INTERNATIONAL JOURNAL OF
PHARMACEUTICAL AND HEALTHCARE INNOVATION
journal homepage: www.ijphi.com

Review Article

OPEN ACCESS

AI-Driven Genomic Mining for Drug Target Discovery and Disease Modelling

Samiksha Mestha[1], Ramdas Bhat*[2]

[1]PG Student, Department of Pharmacy Practice, Srinivas college of Pharmacy, Valachil, Post Farangipete, Mangalore-574253, Karnataka, India

[2]Associate Professor, Department of Pharmacology, Srinivas college of Pharmacy, Valachil,Post Farangipete, Mangalore-574253, Karnataka, India

Article Info

Article history:

Manuscript ID:

IJPHI2127312025
Received: 21- April -2025
Revised :- 27 April -2024
Accepted: 31-May 2025
Available online: June

Keywords:
AI in drug discovery, pharmacogenomics, genomic data mining, personalized medicine, ethical considerations

Corresponding Author:
ramdas21@gmail.com

Abstract

Artificial intelligence (AI) is revolutionizing drug discovery and pharmacogenomics by allowing for the quick analysis of complex genomic data to reveal disease mechanisms and therapeutic targets. Sophisticated machine learning and deep learning algorithms speed up drug development by forecasting drug-target interactions, lead compound optimization, and modeling complex biological pathways with high accuracy. Through the integration of multi-omics data, such as genomic, transcriptomic, and proteomic information, AI systems are able to detect biomarkers and model intricate biological systems, enabling the path for precision medicine to be adapted to personal genetic blueprint. Such technologies have a profound effect in rare disease research, where AI makes the detection of genetic mutations associated with clinical phenotypes easier to enable the design of targeted therapies. In spite of these advances, challenges such as data standardization across diverse genomic datasets, algorithmic bias from underrepresentation in training data, and model interpretability pose important hurdles to clinical adoption. Ethical issues around genetic privacy also underscores the importance of strong frameworks to protect sensitive health data in collaborative analysis. Solving these challenges involves the use of explainable AI architectures to improve model transparency and reliability and multimodal data sources such as CRISPR-edited cellular models and single-cell sequencing results to improve predictive accuracy. Future research directions focus on enhancing computational infrastructure, designing adaptive regulatory policy, and putting equity at the forefront in genomic database curation to make AI-mediated healthcare benefits accessible throughout the world. By overcoming these technical and ethical barriers, AI can empower a revolution in precision medicine that accelerates drug discovery, personalizes patient treatment regimens, and enhances healthcare outcomes in a variety of diverse populations while promoting fairness and inclusivity in its applications. This revolutionary approach can potentially reshape global healthcare systems and drive therapeutic innovation at an unprecedented scale.

1. INTRODUCTION

Genomic data mining is an innovative method in contemporary biomedical research that allows researchers to glean useful patterns and relationships from enormous genomic data. Through intensive gathering, processing, and interpretation of genetic data, scientists are able to detect disease-causing genes, biomarkers, and therapeutic targets that remain invisible using traditional technologies [1]. Systematic investigation of genomic information has revolutionized the comprehension of disease mechanisms and pathways, offering possibilities for intervention at the molecular level.

The incorporation of artificial intelligence tools in genomic data mining has revolutionized analytical power and hastened discovery cycles. AI methods, especially machine learning and deep learning algorithms, can analyze intricate biological associations with unprecedented accuracy and speed [2]. These computational methods allow predictive simulations of biological interactions, detect subtle patterns in heterogeneous datasets, and allow effective drug repurposing through the uncovering of unanticipated associations between diseases and possible treatments. The interaction between genomic data and AI-based analysis is a paradigm shift in disease understanding and therapeutic development.

The role of genomic data mining pervades several aspects of clinical research and healthcare application [3]. By facilitating targeted therapy according to unique genetic profiles, clinicians are able to implement targeted therapy with enhanced effectiveness and minimized side effects. The identification of new drug targets through the study of genomes significantly raises the prospect of successful drug development while lowering drug development failure rates [4]. In addition, computational methods of drug candidate screening and repurposing dramatically shorten research timelines and decrease development expenses. The major aim of this systematic review is to assess the present scenario, approaches, and applications of AI-aided genomic data mining methods in the identification of new disease-causing genes, the identification of possible therapeutic targets, and the creation of personalized treatment strategies that enhance clinical outcomes while shortening drug development timelines and expenses.

