JOURNAL OF PHARMA INSIGHTS AND RESEARCH ISSN NO. 3048-5428

REVIEW ARTICLE

A Review on the Convergence of Artificial Intelligence and Drug Discovery

Komal Dattu Gunjal*[1], Ramdas Balu Darade[2], Vikram Sadashiv Saruk[2], Manoj Chhatrapati Garad[2], Priti Dilip Bhure[1], Anuja Gajanan Ghaywat[1], Swati Rajendra Gaykar[1]

[1] *UG Scholar, Department of Pharmaceutical Chemistry, Jagdamba Education Society's S. N. D. College of Pharmacy, Babhulgaon, Nashik, Maharashtra, India*
[2] *Assistant Professor, Department of Pharmaceutical Chemistry, Jagdamba Education Society's S. N. D. College of Pharmacy, Babhulgaon, Nashik, Maharashtra, India*

Publication history: Received on 14th July 2025; Revised on 22nd Aug 2025; Accepted on 29th August 2025

Article DOI: 10.69613/ej513k09

Abstract: The use of artificial intelligence (AI) and machine learning (ML) has accelerated a fundamental transformation in the field of drug discovery, addressing long-standing challenges of time, cost, and attrition rates. This paper presents a comprehensive analysis of the critical role AI now plays across the pharmaceutical research and development pipeline. Key AI-driven applications are detailed, including genomics- and proteomics-based target identification, high-throughput virtual screening for hit discovery, and generative models for de novo molecular design. Furthermore, the advancements in predictive modeling for absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties are examined, alongside the strategic repurposing of existing drugs for new therapeutic indications. The discussion extends to the foundational elements of this technological shift, such as the crucial data sources, diverse molecular representation techniques, and the spectrum of ML algorithms from classical methods to advanced deep learning architectures like graph neural networks and transformers. Through an examination of recent case studies, the tangible impact of AI in accelerating discovery timelines is highlighted. Persistent challenges, including data quality, model interpretability, and the evolving regulatory landscape, are also critically assessed. The success and integration of AI in medicine discovery depends on the robust benchmarking, transparent validation, and seamless incorporation into experimental workflows, heralding a new era of precision medicine

Keywords: Artificial Intelligence; Drug Discovery; Machine Learning; Generative Models; ADMET Prediction

1. Introduction

The process of discovering and developing novel therapeutics has traditionally been one of the most arduous and capital-intensive endeavors in biomedical science. The journey from an initial biological hypothesis to a market-approved drug typically spans over a decade and incurs costs that can exceed billions of dollars [1, 2]. This protracted timeline is further complicated by exceedingly high attrition rates, with a significant majority of candidate molecules failing during late-stage clinical trials due to unforeseen issues with efficacy or safety [3]. In response to these profound challenges, the pharmaceutical industry has increasingly turned to computational methods to streamline and de-risk the discovery pipeline.

Over the past decade, artificial intelligence (AI), and specifically its subfield of machine learning (ML), has emerged as a disruptive force with the potential to redefine the principles of drug discovery [1]. AI algorithms can identify patterns, generate hypotheses, and make predictions with a speed and scale unattainable through human effort alone by utilizing vast and complex datasets [4]. The application of these technologies is no longer a futuristic concept but a present-day reality, with AI tools being integrated into nearly every phase of the discovery process.

These applications range from the initial identification and validation of novel biological targets to the design of new chemical entities with optimized pharmacological profiles [5, 26]. AI now functions as a powerful augmentation tool, empowering researchers to make more informed decisions, accelerate timelines, and ultimately increase the probability of success for new therapeutic programs. This review work will discuss the methodologies, applications, and challenges that define the current landscape of AI-driven drug discovery.

