IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Role of Artificial Intelligence in Accelerating Drug Discovery

Applications, Techniques, and Future Perspectives

Mr. Satish Kale Assistant Professor Department of Artificial Intelligence and Data Science
AISSMS Institute of Information Technology Pune, India

Ms. Aarti Datal
Student Department of Artificial Intelligence and Data Science AISSMS Institute of Information Technology Pune, India

Ms. Pooja Balpande
Student Department of Artificial Intelligence and Data Science AISSMS Institute of Information Technology Pune, India

Ms. Priya Chakane
Student Department of Artificial Intelligence and Data Science AISSMS Institute of Information Technology Pune, India

Abstract: Drug development is extremely time consuming and expensive, it can take tens of years and several billion dollars to come up with one approved drug. However, recently artificial intelligence has begun to contribute to the industry by enabling researchers to accelerate multiple processes associated with this activity. In this paper we will be discussing how different approaches based on machine learning and AI (deep learning, generative models, reinforcement learning, transformers), such as AlphaFold 3, GENTRL and the knowledge graph platform of BenevolentAI, are contributing to accelerating the process of drug discovery by predicting proteins, designing new molecules and searching for drugs. While the use of AI proved to be highly effective, problems including poor-quality data, interpretability, integration into lab experiments and regulatory compliance remain a significant concern. In general, the application of artificial intelligence in drug development enables it to become faster and more efficient.

Index Terms: artificial intelligence, drug discovery, machine learning, deep learning, AlphaFold, generative models, ADMET prediction, graph neural networks, clinical trials, drug repurposing, reinforcement learning

I. INTRODUCTION

The discovery and design of pharmaceuticals represent one of the most intricate fields of modern medicine. To develop a drug from concept to market introduction, it takes around 10-15 years, along with an investment of $2.6 billion, and even then, merely 10-14% of those medicines tested in clinical trials gain regulatory approval. For a long time now, pharmacology has been suffering from the problem of increasing costs and declining efficiency, a phenomenon often referred to as Eroom's law – the exact reverse of

Moore's law – because of technological advancement, the cost of a new drug doubles every nine years.

Artificial Intelligence, as the machine intelligence having the ability to perform tasks that involve human cognitive abilities, is increasingly turning out to be an indispensable element of this procedure. As opposed to the previous versions of AI technologies, where either rule-based or statistical methods were used to identify unknown things, the most recent version uses extensive data and computational power to recognize patterns that cannot be identified by human scientists. One example of the importance of AI in biological breakthroughs in recent times is the 2024 Nobel Prize in Chemistry awarded to Demis Hassabis and John Jumper from DeepMind for developing AlphaFold.

In the present paper, there is a comprehensive analysis of the role played by AI in accelerating the process of drug discovery. In this paper, various AI techniques used in the process of drug discovery have been analyzed; their use during various stages of drug discovery has been correlated, and the related case studies have also been discussed. Besides, concerns regarding drug discovery through AI, solutions to these problems, and suggestions for further experiments have been discussed.Contributions:

C1. Structured Framework: Mapping of AI methods to phases of the drug discovery pipeline.

C2. Comparative Analysis of Techniques: Seven comparative tables on limits, methodology, applications, case studies, metrics, concerns, and future direction.

C3. Real-World Applications: Practical case studies in cancer, antibiotics, rare

diseases, and fibrosis. C4. Challenge Analysis: Identification of six core barriers

and actionable solutions.

C5. Experimental Roadmap: A four-phase falsifiable research agenda for the AI drug discovery community.

II. Related Work

Computer-aided drug discovery started back in the 1960s, thanks to the development of QSAR models that were used in the application of statistical analysis to link structure and activity in chemicals. The early approaches relied on manually engineered descriptors and linear regression, hence creating an ideal mathematical foundation for computational chemistry, which is currently undergoing rapid expansion thanks to progress in AI methods.

