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

REVIEW ARTICLE

The Evolving Role of Artificial Intelligence and Machine Learning in Drug Discovery and Development

Harika G[1], Govinda Rao Kamala*[2], A V D S Prajna M[1], Syamala B[1], Sindhu Sri Durga N[1]

[1] *UG Scholar, Department of Pharmaceutical Chemistry, Koringa College of Pharmacy, Korangi, Tallarevu, Kakinada, Andhra Pradesh, India*
[2] *Vice-Principal and Professor, Department of Pharmaceutical Chemistry, Koringa College of Pharmacy, Korangi, Tallarevu, Kakinada, Andhra Pradesh, India*

Publication history: Received on 21st July 2025; Revised on 31st Aug 2025; Accepted on 6th September 2025

Article DOI: 10.69613/s1x6av74

Abstract: The traditional pharmaceutical research and development timeline is seriously protracted, costly, and marked by high attrition rates. Artificial intelligence and machine learning (AI/ML) are catalyzing a paradigm shift across this entire pipeline. These computational methods process vast, high-dimensional datasets to uncover novel biological insights and expedite candidate selection. In early-stage discovery, AI models analyze 'omics' data and biological networks to identify and validate novel therapeutic targets. For lead discovery, ML-powered virtual screening and de novo design, utilizing generative models, are creating potent and selective molecules with optimized pharmacokinetic profiles. Predictive algorithms are substantially refining ADMET (absorption, distribution, metabolism, excretion, and toxicity) modeling, reducing late-stage attrition. This transformation extends into clinical development, where AI assists in optimizing trial design, stratifying patient cohorts, and analyzing real-world evidence for post-market surveillance. While significant challenges related to data quality, model interpretability, and regulatory guidelines persist, the integration of AI/ML is remarkably streamlining processes, from initial hypothesis to clinical application. This computational revolution promises to lower development costs and accelerate the delivery of novel, personalized therapies to patients.

Keywords: Artificial Intelligence; Machine Learning; Drug Discovery; Computational Drug Design; Target Identification

1. Introduction

The process of bringing a novel therapeutic agent from initial concept to market is one of the most complex, costly, and time-intensive endeavors in modern science [1]. Spanning over a decade and often costing billions of dollars, the traditional pharmaceutical pipeline is characterized by high attrition rates, with many promising candidates failing in late-stage preclinical and clinical testing [2]. This inefficiency is famously captured by "Eroom's Law" (Moore's Law spelled backward), which observes that the cost of developing a new drug has roughly doubled every nine years since 1950, despite massive technological advances [3]. The industry has been facing a productivity crisis, pressured by patent cliffs, rising R&D costs, and an increasingly stringent regulatory era.

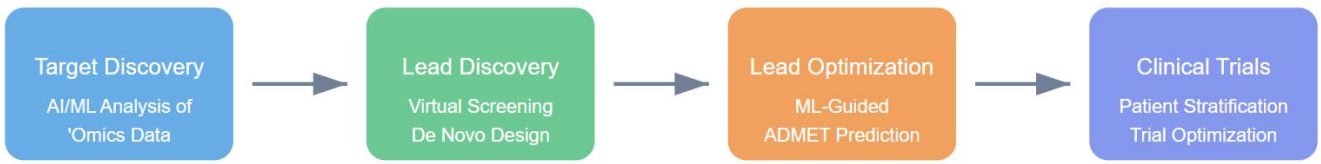

Figure 1. AI-Driven Drug Discovery Pipeline showing the integration of artificial intelligence and machine learning methods

In response to these persistent challenges, the integration of artificial intelligence (AI) and machine learning (ML) has emerged as a transformative force. These computational methods offer powerful new tools to augment and accelerate nearly every stage of drug discovery and development [4]. AI/ML's core strength lies in its ability to identify complex, non-linear patterns within vast, high-dimensional, and heterogeneous datasets—a task that is intractable for human researchers. These datasets include genomic, proteomic, and transcriptomic 'omics' data; large-scale chemical libraries and their bioactivity data; 3D protein structures; digital pathology images; clinical trial results; and real-world evidence from electronic health records (EHRs) [5]. AI/ML models are shifting

* Corresponding author: Govinda Rao Kamala

the R&D paradigm from a sequential, often serendipitous process to a more predictive, efficient, and data-driven discipline by utilizing this data. This review discusses the application of these computational methods, tracking their impact logically from initial target discovery through lead optimization, preclinical analysis, and clinical trial management.

