Cole-Puli International Conference Publishing Group

Science, Engineering and Technology Proceedings

Review

The Role of AI and Machine Learning in Accelerating Drug Discovery and Personalized Medicine

Liang Xu [1] and Xiaoyu Zhang [2,*]

[1] School of Food and Biological Engineering, Zhengzhou University of Light Industry, Zhengzhou, China
[2] College of Biological & Environmental Sciences, Zhejiang Wanli University, Ningbo, China
* Correspondence: Xiaoyu Zhang, College of Biological & Environmental Sciences, Zhejiang Wanli University, Ningbo, China

Abstract: Artificial intelligence (AI) and machine learning (ML) are revolutionizing drug discovery and personalized medicine by offering powerful tools for analyzing complex biological data, predicting drug efficacy, and tailoring treatments to individual patients. This review paper explores the multifaceted roles of AI and ML in these domains, encompassing target identification, drug design, clinical trial optimization, and patient stratification. We begin with a historical overview of AI applications in healthcare, followed by a detailed examination of core themes such as AI-driven drug repurposing and the use of ML in predictive diagnostics. The paper further delves into the challenges and limitations of implementing AI/ML technologies, including data bias, interpretability issues, and regulatory hurdles. Finally, we discuss future perspectives, emphasizing the potential of AI to transform healthcare through enhanced precision, accelerated drug development timelines, and improved patient outcomes. This review synthesizes current research, highlights key advancements, and identifies critical areas for future investigation, aiming to provide a comprehensive understanding of the transformative impact of AI and ML on drug discovery and personalized medicine. The successful integration of these technologies promises to unlock new possibilities for preventing and treating diseases, paving the way for a more proactive and patient-centric healthcare system.

Keywords: artificial intelligence; machine learning; drug discovery; personalized medicine; healthcare; drug repurposing; predictive diagnostics

Received: 02 December 2025
Revised: 11 January 2026
Accepted: 26 January 2026
Published: 30 January 2026

1. Introduction

1.1. The Promise of AI in Pharma

Artificial intelligence (AI) and machine learning (ML) offer transformative potential for the pharmaceutical industry. Traditional drug discovery is a lengthy and expensive process, often taking over a decade and costing billions of dollars per drug. AI/ML algorithms can accelerate target identification, drug design, and clinical trial optimization. Furthermore, these technologies enable personalized medicine approaches by analyzing patient-specific data (x_i) to predict treatment response and tailor therapies, ultimately improving efficacy and reducing adverse effects [1].

1.2. Scope of the Review

This review comprehensively examines the applications of artificial intelligence (AI) and machine learning (ML) in revolutionizing drug discovery and personalized medicine. Our scope encompasses target identification, drug design and development, clinical trial optimization, and patient stratification [2]. We will analyze the algorithms, datasets, and computational infrastructure driving these advancements. The primary objective is to evaluate the current state of AI/ML in these domains, identify key challenges and

opportunities, and explore the potential for future integration to improve healthcare outcomes and reduce the $R\&D$ costs associated with bringing new therapies to market.

2. Historical Overview: AI in Healthcare

2.1. Early Applications of AI in Medicine

Early applications of AI in medicine centered around expert systems, designed to mimic the reasoning of human experts. A prominent example is MYCIN, developed in the 1970s, which aimed to diagnose bacterial infections and recommend appropriate antibiotics. MYCIN employed a rule-based system, using IF-THEN statements to infer diagnoses based on patient data. While MYCIN demonstrated promising diagnostic accuracy in controlled settings, it faced significant limitations. These included difficulties in knowledge acquisition (eliciting rules from experts), limited ability to handle uncertainty beyond predefined probabilities, and challenges in integrating with existing clinical workflows. Table 1 summarizes the timeline of AI in healthcare [3]. Furthermore, MYCIN lacked common-sense reasoning and the ability to explain its conclusions in a way that clinicians found easily understandable, hindering its widespread adoption.

Table 1. Timeline of AI in Healthcare.

Year(s)	Milestone	Description	Limitations
1970s	Development of MYCIN	Expert system designed to diagnose bacterial infections and recommend antibiotics using $IF\text{-}THEN$ rules.	Knowledge acquisition difficulties, limited uncertainty handling beyond predefined probabilities, integration challenges with clinical workflows, lack of common-sense reasoning and explainability.