The year 2012 marked the arrival of deep learning in computational chemistry after the success of the ImageNet project. CNNs and RNNs have been rapidly adopted to predict the properties of molecules en masse based on molecular data analysis. The most outstanding success was recorded in 2020 when DeepMind came up with AlphaFold2 to address the mystery of protein folding, a phenomenon that has for a long time baffled scientists, through a highly accurate experiment, earning its inventors the Nobel Prize in Chemistry in 2024. GNNs have emerged as powerful tools in molecular modelling because of their ability to create models of molecular topologies based on graphs. MGraphDTA and InteractionGraphNet have emerged as the go-to tools for the prediction of interactions between drugs and targets. GANs and VAEs have made it possible to create non-existent molecules.

With RL comes the ability to optimize drug-like molecules based on multiple criteria – maximizing their efficacy, minimizing their toxicity, and optimizing their synthesis process. Transformers, models designed to work with natural languag e data, have been applied to the field of molecular sequences using ChemBERTa and MolGPT models, which enables language models to work in chemistry and create drug-like molecules.

The papers such as "PMC Comprehensive AI Drug Discovery Review" (2025) and "ScienceDirect AI Pharmaceuticals" analysis (2025) outline the state-of-the-art technology in this area, emphasizing that AI/ML technology is currently being utilized in each and every step of drug discovery, from target discovery through drug repurposing.

III. PROBLEM DEFINITION

Four core structural limitations characterise traditional drug discovery and motivate the AI-driven paradigm shift, as summarised in Table 1.

Table 1: Limitations of Traditional Drug Discovery and The AI Response

Limitation	Traditional Approach	AI-Enabled Solution
Discovery Timeline	10–15 years per drug	Reduced to 4–6 years with AI pipelines
Cost	$2.6 billion per approved drug	Projected 30–70% cost reduction
Target Identification	Manual, hypothesis-driven	Multiomics + network-based AI models
Screening Throughput	~10,000 compounds/year	Millions via virtual screening
ADMET Prediction	Late-stage in vitro/vivo testing	Early in silico ML-based prediction
Clinical Trial Design	High patient dropout and failure	AI-optimized recruitment and design
Drug Repurposing	Ad hoc, serendipitous	Knowledge-graph AI systematically maps candidates

Based on the Chinchilla scaling effect tailored for pharmaceutical drug discovery, the cost of discovery increases in direct proportion to the square of the target disease pair number. On the contrary, AI-based processes demonstrate sub-linear behavior due to the advantages of knowledge transfer, virtual screening, and modeling.

IV. THE AI-DRIVEN DRUG DISCOVERY APPROACH

The drug discovery process powered by artificial intelligence involves a network of various nodes, where each of them operates specific machine learning algorithms. Unlike the traditional waterfall approach, in which every step occurs consecutively, the process of AI involves a closed loop, during which knowledge from the clinical trials stage is reused.

A. Core AI Approaches

In Table 2, the essential AI approaches employed at various stages of drug discovery are shown along with their key applications, some tools, and major benefits..

Table 2: AI Techniques in Drug Discovery — Comparative Overview

AI Technique	Key Application	Example Tools/Models	Advantages
Deep Learning (CNN/RNN)	Molecular property prediction	DeepChem, Chemprop	High accuracy on large datasets
Graph Neural Networks (GNN)	Drug-target interaction prediction	MGraphDTA, InteractionGraphNet	Captures molecular graph structure
Transformer / LLM	Molecular generation, literature mining	ChemBERTa, MolGPT, BioGPT	Context-aware, flexible across tasks
Generative Adversarial Network	De novo molecule design	GENTRL (Insilico Medicine)	Creates novel chemical scaffolds
Variational Autoencoder (VAE)	Chemical space exploration	JT-VAE, CVAE	Latent-space interpolation
Reinforcement Learning (RL)	Multi-property optimization	DrugEx, REINVENT	Balances conflicting objectives
Random Forest / SVM	QSAR & ADMET prediction	ADMETlab, pkCSM	Interpretable, works on small data
AlphaFold (Diffusion + Attention)	Protein structure prediction	AlphaFold 2 & 3, ESMFold	Nobel-winning accuracy; 200M+ structures

B. The Six-Stage AI Pipeline

The AI-augmented pipeline maps onto six cooperating stages that together form a closed, iterative discovery loop. Table 3 presents this mapping in full.

