Artificial Intelligence in Drug Discovery: From Computational Target Identification to Clinical Translation

Misba Sheikh, Parijad Sheikh, Ummehani Pathan, Manisha Suryavanshi, Dr. Pankaj Pimpalshende

Hi-Tech College of Pharmacy, Padoli Phata, Morwa, Chandrapur - 442406

Corresponding author Email: misbaharuns1538@gmail.com
Doi:10.5281/zenodo.18835091
Received: 13 February 2026
Accepted: 20 February 2026

Abstract

Artificial intelligence (AI) has become an integral component of modern drug discovery, emerging as a structured and data-driven approach to address long-standing challenges such as rising development costs, extended timelines, and high failure rates. Traditional drug development processes remain inefficient, with many candidates failing in late-stage clinical trials due to inadequate efficacy, safety concerns, or unfavourable pharmacokinetics. Recent progress in computational capabilities, coupled with the rapid growth of large-scale biomedical and clinical datasets, has enabled broader and more reliable application of AI methodologies, including machine learning (ML), deep learning (DL), and natural language processing (NLP), across the drug discovery pipeline. These approaches support the integration and analysis of complex, high-dimensional data generated from genomics, proteomics, transcriptomics, chemical libraries, and real-world clinical sources.

This review outlines the application of artificial intelligence (AI) across the drug discovery pipeline, from target identification and lead optimisation to preclinical evaluation and clinical translation, with emphasis on virtual screening, de novo design, and pharmacokinetic and toxicity prediction.

Keywords: Artificial intelligence; drug discovery; computational target identification; clinical translation;

1. Introduction

Translational research aims to convert fundamental biological insights into clinically effective therapies, ultimately improving patient care. Despite notable advances in molecular biology, genomics, and pharmacological sciences, drug discovery and development remain costly, time- consuming, and associated with high attrition rates [1]. On average, the development of a new drug spans more than a decade and requires substantial financial investment, with failures occurring at multiple stages due to limited therapeutic efficacy, unexpected toxicity, or unfavourable pharmacokinetic and pharmacodynamic properties [2].

In this context, artificial intelligence has attracted increasing attention as a set of computational tools that support data-driven decision-making throughout the drug discovery process.

AI encompasses a broad range of algorithms designed to learn from data, identify patterns, and generate predictive insights. Its growing adoption in pharmaceutical research has been driven by the expansion of biological and chemical datasets, advances in high-throughput experimental technologies, and improvements in algorithms capable of handling complex, nonlinear, and high-dimensional data [3, 4].

Machine learning and deep learning methods offer clear advantages over traditional statistical approaches by enabling the analysis of large, heterogeneous datasets and revealing relationships that may remain undetected using conventional techniques [5]. Natural language processing further strengthens AI applications by allowing automated extraction and synthesis of information from the rapidly expanding biomedical literature, clinical trial databases, and regulatory documents. Recent studies increasingly highlight the benefits of integrating AI tools

across multiple stages of drug discovery and development, particularly in improving efficiency, reducing attrition, and supporting translational decision-making [6].

This review presents a critical and structured examination of AI-based methodologies applied in drug discovery, with emphasis on computational target identification, compound screening and molecular design, preclinical evaluation, and clinical translation. Recent developments are incorporated to reflect current progress and emerging trends toward clinical implementation [7].

Figure 1: The five key role of Generative AI in Drug Discovery

1.1 Computational Target Identification

Identifying and validating suitable biological targets is a fundamental step in the drug discovery process. Conventional experimental approaches, including gene knockdown or overexpression studies and biochemical assays, are often constrained by limited scalability, high costs, and incomplete representation of disease complexity [5, 8]. AI-driven computational strategies help overcome these limitations by enabling large-scale integration and analysis of diverse omics datasets.

Machine learning and deep learning models can simultaneously analyse genomic, transcriptomic, proteomic, metabolomic, and epigenomic data to identify disease-associated molecular signatures and prioritise potential therapeutic targets [9]. Multi-omics integration approaches facilitate a more comprehensive understanding of disease mechanisms and allow identification of functionally relevant targets that may not be evident when datasets are analysed independently. Network-based learning methods further support this process by modelling biological systems as interconnected networks of genes, proteins, and signalling pathways, enabling identification of key regulatory nodes and disrupted pathways [10].

