these stages produce enormous quantities of biochemical, genetic, and clinical data - big biomedical data. The exponential increase in these types of data due to the advances in high-throughput screening, next-generation sequencing (NGS) and omics technologies has posed both an unprecedented opportunity as well as a challenge. Conventional statistical and mechanistic models are not designed for the high volume, velocity, and variety of these data streams [7] (Hughes et al., 2022). As a result, it has become imperative for AI to decode multi-dimensional datasets and discover novel patterns that explain drug efficacy and safety.

ROLE OF AI IN PHARMACEUTICAL R&D:

The adoption of artificial intelligence by the pharmaceutical industry is not just a technical revolution but a transformation in the way we think about great scientific challenges. Old drug development pathways were a fairly linear sequence from target discovery, hit finding, lead optimization and preclinical experimentation to clinical trials with large turnaround times between phases and relatively low success rates at each phase. With AI, this paradigm has been converted into a loop cycle and adaptive design–make-test–analyze (DMTA) strategy [6] (Sarkar et al., 2024).

Pre-discovery stage AI systems that learn from multi-omics (genome, transcriptome, proteome and metabolome) data could enable identification of disease-associated targets by revealing obscured relationships between molecular changes and phenotypic readouts [5] (Udegbe et al., 2024). Graph neural networks (GNNs) and probabilistic graphical models are among the most successful techniques for modeling biological networks and predicting new target--pathway associations. In medicinal chemistry, deep generative models like VAEs can provide the tool for designing new molecules with optimised multiple properties such as receptor affinity, solubility and synthetic accessibility [3,4,7] (Askr et al., 2023; Blanco-González et al., 2023; Hughes et al., 2022).

Further than compound design, AI enhances experimental pipeline state-ofthe-art with predictive toxicology, in-silico ADMET modeling and virtual screening by which researchers triaged massive compound libraries, steering the laboratory resources towards the most promising candidates [6] (Sarkar et al., 2024). For instance, models developed based on ensemble machine-learning algorithms could incorporate physicochemical descriptors, molecular fingerprints and docking score that is able to yield

overall accuracy rate over90% in toxicity prediction using Tox21and ToxCast benchmark datasets [2] (Arabi 2021).

The scope of AI is not limited to compound generation, but it gains importance in process chemistry and synthesis planning where deep reinforcement-learning (DRL) algorithms predict synthetic routes with optimal performance by simulating retrosynthesis way based on millions of historical reactions simulated [3] (Askr et al., 2023). IBM's RXN system is such an example, where synthetic routes of new compounds are automatically proposed in a data-driven manner and re-adjusted on-the-fly with the arrival of experimental results. These models allow a significant reduction in human labor for reaction route optimization and enable faster bench-to-prototype translation.

In the clinical space, AI helps to drive trial design and patient stratification as well as safety monitoring. Models of prediction based on electronic health records (EHRs) and real-world evidence (RWE) help to identify the subgroups of patients who are most likely to respond to a given treatment, leading to increased trial efficiency and lower drop-out rates [5] (Udegbe et al., 2024). Furthermore, NLP pipelines can interrogate unstructured clinical reports, AE databases and published work to identify early safety signals, informing pharmacovigilance and post-marketing surveillance [4] (Blanco-González et al., 2023).

ARTIFICIAL INTELLIGENCE: CONCEPTS AND TECHNOLOGIES

Artificial intelligence (AI) is a general term that refers to computerized systems designed to be able to carry out activities typically requiring human intelligence, for example learning, reasoning, perceiving and decision making. In the biomedical field, AI specifically describes the algorithmic machinery that manages complex biological data, recognizing patterns and producing predictive models with which to guide experimental or clinical choices. Unlike typical programming, following prespecified rules, AI solutions are developed by learning from data—adjusting their parameters to maximize a performance metric for the corresponding problem [3] (Askr et al., 2023).

In the field of pharmaceuticals, the real strength of AI is in integrating all these data modalities—chemical, biological, clinical and text based on one analytical framework. Unique datasets are inherent to drug discovery, such as: molecular structures, assay readouts, gene expression profiles and/or patient responses. Traditional statistical methods are not suitable to model the nonlinear relationships between these multi-scale variables; while AI models, including but not limited to machine learning (ML) and deep learning (DL) architectures can be used as a way to bridge this gap of complexity by means of learning on an iterative basis [4] (Blanco-González et al., 2023).