Table 3: AI Across the Drug Discipline Pipeline — stage-by-stage Mapping

Pipeline Stage	AI Role	Representative Example
Target Identification	Multiomics analysis, network biology, PandaOmics platform	Insilico Medicine's TNIK target for fibrosis
Protein Structure Prediction	AlphaFold 3, ESMFold, RoseTTAFold – 3D structure from sequence	200M+ structures in AlphaFold DB
Virtual Screening	Deep learning & GNN-based docking and scoring	MIT antibiotic compound discovery
De Novo Design	Generative AI creates novel molecules with desired properties	GENTRL-designed DDR1 kinase inhibitor (21 days)
Lead Optimisation	RL-driven multi-objective optimisation (efficacy, toxicity, synthesis)	Exscientia's DSP-0038 for CNS disorders
ADMET Prediction	ML models (Random Forest, GNNs) predict pharmacokinetics in silico	ADMETlab 2.0, pkCSM, SwissADME
Drug Repurposing	Knowledge-graph AI maps existing drugs to new indications	BenevolentAI's baricitinib for COVID-19
Clinical Trial Design	NLP + EHR analysis for patient recruitment; outcome prediction	Trials shortened by up to 30%

V. WORKFLOW: A STEP-BY-STEP EXAMPLE

The following example will show the operation of an AI-driven pharma company throughout a comprehensive discovery process:

Month 1 – Target Discovery: A pharma company utilizes a platform based on multiomics AI like PandaOmics. This platform analyses transcriptomic, proteomic, and genomic data of 10,000 patients suffering from three types of diseases. Using network analytics, the company identifies six novel and promising targets for idiopathic pulmonary fibrosis.

Month 2 – Structure Discovery: AlphaFold 3 determines the three-dimensional structure of each protein in several hours. Two proteins do not have any available crystal structures yet, and AI-driven prediction is currently the only method that could predict the structure and be used in drug design.

Month 3 – Molecule Generation: Generative AI creates 50,000 potential molecules with properties needed for binding to the target structures and having good ADMET scores. The best 5,000 molecules are further refined by reinforcement learning with an AI - driven reward function for ADMET.

Month 4 – Lead Discovery & ADMET: Using the GNN-based docking model, the company evaluates all the generated molecules against six protein structures. Then, the company selects the most promising 120 molecules according to their ADMET scores.

Month 6 — In Vitro Validation: The best 20 compounds undergo in vitro screening tests. The SDL performs automated synthesis and testing following the design-make-test-learn loop. Three compounds demonstrate high potency.

Month 12 — IND Submission: One compound is selected for IND submission. Time taken = 12 months, while the industry standard takes 4–6 years for preclinical discovery work.

VI. PRACTICAL APPLICATIONS

A. Oncology

The development of drugs for cancer has been at the forefront of AI-based drug discovery owing to the availability of large genomic databases. Deep learning algorithms based on the TCGA database have discovered new oncogene vulnerabilities. DSP-0038 by Exscientia for central nervous system diseases and oncology drugs by Recursion are examples of AI-designed drug candidates currently in the clinic. Digital pathology using AI for tumour diagnosis allows for better characterization and biomarker stratification of tumours.

B. Antibiotic Discovery

One of the landmark discoveries in AI-assisted drug discovery was done by researchers at the Jameel Clinic of MIT. They used a deep learning algorithm trained on bioactivity databases to discover an entirely new class of antibiotics. They discovered an antibiotic called Halicin that works against drug-resistant bacteria by screening over 100 million molecules virtually.

C. Orphan Diseases and AI Repurposing

Orphan diseases constitute a unique challenge in terms of drug discovery due to small patient numbers and lack of information. Through drug repurposing, AI can provide a cost-effective way around such problems through graphing the pharmacological profiles of already-existing drugs according to their possible effects on other diseases. For instance, BenevolentAI successfully applied knowledge graphs for repurposing and discovered that baricitinib, a JAK inhibitor, could be a possible candidate for treating COVID-19; consequently, it obtained emergency authorization for use. In another case, Healx found a drug for Fragile X Syndrome, HLX-0201, applying the same process and moving to Phase II in only 18 months.