* Corresponding author: Komal Dattu Gunjal

2. The Drug Discovery Pipeline: Integrating Artificial Intelligence

The conventional drug discovery pipeline is a linear, multi-stage process characterized by significant investment and a high probability of failure. It traditionally begins with target identification and validation, progresses through hit discovery and lead optimization, and culminates in extensive preclinical and clinical testing [1, 2]. Each stage acts as a filter, with the number of candidate compounds decreasing exponentially. This process is inherently inefficient; for instance, high-throughput screening (HTS) may assess millions of compounds to identify a handful of promising "hits" [6]. The subsequent optimization of these hits into clinical candidates is a multi-parameter challenge that often fails due to unforeseen issues in pharmacokinetics or safety, which may only become apparent in late-stage development [3, 22].

Artificial intelligence provides a suite of tools to address these inefficiencies at every stage. The integration of AI aims to transform the pipeline from a high-risk, sequential process into a more predictive, data-driven, and iterative cycle.

2.1. Target Identification and Validation

The selection of a viable biological target is the foundational step of any drug discovery campaign. AI algorithms can systematically analyze vast, multimodal datasets—including genomics, proteomics, and extensive scientific literature—to identify and prioritize novel targets with a strong link to disease pathology [28, 31]. Knowledge graphs, for example, can construct complex networks of biological relationships, uncovering non-obvious connections between genes, proteins, and diseases that might be missed by human researchers [4]. Large language models (LLMs) are also being trained on biomedical corpora to accelerate hypothesis generation by rapidly synthesizing information from millions of research articles [26].

2.2. Hit Discovery and Lead Optimization

Following target identification, the goal is to find molecules that can modulate the target's activity. AI-powered virtual screening can computationally evaluate libraries of billions of chemical compounds far more efficiently than physical HTS [22]. Predictive models, particularly those based on deep learning, can forecast the binding affinity of a molecule to a target with increasing accuracy [6]. During lead optimization, the challenge shifts to refining a hit's properties to achieve a balance of potency, selectivity, and favorable ADMET characteristics. AI excels at this multi-objective optimization, concurrently predicting a suite of properties and suggesting structural modifications to improve the overall profile of a candidate molecule, thereby accelerating the design-make-test-analyze cycle [3].

3. Molecular Representations in AI-Driven Discovery

The efficacy of any AI model is fundamentally dependent on the quality and volume of the data used for its training. The field of drug discovery benefits from several large, publicly accessible databases that have become indispensable resources. Repositories such as ChEMBL, PubChem, and DrugBank contain millions of data points on chemical structures and their associated biological activities [16, 20]. The Protein Data Bank (PDB) provides essential 3D structural information for biological macromolecules [21]. Furthermore, specialized resources like the Genomics of Drug Sensitivity in Cancer (GDSC) offer crucial links between molecular features and phenotypic responses, enabling the development of models for personalized medicine [35].

A critical determinant of model performance is the method used to represent a molecule in a machine-readable format. Several distinct approaches have been developed:

3.1. 1D Representations

Simplified Molecular-Input Line-Entry System (SMILES) strings encode the structure of a molecule as a sequence of characters. This format has enabled the application of natural language processing (NLP) models, such as transformers, to chemical data [16, 26].

3.2. 2D Representations

These include molecular fingerprints, which are bit vectors that encode the presence or absence of specific substructural features (e.g., Extended-Connectivity Fingerprints or ECFPs), and calculated physicochemical descriptors (e.g., molecular weight, logP). These have been mainstays of quantitative structure-activity relationship (QSAR) modeling for decades [22].

3.3. Graph-Based Representations

Increasingly, molecules are treated as graphs, where atoms are nodes and chemical bonds are edges. This is a more natural representation of molecular topology and is the ideal input for graph neural networks (GNNs), which have demonstrated state-of-the-art performance in property prediction [6].

3.4. 3D Representations

To capture the spatial arrangement of atoms, molecules can be represented by their 3D coordinates, as point clouds, or on a 3D grid (voxels). These representations are vital for structure-based drug design tasks where the geometry of the protein-ligand interaction is paramount [35].