2.2. Evolution of Machine Learning Techniques

The application of machine learning (ML) in healthcare, particularly in drug discovery and personalized medicine, has been shaped by the evolution of various algorithms. Early efforts utilized rule-based systems and simple statistical models. The rise of support vector machines (SVMs) and artificial neural networks (ANNs) in the late 20th century offered improved predictive capabilities for tasks like target identification and toxicity prediction. More recently, deep learning architectures, including convolutional neural networks (CNNs) for image analysis and recurrent neural networks (RNNs) for sequential data like genomic sequences, have demonstrated significant advancements. These methods allow for the analysis of complex, high-dimensional datasets, enabling the identification of subtle patterns and relationships that were previously inaccessible. The ongoing development of techniques like graph neural networks (GNNs) to model molecular structures promises further breakthroughs [4].

3. AI in Drug Target Identification and Validation

3.1. Network Biology and Target Prediction

Network biology offers a powerful framework for understanding complex biological systems and identifying potential drug targets. By representing cellular components and their interactions as networks, we can leverage AI and machine learning to uncover hidden relationships and predict novel targets. These networks, often constructed from diverse omics data such as genomics, proteomics, and metabolomics, capture the intricate interplay of genes, proteins, and metabolites within a cell or organism [5].

AI algorithms, particularly graph neural networks (GNNs) and network embedding techniques, excel at analyzing these complex network structures. GNNs can learn node representations that capture both the local and global context of each node within the network, allowing for the identification of key proteins or genes that are central to disease

pathways. Network embedding methods, on the other hand, aim to represent the entire network in a lower-dimensional space, preserving the network's topology and enabling efficient similarity searches for potential drug targets.

Furthermore, machine learning models can be trained to predict the effect of perturbing specific nodes in the network, simulating the impact of a drug on the system. For instance, if a node i represents a protein and its expression level is denoted by x_i, we can use AI to predict how changing x_i will affect the expression levels of other proteins in the network, denoted as x_j for j≠i. Table 2 provides a comparison of target identification methods. This allows researchers to prioritize targets that are likely to have a significant therapeutic effect with minimal off-target effects [6]. The integration of AI with network biology provides a promising avenue for accelerating drug discovery by identifying and validating novel drug targets in a systematic and data-driven manner.

Table 2. Comparison of Target Identification Methods.

Method	Description	Advantages	Disadvantages
Graph Neural Networks (GNNs)	Learn node representations capturing local and global context within a biological network.	Identifies key proteins/genes central to disease pathways by analyzing network structure.	Computationally intensive; requires significant training data.
Network Embedding	Represents the entire network in a lower-dimensional space, preserving topology.	Enables efficient similarity searches for potential drug targets.	May lose fine-grained information about individual nodes.
Perturbation Prediction using AI	Trains machine learning models to predict the effect of perturbing nodes, such as proteins, in the network. For example, predicting the effect of changing x_i on x_j.	Simulates drug impact on the system; helps prioritize targets with therapeutic effect and minimal off-target effects.	Accuracy depends on the quality and completeness of the network data; may not capture all biological complexities.

3.2. Genomics and Proteomics Data Analysis

AI's ability to process and interpret vast datasets has revolutionized target validation using genomics and proteomics data. Traditional methods often struggle with the sheer volume and complexity of these datasets, hindering the identification of promising drug targets. AI algorithms, particularly machine learning models, offer a powerful alternative by identifying patterns and correlations that might be missed by conventional approaches.

For example, deep learning models can be trained on large-scale genomic datasets, such as those generated by genome-wide association studies (GWAS), to predict the effect of genetic variations on disease risk. These models can identify non-coding regions of the genome that regulate gene expression and contribute to disease pathogenesis. By integrating these genomic insights with proteomic data, AI can pinpoint specific proteins that are dysregulated in disease states and are therefore potential drug targets.

Furthermore, AI can analyze protein-protein interaction (PPI) networks derived from proteomics experiments. By applying network analysis techniques, AI can identify key hub proteins that play a central role in disease-related pathways. Targeting these hub proteins can have a broader therapeutic effect than targeting individual proteins in the pathway. For instance, AI-driven analysis of PPI networks in cancer cells has revealed novel therapeutic targets by identifying proteins that are essential for tumor growth and survival, even if they were not previously known to be directly involved in cancer development. This integrated approach, leveraging both genomic and proteomic data

through AI, significantly accelerates the target validation process and increases the likelihood of identifying effective drug targets [7].

