IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

Comparative Analysis of AI-Driven Drug Discovery and Drug Repurposing

Saniya Shaikh[1], Sumaiyya Kamble[2], Hasan Phudinawala[3]
[1,2] P.G. Students, Department of Data Science
[3]Coordinator, Department of Data Science
Royal College of Arts, Science and Commerce (Autonomous), Mira Road(E)

Abstract: *Artificial intelligence (AI) is essentially remodelling pharmaceutical research and development, which calls for a comparative analysis of its use in de novo drug discovery (DNDD) and drug repurposing (DRP). This review of the literature compares the strategic and methodological divergence of these two avenues. DNDD uses generative AI to canvass immense chemical space to produce high-risk, high-reward new entities but takes 10–17 years and more than $2 billion. In contrast, DRP utilizes predictive AI and network-based approaches (e.g., Graph Neural Networks, multi-omics integration) to predict novel therapeutic applications for known, safety-tested compounds. DRP provides shortened timelines (3–12 years), significantly lower costs, and triple the success rate. Success in DNDD is dependent on fidelity of generative models and synthesizability, whereas DRP depends upon the ability to perform scalable, explainable multi-omics data integration. Ultimately, the two plans are complementary, with DRP offering clinical speed and DNDD guaranteeing long-term pipeline invention.*

Keywords: Artificial Intelligence (AI), Drug Discovery, Drug Repurposing, Generative AI, Predictive Modeling, Multi-omics, Graph Neural Networks, Precision Medicine

I. INTRODUCTION

The traditional process of pharmaceutical research and development (R&D) is currently defined by significant capital expenditure and protracted timelines. It is estimated that bringing a novel chemical entity (NCE) to market requires an investment exceeding $2 billion and a development period typically spanning 10 to 17 years. [1] This long duration and immense cost are exacerbated by exceptionally high attrition rates, with failure often occurring in the late stages of clinical development, such as Phase II or Phase III, particularly when targeting complex diseases like neurodegeneration. [1] The escalating complexity of biological systems and drug targets, combined with increasing demands for regulatory rigor, mandates the adoption of disruptive technological solutions to maintain the economic viability of the R&D pipeline. [1] Artificial intelligence (AI), characterized as a machine-based system capable of automated analysis and inference, offers a solution to this crisis. [3] AI methodologies, leveraging subsets like machine learning (ML), provide rational, high-efficiency tools for assessing complex parameters and selecting high-value entities across all phases of drug development. [4] The adoption of AI is driven primarily by the need to enhance the efficacy of the targeted approaches and directly mitigate the critical risk associated with expensive late-stage clinical failures. Within the AI-accelerated pharmaceutical landscape, two primary, structurally divergent strategies have emerged: *de novo* drug discovery (DNDD) and drug repurposing (DRP), often referred to as the repositioning. DNDD involves the use of computational methods, specifically generative AI algorithms, to systematically explore vast chemical spaces and construct entirely new molecules.[5]

The objective is to design compounds with optimized properties, including targeted molecular interactions, absorption, distribution, metabolism, excretion, and toxicity (ADMET) characteristics. [4] Drug repurposing (DRP) focuses on identifying novel therapeutic indications for compounds that already possess established safety and pharmacokinetic profiles. [6] This strategic approach, coined in 2004, is defined as finding new uses for existing drugs outside their original medical scope. [5] Because DRP capitalizes on pre-existing data, it offers a crucial advantage by often bypassing the early, high-risk, and costly stages of development, potentially qualifying compounds directly for Phase II clinical trials.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

[6] AI methodologies are now integrated throughout the entire drug product life cycle, supporting nonclinical, clinical, post marketing, and manufacturing phases for both DNDD and DRP. [6]

The Scope of the paper is to broadly explore the applications of AI in pharmaceutical R&D, there is a strategic need for a distinguished, in-depth comparative analysis focused specifically on the methodological, strategic, and translational trade-offs between AI-driven DNDD and AI-driven DRP. [4] This report fulfils that need by critically evaluating the core divergence in algorithmic requirement generative models for DNDD versus predictive, network-based models for DRP. The analysis utilizes a synthesized review of 14 peer-reviewed research papers, contrasting their technical implementations, strategic implications (economic and regulatory), and inherent limitations, thereby providing a nuanced perspective on their respective strengths and long-term utility in the precision medicine era.