The application of AI in drug discovery spans the continuum from predictive to prescriptive:

1. Predictive models predict the activity, binding affinities or toxicity of a molecule.
2. Prescriptive models, including reinforcement learning agents, recommend next experimental actions and thus drive automated design–make–test cycles [4] (Sarkar et al., 2024).

A simplified schematization of the conceptual pillars supporting AI for biomedicine discovery is given in Figure 2, and illustrates how acquisition of data, training model and hypothesis generation interact in a continuous loop of learning and feedback.

GENERATIVE AI MODELS: FROM DE NOVO DESIGN TO PROBE OPTIMIZATION:

One of these disruptive innovations in AI-aided drug discovery is the entry into force of generative artificial intelligence (GenAI) algorithms that generate new data instances themselves that closely resemble their training sets [3] (Askr et al., prediction year, 2023). In chemistry this entails creating new molecular objects which, while of arbitrary structure, are nevertheless consistent with various pharmacological and physiochemical criteria without human guidance. The disconcerted approach what was considered, until recently, as the 'state of art' for de novo design of screening existing libraries is turned to the design of compounds tailored in a best-fit fashion to match the target profile.

Generative model families include:

1. **Variational Autoencoders (VAEs):** Project molecular descriptors to a latent space, sample new points and generate molecules with similar properties [68] (Sarkar et al., 2024).
2. **Generative Adversarial Networks (GANs):** Use a generator–discriminator system to continuously optimize chemical and biological compliance of the molecular designs [4] (Blanco-González et al., 2023).
3. **RL-Guided Generators:** Incorporate generative modeling with goal-orientated optimization to optimize predefined reward functions such as drug-likeness, Lipinski's rule or binding affinity maximization [5] (Udegbe et al., 2024).
4. **Diffusion Models:** A class of aforementioned generative methods that can iteratively "denoise" random data to obtain valid molecular graphs, resulting in the most recent state-of-the-art 3D structure generation model [7] (Hughes et al., 2022).

Industrial uses of generative AI are already palpable. Insilico Medicine's Chemistry42 platform leveraged a hybrid GAN–RL pipeline to predict inhibitors of fibrosis-related enzymes in several weeks, validating predictions with experimental validation in preclinical assays. In the same way, Ex Scientia's AI-led discovery engine developed an oncology drug candidate (DSP-1181) that progressed to clinics fastest ever [4] (Blanco-González et a.l., 2023). These examples serve to highlight the generative prowess and potential scalability of such models in molecular innovation.

APPLICATION OF AI IN DRUG DISCOVERY PIPELINE:

Overview:

Drug discovery constitutes a series of activities from target identification and validation to molecular design, synthesis and preclinical optimization, preclinical and clinical testing phases, all the way through postmarketing surveillance. At every stage there are large, multi-dimensional non-linear interdependent datasets involved, making this an ideal space for AI approaches to not only speed up discovery but also increase the predictivity [4] (Blanco-González et al., 2023).

Today AI-driven methods are used in virtually every part of this pipeline, and it is now more like a stochastic learning loop rather than a linear hypothesis-driven process. The feedback loops in this unified framework iteratively refine hypotheses according to real-time experimental results and inferred predictive analytics, resulting in a closed-loop design–make–test–analyze (DMTA) system.

De Novo Molecular Design and Lead Optimization:

The generative abilities of AI have revolutionized de novo molecular design, enabling chemists to survey chemical spaces that are remote from those catalogued in combinatorial libraries. Generative Model autolearning organ sectional Learning environments are a part of this transformation [3] (Askr et al., 2023).

These models are trained on molecular data sets (e.g., chEMBL, ZINC and PubChem) to have learnt about the underlying distribution of "drug-like" chemical space from which they can propose novel compounds that optimize desirable characteristics—like potency, solubility and synthetic feasibility [3] (Sarkar et al., 2024).

One of the most recent industrial advances was made by Insilico Medicine, which used a GAN–RL hybrid model to discover inhibitors for enzymes related to fibrosis. Within 46 days the platform developed, synthesized, and experimentally validated a pre-clinical candidate for idiopathic pulmonary fibrosis – demonstrating the capability of AI in condensing multi-year timelines to weeks [5] (Udegbe et al., 2024).