4. AI in Drug Design and Repurposing

4.1. De Novo Drug Design with AI

De novo drug design, the creation of novel molecules from scratch, represents a paradigm shift in pharmaceutical research. AI is revolutionizing this field by enabling the generation of drug candidates with pre-defined characteristics, bypassing the limitations of traditional library screening or chemical modification of existing compounds. Generative models, a class of AI algorithms, are at the forefront of this innovation. These models learn the underlying patterns and rules governing chemical structures and their properties from vast datasets of known molecules [8].

One prominent example is the use of Generative Adversarial Networks (GANs). GANs consist of two neural networks: a generator and a discriminator. The generator creates new molecular structures, while the discriminator evaluates their "realness" based on the training data. Through an iterative adversarial process, the generator learns to produce increasingly realistic and desirable molecules, effectively exploring the chemical space far beyond what is currently known. The generator aims to minimize the probability that the discriminator can distinguish its generated samples, $p(D(G(z)))$, while the discriminator aims to maximize this probability. This can be represented as $\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]$, where x represents real data,z represents random noise,G is the generator, and D is the discriminator.

Beyond GANs, other generative models like variational autoencoders (VAEs) and recurrent neural networks (RNNs) are also employed. VAEs learn a latent representation of molecular structures, allowing for smooth interpolation and generation of novel compounds with desired properties. RNNs, particularly well-suited for sequential data, can generate molecules atom-by-atom, ensuring syntactical validity and chemical feasibility. Table 3 illustrates the AI-driven drug design workflow. The application of these AI techniques significantly accelerates the early stages of drug discovery, offering a promising avenue for identifying novel therapeutic agents.

Table 3. AI-Driven Drug Design Workflow.

Stage	AI Technique	Description	Benefits
Data Acquisition & Preparation	Data Mining, Data Cleaning	Gathering and curating large datasets of chemical structures, properties, and biological activities.	Ensures high-quality training data for AI models, leading to more accurate predictions and effective drug design.
Model Training & Validation	GANs, VAEs, RNNs	Training generative models on chemical datasets to learn the rules and patterns of molecular structures and their properties. Includes validation to assess model performance.	Enables the generation of novel molecules with desired characteristics. Specifically for GANs: The generator aims to minimize the probability that the discriminator can distinguish its generated samples, $p(D(G(z)))$, while the discriminator aims to maximize this probability. This can be represented as $\min_G \max_D V(D,G) = E_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]$, where x represents real data,z

			represents random noise, G is the generator, and D is the discriminator.
De Novo Molecule Generation	GANs, VAEs, RNNs	Using trained generative models to create new molecular structures from scratch, potentially with pre-defined properties or target activities.	Expands the search space beyond existing compounds, enabling the discovery of novel drug candidates.
Property Prediction & Optimization	Machine Learning Models (e.g., regression models, classification models)	Predicting the properties (e.g., binding affinity, solubility, toxicity) of generated molecules using machine learning models. Optimizing molecular structures based on predicted properties.	Streamlines the drug design process by prioritizing molecules with desirable properties and reducing the need for extensive experimental testing. Allows for "fine-tuning" of molecules.
Virtual Screening & Ranking	Scoring functions, Docking simulations	Evaluating the generated molecules for their potential efficacy against a specific biological target using virtual screening techniques. Ranking molecules based on their predicted activity and other relevant properties.	Focuses experimental efforts on the most promising drug candidates, reducing the cost and time of drug discovery.
Experimental Validation	In vitro and in vivo assays	Synthesizing and experimentally testing the most promising molecules identified through AI-driven methods to validate their activity, safety, and efficacy.	Confirms the predictions made by AI models and identifies lead compounds for further development.

4.2. Drug Repurposing Strategies

AI-driven drug repurposing offers a powerful approach to accelerate drug discovery by identifying novel applications for existing drugs. This strategy leverages the wealth of existing data on drug properties, mechanisms of action, and clinical effects to predict potential new uses, significantly reducing the time and cost associated with traditional drug development. Machine learning models can analyze diverse datasets, including genomic data, proteomic profiles, and electronic health records, to uncover hidden relationships between drugs and diseases [9].

One successful case study involves the identification of baricitinib, an FDA-approved drug for rheumatoid arthritis, as a potential treatment for COVID-19. AI algorithms analyzed gene expression data from COVID-19 patients and identified baricitinib as a candidate due to its ability to inhibit the AP2-associated protein kinase 1 (AAK1), which is involved in viral entry. Subsequent clinical trials demonstrated the efficacy of baricitinib in reducing mortality in hospitalized COVID-19 patients. Another example is the use of AI to repurpose sildenafil, originally developed for hypertension, as a treatment for erectile dysfunction. Analysis of clinical trial data revealed an unexpected side effect of vasodilation in the penis, leading to its successful repurposing. Table 4 provides examples of successful drug repurposing. Furthermore, AI has been employed to identify potential treatments for rare diseases by matching drug profiles with disease-specific gene expression signatures. The ability of AI to analyze vast amounts of data and identify subtle patterns makes it a valuable tool for drug repurposing, offering the potential to rapidly address unmet medical needs.