II. COMPARATIVE ANALYSIS

2.1 AI in De Novo Drug Discovery (DNDD)

AI in De Novo Drug Discovery (DNDD) is characterized by the use of algorithms designed to generate novel data points, specifically focusing on generating molecular structures that possess desired biological and physicochemical profiles. The primary methodologies employed include Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), reinforcement learning (RL) agents, and flow models. [8] These generative algorithms are designed to navigate the complex, high-dimensional latent space of chemistry, allowing for the rapid, semi-automatic design and optimization of drug-like molecules. [5] A landmark success demonstrating the practical utility of these architectures is the work by Zhavoronkov et al. (2019), who employed a deep generative model called Generative Tensorial Reinforcement Learning (GENTRL). [9] The study aimed to rapidly identify potent DDR1 kinase inhibitors, a key target implicated in fibrosis. GENTRL successfully optimized molecular properties, including novelty, synthetic feasibility, and biological activity, resulting in the discovery of a lead candidate in just 21 days. [9] This rapid turnaround, validated through cell-based assays and *in vivo* pharmacokinetic (PK) testing, illustrates the capacity of generative models to design novel, highly optimized chemical matter with unprecedented speed. Furthermore, Graph Neural Networks (GNNs) are integral to enhancing DNDD efforts. GNNs effectively process complex molecular graphs, which are crucial for high-fidelity molecular property prediction. [11] The deployment of Graph Attention Mechanisms, as detailed by Xiong et al. (2020), enhances the accuracy of prediction by capturing nuanced structural relationships, thereby strengthening the quality of lead optimization. [12] However, the utility of generative AI is subject to a credibility challenge: to be truly transformative for chemistry, the generated compounds must be capable of predicting phenomena not previously observed. [8] Achieving this level of predictive utility necessitates that future AI models incorporate core chemical principles, such as statistical mechanics, ensuring that generated structures are robust and synthetically tractable, rather than simply replicating patterns within existing data. [8]

2.2 AI in Drug Repurposing (DRP)

AI in Drug Repurposing is fundamentally an inferential process, relying on algorithms to predict hidden associations between existing compounds and novel diseases or targets. This strategy utilizes both supervised and unsupervised machine learning, including deep learning models optimized for association and classification tasks. [6] Examples include the unsupervised ML techniques MANTRA and PREDICT, which forecast therapeutic efficacy by analysing gene expression profiles from extensive datasets like Connectivity Map (CMap). [7] These models cluster compounds based on shared mechanisms of action and common biological pathways, thereby suggesting new indications based on inferred target perturbation. [7] The success of DRP hinges upon the integration of multimodal, multi-omics data, including genomics, transcriptomics, and proteomics. [6] This integration provides an unparalleled systems-level context required for inferring complex drug-disease associations. [14] Transcriptomics data, for example, allows researchers to navigate drug potential based on expression patterns, offering mechanistic insights into the effects of known drugs. [6] AI algorithms excel at analysing these large-scale datasets, identifying intricate patterns of drug responses that evade traditional detection methods. [6] Network-based approaches, particularly GNNs used for predicting drug-target interactions (DTI) and drug-drug interactions (DDI), are the dominant analytical tool in DRP. [12] These methods are crucial for harmonizing complex biological data, organizing molecular interactions into coherent systems biology

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

frameworks that enable the prediction of new therapeutic targets. [14] However, the use of vast, multimodal data also introduces the greatest technical difficulty: the challenge of data heterogeneity and computational scalability. [14] Successfully integrating and harmonizing disparate data sources such as merging genomics data with complex image assay data or clinical records is a critical bottleneck that defines the practical limits of predictive DRP models. [6]