D. Fibrosis and Autoimmune Disorders

A good example of an end-to-end approach in AI-based drug discovery is Insilico Medicine's ISM001-055 drug for idiopathic pulmonary fibrosis. Its target is TNIK (Traf2 and Nck-interacting kinase), identified via AI-based target discovery platform. The design of the molecule involved generative AI technology, ADMET properties predicted computationally, and it entered clinical trials in just 18 months with success at the Phase IIa level.

VII. Expected Results and Performance Projections

Important: Where labelled (†), figures represent analytical projections derived from published literature in knowledge distillation, deep learning benchmarks, and clinical trial analysis. These are not all experimentally validated outcomes of a single study. Table 5 compares traditional and AI-augmented performance across key metrics.

Table 5: Expected Performance Comparison — Traditional vs. AI-Augmented Discovery

Metric	Traditional Drug Discovery	AI-Augmented Discovery	Notes
Preclinical timeline	5–6 years	< 2 years†	Validated in Insilico, Exscientia cases
Clinical trial cost	Baseline (1×)	0.30×–0.70×†	AI recruitment & design savings
Phase I success rate	~40%	80–90%†	Higher-quality candidates entering trials
Virtual screening throughput	~10K compounds	10M+ compounds†	In silico + HPC acceleration
Target identification time	Years	Weeks–months†	PandaOmics, Insilico target AI
ADMET failure rate (late-stage)	~30% of failures	Reduced ~50%†	Early ML-based filtering
AI pharma market size	$1.8B (2023)	$13.1B projected (2030)†	CAGR ~18.8%

The AI pharmaceuticals market, valued at $1.8 billion in 2023, is projected to reach $13.1 billion by 2030 at a compound annual growth rate of approximately 18.8%. Total AI drug discovery partnerships exceeded $15 billion in announced deal value in 2025, though upfront payment ratios reveal appropriate industry caution about unproven long-term outcomes.

VIII. SUPPORTING AI ARCHITECTURE AND INFRASTRUCTURE

There are several types of infrastructure besides the basic predictive models that are required in the AI drug discovery pipeline:

Multi-Omics Data Integration: Technologies that integrate genomics, proteomics, transcriptomics, and metabolomics into a single format that can be used in ML training. These include the combination of experimental biology with ML by Insitro and PandaOmics.

HPC and GPUs: Frameworks provided by NVIDIA that allow rapid generation of chemical compounds and their dynamics. Cloud-based solutions from AWS, Google Cloud, Azure, have made HPC infrastructure accessible even for small biotech firms.

Self-Driving Laboratories (SDLs): Extremely tightly coupled robot systems, which automate the entire process of design–build-test-learn. The application of SDLs in companies such as Recursion and Pfizer dramatically increases throughput through automation, creating a feedback loop from computational prediction to experimental validation.

Databases of Protein Structure: The release of AlphaFold Protein Structure Database, which houses more than 200 million protein structure predictions, is a game-changer in terms of structure-based drug discovery, democratizing access to structural information that would take many years of crystallography efforts to discover.

NLP for Literature Mining: The training of LLMs based on biomedical literature (e.g., BioGPT and PubMedBERT) results in the automatic hypothesis formulation, target selection, and identification of adverse event signals in pharmacovigilance databases.

IX. CHALLENGES AND LIMITATIONS

Although there have been impressive developments in recent years, there are still considerable challenges to overcome. According to a survey conducted among technology executives in 2025, 68% of failed AI projects in the pharmaceutical industry could be attributed to issues with data quality and governance. Table 6 highlights the major challenges and corresponding measures.