2. Target Identification and Validation

The foundational step in any drug discovery program is the identification and validation of a biological target, typically a protein or nucleic acid, that is critically involved in a disease pathway [6].

2.1. AI in Target Identification

AI/ML excels at navigating the enormous datasets required for this task. Deep learning models can sift through genomic, proteomic, and transcriptomic ('omics') data from patient samples to identify novel gene signatures or protein expression patterns associated with a specific disease state [7]. For instance, Graph Neural Networks (GNNs) can model complex protein-protein interaction (PPI) networks, identifying "hub" or "bottleneck" proteins whose modulation would have a cascading effect on a disease pathway [8]. Concurrently, Natural Language Processing (NLP) models contribute by scanning millions of scientific publications, patent databases, and clinical trial registries. These models go beyond simple keyword matching to extract latent relationships between genes, proteins, and diseases, distinguishing between mere co-occurrence and potential causal links, thereby highlighting promising new avenues for intervention [9].

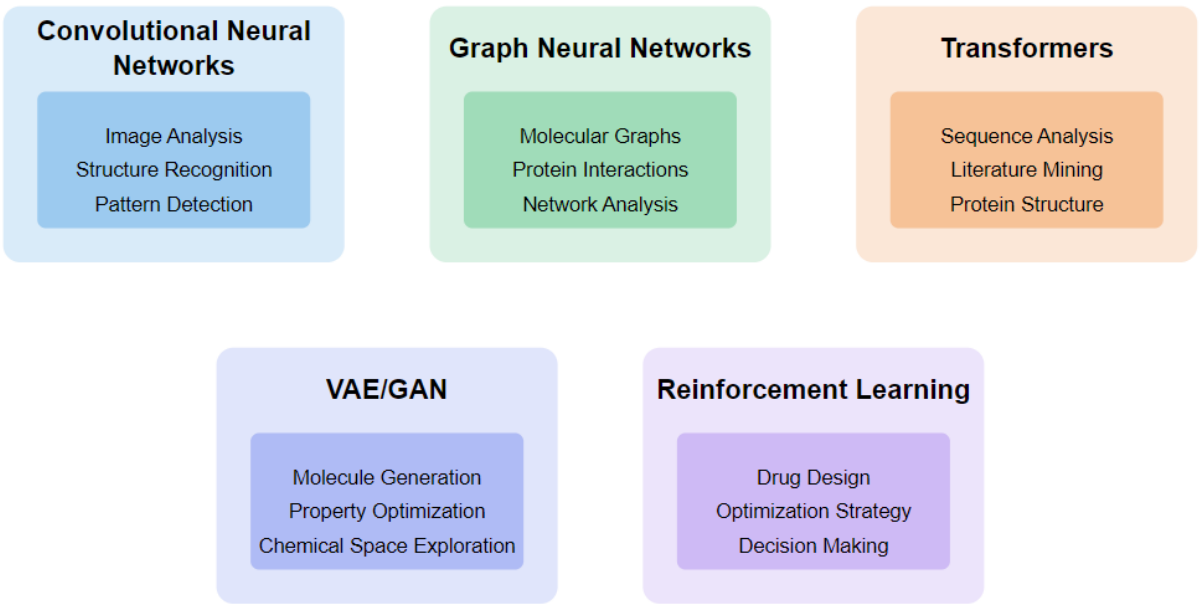

Figure 2. Major AI Model Architectures used in pharmaceutical research and development

2.2. Computational Target Validation

Identifying a target is insufficient; it must also be validated. Validation involves confirming the target's role in the disease and, crucially, assessing its 'druggability'—the likelihood that it can be modulated by a small-molecule drug or biologic [10]. ML models, trained on known protein structures and ligand-binding data, can predict the presence and characteristics of both orthosteric (active) and allosteric (regulatory) binding pockets on a protein's surface. The latter is particularly valuable for traditionally "undruggable" targets. This *in silico* validation is powerfully enabled by structural prediction tools like AlphaFold and RoseTTAFold, which provide high-accuracy 3D models for proteins without experimental structures [11]. These predicted structures serve as the direct input for molecular dynamics simulations and docking studies to probe druggability, helping prioritize targets that are not only biologically relevant but also chemically tractable.