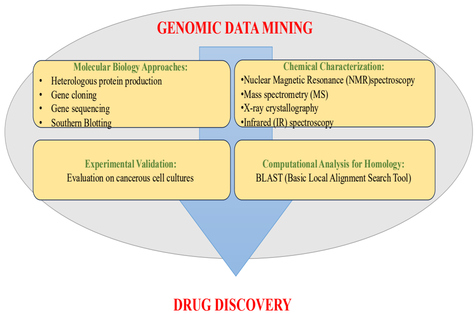

Fig. 1: Overview of genomic data mining process and its applications in drug discovery. This figure was created by the author and does not require external referencing

2. PHARMACOGENOMICS APPLICATIONS

Pharmacogenomics examines the influence of genetic differences on personal responses to individual drugs. AI-based genomic data mining strengthens applications of pharmacogenomics by predicting drug action in particular patient groups based on genetic profiles [5], gaining insight into disease pathophysiology and disease progression [6], and tailoring dosing regimens to enhance therapeutic outcomes and reduce side effects [7]. The convergence of AI and pharmacogenomics has resulted in unprecedented progress in precision medicine, where more precise and effective therapeutic strategies are made possible [8]. Machine learning techniques can recognize drug response-associated genetic markers, and thus, make more educated decisions regarding treatment [9].

Table 1: Key pharmacogenomic applications of AI in different therapeutic areas.

Therapeutic Area	AI Approach Used	Genetic Markers Identified	Clinical Implications
Oncology	Machine Learning (ML) for biomarker discovery	BRCA1, BRCA2, TP53	Personalized cancer therapies, risk prediction
Cardiology	Deep Learning (DL) for genomic risk assessment	APOE, LDLR, PCSK9	Precision medicine for cardiovascular diseases

Neurology	Natural Language Processing (NLP) for genetic variant analysis	APOE4, MAPT, SNCA	Tailored treatment for Alzheimer's, Parkinson's
Psychiatry	AI-driven Polygenic Risk Score (PRS) models	DRD2, SLC6A4, COMT	Personalized antidepressant and antipsychotic selection
Endocrinology	AI-based multi-omics integration	HNF1A, TCF7L2, INS	Individualized diabetes management
Infectious Diseases	AI-assisted genomic epidemiology	HLA-B*57:01, IFNL3	Optimized antiviral drug response
Autoimmune Disorders	Predictive modeling using AI	PTPN22, HLA-DRB1	Personalized immunosuppressive therapy

3. ARTIFICIAL INTELLIGENCE IN EXTRACTION OF GENOMICS DATA

Several AI technologies are utilized in genomic mining analyses, such as deep learning methods for multi-omics data analysis (genomics, proteomics, transcriptomics) to study potential drug targets [10]. Natural language processing is used for literature mining and knowledge extraction from biomedical literature [11], while reinforcement learning is used to optimize experimental design within genomic research [12]. Advanced disease modeling is achieved through genomic data integration with electronic health records [13], and automated feature selection and annotation enhance predictive performance [14]. These strategies have greatly enhanced the accuracy and efficiency of analyzing genomic data, allowing researchers to yield valuable information from intricate biological data sets [15].

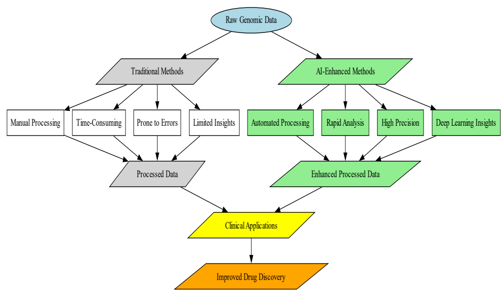

Fig. 2: Comparison of traditional vs. AI-enhanced genomic data extraction methods. This figure was created by the author and does not require external referencing.

4. ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL RESEARCH AND DEVELOPMENT

AI supports pharma research and drug development via several mechanisms. The machine learning methods accurately forecasted drug-target relationships with fewer labor-intensive experimentational screenings [16]. Generative models enable novel pharmacologic compounds with wanted behaviors [17] and enhanced AI systems model life systems for discerning the understanding of pathogeneses and therapies of disease processes [18]. These software applications have sped up the drug discovery pipeline, lowered drug development expenditure, and enhanced candidate drug success rate [19]. AI-based methodologies have proven to be very useful in the identification of repositioning opportunities for reusing approved drugs to cure novel indications [20]

Table 2: Success cases of AI application in drug discovery.