Natural language processing also plays a significant role in computational target identification by extracting relationships among genes, proteins, diseases, and drugs from scientific literature and curated databases. Domain-specific biomedical language models allow rapid synthesis of existing knowledge, support hypothesis generation, and reduce reliance on manual literature curation. Recent studies demonstrate that AI-assisted literature mining can accelerate target discovery by revealing previously underexplored biological associations [11, 12].

Deep learning architectures, including convolutional and graph-based neural networks, have been applied to genomic and proteomic datasets to identify disease-driving mutations and protein interaction patterns. By learning hierarchical representations of complex biological features, these models enhance predictive consistency and accuracy. Overall, AI-enabled computational approaches improve analytical efficiency and expand the scope of target discovery, providing a robust foundation for downstream drug development [13].

Figure 2: Computational Target Identification

1.2 AI-Enhanced Compound screening and Molecular Design

Following target identification, discovering chemical entities capable of effectively modulating target activity remains a major challenge. Traditional high-throughput screening approaches are labour-intensive and often yield low hit rates. AI-assisted virtual screening has therefore emerged as a valuable complementary strategy for prioritising promising compounds [14].

Machine learning models trained on chemical structure descriptors, physicochemical properties, and bioactivity data can predict binding affinity, selectivity, and drug-likeness of candidate molecules. These models enable large-scale in silico screening of chemical libraries, substantially reducing the number of compounds requiring experimental evaluation. Deep learning techniques further improve predictive performance by capturing complex structure–activity relationships and nonlinear interactions between molecular features and biological targets [15, 16].

Generative modelling has significantly advanced AI-driven molecular design. Techniques such as variational autoencoders, generative adversarial networks, and reinforcement learning enable de novo generation of novel chemical structures optimised for predefined objectives, including potency, selectivity, and favourable pharmacokinetic profiles. These models allow iterative optimisation of candidate molecules while balancing multiple design constraints, thereby accelerating lead discovery and optimisation [17].

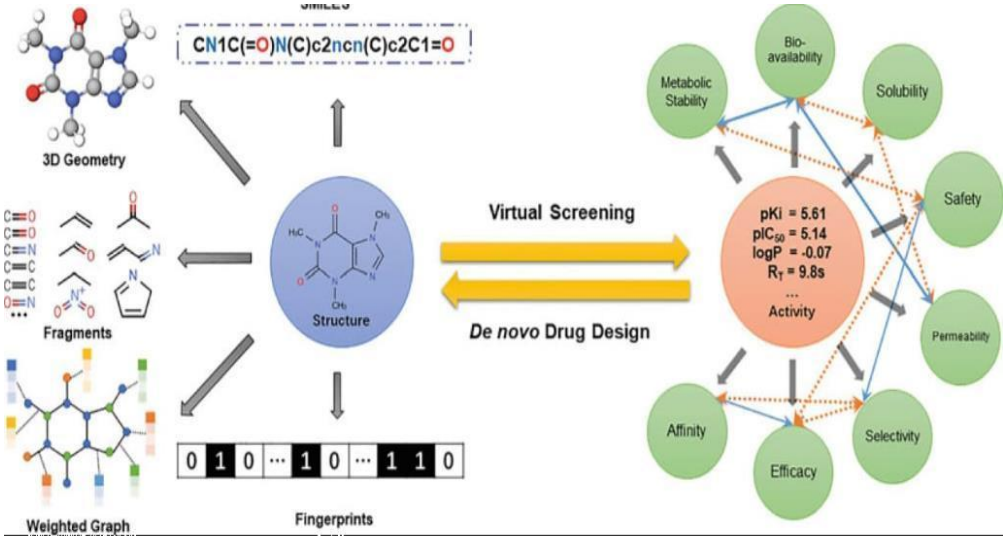

Figure 3: De Novo Drug Design

AI-based approaches are also widely applied in drug repurposing, where existing pharmacological, genomic, and clinical data are leveraged to identify new therapeutic indications for approved or investigational drugs. Given the availability of prior safety and pharmacokinetic information, AI-supported repurposing strategies offer a practical route to reduce development timelines and translational risk. Recent studies highlight the increasing role of AI in integrating chemical and biological data to support rational compound selection and design [18, 19].