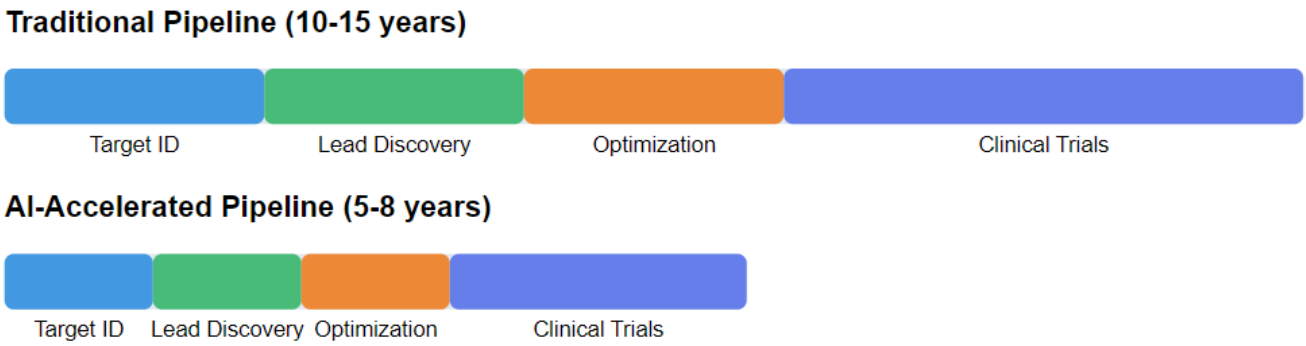

Figure 3. Traditional vs AI-Accelerated Drug Development Timelines

Table 1. AI/ML Applications in the Pharmaceutical R&D Pipeline

R&D Stage	Objective	Examples of AI/ML Applications	Model Types
Target Discovery	Identify novel, "druggable" biological targets.	Analyzing 'omics' data (genomics, proteomics) to find disease-specific biomarkers. Mining scientific literature (NLP) for gene-disease associations. Modeling protein-protein interaction (PPI) networks.	Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Natural Language Processing (NLP)
Lead Discovery	Find and design molecules ('hits' or 'leads') that modulate the target.	High-throughput virtual screening (VS) of vast chemical libraries. *De novo* design of novel molecules with desired properties. Pharmacophore modeling and QSAR.	Support Vector Machines (SVMs), Random Forests, CNNs, VAEs, GANs, Transformers
Preclinical	Assess safety and efficacy before human testing.	*In silico* ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) prediction. Digital pathology (CNNs) for analyzing tissue slides from *in vivo* studies. QST modeling using organ-on-a-chip data.	QSAR Models, Deep Learning (CNNs), Systems Biology Models
Clinical Trials	Evaluate safety and efficacy in humans.	Patient stratification and cohort selection using EHR and genomic data. Optimizing trial design (e.g., adaptive trials). Analyzing real-world data (RWD) from wearables.	Clustering Algorithms (e.g., K-Means), Bayesian Models, NLP
Post-Market	Monitor long-term safety and real-world effectiveness.	Pharmacovigilance: Detecting adverse drug reactions (ADRs) from EHRs and social media. Analyzing real-world evidence (RWE) to confirm efficacy.	Natural Language Processing (NLP)

3. Lead Discovery and Optimization

Once a target is validated, the search begins for 'hits'—small molecules that interact with the target—which are then optimized into 'leads' with drug-like properties.

3.1. Accelerating Hit Discovery

While high-throughput screening (HTS) remains a workhorse, AI is refining this process. ML algorithms facilitate "active learning," where the model iteratively guides the screening campaign. A small, diverse subset of the library is screened, the data is used to train a model, and the model then predicts the most promising compounds to screen in the next round [12]. This "smart screening" can lead to high-quality hits while testing a fraction of the full library, optimizing time and reagents. Moreover, AI-based image analysis, particularly using convolutional neural networks (CNNs), automates high-content screening (HCS). CNNs can extract subtle, multiparametric morphological features (phenotypes) from cell-based assays—such as changes in organelle shape, protein

translocation, or cell-cell interactions—that correlate with compound activity, providing far richer data than simple live/dead readouts [13].

