Global Journal of Research in
Biology and Pharmacy
Journal homepage: https://gsjournals.com/gjrbp/
ISSN: 2980-4221 (Online)

(REVIEW ARTICLE)

Check for updates

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN DRUG DISCOVERY, REPURPOSING, AND CLINICAL TRIAL OPTIMIZATION

Abdulrahman Nabillah *

Independent Researcher, Nigeria.
* Corresponding Author

Global Journal of Research in Biology and Pharmacy, 2023, 02(03), 001–005

Publication history: Received on 06 May 2023; revised on 21 July 2023; accepted on 25 July 2023

Article DOI: https://doi.org/10.58175/gjrbp.2023.2.3.0092

Copyright © 2023 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Artificial intelligence and machine learning are transforming pharmaceutical research by enabling faster, cheaper, and more informed decision-making across the drug development pipeline. This review examines the applications of AI and machine learning in three major domains: drug discovery, drug repurposing, and clinical trial optimization. Deep learning architectures, including convolutional neural networks, graph neural networks, transformers, and generative models, have been applied to molecular property prediction, de novo drug design, and virtual screening. AlphaFold2 has revolutionized protein structure prediction, providing new opportunities for structure-based drug discovery. Machine learning has also accelerated drug repurposing efforts, particularly during the COVID-19 pandemic, by identifying candidate compounds from existing drug libraries. In clinical trials, AI tools are being used for patient recruitment, eligibility assessment, adaptive trial design, and synthetic control arms, with the potential to reduce cost and improve success rates. However, significant challenges remain, including data quality and availability, model interpretability, regulatory acceptance, and the risk of bias. This review critically discusses recent progress, limitations, and future directions, drawing primarily on literature published through early 2023. The evidence suggests that while AI has not yet produced a paradigm-shifting number of approved drugs, its influence is expanding rapidly and is likely to become integral to drug development.

Keywords: Artificial Intelligence; Machine Learning; Drug Discovery; Drug Repurposing; Clinical Trial Optimization; Deep Learning

1 INTRODUCTION

Pharmaceutical research and development is characterized by high costs, long timelines, and high failure rates. The average cost of bringing a new drug to market is estimated to exceed one billion dollars, and clinical attrition remains a major problem. Artificial intelligence offers the potential to reduce costs and timelines by improving the efficiency of target identification, lead optimization, preclinical safety assessment, and clinical development.

The application of machine learning to chemistry and biology is not new, but recent advances in deep learning, the availability of large datasets, and increased computing power have accelerated progress. Deep learning models can learn complex patterns from molecular structures, genomic data, and clinical records, enabling predictions that were

previously impossible. Generative models can design novel molecules with desired properties, while natural language processing can extract information from scientific literature and electronic health records.

The COVID-19 pandemic highlighted the value of computational approaches for rapid drug repurposing. Machine learning models were used to screen existing drugs for antiviral activity and to analyze protein–protein interaction networks. At the same time, clinical trial disruptions prompted interest in decentralized and AI-assisted trial designs.

This review provides a critical overview of AI and machine learning in drug discovery, repurposing, and clinical trial optimization, focusing on developments through early 2023.

2 AI AND MACHINE LEARNING IN DRUG DISCOVERY

2.1 Key Machine Learning Methods

Supervised learning methods, including random forests, support vector machines, and neural networks, are commonly used for quantitative structure–activity relationship modeling and property prediction. Deep learning architectures such as graph neural networks operate directly on molecular graphs, capturing atom and bond features. Recurrent neural networks and transformers process sequential data, including SMILES strings and amino acid sequences. Generative adversarial networks and variational autoencoders generate novel molecular structures with optimized properties. Reinforcement learning has been used for de novo molecule generation and synthesis planning.

2.2 Target Identification and Validation

AI can identify disease-relevant targets by integrating multi-omics data, protein–protein interaction networks, and text mining. Knowledge graphs connect genes, diseases, drugs, and pathways, enabling target prioritization. AlphaFold2, a deep learning system developed by DeepMind, can predict protein structures with near-experimental accuracy. This breakthrough facilitates structure-based drug design for previously intractable targets and supports the identification of binding pockets and allosteric sites.

2.3 Hit Identification and Lead Optimization

Virtual screening using machine learning models can rapidly evaluate large chemical libraries to identify active compounds. Deep learning models trained on bioactivity data can predict binding affinities and selectivity. Generative models have been used to design novel inhibitors. A notable example is the discovery of potent DDR1 kinase inhibitors using a generative reinforcement learning approach. Another study used a deep neural network to identify halicin, an antibacterial compound with a novel mechanism, demonstrating the potential for antibiotic discovery.

2.4 ADMET Prediction

Absorption, distribution, metabolism, excretion, and toxicity prediction is essential for reducing late-stage failures. Machine learning models trained on historical ADMET data can flag problematic compounds early. Deep learning models have been applied to predict hERG channel inhibition, cytochrome P450 metabolism, and hepatotoxicity. These models are increasingly integrated into multi-parameter optimization workflows.

3 DRUG REPURPOSING

Drug repurposing identifies new uses for existing approved drugs, offering shorter development timelines and lower costs. AI approaches can analyze gene expression signatures, molecular structures, and biomedical literature to suggest candidate drugs for a given disease.

During the COVID-19 pandemic, machine learning models were used to screen FDA-approved drugs for potential activity against SARS-CoV-2. A deep learning drug–target interaction model predicted several existing antiviral agents, including atazanavir and remdesivir, as potential candidates. Protein–protein interaction maps and network-based approaches identified host factors and suggested repurposable drugs. Machine learning was also used to analyze electronic health records to identify drugs associated with improved outcomes in COVID-19 patients.