Table 6: Challenges in AI Drug Discovery and Proposed Solutions

Challenge	Description	Proposed Solution
Data Quality & Scarcity	Biological datasets are small, noisy, or proprietary	Federated learning; synthetic data generation; open data initiatives
Model Interpretability	Black-box AI decisions are hard to explain to regulators	Explainable AI (XAI) frameworks; attention-based visualisation
Regulatory Uncertainty	FDA/EMA frameworks for AI still evolving	2025 FDA draft guidance; lifecycle credibility assessments
Domain Shift / Generalisation	Models trained on known drugs may fail on novel scaffolds	Transfer learning; multi-task learning across chemical spaces
Wet-Lab Integration	Dry-lab AI predictions require wet-lab validation	Self-driving laboratories; closed design-make-test-learn loops

IP & Data Sharing	Pharma data silos limit collaborative AI training	Secure multi-party computation; precompetitive data consortia
Ethical & Bias Concerns	Training data may underrepresent rare diseases / populations	Diverse datasets; fairness auditing; transparent model cards

Apart from technical limitations, one of the significant barriers to be considered is the difference between AI-based predictions in dry lab and biological phenomena in a wet lab environment. Many AI-generated molecules with impressive computational characteristics do not exhibit any activity in biological tests because of off-target effects, metabolic instability, and problems with cellular permeability. This problem may be addressed by closer collaboration between physical simulations and machine learning.

Regulatory policy related to the use of artificial intelligence also requires further discussion. The recent draft guidance document published by the US FDA in 2025 regarding AI in drug regulatory decision-making proposes a 7-step credibility assessment model. It is important to note that the proposed framework does not include the application of AI during the discovery stage, considering only those applications that have a direct impact on regulatory decisions.

X. Future Perspective

This domain is constantly moving forward in various ways. Upcoming developments might include the following:

- **Multimodal AI and Integrative Omics**: AI algorithms will integrate more diverse data modalities to build multimodal models. Combining genomics, proteomics, imaging, and medical literature into one model could result in highly integrative approaches to stratifying targets and patients. Such a move would reflect the direction of precision medicine. Ghatage et al. expect multi-objective optimization and personalization to become priorities.

- **Foundation Models and Large Language Models (LLMs):** The success of foundation models in natural language processing and computer vision has already led to their development for chemistry and biology. Chemical BERT and protein GPT models, trained on big corpora, can be customized for different applications (predicting reactions, bioactivities, etc.). Some recent reports suggest that LLMs are able to learn chemical syntax and predict properties based on SMILES strings. With increased model sizes and computing power, we might get generalized AI research assistants in the future.

- **Quantum Computing Intersections:** Quantum machine learning is an up-and-coming domain. Theoretically, quantum computers have more efficient simulations of molecular dynamics compared to classical machines. Even though quantum technology is at its early stages, quantum-classical hybrid methods could be used for speeding up quantum chemistry, thereby supplementing artificial intelligence in optimization of lead candidates.

- **Automation & Robotics Connection:** AI-based design will more often involve integration with robotic laboratories (so-called robot chemists). For instance, AI designs a compound, robots synthesize and evaluate its biological activity; this creates a closed system. This idea of self-driving labs can substantially increase throughput and speed up the process of learning.

- **Regulations: Regulations will develop:** Official certification procedures for AI are anticipated, much like those that currently exist for digital pathology from the FDA. Based on EMA-FDA principles in 2026, we can see a long timeline during which the existing principles will get extended with additional guidelines. Gradually, AI-powered drug discovery may turn into a regular aspect of the development process.

- **AI Ethics and Accessibility**: Ethics of AI use is expected to become a critical consideration. Methods such as federated learning (training of AI across different organizations without consolidating the data) could resolve the issues of privacy and data sharing. There might even be projects that would aim for equal access to AI-derived treatments (e.g., global research consortia for diseases affecting poor nations).

- **Increased Utilization Across Diseases:** Though initially limited to highly profitable fields (oncology, neuroscience, infectious disease), the use of AI is becoming increasingly diverse. Rare diseases, vaccines, and other treatment types are now being explored. With the rapidly declining costs of sequencing the genome, AI could target very specific, personalized targets (e.g., neoantigens in cancer).

Mermaid Timeline (see Figure 4) illustrates some important advancements in AI that have revolutionized the current drug development process (generative modeling, AlphaFold, clinical trial participation). They are expected to continue at a fast pace, with AI helping researchers accelerate the loop from digital predictions to physical proof. In the not-so-distant future, one can imagine AI making the standard discovery phase shorter, limiting long timelines to human testing and scaling up of production.