Drug/Compound Name	Original Application	AI Method Used	New Target/Indication Identified	Development Status
Baricitinib	Rheumatoid Arthritis	Deep Learning (Benevolent AI)	COVID-19 (Anti-inflammatory & Antiviral Properties)	Approved for Emergency Use
Halicin	Experimental Antibiotic	Machine Learning (MIT's Deep Learning Model)	Broad-spectrum antibiotic against drug-	Preclinical Stage

			resistant bacteria	
DSP-1181	N/A	Reinforcement Learning (Exscientia & Sumitomo Dainippon)	OCD Treatment	Phase I Trials
Insilico Compound (INS018_055)	N/A	Generative AI (Insilico Medicine)	Idiopathic Pulmonary Fibrosis	Phase I Trials
Alpelisib	Cancer Treatment (PI3K Inhibitor)	AI-driven Target Identification	Treatment for PIK3CA-related overgrowth spectrum (PROS)	FDA Approved

5. CHALLENGES AND LIMITATIONS

As promising as they are, however, AI-driven genomic data mining is still beset with several key challenges, especially with respect to drug discovery.

Data quality and standardization problems are a significant stumbling block. Genomic data tend to originate from multiple labs, with diverse platforms, protocols, and annotation schemes [21]. Such variability introduces noise and bias, which hinders AI models from learning useful patterns. In drug discovery, where the discovery of a defective gene or pathway might result in a new therapeutic target, substandard data quality may lead to inaccurate predictions, wasted resources, and lost opportunities.

Computational complexity is another major issue. Genomic information is very high-dimensional — consider millions of genetic variants across thousands of patients [22]. To run AI algorithms on such huge data takes huge amounts of computational horsepower and very well-tuned code. Without it, analyses will be agonizingly slow or worse, fail outright. In drug discovery timelines, where speed is paramount, computational bottlenecks can hold back the identification of candidate molecules or biomarkers.

Model interpretability also becomes a serious issue. Most advanced AI models, particularly deep learning structures, are like "black boxes" — they predict, but it's difficult to know why they did so [23]. In drug discovery, knowing the "why" is important. Scientists require transparent explanations of why certain genes or pathways are being identified as drug targets. Without transparency, therapeutic leads with promise could be discarded because of the absence of scientific rationale.

Concerns about data privacy add yet more complexity. Genomic data is extremely sensitive and uniquely identifiable [24]. Its improper handling can result in severe ethical and legal problems. Laws such as GDPR and HIPAA strictly regulate how such data are stored, analyzed, and transmitted. In drug development, this can restrict researchers' access to the heterogeneous, large-scale datasets they require, ultimately hindering the production of effective, broadly useful drugs.

Overcoming these hurdles requires collaborative work across disciplines. Scientists in AI, genomics, and ethics need to collaborate to construct standardized protocols, increase model explainability, design computational tools, and enforce rigid compliance with privacy laws [25]. It is only through inter-disciplinary collaboration that we can maximize the potential of AI-based genomic data mining for drug discovery.

6. ETHICAL ISSUES IN AI FOR GENOMIC RESEARCH

The application of AI in genomic research raises important ethical considerations. Informed consent processes must ensure participants understand the scope and implications of AI-driven genomic analysis [26]. Algorithmic bias is a significant concern as genomic research often suffers from inadequate representation of diverse populations, leading to biased AI models [27]. Data security measures must protect genetic data from unauthorized access and cyberattacks [28]. Equity in access to AI-driven genomic medicine must be ensured across all populations [29]. Robust ethical frameworks and regulations are essential to guide the responsible development and application of AI in genomic research [30].

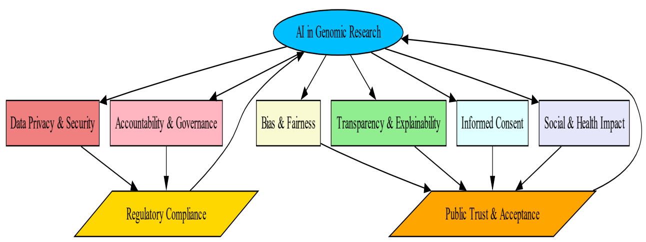

Fig. 3: Ethical framework for AI implementation in genomic research. This figure was created by the author and does not require external referencing.