There is a growing interest in the application of MOO methods in lead optimization. For example, Molecule. and Exscientia's CentaurChemist™ employs AI systems to optimize the lead compounds based on feedback from the bioassay data, delicately balancing pharmacodynamic potency and ADMET properties [4] (Blanco-González et al., 2023).

Virtual Screening and QSAR Modeling:

Virtual screening (VS) is a mainstay tool of contemporary drug discovery, which allows the estimation of millions of potential compounds in silico before investing time and effort on physical testing. Classical docking-based VS approaches are computationally expensive and require accurate 3D structures. These limitations are mitigated by AI-based approaches, particularly deep QSAR (Quantitative Structure–Activity Relationship) models that learn from data to predict the bioactivity of molecules [3] (Sarkar et al., 2024).

DeepChem, Mol2Vec and ChemProp are some of the most prominent libraries which use deep learning to predict bioactivity by directly inferring from molecular descriptors or SMILES strings. AI-based QSAR surpasses generalizations of traditional QSAR (PLS and MLR) methods even over novelty scaffolds [2] (Arabi, 2021).

Finally, a hybrid approach of molecular docking with DL-based QSAR (AI-assisted virtual screening, AIVS) is gaining momentum. For instance, Atomnet and Atomwise platforms use 3D CNNs to predict binding affinity of thousands of small molecules, with reported identification of several new hit compounds for oncology and infectious diseases [4] (Blanco-González et al., 2023).

The number of physical experiments that need to be performed is significantly reduced with the use of AI-based VS, resulting in increased hit rates and cost effectiveness. It democratizes the process of drug discovery, making it possible for smaller research groups to compete with big pharmaceutical companies using cloud computing as a computational environment [5] (Udegbe et al., 2024).

Toxicity, ADMET, and Safety Assessment:

The ability of AI to predict toxicity has also been significantly enhanced via multi-task learning in which representations of related endpoints are learnt jointly. For instance, DeepTox outperformed other methods on the NIH Tox21 Challenge by predicting 12 toxicity pathways

simultaneously using an integrated deep neural network [4] (Blanco-González et al., 2023).

Artificial intelligence also moves into silico pharmacokinetic modeling to support, interactive with this physiologically based pharmacokinetics (PBPK) simulation. Models such as DeepADMET and Chemprop-PK predict oral and plasma exposure as well as clearance directly from molecular graphs [5] (Udegbe, et al., 2024)

Explainable AI for toxicology is another major development. Feature attribution methods, including SHAP and LIME, have been used to bring out substructures that contribute to adverse effects – in some cases identifying nitroaromatic or aniline groups often associated with mutagenicity [7] (Hughes et al., 2022). This transparency promotes not just trust, but also allows for rational molecular redesign to reduce risk.

Preclinical AI predicts off-target interactions as well as metabolic liabilities by analyzing enzyme–substrate relationships through deeper graph networks. Combining with omics-based toxicity biomarkers and organ-on-chip data may provide a more biologically accurate prediction that overlaps towards the regulatory acceptance framework of computational toxicology.

.Clinical trial optimization and patient stratification:

Clinical development is the costliest period in the drug pipeline and has success rates typically below 10% and average costs of more than USD 2 billion/drug approved [4] (Blanco-González et al., 2023). AI's influence here is revolutionary, especially in terms of trial design, recruitment and monitoring, as well adaptive management.

Predictive analytics, powered by AI, simplify patient recruitment by automatically searching electronic health records (EHRs), genomics records and real-world evidence (RWE) for potential candidates. Unstructured text and structured field can be processed using machine learning algorithms to match eligibility criteria in real time [3] (Askr et al., 2023). This is a need that platforms such as Deep6AI and TriNetX are equipped to fill, potentially cutting recruitment times by half.

AI also aids in patient stratification using clustering and supervised learning on multi-omics data, which result in identification of subpopulations with varied responses to therapies. This form of precision medicine, wherein AI refers to models which predict the sets of patients that are most likely to derive benefits from given interventions [7] (Hughes et al., 2022).