Table 1. Public Databases for Drug Discovery

Database Name	Primary Data Type	Example Content	URL
ChEMBL	Bioactivity	SAR data, IC50/EC50 values, drug targets	www.ebi.ac.uk/chembl
PubChem	Chemical Compounds	Small molecules, structures, chemical properties	pubchem.ncbi.nlm.nih.gov
DrugBank	Drug & Target Info	FDA-approved drugs, mechanisms, interactions	go.drugbank.com
Protein Data Bank (PDB)	3D Structures	Experimentally determined protein & nucleic acid structures	www.rcsb.org
TCGA / GDSC	Genomics/Phenotypes	Cancer genomics, drug sensitivity in cell lines	portal.gdc.cancer.gov
ClinVar	Genetic Variation	Relationships between human variations and phenotypes	www.ncbi.nlm.nih.gov/clinvar

The recent emergence of foundation models trained on vast unlabeled chemical datasets is also creating powerful, pre-trained molecular embeddings that can be fine-tuned for a wide range of predictive tasks, enhancing model generalization and reducing the need for extensive task-specific data [35].

4. Core AI and Machine Learning Techniques

A diverse array of AI and ML algorithms is employed in drug discovery, ranging from established statistical methods to sophisticated deep learning architectures.

4.1. Classical Machine Learning

Traditional ML algorithms such as Random Forests, Support Vector Machines (SVMs), and Gradient Boosting methods remain widely used, particularly for QSAR and ADMET modeling [24]. Their continued relevance stems from their robustness, especially with smaller or less complex datasets, and their relatively higher degree of interpretability compared to deep learning models [14].

4.2. Deep Learning Architectures

Deep learning has been responsible for many of the recent breakthroughs in the field. The main architectures include:

4.2.1. Graph Neural Networks (GNNs)

As the current standard for many molecular modeling tasks, GNNs operate directly on the graph structure of molecules, enabling them to learn intricate structure-property relationships and achieve superior predictive accuracy for tasks like virtual screening and interaction modeling [6, 22].

4.2.2. Transformers and Recurrent Neural Networks (RNNs)

These sequence-based models are primarily applied to SMILES strings for de novo molecular design and chemical reaction prediction. They excel at learning the grammatical and syntactic rules governing valid chemical structures [17, 18].

4.2.3. Convolutional Neural Networks (CNNs)

CNNs are well-suited for processing grid-like data. In drug discovery, they are often used to analyze 3D representations of protein-ligand binding pockets or to classify cellular images from high-content screening assays [25].

4.3. Generative and Reinforcement Learning Models

Generative models have revolutionized the design of novel molecules. Early approaches utilized Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) [9]. More recently, advanced methods like diffusion models and transformer-based generators have demonstrated superior performance in generating valid, novel, and diverse chemical structures [35]. To steer the generation process towards molecules with a specific desired profile, these models are often coupled with Reinforcement Learning (RL). In this paradigm, an "agent" (the generative model) is rewarded for designing molecules that meet predefined multi-objective criteria, such as high target potency, low predicted toxicity, and high synthetic accessibility [17, 18].

Table 2. Comparison of Core AI/ML Methodologies in Drug Discovery

Methodology	Primary Strength	Common Applications	Limitations
Graph Neural Networks (GNNs)	Capturing molecular topology and relational data	Property prediction, virtual screening, protein interaction	Can be computationally expensive; performance depends on graph quality
Transformers	Processing sequential data and learning long-range dependencies	*De novo* design (SMILES), reaction prediction, literature mining	Requires large datasets for training; less inherently suited for 3D structures
Generative Adversarial Networks (GANs)	Generating novel and diverse molecular structures	*De novo* design, scaffold hopping	Prone to training instability (mode collapse); ensuring validity can be difficult
Reinforcement Learning (RL)	Optimizing molecule generation towards specific goals	Multi-objective lead optimization, synthesis planning	Defining appropriate reward functions is challenging; can be sample-inefficient
Support Vector Machines (SVMs) / Random Forests	Robust on smaller datasets; good interpretability	QSAR modeling, toxicity classification, ADMET prediction	Less effective on complex, high-dimensional data compared to deep learning

4.4. Explainable AI (XAI)

A significant limitation of many deep learning models is their "black box" nature. To address this, XAI techniques are being developed to provide insights into model predictions. Methods such as saliency mapping and counterfactual reasoning can help medicinal chemists understand which parts of a molecule are contributing to a predicted property, thereby facilitating more rational and trust-based drug design [14, 15].