2.3 Strategic and Economic Comparative Analysis

The strategic choice between DNDD and DRP is often dictated by economic viability and acceptable risk tolerance. The quantifiable benefits of DRP are substantial: the average cost to bring a repurposed drug to market is approximately $300 million, potentially reducing the baseline cost of R&D by up to 60% compared to a *de novo* agent. [1] This immense cost saving stems primarily from leveraging existing safety and toxicity data, which often eliminates the need for full preclinical studies and Phase I safety trials, allowing compounds to move directly into Phase II trials. [6] The reduced timeline of DRP, typically ranging from 3 to 12 years, represents a significant acceleration compared to the 10 to 17 years required for DNDD. [1] Furthermore, the probability of approval for a repurposed drug is reported to be around 30%, constituting a threefold increase over the success rate for new drug applications. [15] AI amplifies these strategic advantages by improving candidate selection and optimizing clinical trial design, thereby driving R&D costs down by an additional 40%. [2] This heightened capital efficiency and clinical agility position DRP as a crucial strategic hedge, particularly for pharmaceutical companies seeking quicker returns on investment or addressing diseases with immediate therapeutic gaps. DNDD, while essential for long-term therapeutic breakthrough, represents a high-risk, long-term investment.

Table 1: Comparative Summary of Traditional Drug Discovery and AI-Accelerated Repurposing

Metric	**Traditional De Novo Discovery (DNDD)**	**Drug Repurposing (DRP) Baseline**	**AI-Accelerated Strategy**
Average Cost to Market	Less than $2 Billion [2]	Million [15]	Potential for further 40% R&D cost reduction via AI optimization [2]
Development Timeline	10–17 Years [1]	3–12 Years [15]	Accelerated through high-throughput computational analysis [4]
Initial Clinical Stage	Phase I Safety Trials Required	Often bypasses Phase I (starts at Phase II) [7]	Depends on new dosage/route [1]

III. LITERATURE REVIEW

The comparative overview in **Table 1** highlights how AI-accelerated strategies are transforming the economics and timelines of drug discovery and repurposing. To support these trends with empirical evidence, **Table 2** summarizes major studies that demonstrate how specific AI models and computational techniques have achieved the cost reductions, time efficiency, and clinical advantages outlined in **Table 1.** Together, these tables illustrate the shift from traditional experimental discovery toward data-driven, computationally optimized pipelines in modern pharmaceutical research.

Table 2: Empirical Evidence Supporting AI-Accelerated Drug Discovery and Repurposing

Author and Title	**Objective**	**Key Findings**	**Advantages**	**Challenges**	**Linked Metric (from Table 1)**
Zhavoronkov et al. (2019): Deep learning enables rapid	Develop a deep generative model (GENTRL) for	GENTRL discovered potent DDR1 kinase inhibitors in 21	Ultra-rapid discovery timeline (21 days); ability to	Requires high-quality structural data; risk of	**Reduced cost and development time through**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

identification of potent [2]	*de novo* small molecule design, optimizing synthetic feasibility, novelty, and biological activity. [8]	days; four compounds were biochemically active, and one demonstrated favorable pharmacokinetics in mice. [8]	generate novel, optimized, and validated molecules. [8]	failure in late-stage validation.	**AI-based molecular generation.**
Zitnik et al. (2018): Modelling polypharmacy side effects with Graph Convolutional Networks [3]	Model polypharmacy side effects and drug–drug interactions (DDI) using Graph Convolutional Networks (GCNs). [10]	Developed a GCN model that captured complex DDI relationships and predicted side effects accurately. [10]	Improves safety prediction and understanding of multi-drug interactions.	High computational cost; need for explainable AI (XAI).	**Improved preclinical safety and reduced attrition risk.**
Xiong et al. (2020): Pushing the Boundaries of Molecular Representation with Graph Attention Mechanisms [10]	Enhance molecular property prediction and representation using Graph Attention Networks (GATs). [10]	Demonstrated higher accuracy in predicting molecular properties and lead optimization. [10]	Increases quality and efficacy of candidate molecules during lead optimization.	Must predict unseen molecular behaviors to be transformative.	**Accelerated lead optimization; better compound selection.**
Altae-Tran et al. (2017): Low-Data Drug Discovery with One-Shot Learning [10]	Apply One-Shot Learning for effective drug discovery using minimal training data. [10]	Successfully predicted drug properties even with limited datasets. [10]	Reduces reliance on large datasets; useful for rare diseases and novel targets.	Limited generalizability; data bias due to few known structures.	**Reduces cost and data dependency in early-stage discovery.**
Simm et al. (2018): Repurposing High-Throughput Image Assays Enables Biological Activity Prediction [11]	Combine image-based assays with AI to predict compound biological activity for repurposing. [11]	Showed that image-based data can effectively predict compound activity. [11]	Uses rich phenotypic data; enables screening of existing compounds for new uses.	Challenges in integrating and standardizing image and molecular data.	**Supports rapid drug repurposing and re-validation.**
Stokes et al. (2020): Halicin – A Deep Learning	Use deep learning to identify molecules with	Discovered *Halicin*, a non-antibiotic molecule acting as a	Rapid identification of new drug uses; success in	Requires extensive in vivo validation for new applications.	**Cost and time reduction via AI-based drug repurposing**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