Table 2. Comparison of Machine Learning Model Architectures in Drug Design

Model Type	Primary Application in Drug Design	Strengths	Limitations
Random Forest (RF) / Support Vector Machine (SVM)	QSAR models. ADMET prediction. Virtual screening scoring.	Work well with smaller, tabular datasets. Highly interpretable (especially RF). Robust and well-understood.	Less effective on high-dimensional data (e.g., images, sequences). Cannot generate novel data (non-generative).
Convolutional Neural Network (CNN)	Analyzing medical/pathology images. High-content screening analysis. Predicting bioactivity from 2D molecular structures.	State-of-the-art for image-based tasks. Can learn spatial hierarchies of features.	Requires large labeled datasets. Less intuitive for non-image data (like sequences or graphs).
Graph Neural Network (GNN)	Predicting molecular properties (molecules as graphs). Modeling protein-protein interaction networks. *De novo* graph-based generation.	Natively handles 3D molecular structures and relationships. Captures topological and relational information.	Computationally intensive. Field is newer; best practices are still evolving.
Variational Autoencoder (VAE)	*De novo* molecule generation. "Chemical space" dimensionality reduction. Generating molecules with optimized properties.	Learns a smooth and continuous latent space. Good for optimization and property-guided generation.	Can be difficult to train. May generate less valid or "drug-like" structures than other models.
Generative Adversarial Network (GAN)	*De novo* molecule generation. Generating realistic medical images (e.g., for data augmentation).	Can produce highly novel and realistic-looking data (molecules).	Notoriously unstable and difficult to train. Prone to "mode collapse" (low diversity of outputs).
Transformer	NLP for literature mining. Processing "SMILES" strings for *de novo* design. Protein sequence analysis (e.g., AlphaFold).	State-of-the-art in sequence-based tasks (text, genes, proteins). Captures long-range dependencies via attention mechanisms.	Requires massive datasets and significant computational power.

3.2. Virtual Screening

Virtual screening (VS) represents a primary application of computational power to reduce the search space from billions of potential compounds to a manageable number for *in vitro* testing.

3.2.1. Structure-Based Virtual Screening (SBVS)

When the 3D structure of the target is known (either experimentally or via prediction), SBVS methods like molecular docking are used to computationally "fit" molecules into the target's binding site [14]. AI is enhancing this process significantly. While traditional docking relies on physics-based scoring functions, ML models (e.g., RF-Score, NNScore) trained on experimental binding affinity data can develop more accurate, data-driven scoring functions. These functions learn to recognize complex, non-linear patterns related to solvation, entropy, and specific atomic interactions, allowing them to better rank and identify true binders from decoys [15].

3.2.2. Ligand-Based Virtual Screening (LBVS)

In the absence of a reliable target structure, LBVS methods are employed. These models rely on the principle that structurally similar molecules often have similar biological activities. ML techniques, such as support vector machines (SVMs) and deep neural networks, can build robust pharmacophore models or quantitative structure-activity relationship (QSAR) models from a small set of known

active ligands [16]. These models, which capture 3D electronic and steric features beyond simple 2D structural similarity, then screen large databases to find novel chemotypes that match the key chemical features required for binding.

3.3. De Novo Drug Design

Perhaps the most disruptive application of AI in drug design is *de novo* design, where algorithms generate novel molecular structures from scratch rather than simply screening existing ones.

3.3.1. Generative Models

Generative models, such as Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Transformers, learn the underlying "rules" and patterns of chemical space (e.g., valency, aromaticity, 3D conformation) from large chemical databases [17]. These models can operate in different ways: some generate 1D text-based SMILES strings, which are then converted to 2D or 3D structures, while more recent graph-based models construct the 3D molecular graph directly. They can be conditioned to generate entirely new molecules optimized for a suite of specific properties, such as high predicted affinity, synthetic accessibility, and favorable drug-like characteristics [18].