5. Applications of Artificial Intelligence in Drug Discovery

The theoretical methodologies of AI translate into a wide range of practical applications that are actively reshaping the drug discovery landscape. These tools are being deployed to address specific bottlenecks at various stages of the R&D pipeline, leading to significant gains in efficiency and novel scientific insights.

5.1. Target Identification and Validation

Identifying a valid biological target is a critical first step. AI platforms accelerate this process by integrating and analyzing diverse, large-scale biological data. AI models can uncover novel correlations between biological entities and disease states by mining genomic, proteomic, and transcriptomic data alongside clinical information and scientific literature [26]. For example, knowledge graphs can map intricate biological pathways, revealing previously unknown proteins that may play a causal role in a disease, thus presenting them as potential therapeutic targets [4]. This data-driven approach moves beyond traditional, hypothesis-limited methods to systematically identify targets with a higher probability of clinical relevance [28].

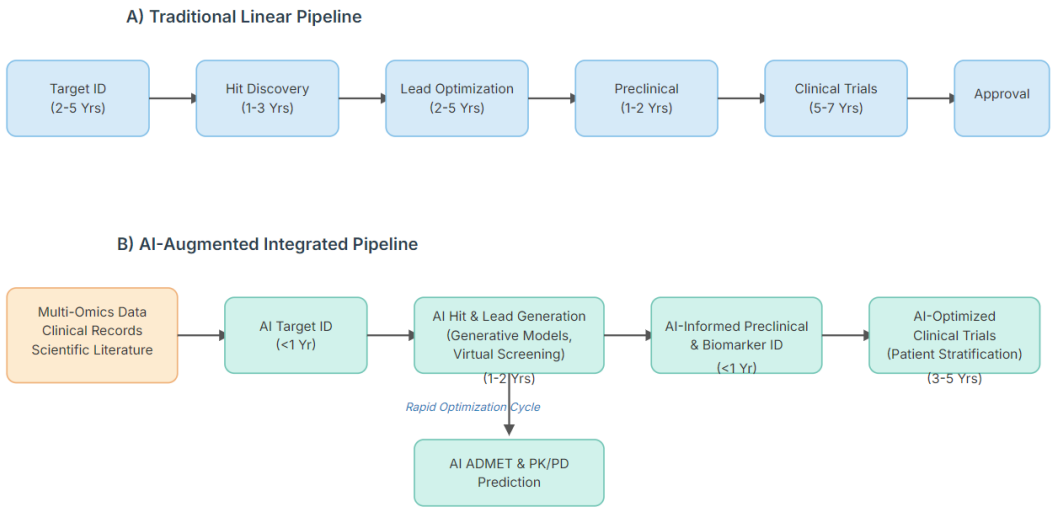

Figure 1. Comparison of Traditional and AI-Augmented Drug Discovery

5.2. Hit Discovery and Virtual Screening

Once a target is validated, the search for molecules that can interact with it begins. Traditional high-throughput screening (HTS) is a resource-intensive process limited by the physical size of compound libraries. AI-driven virtual screening offers a powerful alternative, enabling the rapid computational evaluation of vast chemical spaces containing billions or even trillions of potential molecules [22]. Deep learning models, particularly GNNs, are trained on large bioactivity datasets to predict the binding affinity of a compound to a specific target with remarkable accuracy [6]. This allows for the prioritization of a small, diverse set of high-potential compounds for subsequent experimental testing, dramatically increasing the hit rate and reducing the time and cost associated with this discovery phase [7].