Beyond COVID-19, AI-based repurposing has been applied to rare diseases, oncology, and neurological disorders. Knowledge graph approaches integrate diverse biological data to predict novel drug–disease associations. However, many predictions require experimental validation, and the translation of computational hits to clinical benefit remains limited.

4 CLINICAL TRIAL OPTIMIZATION

4.1 Patient Recruitment and Eligibility

Patient recruitment is a major bottleneck in clinical trials. Machine learning models can analyze electronic health records to identify eligible patients and predict enrollment rates. Natural language processing can extract inclusion and exclusion criteria from trial protocols and match them to patient records. AI-powered platforms have been shown to improve recruitment efficiency and reduce screen failure rates.

4.2 Adaptive Trial Design and Synthetic Control Arms

AI can support adaptive trial designs by predicting treatment effects and suggesting modifications to dosing or patient allocation in real time. Synthetic control arms, built using historical data or real-world evidence, can reduce the number of patients assigned to placebo and accelerate trials. Machine learning models can generate synthetic patient outcomes based on baseline characteristics, although regulatory acceptance of these approaches remains limited.

4.3 Monitoring and Safety

AI can analyze adverse event reports, laboratory data, and wearable device data to detect safety signals early. Machine learning models can identify patients at high risk of adverse events and support risk-based monitoring. Natural language processing can process unstructured clinical notes to extract safety information more efficiently.

4.4 Real-World Evidence and Digital Twins

Real-world data from electronic health records, claims databases, and registries are increasingly used to complement clinical trial evidence. AI can create digital twins—computational models of individual patients—to simulate treatment outcomes and inform personalized trial designs. These approaches may improve external validity and reduce trial costs.

5 CHALLENGES AND LIMITATIONS

5.1 Data Quality and Availability

AI models require large, high-quality, well-annotated datasets. Public databases such as ChEMBL, PubChem, and the Protein Data Bank are valuable but have biases and incompleteness. Clinical data are often unstructured, heterogeneous, and subject to privacy regulations. Data sharing remains a barrier.

5.2 Interpretability and Trust

Many deep learning models are black boxes, making it difficult to understand why a prediction was made. This limits acceptance by medicinal chemists, toxicologists, and regulators. Explainable AI methods are being developed, but they often lag behind model performance.

5.3 Validation and Generalization

Models trained on one chemical space or patient population may not generalize to others. Overfitting and data leakage can produce overly optimistic performance. Prospective validation is essential but rare. Many AI-discovered molecules have not progressed to advanced clinical trials.

5.4 Regulatory and Ethical Considerations

Regulatory agencies have begun to issue guidance on AI in drug development, but standards are evolving. Bias in training data can lead to inequitable outcomes. The use of synthetic control arms and digital twins raises methodological and ethical questions. Transparency, reproducibility, and data governance are critical.

6 FUTURE PERSPECTIVES

The integration of AI into drug discovery and development is likely to deepen. Foundation models for chemistry and biology, analogous to large language models, may enable transfer learning across tasks. AlphaFold2 and its successors will continue to advance structural biology, potentially enabling structure-based drug design for a broader range of targets. Multimodal models that combine chemistry, biology, and clinical data may improve target validation and patient stratification.

In clinical trials, AI may enable more patient-centric, decentralized, and adaptive designs. Wearable devices and digital biomarkers will generate high-frequency data for real-time monitoring. Synthetic control arms and digital twins may become more accepted as regulatory frameworks mature.

However, the ultimate test is whether AI can reduce the time and cost of bringing new drugs to patients. This will require prospective validation, rigorous benchmarking, and collaboration among academia, industry, and regulators.

7 CONCLUSION

Artificial intelligence and machine learning are making significant contributions to drug discovery, repurposing, and clinical trial optimization. Deep learning models can predict molecular properties, generate novel compounds, and identify new uses for existing drugs. AlphaFold2 has transformed structural biology. In clinical development, AI tools are improving recruitment, enabling adaptive designs, and supporting safety monitoring. Despite these advances, challenges related to data quality, interpretability, validation, and regulatory acceptance remain. The field is evolving rapidly, and while AI has not yet revolutionized drug development, its influence is growing. Continued interdisciplinary research and rigorous evaluation will be essential to realize the full potential of AI in pharmaceutical innovation.

REFERENCES

[1] Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463–477.

[2] Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702.

[3] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–589.

[4] Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038–1040.

[5] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Drug Discov Today. 2021;26(2):511–524.

[6] Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80–93.

[7] Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583:459–468.

[8] Beck BR, Shin B, Choi Y, et al. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J. 2020;18:784–790.

[9] Harrer S, Shah P, Antony B, et al. Artificial intelligence for clinical trial design. Trends Pharmacol Sci. 2019;40(8):577–591.

[10] Woo M. An AI boost for clinical trials. Nature. 2019;573:S100–S102.

[11] Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241–1250.

[12] Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268–276.

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

[14] Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. JAMA. 2020;323(9):844–853.

[15] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347–1358.

[16] Mamoshina P, Vieira A, Putin E, et al. Applications of deep learning in biomedicine. Mol Pharm. 2016;13(5):1445–1454.

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

[18] Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inform. 2016;35(1):3–14.

[19] Brown N, Ertl P, Lewis R, et al. Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des. 2020;34(7):709–715.

[20] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.