3.3.2. Reinforcement Learning for Molecule Optimization

This generative process is often coupled with reinforcement learning (RL). An RL agent can be tasked with "designing" a molecule atom by atom or fragment by fragment, receiving "rewards" for improving desired parameters [19]. The reward function is key, as it is typically a multi-objective optimization problem: the agent is rewarded for maximizing target affinity while *simultaneously* minimizing predicted toxicity, minimizing synthetic complexity, and maximizing novelty. This iterative, goal-directed optimization loop allows the algorithm to navigate the vast chemical space and discover novel, high-quality leads that a human chemist might never conceive of.

3.4. AI-Driven ADMET Prediction

A primary cause of late-stage drug failure is poor pharmacokinetics or unforeseen toxicity [20]. ML models are now integral to predicting Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties *in silico*. These models can flag problematic compounds early in the design phase by training on large datasets of experimental ADMET data. Specific models predict key failure points like hERG channel blockage (cardiotoxicity), drug-induced liver injury (DILI, hepatotoxicity), or mutagenicity (Ames test) [21]. This 'fail-fast' approach allows medicinal chemists to prioritize compounds that are not only potent but also have a high probability of being safe and bioavailable in humans.

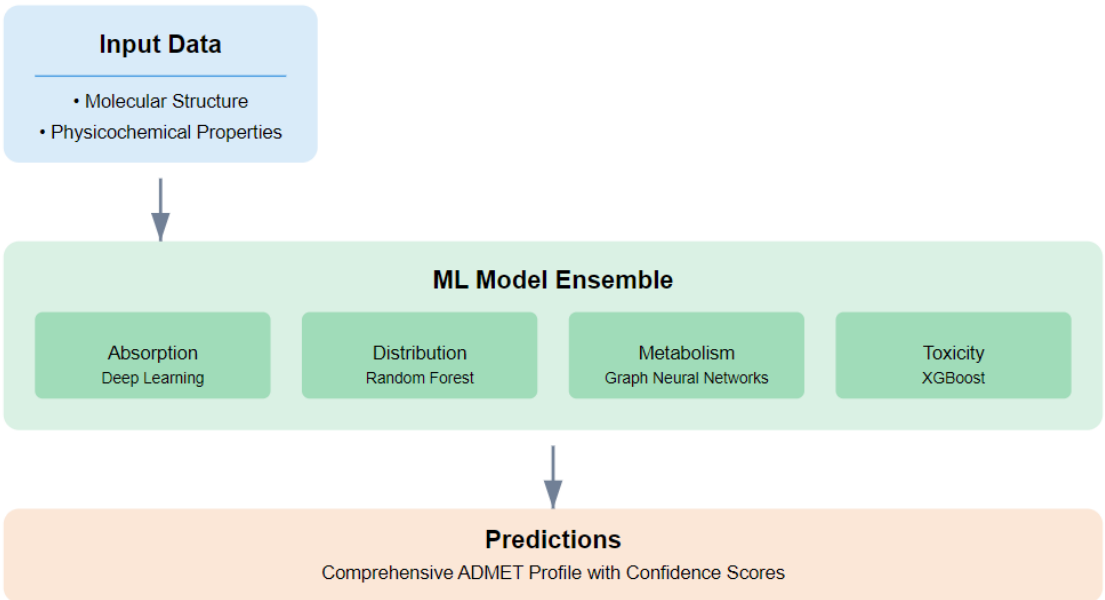

Figure 4. AI-powered ADMET prediction

4. Preclinical and Clinical Development

The predictive power of AI is not limited to early discovery; it is also streamlining the costly and complex development phases.

4.1. Enhancing Preclinical Studies

In the preclinical stage, AI is helping to refine and, in some cases, reduce reliance on traditional animal models. ML models, particularly in the field of "quantitative systems toxicology" (QST), can integrate *in vitro* assay data from microphysiological systems (e.g., organ-on-a-chip) with systems biology pathway modeling [22]. This approach can create "virtual organs" to better predict organ-level toxicity in humans, offering more relevant data than animal models. In parallel, AI-driven analysis of digital pathology images from *in vivo* efficacy studies provides more objective, quantitative, and reproducible endpoints. CNNs can segment tumor regions, count mitotic figures, or quantify tissue-specific biomarkers, replacing or augmenting subjective, manual inspection by a pathologist [23].