Table 3. AI Applications Across the Drug Discovery Pipeline

Pipeline Stage	AI-Driven Task	Methodologies Used	Desired Outcome
Target Identification	Novel target discovery from multi-omics data & literature	Knowledge Graphs, NLP, GNNs	Identification and prioritization of therapeutically relevant biological targets
Hit Discovery	High-throughput virtual screening of large compound libraries	GNNs, Deep Neural Networks, SVMs	Rapid identification of initial compounds with desired bioactivity ("hits")
Lead Optimization	Multi-objective optimization of hit compounds	Reinforcement Learning, Generative Models	Refinement of hits into lead candidates with improved potency, selectivity, and ADMET
ADMET Prediction	Early-stage prediction of pharmacokinetics and toxicity	GNNs, Random Forests, QSAR models	Reduction of late-stage attrition by filtering out candidates with poor safety profiles
Drug Repurposing	Identifying new indications for existing approved drugs	Network Medicine, NLP on EHRs, GNNs	Faster and lower-cost development of new therapies for unmet needs

5.3. De Novo Molecular Design

Beyond screening existing libraries, generative AI models can design entirely novel molecules optimized for a specific biological target and a desired set of properties. Using architectures like VAEs, GANs, and, more recently, diffusion models, these systems can generate chemical structures that are both novel and tailored to a predefined therapeutic profile [9, 35]. When combined with reinforcement learning, the generation process can be guided to simultaneously optimize for multiple parameters, such as high potency, metabolic stability, low toxicity, and synthetic feasibility [17, 18]. This capability has fundamentally transformed medicinal chemistry, allowing for the exploration of new regions of chemical space and the creation of highly specialized lead compounds.

5.4. Prediction of ADMET and Toxicity

A primary cause of late-stage drug failure is an unacceptable pharmacokinetic or safety profile. The ability to predict Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties early in the discovery process is therefore of immense value. AI models trained on historical in vitro and in vivo experimental data, as well as adverse event reports, can provide early warnings for potential liabilities [23, 25]. These predictive tools allow for the early deselection of compounds with a high risk of failure, enabling research teams to focus resources on candidates with a greater chance of clinical success and significantly reducing downstream attrition [24].

5.5. Drug Repurposing and Clinical Trial Optimization

AI also offers powerful strategies for finding new therapeutic uses for existing, approved drugs—a process known as drug repurposing. By analyzing the relationships between drug structures, target profiles, and disease signatures from multi-omics data and electronic health records, AI can identify unexpected connections and generate hypotheses for repurposing [28, 29]. This approach offers a significantly faster and less expensive path to new treatments, as the safety profiles of these drugs are already well-established. Furthermore, in the clinical phase, AI can optimize trial design by improving patient stratification. Predictive models can identify patient subgroups most likely to respond to a treatment, leading to smaller, more efficient trials with a higher probability of success [30].