Approach to Antibiotic Discovery [11]	new or unrecognized therapeutic effects. [11]	broad-spectrum antibiotic. [11]	antibiotic discovery.		
Jin et al. (2024): Multi-omics Model Prioritizing SPP1 and SMAD3 for COPD [12]	Integrate genome, transcriptome, proteome, and metabolome data to find repurposing targets for COPD. [12]	Identified SPP1 and SMAD3 as high-potential drug targets using a distance-based model. [12]	Combines multiple biological layers; strong potential for novel target validation. [12]	Limited population diversity; computational scalability issues. [12]	**Multi-omics integration for target discovery.**
Simm et al. (2018): Unsupervised ML (MANTRA, PREDICT) for Drug Repurposing [7]	Use unsupervised ML to predict therapeutic efficacy and mechanisms for DRP. [7]	MANTRA grouped compounds by gene-expression profiles, revealing shared mechanisms. [7]	Generates hypotheses without prior knowledge; efficient large-scale screening. [7]	Needs experimental validation; possible dataset bias.	**Accelerated hypothesis generation and mechanism identification.**
Sun et al. (2024): Network-Based Multi-Omics Integration for Drug Discovery [13]	Review and categorize network-based multi-omics integration approaches. [13]	Highlighted GNNs and propagation models as promising methods. [13]	Provides a systems-level understanding of complex biological data. [13]	Computationally demanding; data harmonization challenges. [13]	**Data integration supporting AI-driven analysis (cost & time efficiency).**
Xu, Li & Lin (2024): Computational Drug Repurposing using Deep Learning [6]	Review deep learning for drug repurposing, focusing on omics applications. [6]	Demonstrated transcriptomics-based prediction of drug potential and mechanism. [6]	Reuses safe drugs efficiently; enhances accuracy of repurposing workflows. [6]	Integration across databases is challenging; computational intensity. [6]	**Improved repurposing speed; reduced R&D cost.**
Saha, Manna & Bhattacharya (2024): Comparative Analysis of AI in Drug Discovery [4]	Compare AI methodologies across all stages of the drug discovery pipeline. [4]	Confirmed AI's role in rational parameter selection for high-value drug candidates. [4]	Transforms researcher's role into "hypothesis strategist." [2]	Requires transparency and explainable AI (XAI). [4]	**Applies AI across all phases to shorten timelines and improve success.**
Vamathevan et al. (2019): Artificial Intelligence in Drug Discovery and	Review AI's integration across the R&D process, including	Found AI integrated into non-clinical, clinical, and post-marketing phases. [3]	Improves R&D efficiency and market delivery speed. [1]	Needs clearer regulatory frameworks and expert training. [3]	**Overall R&D acceleration; supports regulatory AI adoption.**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

Development [14]	DNDD and DRP. [3]				
Ashburn & Thor (2004): Drug Repositionig Identifying and Developing New Uses for Existing Drugs [1]	Define and formalize the concept of drug repositioning. [1]	Introduced drug repositioning as finding new uses for existing drugs. [1]	Low cost and risk; avoids early safety trials. [1]	Limited patent rights; may require new trials. [1]	**Baseline model for AI-enhanced repurposing efficiency.**
(GNN Review – 2021+): Role of Graph Neural Networks in Drug Discovery [15]	Review GNNs for molecule generation, property prediction, and DDI modeling. [18]	GNNs captured complex molecular patterns, enabling *de novo* design. [18]	Automates molecule design and reduces lab effort. [18]	Must integrate chemical rules for realistic predictions. [16]	**Automation of early discovery stages; reduced R&D workload.**