1.3 Preclinical Evaluation and Toxicity Prediction

Preclinical evaluation assesses the safety, efficacy, and pharmacokinetic behaviour of drug candidates before clinical testing. Failures at this stage contribute substantially to development costs and delays. AI-based predictive models provide important tools to improve preclinical decision-making and minimise late-stage attrition [20].

Machine learning algorithms are commonly used to predict absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties by integrating chemical descriptors with biological assay data. These predictions support early identification of safety concerns and unfavourable pharmacokinetic profiles, guiding compound optimisation. Graph-based deep learning models, such as graph neural networks, offer improved molecular representations and have demonstrated enhanced performance in predicting toxicity and off-target effects [21].

AI-driven data analysis also complements advanced experimental systems, such as organ-on- a-chip and micro physiological models, by facilitating the interpretation of complex datasets generated under physiologically relevant conditions. Integrating computational predictions with advanced in vitro platforms improves translational relevance and supports more informed progression decisions. Recent research emphasises that AI-based predictions must be combined with experimental validation to ensure biological relevance and reliability [22].

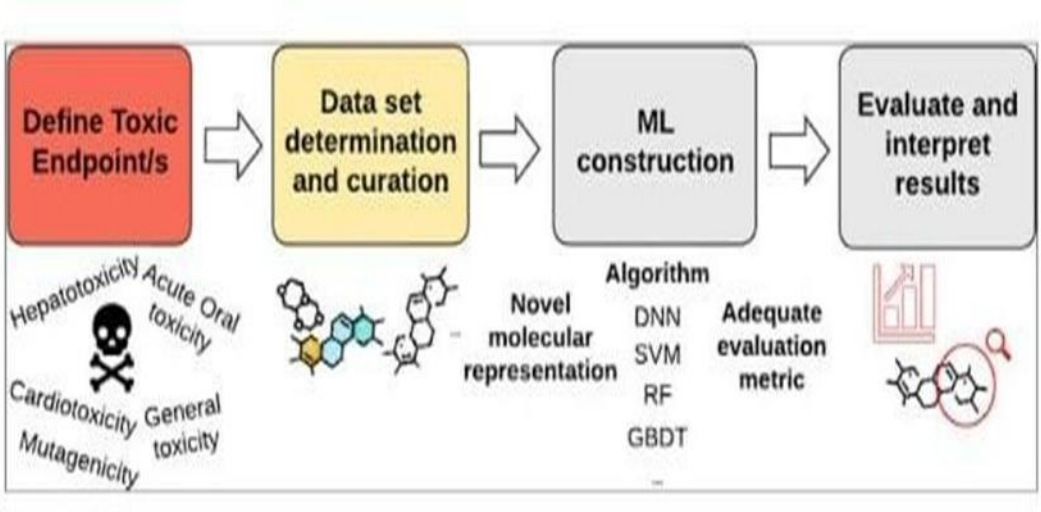

Figure 4: AI/ML – based drug toxicity prediction workflow

1.4 Clinical Translation and Trial Optimisation

Clinical translation remains one of the most challenging and resource-intensive stages of drug development. Despite encouraging preclinical findings, many candidates fail in clinical trials due to insufficient efficacy, unexpected safety issues, or limitations in trial design. AI offers data-driven strategies to address these challenges by improving trial planning, patient selection, outcome prediction, and real-time decision-making [23].

AI-based methods enable systematic analysis of heterogeneous clinical datasets, including electronic health records, genomic data, imaging results, laboratory measurements, and real- world evidence. Machine learning models can identify clinically relevant patterns that are difficult to detect using traditional analytical approaches. These insights support improved patient stratification by grouping individuals based on molecular features, disease subtypes, or predicted treatment response, thereby enabling more targeted and efficient clinical trials [24].