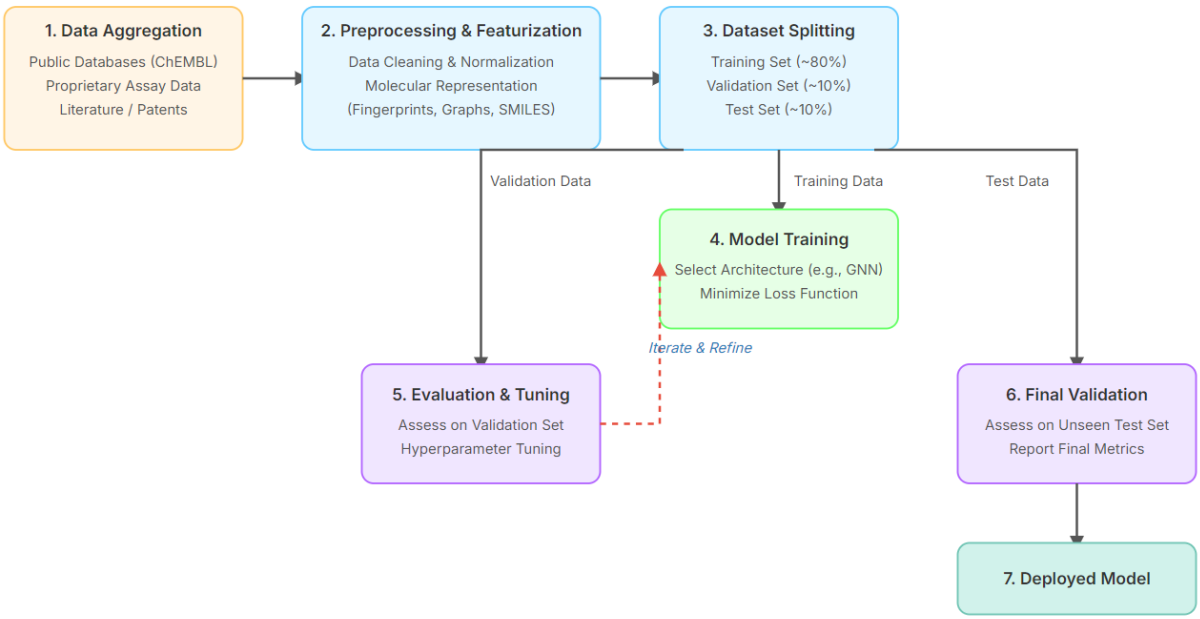

Figure 2. Workflow for Building Drug Discovery Models

6. Industrial Adoption

6.1. Case Studies

The pharmaceutical and biotechnology sectors have rapidly embraced AI, moving from exploratory research to active integration into discovery workflows. A growing number of biotechnology startups are founded on AI-native platforms, while major pharmaceutical companies are establishing internal AI teams and forming strategic partnerships [19, 34]. Several case studies have demonstrated the potential for significant timeline compression. For instance, AI-driven platforms have been reported to advance programs from target identification to a preclinical candidate in under two years—a process that traditionally takes four to five years [3, 4]. Notable successes include the rapid identification of novel kinase inhibitors and the discovery of a new class of antibiotics using deep learning [4, 27]. While a drug designed entirely by AI has yet to achieve regulatory approval, these early successes validate AI's role as a powerful accelerator and augmentation tool in drug discovery.

Table 4. Industry Case Studies in AI-Driven Drug Discovery

Company / Collaboration	Therapeutic Area	AI Approach	Reported Achievement
Insilico Medicine	Idiopathic Pulmonary Fibrosis (IPF)	Generative models for *de novo* design & target ID	Advanced an AI-designed drug candidate (INS018_055) to Phase II clinical trials.
Exscientia & Sumitomo Dainippon Pharma	Psychiatry / CNS	Generative models with active learning for lead optimization	Identified a novel serotonin 5-HT1A receptor agonist (DSP-1181) for OCD, advancing it to clinical trials in under 12 months.
Recursion Pharmaceuticals	Rare Diseases / Oncology	Image-based phenotypic screening using CNNs	Built a large-scale platform to model thousands of diseases and identify potential therapeutics; multiple candidates in clinical stages.
AbCellera & Eli Lilly	Infectious Disease (COVID-19)	AI-powered analysis of immune responses to find antibodies	Discovered bamlanivimab, an antibody therapeutic for COVID-19, which received Emergency Use Authorization.

6.2. Validation and Reproducibility

Model performance is typically assessed using standard metrics—for regression tasks like affinity prediction, Root Mean Square Error (RMSE) is common, while the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) is used for classification tasks [7, 8]. For generative models, evaluation is more complex, involving metrics that assess the novelty, diversity, and chemical validity of the generated molecules [7, 8].

Ultimately, computational predictions must be validated through experimental testing. However, to ensure computational rigor, the community has developed standardized benchmark suites. Platforms like MoleculeNet and Therapeutics Data Commons provide curated datasets and standardized evaluation protocols, allowing for the fair and direct comparison of different models [20, 21]. These initiatives are crucial for addressing issues of data leakage and inconsistent evaluation that have previously led to inflated performance claims, thereby promoting a more transparent and reproducible scientific culture [7].

7. Challenges and Enduring Limitations

Despite the rapid progress and demonstrated successes, the widespread implementation of AI in drug discovery faces several significant hurdles that must be addressed for the technology to realize its full potential.

7.1. Data Quality and Accessibility

The adage "garbage in, garbage out" is particularly resonant for AI in drug discovery. The performance of any model is inextricably linked to the quality of its training data. Publicly available datasets, while invaluable, can suffer from inconsistencies, errors, and experimental biases that can mislead model training and lead to poor generalization [10, 11]. Conversely, high-quality, curated datasets generated by pharmaceutical companies are often proprietary and inaccessible to the broader research community, which can stifle innovation and independent validation [14]. The scarcity of high-quality data for novel biological targets or rare diseases further compounds this challenge, limiting the applicability of AI in these areas.