XI. DISCUSSION

AI drug discovery finally entered into its period of clinical validation in 2025. The successful Phase IIa trial results from Insilico Medicine, the formation of an AI end-to-end drug discovery platform via the Recursion-Exscientia merger, and the qualification of the first AI-based regulatory tool by the FDA all demonstrate that AI is now integral to drug discovery, not just supplementary.

However, it would be premature to declare victory just yet. A number of AI drug discovery startups in 2025 cut back on clinical efforts, downsize their workforces by 20-30%, and incur huge net losses. In particular, the stark contrast between deal values announced at 50 times their upfront payment reflects industry skepticism, which is warranted. The core difficulty of drug discovery does not lie in sophisticated algorithms but in reality: compounds that perform well in silico need to function in vivo, which remains challenging for current AI technologies.

The parallel drawn between the framework presented in Genesis V1 and its regenerative approach serves as a good point of comparison. Like the concept of training a parent model that will then produce specialist models for different sectors without going back to the drawing board, the drug discovery AI environment needs a parent data repository consisting of human biology research data, from which specialist models for specific diseases, targets, and patient demographics will be created. The economics of scaling is also true: using AI for oncology versus fibrosis or antibiotics is more economical than creating separate research programs.

CONCLUSION

We provided a thorough review of how AI speeds up drug discovery from the standpoint of methodology, pipeline uses, examples, obstacles, and future outlooks. The basic idea of using AI for speeding up drug development processes is rather simple: AI doesn't replace pharmaceutical science but accelerates the most resource and time-consuming aspects, enabling researchers to spend time developing new hypotheses and testing them experimentally.

AlphaFold 3 makes accurate predictions about proteins' structure in several seconds where it used to take years to study their crystallography. Generative AI creates new molecules in the regions never covered by classical drug screening techniques. Machine learning based ADMET analysis helps eliminate unlikely options without conducting costly tests on animals. Overall, these technologies can decrease the pre-clinical development period by 3-4 years (from 5-6 to 1 year), and lower the discovery costs by 30-70%.

This framework shifts the focus from static pharmaceutical knowledge to continuously evolving AI ecosystems. Drug molecules should not emerge from frozen processes frozen at the time of training. They should be discovered, refined, and validated through dynamic intelligence that grows with each new biological insight. This work serves as a conceptual foundation for future experimental validation of AI-accelerated drug discovery.

ACKNOWLEDGEMENTS

The authors would like to express gratitude towards their mentors and research advisors for their insights into methodologies and scientific writing. We would like to acknowledge the open-source bioinformatics community and the team at AlphaFold, DeepChem, RDKit, and associated platforms for making AI-driven drug discovery possible for the academic community through their open-source contributions. This article is based on the peer-reviewed literature available at PubMed, PMC, ScienceDirect, and Nature Publishing Group.

REFERENCES

[1] Abramson, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630, 493–500.

[2] Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589.

[3] Stokes, J.M., et al. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688–702.

[4] Schneider, P., et al. (2020). Rethinking drug design in the artificial intelligence era. Nature Reviews Drug Discovery, 19(5), 353–364.

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

[6] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv:1503.02531.

[7] Hu, E.J., et al. (2022). LoRA: Low-rank adaptation of large language models. ICLR 2022.

[8] Chen, L., et al. (2023). FrugalGPT: How to use large language models while reducing cost and improving performance. arXiv:2305.05176.

[9] Jiao, X., et al. (2020). TinyBERT: Distilling BERT for natural language understanding. Findings of EMNLP 2020.

[10] Ren, F., et al. (2025). A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. Nature Biotechnology, 43, 63–75.

[11] Fedus, W., Zoph, B., & Shazeer, N. (2022). Switch Transformers: Scaling to trillion parameter models. JMLR, 23(120), 1–39.

[12] Kirkpatrick, J., et al. (2017). Overcoming catastrophic forgetting in neural networks. PNAS, 114(13), 3521–3526.