Table 4. Examples of Successful Drug Repurposing.

Drug Name	Original Use	Repurposed Use	AI's Role
Baricitinib	Rheumatoid Arthritis	COVID-19 Treatment	AI analyzed gene expression data to identify baricitinib's potential to inhibit AAK1, involved in viral entry.
Sildenafil	Hypertension	Erectile Dysfunction	AI analysis of clinical trial data revealed an unexpected side effect of vasodilation.
		Rare Disease Treatment	AI matches drug profiles with disease-specific gene expression signatures.

5. Challenges and Limitations

5.1. Data Bias and Generalizability

Data bias presents a significant hurdle in applying AI to drug discovery and personalized medicine. Machine learning models are only as good as the data they are trained on; if the training data disproportionately represents certain demographics or disease subtypes, the resulting models will exhibit skewed performance. This can lead to inaccurate predictions and potentially harmful treatment recommendations for underrepresented groups [10].

Generalizability is further compromised by the inherent heterogeneity of human populations. Factors such as genetics, lifestyle, and environmental exposures contribute to variations in drug response [11]. Models trained on specific cohorts may fail to accurately predict outcomes in individuals from different backgrounds. Addressing this requires careful consideration of data diversity and the development of robust algorithms that can account for inter-individual variability, potentially through techniques like transfer learning and domain adaptation, to ensure equitable and effective application of AI across diverse patient populations. The impact of bias can be quantified using metrics like F_1-score and disparate impact ratio, $DIR = \frac{P(Y=1|A=a)}{P(Y=1|A=a')}$, where A is a sensitive attribute [12].

5.2. Interpretability and Explainability

A significant hurdle in applying AI to drug discovery lies in the 'black box' nature of many advanced models, particularly deep learning architectures. While these models can achieve high accuracy in predicting drug-target interactions or patient responses, understanding *why* they arrive at a specific conclusion remains a challenge [13]. This lack of interpretability hinders trust and adoption, as researchers need to validate the underlying biological rationale. Explainability is crucial for identifying potential biases in

the training data, uncovering novel biological mechanisms, and ensuring that AI-driven predictions are not simply based on spurious correlations. The development of methods to open the 'black box' and provide transparent, understandable results is essential for responsible and effective AI-driven drug discovery. Furthermore, the ability to explain AI predictions is vital for regulatory approval and clinical implementation, where understanding the basis for a treatment recommendation is paramount. The variable x may represent a specific gene [14].

6. Future Perspectives

6.1. Integration with Multi-Omics Data

The future of drug discovery and personalized medicine hinges on the comprehensive integration of AI with multi-omics data. This involves leveraging machine learning algorithms to analyze vast datasets encompassing genomics, transcriptomics, proteomics, and metabolomics, offering a holistic view of disease mechanisms [15]. AI can identify complex correlations between these omics layers and patient phenotypes, leading to more accurate disease subtyping and the identification of novel drug targets. Furthermore, AI-driven models can predict individual patient responses to therapies based on their unique omics profiles, paving the way for truly personalized treatment strategies. The challenge lies in developing robust AI methods capable of handling the high dimensionality and inherent noise within multi-omics datasets, requiring sophisticated feature selection and data integration techniques. Ultimately, this integration promises to accelerate the drug development pipeline and improve patient outcomes by tailoring treatments to individual biological characteristics, where x represents the omics data and y the predicted outcome [16].

6.2. AI-Driven Clinical Trials

AI holds immense promise in revolutionizing clinical trials, addressing long-standing challenges in efficiency and efficacy. Machine learning algorithms can optimize patient recruitment by identifying suitable candidates based on diverse datasets, including genomic profiles, medical history, and lifestyle factors, thereby reducing screening failures and accelerating enrollment. Furthermore, AI can predict trial outcomes with greater accuracy by analyzing complex interactions between variables such as drug dosage (d), patient characteristics (p), and biomarkers (b), potentially leading to better-designed trials with increased probability of success. This predictive capability allows for adaptive trial designs, enabling modifications based on real-time data analysis and improved resource allocation [17].