7.2. Model Interpretability

Many of the most powerful deep learning models function as "black boxes," making it difficult to understand the reasoning behind their predictions [14]. This lack of interpretability is a major barrier to adoption, as medicinal chemists and biologists are often hesitant to trust predictions without a clear, mechanistically plausible rationale. It also poses a significant challenge for regulatory agencies, who require a clear understanding of a model's decision-making process to approve AI-generated candidates for clinical trials [15, 32]. While explainable AI (XAI) is an active area of research, developing methods that are both robust and intuitive remains a key challenge [14].

7.3. Synthetic Feasibility and Real-World Translation

Generative models can design molecules with excellent predicted properties but may neglect the practicalities of chemical synthesis [10]. A molecule that is promising in silico is of little value if it cannot be synthesized efficiently and scalably in a laboratory. Integrating synthetic accessibility scores and retrosynthesis prediction models directly into the generative loop is an ongoing effort to bridge this gap [10, 11]. Furthermore, a significant challenge is the "domain shift" between preclinical data and clinical outcomes. Models trained on in vitro assays or animal models may not accurately predict efficacy and safety in humans, a translational gap that AI has yet to fully overcome [33].

7.4. Ethical Guidelines

The combination of AI into a highly regulated field like medicine brings a host of non-technical considerations. Regulatory bodies such as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) are actively developing frameworks for the evaluation of AI-driven technologies [32, 33]. These frameworks emphasize the need for transparency in model development, rigorous validation, and clear documentation of data provenance to ensure the safety and efficacy of AI-influenced therapeutics [32].

From an ethical standpoint, biases embedded in training data could lead to the development of drugs that are less effective for underrepresented patient populations, thereby exacerbating health inequities [30]. The use of patient data from electronic health records and genomic databases also raises significant privacy concerns that must be managed through robust data governance and anonymization techniques [14, 15]. Societally, the automation of tasks traditionally performed by scientists may also lead to shifts in the pharmaceutical workforce, requiring new skill sets and roles that blend computational and domain expertise.

8. Current Trends

The field of AI in drug discovery is continually evolving, with several exciting frontiers poised to further enhance its impact.

8.1. Foundation Models and Multimodal Use

Large-scale models pre-trained on vast, diverse datasets spanning chemistry, biology, and clinical text are emerging as powerful platforms. These "foundation models" can be fine-tuned for a wide array of specific tasks, reducing the data requirements for niche applications and enabling a more holistic approach to drug design that integrates multiple data modalities simultaneously [35].

8.2. Closed-Loop Discovery and Laboratory Automation

The synergy between AI and robotic automation is enabling the creation of "self-driving" laboratories. In this paradigm, AI algorithms design novel molecules and experiments, which are then physically executed by automated synthesis and testing platforms. The experimental results are fed back to the AI in real-time, creating a closed design-make-test-learn loop that can operate with minimal human intervention, dramatically accelerating the pace of discovery [27, 31].

8.3. Precision Medicine and Biomarker Discovery

AI is set to play a pivotal role in the advancement of precision medicine by combining patient-level multi-omics data. Models can identify novel biomarkers to predict patient responses to treatment, enabling the design of therapies tailored to an individual's specific genetic and molecular profile [26, 31].

8.4. Quantum Computing and Molecular Simulation

While still in its nascent stages, the intersection of AI and quantum computing holds the promise of revolutionizing molecular simulation. Quantum computers could one day model molecular interactions with an accuracy that is intractable for classical computers, providing unprecedented insights into drug-target binding that could guide AI-driven design [34].