Table 3. AI-Driven Drug Discovery Platforms and Success Stories

Company / Platform	Core AI/ML Technology	Application	Impact
DeepMind (AlphaFold)	Deep Learning (Transformer-based)	Solved the protein-folding problem, providing high-accuracy 3D structures for millions of proteins.	Target Identification & Validation
BenevolentAI	Knowledge Graphs, NLP, Deep Learning	Identified Baricitinib (an existing arthritis drug) as a potential treatment for COVID-19.	Drug Repurposing
Insilico Medicine	Generative AI (GANs, Transformers), Reinforcement Learning	Designed a novel drug candidate (for Idiopathic Pulmonary Fibrosis) from target discovery to preclinical candidate in 18 months.	Target ID, *De Novo* Design
Exscientia	Generative AI, Active Learning	Developed the first AI-designed molecule (for OCD, with Sumitomo Dainippon Pharma) to enter Phase 1 clinical trials.	*De Novo* Design, Lead Optimization
Atomwise (AtomNet)	Deep Learning (CNNs) for SBVS	Used its platform to screen billions of compounds for an Ebola virus target, identifying promising candidates.	Virtual Screening
PathAI	Deep Learning (CNNs) for Digital Pathology	Develops AI-powered pathology models to improve accuracy and efficiency in clinical trials and diagnostics.	Preclinical & Clinical Trials

4.2. Optimizing Clinical Trials

The clinical trial phase is the most expensive and high-risk component of drug development. AI is being applied to optimize trial design and execution. ML algorithms can analyze electronic health records (EHRs), genomic data, and medical images from millions of patients to identify and recruit the most suitable patient cohorts for a given trial [24]. This precision patient stratification improves the chances of detecting a therapeutic signal, especially for targeted therapies. AI platforms can also help in designing adaptive trials, where Bayesian statistics and ML models re-allocate patient arms to more promising treatments mid-trial based on incoming data. During the trial, AI-powered digital health tools and wearable sensors can monitor patient adherence and collect real-world data (RWD), providing more continuous and objective endpoints [25].

4.3. Drug Repurposing

AI has also proven exceptionally powerful in drug repurposing (or repositioning). ML models can identify existing, approved drugs that may be effective against new indications by analyzing vast knowledge graphs of drug-target-disease interactions from public databases and literature [26]. This strategy was famously highlighted during the COVID-19 pandemic, where AI models rapidly identified baricitinib, an existing rheumatoid arthritis drug, as a potential treatment for severe COVID-19, a finding later confirmed in clinical trials [27]. Because the safety profiles of these drugs are already known, the development timeline can be drastically shortened.

5. Manufacturing and Post-Market Surveillance

The influence of AI extends beyond approval and into the manufacturing and long-term monitoring of therapeutics.

5.1. Smart Manufacturing and Quality Control

In pharmaceutical manufacturing, AI systems are used to monitor complex biologic production processes in real-time, predicting equipment failures and optimizing yields. This is a core component of "Process Analytical Technology" (PAT) [28]. For instance, ML "soft sensors" can analyze real-time data from temperature, pH, and dissolved oxygen sensors in a bioreactor to predict final product quality (e.g., protein titer or glycosylation patterns) hours or days in advance. AI-powered computer vision can also automate quality control, inspecting vials for particulate matter, ensuring correct labeling, and verifying the integrity of lyophilized (freeze-dried) cakes with greater speed and accuracy than human inspectors [29].

5.2. Pharmacovigilance and Real-World Evidence (RWE)

After a drug is approved, AI continues to play a critical role in pharmacovigilance (drug safety monitoring). Traditional methods rely on "spontaneous reporting systems" (like the FDA's FAERS), which suffer from significant under-reporting. NLP tools can scan millions of unstructured EHR notes, insurance claims databases, and even social media platforms to detect adverse drug reaction (ADR) signals much earlier [30]. These tools use named entity recognition (NER) to identify drugs and symptoms, and relation extraction to determine if they are linked. This ongoing analysis of RWE is critical for monitoring long-term safety and confirming a drug's effectiveness in a broad, real-world population outside the controlled setting of a clinical trial.

6. Challenges and Ethical Considerations

Despite the immense potential, significant hurdles remain for the widespread adoption of AI in pharmaceutical R&D.