Predictive analytics also play a key role in trial optimisation by identifying and validating biomarkers associated with therapeutic response and safety. AI models can link molecular and clinical features to outcomes, supporting adaptive trial designs that allow protocol modifications based on interim analyses. Such designs reduce unnecessary exposure, improve resource allocation, and enhance overall trial efficiency.

In addition, AI-driven tools support patient recruitment and retention by improving eligibility identification and predicting factors associated with dropout or non-adherence. Real-time monitoring of patient data further enables early detection of adverse events, contributing to proactive safety management. Beyond trial execution, AI-supported clinical decision tools integrate pharmacogenomic information with therapeutic outcomes, advancing personalised medicine. However, successful clinical implementation requires rigorous validation, transparency, and alignment with regulatory expectations. When integrated with clinical expertise and ethical oversight, AI-enabled clinical translation can significantly improve trial success and accelerate access to effective therapies [25].

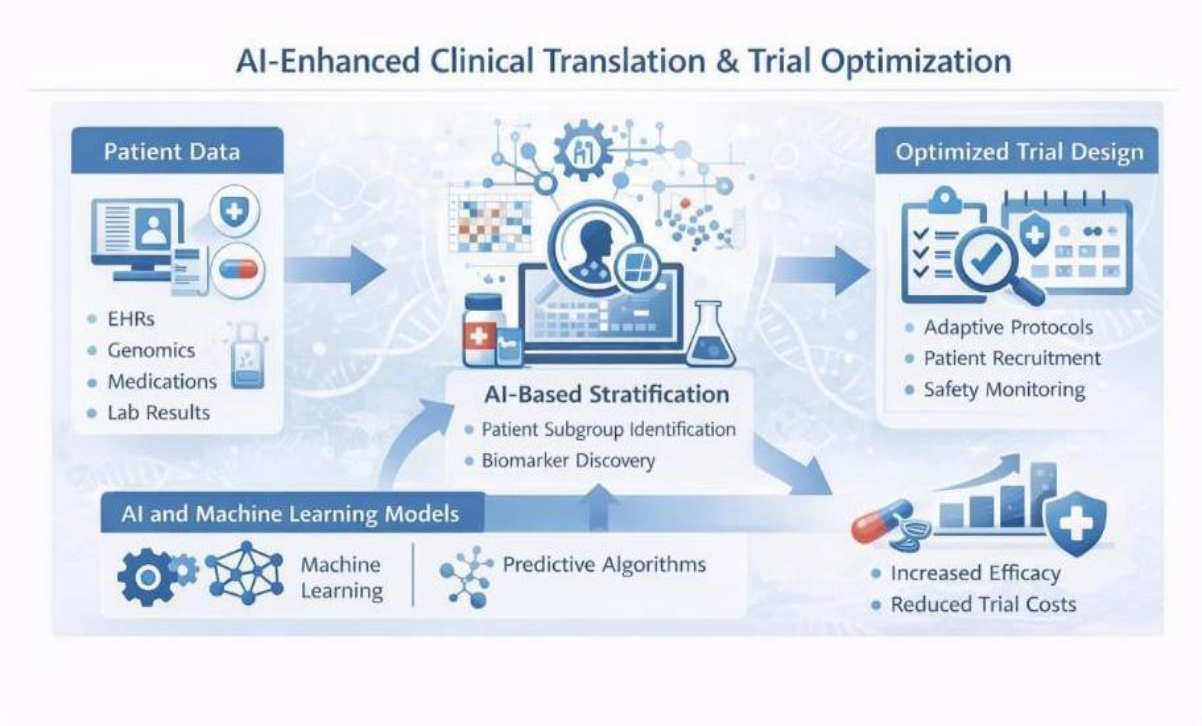

Figure 5: AI- Enhanced Clinical Translation and Trial Optimization

2. Discussion and Future Perspectives

Despite significant progress, several challenges continue to limit the widespread adoption of AI in drug discovery. Model performance remains highly dependent on the quality, completeness, and representativeness of available data. Variability across data sources, limited data sharing, and inherent biases can compromise generalizability. In addition, the limited interpretability of complex deep learning models poses challenges for regulatory review and clinical acceptance.