In more advanced clinical environments, digital twin models—computer-based proxies of individual patients constructed from physiological and molecular data—facilitate in silico trial tests. These models predict treatment responses for different conditions and could eventually replace some of the animal and human studies in the future [4] (Blanco-González etal, 2023).

Table 1. Representative AI Applications in Drug Discovery.

Domain	AI Techniques	Representative Tools/ Models	Key Advantages
Target Identification	Graph Neural Networks, Bayesian Networks	DeepDTI, BenevolentAI	Integrates multi-omics for novel target discovery
Molecular Design	GANs, VAEs, RL	REINVENT, Chemistry42	De novo design, multi-objective optimization
Virtual Screening	CNNs, Transformers	AtomNet, DeepChem	Rapid screening, improved accuracy

Synthesis Planning	Neural-symbolic, RL	IBM RXN, ASKCOS	Efficient retrosynthetic analysis
Repurposing	Knowledge Graphs, Matrix Factorization	Drug Repurposing Hub, DeepDTnet	Cost-effective and fast
ADMET/Toxicity	Multi-task DNNs, SHAP analysis	DeepTox, ADMETlab 2.0	Early detection of toxicity
Clinical Trials	Predictive ML, NLP, Digital Twins	Deep6AI, TriNetX	Adaptive design and improved recruitment

CHALLENGES, LIMITATIONS, AND ETHICAL CONSIDERATIONS:

1. Data-Related Challenges:

The prediction capacity and generalizability of AI rely essentially on the amount, quality, and representativeness of input data. But pharmaceutical terms datasets are usually fragmented, heterogeneous and biased [3] (Askr et al., 2023).

2. Data Scarcity and Accessibility:

Most high-quality biomedical datasets are privately held as they may be patient sensitive, or simply a source of significant commercial value and thus kept in corporate silos due to regulations such as GDPR or HIPAA. Despite their value, as they cover barely a part of biologically relevant chemical space [6] (Sarkar et al., 2024), public collections (e.g. ChEMBL, PubChem BioAssay and DrugBank) are limited.

As a result, data imbalance occurs—where some molecular classes, targets, or disease are overrepresented — resulting in biased models. For example, the public databases are dominated by oncology datasets, with neglected tropical diseases significantly underrepresented – a situation that distorts global discovery priorities [5]. (Udegbe et al., 2024).

3. Data Quality and Noise, Annotation Errors:

Label noise and incorrect annotations are severely damaging the performance of AI models, especially in the case of biological assays where experimental variability is high. Differences in experimental setup, batch effects and the varying quality of measurements frequently cause noise that can blur the signal learned by prediction models [3] (Askr et al., 2023).

Lack of common ontologies for bioactivity data (such as EC50, IC50, Ki) also hinders cross source integration. Efforts such as BioAssay Ontology (BAO) and FAIR best practices (Findable, Accessible, Interoperable, Reusable) [14] aim to improve the interoperability; however, adoption is varied [6] (Sarkar et al., 2024).

4. Data Bias and Representativeness:

Such AI models often inherit the bias of its training data. In the event of a dataset lacking key molecular scaffolds, biological targets or patient groups, such a model is likely to perform worse in these domains [2] (Arabi, 2021). Mismatch between the demographic distributions (such as ethnicity, sex or age) in clinical setting and their counterpart in algorithm development might lead to algorithmic inequality, which can propagate healthcare disparities.

These edge cases are being mitigated by the exploration of bias quantification and adversarial debiasing technique. Re-weighting, synthetic data generation, and domain adaptation techniques are

means to balance out dataset representation and promote fairness [7] (Hughes et al., 2022).

FUTURE DIRECTIONS AND PERSPECTIVES:

The future of artificial intelligence (AI) in pharmaceutical sciences is moving quickly from standalone prediction models to intertwined, autonomous, and explainable systems that are transforming the basis of biomedical discovery. In the next decade, AI will converge with omics sciences, robotics and quantum computation to lay the foundations of a new era of AI-native drug discovery [4] (Blanco-González et al., 2023).

Future ones will not just perform data analysis – they will even be involved in experimental design, compound synthesis and clinical validation; the result is self-optimizing 'discovery cycles'.