6.1. Data Quality and Accessibility

The adage "garbage in, garbage out" is paramount in AI; models are only as good as the data they are trained on. High-quality, large-scale, and well-annotated biological and chemical datasets are difficult and expensive to produce. Moreover, much of this valuable data remains siloed within competing pharmaceutical companies or academic institutions, hindering the development of robust models that can generalize well [31]. As a potential solution, "federated learning" is emerging. This approach allows an ML model to be trained on decentralized data sources (e.g., at different hospitals) without the sensitive private data ever leaving its source, thus preserving privacy while enabling collaborative model building [32].

Table 4. Public Datasets Fueling AI in Pharmaceutical Research

Database Name	Data Type	Primary Use in AI/ML
PubChem	Chemical structures, bioactivity data.	Training QSAR models. Virtual screening library source.
ChEMBL	Curated bioactivity data from literature.	Training target-specific predictive models. ADMET model training.
Protein Data Bank (PDB)	3D experimental structures of proteins and macromolecules.	Training structure-based models (e.g., docking scoring). Validating predicted structures (e.g., AlphaFold).
UK Biobank	Deep phenotypic and genomic data from 500,000 participants.	Identifying novel gene-disease associations. Building patient stratification models.
The Cancer Genome Atlas (TCGA)	Comprehensive 'omics' and clinical data for various cancers.	Identifying new cancer targets. Subtyping tumors and predicting drug response.
ClinicalTrials.gov	Registry of clinical trials.	NLP analysis of trial protocols, endpoints, and results. Optimizing trial design.
FAERS / VAERS	Spontaneous adverse event reporting systems.	Training NLP models for pharmacovigilance and ADR detection.

6.2. Model Interpretability and the "Black Box" Problem

Many of the most powerful AI models, especially in deep learning, are "black boxes," meaning their internal decision-making processes are not easily interpretable by humans. This is a significant problem in a highly regulated field like medicine, where researchers and regulators must be able to justify *why* a model prioritized a specific compound or patient-stratification strategy [33]. A lack of interpretability erodes trust and creates barriers to regulatory acceptance. In response, the field of "explainable AI" (XAI) is developing techniques like SHAP (SHapley Additive exPlanations), which can highlight the specific molecular fragments or patient features that a model weighed most heavily in its prediction.

6.3. Regulatory and Intellectual Property Challenges

Regulatory agencies like the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) are still developing clear frameworks for validating and approving drugs discovered using novel AI-driven methods [34]. Questions abound regarding the level of *in silico* validation required before progressing to *in vitro* or *in vivo* testing. The FDA's "AI/ML-based Software as a Medical Device (SaMD) Action Plan" is one example of agencies adapting to this new technology. Moreover, complex intellectual property questions arise. In recent legal test cases (e.g., the DABUS system), AI algorithms have been listed as inventors on patent applications, only to be rejected by major patent offices. This highlights the profound legal uncertainty over who owns an AI-generated discovery [35].

Table 5. Challenges and Mitigation Techniques

Challenge	Description	Potential Mitigation Techniques
Data Quality & Accessibility	AI models require vast, high-quality, labeled data, which is often siloed, sparse, or noisy.	Federated Learning: Training models on decentralized data without data sharing. Data Augmentation: Using generative models to create synthetic, realistic data.
Model Interpretability ("Black Box")	Inability to understand *how* a complex model (e.g., deep learning) arrived at a prediction, hindering trust and regulatory approval.	Explainable AI (XAI): Using techniques like SHAP or LIME to identify key features driving a prediction. Simpler Models: Using inherently interpretable models (e.g., logistic regression, decision trees) where feasible.
Generalizability & Domain Shift	A model trained on one data distribution (e.g., *in vitro* assays) fails when applied to a new one (e.g., *in vivo* or human data).	Transfer Learning: Fine-tuning a pre-trained model on a smaller, domain-specific dataset. Prospective Validation: Rigorously testing models on new, unseen data, not just historical data.
Regulatory & Validation Frameworks	Regulatory agencies are still developing clear guidelines for validating and approving AI-generated *in silico* data.	Collaboration between AI developers and regulatory bodies (e.g., FDA SaMD action plan). Establishing clear "good machine learning practice" (GMLP) guidelines.
Intellectual Property (IP)	Ambiguity over inventorship and patent rights for a molecule designed by a generative AI.	Legal test cases (e.g., DABUS). Developing new legal frameworks that define "inventorship" in the context of AI-assisted discovery.