Future research should emphasise the development of interpretable and explainable AI approaches, standardised validation protocols, and collaborative data-sharing frameworks. Integration of multi-omics datasets, application of transfer learning, and adoption of federated learning strategies may enhance model robustness while preserving data privacy. Importantly, AI-based predictions must be evaluated alongside experimental and clinical evidence to ensure translational relevance and patient safety [26–28].

3. Conclusion

Artificial intelligence is increasingly influencing drug discovery by providing data-driven support across target identification, compound design, preclinical evaluation, and clinical translation. When applied appropriately, AI-based methodologies offer clear advantages in improving efficiency, reducing attrition, and strengthening decision-making throughout the development pipeline. Nevertheless, successful clinical translation of AI-enabled insights will depend on sustained attention to data quality, model transparency, validation, and regulatory alignment. With careful implementation and interdisciplinary collaboration, AI-supported approaches are well-positioned to play a transformative role in future therapeutic development and ultimately improve patient outcomes.

References

[1] Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery. *Drug Discov Today* 2018; 23: 1241–1250.

[2] Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* 2019; 18: 463–477.

[3] Mak K-K, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today* 2019; 24: 773–780.

[4] Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. *Drug Discovery Today* 2021; 26: 80–93.

[5] Ekins S, Puhl AC, Zorn KM, et al. Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater* 2019; 18: 435–441.

[6] Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. *Molecular Informatics* 2016; 35: 3–14.

[7] Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today* 2015; 20: 318–331.

[8] Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol* 2019; 37: 1038–1040.

[9] Stokes JM, Yang K, Swanson K, et al. A Deep Learning Approach to Antibiotic Discovery. *Cell* 2020; 180: 688-702.e13.

[10] Brown N, Fiscato M, Segler MHS, et al. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J Chem Inf Model* 2019; 59: 1096–1108.

[11] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021; 596: 583–589.

[12] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 2018; 555: 604–610.

[13] Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020; 577: 706–710.

[14] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell* 2020; 2: 573–584.

[15] Fernandes DC, Reis RL, Oliveira JM. Advances in 3D neural, vascular and neurovascular models for drug testing and regenerative medicine. *Drug Discovery Today* 2021; 26: 754–768.

[16] Paray BA, Ahmad A, Khan JM, et al. The role of the multifunctional antimicrobial peptide melittin in gene delivery. *Drug Discovery Today* 2021; 26: 1053–1059.

[17] Krenn M, Häse F, Nigam A, et al. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach Learn: Sci Technol* 2020; 1: 045024.

[18] Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019; 18: 41–58.

[19] Harrer S, Shah P, Antony B, et al. Artificial Intelligence for Clinical Trial Design. *Trends in Pharmacological Sciences* 2019; 40: 577–591.

[20] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; 25: 44–56.

[21] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *New England Journal of Medicine* 2019; 380: 1347–1358.

[22] Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA* 2018; 319: 1317–1318.

[23] He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in

medicine. *Nat Med* 2019; 25: 30–36.

[24] Regulatory science strategy | European Medicines Agency (EMA), https://www.ema.europa.eu/en/about-us/how-we-work/regulatory-science-strategy (2018, accessed 4 February 2026).

[25] Arrowsmith J, Miller P. Phase II and Phase III attrition rates 2011–2012. *Nature Reviews Drug Discovery* 2013; 12: 569–569.

[26] Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* 2020; 323: 844–853.

[27] Van Norman GA. Drugs, Devices, and the FDA: Part 1. *JACC: Basic to Translational Science* 2016; 1: 170–179.

[28] Health C for D and R. Artificial Intelligence in Software as a Medical Device. *FDA*, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence- software-medical-device (2025, accessed 4 February 2026).