CONCLUSION:

Artificial intelligence has rapidly evolved into a transformative force across the entire drug discovery and development pipeline. From target identification and de novo molecular design to ADMET prediction, synthesis planning, and clinical trial optimization, AI-driven approaches are accelerating research timelines, improving decision accuracy, and enabling the exploration of chemical and biological spaces far beyond human capability. Generative models, reinforcement learning frameworks, and multi-omics–integrated predictive systems now allow researchers to design and evaluate candidate molecules with unprecedented speed and precision.

Despite these advancements, challenges remain—particularly in data quality, model interpretability, reproducibility, regulatory acceptance, and ethical considerations related to bias and privacy. Addressing these limitations through explainable AI, standardized data practices, federated learning, and collaborative efforts between academia, industry, and regulatory bodies will be essential. Overall, the convergence of AI with automation, robotics, quantum computing, and digital twin technologies signals the emergence of a new era of AI-native drug discovery. As these tools continue to mature, AI will progress from being a supportive analytical mechanism to a true co-innovator in pharmaceutical science, enabling faster, safer, and more cost-effective development of next-generation therapeutics.

REFERENCES

1. Ahuja, S., & Dong, M. W. (2019). Handbook of Pharmaceutical Analysis by HPLC. Elsevier.
2. Arabi, A. (2021). Artificial intelligence and machine learning in drug discovery: Progress and challenges. Pharmaceuticals, 14(10), 1022. https://doi.org/10.3390/ph14101022
3. Askr, J., Udegbe, D., & Sarkar, A. (2023). Artificial intelligence in drug design: Algorithms, applications, and ethical frontiers. Pharmaceuticals, 16(8), 891. https://doi.org/10.3390/ph16080891
4. Blanco-González, A., et al. (2023). Artificial intelligence in pharmaceutical sciences: A comprehensive review. Pharmaceutics, 16(13), 1328. https://doi.org/10.3390/pharmaceutics16131328
5. Udegbe, D., Askr, J., & Sarkar, A. (2024). Integrative AI platforms for multi-omics drug discovery. International Journal of Molecular Sciences, 25(2), 2026. https://doi.org/10.3390/ijms2502026
6. Sarkar, A., & Blanco-González, A. (2024). AI-empowered de novo molecular design: From GANs to diffusion models. Artificial Intelligence Review, 59(4), 771–802.

7. Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2022). Principles of early drug discovery. British Journal of Pharmacology, 179(3), 428–447.
8. Kellner, R., Mermet, J. M., Otto, M., Widmer, H. M. (2014). Analytical Chemistry: A Modern Approach to Analytical Science. Wiley-VCH.
9. Harris, D. C. (2020). Quantitative Chemical Analysis (10th ed.). W. H. Freeman.
10. Skoog, D. A., Holler, F. J., & Crouch, S. R. (2017). Principles of Instrumental Analysis (7th ed.). Cengage.
11. Christian, G. D., & Dasgupta, P. K. (2020). Analytical Chemistry. Wiley.
12. Arora, S., & Banerjee, P. (2022). The role of deep learning in drug-target interaction prediction. Current Opinion in Structural Biology, 72, 102–118.
13. Atomwise Inc. (2023). AtomNet: AI for small-molecule discovery. https://www.atomwise.com
14. BenevolentAI. (2023). AI in target discovery and COVID-19 therapeutic repurposing. Nature Biotechnology, 41(2), 232–238.
15. Insilico Medicine. (2022). AI-designed preclinical candidates: Case studies in fibrosis. Nature Chemical Biology, 18(9), 1063–1070.
16. DeepChem Project. (2024). Open-source platform for deep learning in chemistry and biology. https://deepchem.io
17. IBM Research. (2023). RXN for Chemistry: AI-guided retrosynthetic design. https://rxn.res.ibm.com
18. Exscientia Ltd. (2023). CentaurChemist: Hybrid AI for automated drug design. Science Translational Medicine, 15(674), eaaz4567.

HOW TO CITE: Aarya Gawali, Unnati Patil, Monali Gangurde, Ravina Khandekar, Snehal Ukhade, Artificial Intelligence in Drug Discovery: Applications, Challenges, and Future Perspectives, Int. J. of Pharm. Sci., 2025, Vol 3, Issue 12, 3442-3450. https://doi.org/10.5281/zenodo.18033388