7. Conclusion

The incorporation of artificial intelligence and machine learning is fundamentally reshaping the landscape of drug discovery and development. These computational tools are moving the pharmaceutical industry from a reliance on empirical screening and incremental advances toward a more predictive, efficient, and rational design paradigm. AI/ML is directly addressing the core challenges of cost and time that have long plagued the field by identifying novel targets, designing superior molecules, and streamlining clinical trials. While substantial technical, ethical, and regulatory challenges must still be overcome, the momentum is undeniable. The continued refinement and adoption of these technologies hold the promise of accelerating the delivery of novel therapeutics and enabling a new era of personalized medicine.

References

[1] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.

[2] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9(3):203-14.

[3] Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11(3):191-200.

[4] 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):463-77.

[5] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170387.

[6] Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. Nat Rev Drug Discov. 2006;5(10):821-34.

[7] Gligorijević V, Gligorijević D, Pavlović M, Obradović Z. Deep learning for protein function prediction from omics data. Brief Bioinform. 2021;22(1):207-25.

[8] Yue L, Wang H, Smedley D, Gao Z, Li Z, Liu H, et al. Graph neural network for the prediction of protein interaction sites. Brief Bioinform. 2021;22(5):bbab029.

[9] Crichton G, B K, G S. NLP in drug discovery: A comprehensive review of the state-of-the-art. Brief Bioinform. 2021;22(3):bbab005.

[10] Hann MM, Keserü GM. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. Nat Rev Drug Discov. 2012;11(5):355-65.

[11] 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):583-9.

[12] Kalliokoski T, Kramer C, Vulpetti A, Gedeck P. Iterative screening with active learning. J Chem Inf Model. 2013;53(1):43-52.

[13] Caicedo JC, Cooper S, Heigwer F, Warchal S, Warchal L, Hupfeld K, et al. Data-analysis strategies for image-based cell profiling. Nat Methods. 2017;14(7):679-88.

[14] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3(11):935-49.

[15] Ballester PJ, Mitchell JB. A machine learning approach to predicting protein-ligand binding affinity with applications to virtual screening. Bioinformatics. 2010;26(9):1169-75.

[16] Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure-activity relationships. J Chem Inf Model. 2015;55(2):263-74.

[17] Segler MH, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120-31.

[18] Merk D, Grisoni F, Friedrich L, Schneider G. De novo design of bioactive small molecules with recurrent neural networks and reinforcement learning. J Chem Inf Model. 2018;58(9):1863-71.

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

[20] Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov. 2015;14(7):475-86.

[21] Yang H, Sun L, Li W, Liu G, Tang Y. In silico prediction of ADMET properties for drug discovery. Expert Opin Drug Discov. 2018;13(8):717-36.

[22] Edington CD, Chen WL, Geishecker E, Kassis T, Soenksen LR, Wijnand BM, et al. Interconnected microphysiological systems for quantitative biology and pharmacology studies. Sci Rep. 2018;8(1):4530.

[23] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J Pathol Inform. 2016;7:29.

[24] Mamitsuka H. Artificial intelligence in clinical trial design. Nat Rev Drug Discov. 2018;17(10):683-85.

[25] Izmailova ES, Wagner JA, TMF. A review of the applications of wearable technology in clinical trials. Clin Pharmacol Ther. 2018;104(3):414-24.

[26] Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58.

[27] Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet. 2020;395(10223):e30-e31.

[28] Rathore AS, B S, G A. Process analytical technology (PAT) for biopharmaceutical products. Anal Bioanal Chem. 2018;410(27):6885-900.

[29] Tziotzios G, B K, G S. A review of machine vision in pharmaceutical manufacturing. J Pharm Innov. 2019;14(2):97-111.

[30] Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform. 2015;54:202-12.

[31] Vayena E, D K, W L. Data-driven health research: The challenge of data silos. Sci Transl Med. 2018;10(451):eaau2468.

[32] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3:119.

[33] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-e750.

[34] US Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML) in Software as a Medical Device (SaMD) Action Plan. Silver Spring, MD: FDA; 2021.

[35] Hvizdak V, K C. AI as an inventor: The DABUS case and its implications for patent law. AI and Ethics. 2022;2:591–599.