9. Conclusion

Artificial intelligence has changed from a novelty to an indispensable component of the modern drug discovery ecosystem. Methodological improvements, particularly in generative modeling and graph neural networks, have equipped researchers with powerful tools to enhance and accelerate nearly every stage of the R&D pipeline, from initial target identification to the optimization of clinical trials. The impact is already tangible, with AI-driven approaches demonstrably shortening discovery timelines and uncovering novel therapeutic candidates. However, the path to fully realizing AI's transformative potential is not without significant obstacles. Issues of data quality and accessibility, the inherent "black box" nature of many advanced models, and the challenge of translating in silico predictions into real-world clinical success remain paramount. The future of the field will be defined not only by continued algorithmic innovation but also by the successful integration of AI with laboratory automation, the establishment of clear regulatory pathways, and an unwavering commitment to ethical principles. AI is poised to usher in a new era of pharmaceutical research, characterized by greater efficiency, reduced costs, and the development of highly personalized medicines that were previously beyond our reach. The coming decade will be critical in determining the extent to which this promise is translated into clinical and commercial reality.

References

[1] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Nat Rev Drug Discov. 2018;17(10):683-685.

[2] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477.

[3] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038-1040.

[4] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688-702.e13.

[5] Walters WP, Stahl MT, Murcko MA. Virtual screening – an overview. Drug Discov Today. 1998;3(4):160-178.

[6] Gaudelet T, Day B, Jamasb AR, Soman J, Regep C, Liu G, et al. Utilising graph machine learning within drug discovery and development. Drug Discov Today. 2021;26(11):2653-2666.

[7] Brown N, Fiscato M, Segler MHS, Vaucher AC. GuacaMol: Benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59(3):1096-1108.

[8] Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, Golovanov S, Tatanov O, Belyaev S, et al. Molecular Sets (MOSES): A benchmarking platform for molecular generation models. Front Pharmacol. 2020;11:565644.

[9] Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science. 2018;361(6400):360-365.

[10] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555(7698):604-610.

[11] Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci. 2019;5(9):1572-1583.

[12] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

[13] Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 2021;373(6557):871-876.

[14] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to apply AI and network analysis. Drug Discov Today. 2021;26(4):1001-1012.

[15] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: A discussion of case histories and lessons learned. Drug Discov Today. 2021;26(5):1045-1055.

[16] Yang K, Swanson K, Jin W, Coley C, Eickhoff C, Jensen KF, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370-3388.

[17] Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. J Cheminform. 2017;9(1):48.

[18] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):eaap7885.

[19] Walters WP, Barzilay R. Applications of deep learning in molecule generation and molecular property prediction. J Med Chem. 2021;64(18):13245-13255.

[20] Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-530.

[21] Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, et al. Therapeutics Data Commons: Machine learning datasets and tasks for drug discovery and development. Proc Neural Inf Process Syst. 2021;34:22378-22395.

[22] Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem. 2020;63(16):8749-8760.

[23] Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: Toxicity prediction using deep learning. Front Environ Sci. 2016;3:80.

[24] Xu Y, Pei J, Lai L. Deep learning in drug discovery. Expert Opin Drug Discov. 2017;12(12):1209-1220.

[25] Pires DEV, Blundell TL, Ascher DB. pkCSM: Predicting small-molecule pharmacokinetic properties using graph-based signatures. J Med Chem. 2015;58(9):4066-4072.

[26] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev. 2019;119(18):10520-10594.

[27] Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18(5):435-441.

[28] Zhou Y, Wang F, Tang J, Nussinov R, Cheng F. Artificial intelligence in COVID-19 drug repurposing. Lancet Digit Health. 2020;2(12):e667-e676.

[29] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13):i457-i466.

[30] Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-1318.

[31] Shankar V, Huang K, Subramanian S, Cofer EM, Leskovec J, Plevritis SK, et al. Evaluating the clinical utility of machine learning for treatment effect estimation. Nat Med. 2022;28(1):122-131.

[32] U.S. Food and Drug Administration, Health Canada, Medicines and Healthcare products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles. Published online October 27, 2021.

[33] European Medicines Agency. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle to marketing authorisation. Published online July 31, 2023. Report No.: EMA/CHMP/CVMP/812235/2023.

[34] Walters WP, Murcko M. Assessing the impact of generative AI on medicinal chemistry. Nat Biotechnol. 2023;41(3):305-307