IV. CONCLUSION

In summary, the integration of Artificial Intelligence within De Novo Drug Discovery (DNDD) and Drug Repurposing (DRP) represents a transformative convergence rather than a competition between two approaches. DNDD fuels long-term pharmaceutical innovation by generating entirely new molecular entities to address complex biological challenges, while DRP offers rapid, cost-effective, and lower-risk solutions by identifying new therapeutic applications for existing drugs. Together, they create a balanced ecosystem of innovation and efficiency within the modern drug development pipeline.

The success of both strategies depends heavily on the nature and quality of available data. DNDD relies on high-fidelity structural and biochemical information to train generative models capable of producing novel compounds, whereas DRP depends on integrating vast, heterogeneous, and often imperfect datasets drawn from multi-omics and real-world evidence. Bridging these data-driven approaches are Graph Neural Networks (GNNs), Linking molecular design with biological interpretation to generate deeper, more meaningful insights across both domains.

Looking ahead, the continued evolution of Explainable AI (XAI) will be essential to ensure transparency, regulatory compliance, and scientific credibility in AI-driven research. As these technologies advance, the role of pharmaceutical scientists will continue to evolve, from performing manual experiments to acting as strategic data curators, hypothesis architects, and validators of AI-generated outcomes. Ultimately, the synergy between DNDD and DRP underscores a new era in pharmaceutical innovation that one where human expertise and artificial intelligence work in tandem to accelerate discovery, reduce uncertainty, and deliver safer, more effective therapies to patients worldwide.

REFERENCES

[1] Ashburn T. T., Thor K. B. (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov.

[2] Zhavoronkov A, Ivanenkov YA, Aliper A, et al (2019) Deep learning enables rapid identification of potent. Nature Biotechnology.

[3] Zitnik M, Agrawal M, Leskovec J (2018) Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics.

[4] Xiong Zp, Chen H, Tan X (2020) Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. J Med Chem.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, October 2025
Impact Factor: 7.67

[5] Altae-Tran H, Ramsundar B, Pappu AS (2017) Low-Data Drug Discovery with One-Shot Learning. ACS Central Sci.
[6] Simm J, Klambauer G, Arany A (2018) Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. Cell Chem Biol.
[7] Stokes JM, Lopatkin AJ, Luthra P (2020) A Deep Learning Approach to Antibiotic Discovery. Cell.
[8] Saha K, Manna P, Bhattacharya M. (2024) Comparative analysis of AI methodologies across the entire drug discovery pipeline: Practical challenges and future directions. Frontiers in Drug Discovery.
[9] Sun J, Ma X, Huang P. Network-based approaches for multi-omics data integration in drug discovery: A comprehensive review (2015–2024). Briefings in Bioinformatics. (2024)
[10] Ozen M., Emamian E. S., Abdi A (2023) From data to knowledge: A mini-review on molecular network modeling and analysis for therapeutic target discovery. Archives of Pharmacology and Therapeutics.
[12] Vamathevan JJ, Clark D, Czodrowski P (2019) Artificial intelligence in drug discovery and development. Nature Reviews Drug Discovery.
[13] Xu C., Li Z., Lin H (2024) Computational drug repurposing using deep learning approaches: advances and challenges. Adv Sci (Weinh).
[14] Simm J, Klambauer G, Arany A (2018) Unsupervised machine learning methods MANTRA and PREDICT for drug repurposing. Mol Cell Proteomics.
[15] Mohanambigai A (2023) AI and Machine Learning in Drug Repurposing: A Review. Journal of Pharmaceutical Research International.
[16] Jin Y (2024-2025) Multi-omics analysis and computational model prioritize SPP1 and SMAD3 as potential drug-repurposing targets for COPD. Int J Mol Sci. (2024-25)