7. Conclusion

7.1. Summary of Key Findings

This review has highlighted the transformative potential of AI and machine learning in revolutionizing drug discovery and personalized medicine. We demonstrated how AI algorithms accelerate target identification, predict drug efficacy, and optimize drug design, significantly reducing the time and cost associated with traditional methods. Furthermore, the application of machine learning in analyzing patient data, including genomics (x_i), proteomics (y_i), and clinical records (z_i), enables the development of tailored treatment strategies. The integration of these technologies promises a future where drug development is faster, more efficient, and more personalized, ultimately improving patient outcomes.

7.2. Concluding Remarks

AI and machine learning hold immense promise for revolutionizing drug discovery and personalized medicine. While challenges remain in data quality, model

interpretability, and regulatory frameworks, the trajectory is undeniably positive. Future advancements will likely see AI playing an increasingly central role in identifying novel drug targets, predicting patient responses to therapies, and optimizing treatment strategies. The convergence of AI with multi-omics data, coupled with sophisticated algorithms, will pave the way for more effective and tailored healthcare solutions, ultimately improving patient outcomes and reducing healthcare costs. The future hinges on responsible development and ethical implementation of these powerful technologies.

References

1. J. Jin, T. Zhu, and C. Li, "Graph Neural Network-Based Prediction Framework for Protein-Ligand Binding Affinity: A Case Study on Pediatric Gastrointestinal Disease Targets," Journal of Medicine and Life Sciences, vol. 1, no. 3, pp. 136–142, 2025.
2. P. P. Parvatikar, S. Patil, K. Khaparkhuntikar, S. Patil, P. K. Singh, R. Sahana, A. V. Raghu, "Artificial intelligence: Machine learning approach for screening large database and drug discovery," Antivir. Res., vol. 220, 105740, 2023.
3. C. L. Cheong, "Research on AI security strategies and practical approaches for risk management," Journal of Computer, Signal, and System Research, vol. 2, no. 7, pp. 98–115, 2025.
4. A. V. Joshi, "Machine learning and artificial intelligence."
5. S. Li, K. Liu, and X. Chen, "A context-aware personalized recommendation framework integrating user clustering and BERT-based sentiment analysis," Journal of Computer, Signal, and System Research, vol. 2, no. 6, pp. 100–108, 2025.
6. J. M. Helm, A. M. Swiergosz, H. S. Haeberle, J. M. Karnuta, J. L. Schaffer, V. E. Krebs, P. N. Ramkumar, "Machine learning and artificial intelligence: definitions, applications, and future directions," Curr. Rev. Musculoskelet. Med., vol. 13, no. 1, pp. 69-76, 2020.
7. R. Luo, X. Chen, and Z. Ding, "SeqUDA-Rec: Sequential user behavior enhanced recommendation via global unsupervised data augmentation for personalized content marketing," arXiv preprint arXiv:2509.17361, 2025.
8. L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, "Machine learning and artificial intelligence in research and healthcare," Injury, vol. 54, pp. S69-S73, 2023.
9. P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in 2017 15th Int. Conf. ICT Knowl. Eng. (ICT&KE), 2017, pp. 1-6.
10. N. Kühl, M. Schemmer, M. Goutier, and G. Satzger, "Artificial intelligence and machine learning," Electron. Markets, vol. 32, no. 4, pp. 2235-2244, 2022.
11. R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar, "Artificial intelligence to deep learning: machine intelligence approach for drug discovery," Mol. Divers., vol. 25, no. 3, pp. 1315-1360, 2021.
12. T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," J. Glob. Health, vol. 8, no. 2, 020303, 2018.
13. Y. S. Cai, "Organizational restructuring of fintech enterprises: A strategic study balancing compliance and innovation," Financial Economics Insights, vol. 3, no. 1, pp. 1–10, 2026.
14. L. D. Jones, D. Golan, S. A. Hanna, and M. Ramachandran, "Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern?," Bone & Joint Res., vol. 7, no. 3, pp. 223-225, 2018.
15. A. Karthikeyan and U. D. Priyakumar, "Artificial intelligence: machine learning for chemical sciences," J. Chem. Sci., vol. 134, no. 1, 2, 2022.
16. J. L. Zhao, "Graph-based deep dive on AI startup revenue composition and venture capital network effect," Economics and Management Innovation, vol. 3, no. 1, pp. 27–36, 2026.
17. P. Mathur, S. Srivastava, X. Xu, and J. L. Mehta, "Artificial intelligence, machine learning, and cardiovascular disease," Clin. Med. Insights Cardiol., vol. 14, 1179546820927404, 2020.