7. FUTURE DIRECTIONS

To overcome the challenges facing us today and continue to move forward on AI-embedded genomic research, some crucial avenues need to be pursued. Development of explainable AI (XAI) models is critical to enable transparent and interpretable outcomes so that researchers and clinicians can gain greater insight into the biological significance of AI predictions. Increasing model transparency will foster increased trust in AI-based decisions, particularly in areas of high stakes like drug discovery and personalized medicine. Encouraging multi-omics integration, through the integration of genomics with proteomics, transcriptomics, metabolomics, and epigenomics, will provide a more holistic understanding of disease processes. This can enable the identification of new therapeutic targets and facilitate the development of multi-targeted drug therapies. Enhancing data security and privacy measures is also imperative. Methods such as federated learning, where AI models can learn from decentralized data without violating personal privacy, and blockchain technology, which provides transparent and secure handling of sensitive data, can safeguard genetic data while promoting international research collaborations. Greater diversity in genomic datasets is another necessary step. Existing genomic studies over-represent European ancestry populations and thus constrain the generalizability of findings to other ethnic groups. Greater diversity will ensure that genomic findings and AI models equally benefit all populations. Progressing adaptive AI systems that can learn continuously from new genomic and clinical information will further improve predictive power. Dynamic systems will enable real-time updating of disease models and drug discovery processes, making them more robust and adaptable to new knowledge. Developing appropriate ethical and regulatory guidelines is also essential to govern the proper application of AI technologies in genomics. Ensuring robust policies that safeguard human rights, maintain equity, and preserve public trust will be key to the long-term success of AI-based healthcare innovation. Following these paths will place AI-conceptualized genomic science on a path to greatly speed up drug discovery and radically change disease treatment and healthcare provision

CONCLUSION

AI-driven genomic data mining has immensely revolutionized the field of drug target discovery and disease modeling, propelling dramatic innovation in personalized medicine. By revealing intricate genetic interactions and discovering new therapeutic targets, AI technologies have empowered more accurate, patient-specific treatment regimens. To realize the full potential of AI in genomics, though, it is critical to overcome long-standing challenges involving data quality, ethical issues, and interpretability of models. Enhancing the reliability and consistency of genomic data sets, ensuring transparency of predictions made using AI, and enforcing strict ethical controls will improve the credibility, justice, and replicability of genomics. Progress in the near future is anticipated to concentrate on the combination of AI with multi-omics, the union of genomics and proteomics, transcriptomics, and metabolomics, to get a more thorough insight into mechanisms of disease. These developments, while speeding the discovery of medicines, will also provide the building blocks for the emergence of novel forms of therapeutics, paving the way towards improved treatments, enhanced patient health outcomes, and better overall health efficiency in health provision. Through continuous development of the sciences of medicine, such integration of genomic science and AI technology promises to significantly redefine medicine in the years ahead.

Ethical Approval

NA

Informed Consent

Not Applicable.

Funding

No funding was received for conducting this study.

Conflict of Interest

There are no apparent conflicts of interest between the authors' personal relationships or financial interests that may have affected the results of this study, the authors state. There is no conflict of interest, according to the writers. All ideas and opinions expressed in this article are those of the authors.

Financial Interests

The authors declare they have no financial interests.

REFERENCES

1. Yang W, Yoshigoe K, Qin X, Liu JS, Yang JY, Niemierko A, et al. Identification of genes and pathways involved in kidney renal clear cell carcinoma. BMC Bioinformatics. 2014;15(17):1-15.

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):463-77.

3. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet. 2015;16(6):321-32.

4. 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.

5. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015;526(7573):343-50.

6. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. N Engl J Med. 2015;372(23):2229-34.

7. Weinshilboum RM, Wang L. Pharmacogenomics: precision medicine and drug response. Mayo Clin Proc. 2017;92(11):1711-22.

8. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet. 2015;16(2):85-97.

9. Saito Y, Stamp LK, Caudle KE, Hershfield MS, McDonagh EM, Callaghan JT, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for human leukocyte antigen B (HLA-B) genotype and allopurinol dosing. Clin Pharmacol Ther. 2016;99(1):36-7.

10. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nat Genet. 2019;51(1):12-8.

11. Lever J, Barbarino JM, Gong L, Huddart R, Sangkuhl K, Whaley R, et al. Text-mining for precision medicine: automating disease-mutation relationship extraction from biomedical literature. J Am Med Inform Assoc. 2019;26(7):637-45.

12. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. Mol Syst Biol. 2016;12(7):878.

13. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102-10.

14. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform. 2017;18(5):851-69.

15. 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.

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

17. 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):1038-40.

18. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. Nat Biotechnol. 2018;36(9):829-38.

19. Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-64.

20. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. Nat Med. 2017;23(4):405-8.

21. Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. Nat Rev Genet. 2010;11(9):647-57.

22. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. Biomed Inform Insights. 2016;8:1-10.

23. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923. 2017.

24. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. Eur J Hum Genet. 2018;26(2):149-56.

25. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med. 2019;25(1):37-43.

26. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med. 2018;378(11):981-3.

27. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91.

28. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics. 2016;22(2):303-41.

29. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161-4.

30. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. PLoS Med. 2018;15(11):e1002689.