DOI: 10.21088/ijgmr.2319.4782.14225.1

Indian Journal of Genetics and Molecular Research
Volume 14 Number 2, July–December 2025

REVIEW ARTICLE

Artificial Intelligence in Human Genetics: Current Applications and Future Directions

Nitin Kumar[1], Abhishek Basak[2], Hiyam Hamel Mohammed[3], Rajaneesh Kumar Gupta[4]

HOW TO CITE THIS ARTICLE:

Nitin Kumar, Abhishek Basak, Hiyam Hamel Mohammed et. al, Artificial Intelligence in Human Genetics: Current Applications and Future Directions. Ind J Genet Mol Res. 2025; 14(2): 43-51.

ABSTRACT

Artificial intelligence (AI) has emerged as a powerful tool in human genetics, enabling the analysis and interpretation of complex datasets generated by next-generation sequencing, single-cell profiling, and large population biobanks. Traditional statistical and computational methods struggle with the scale, noise, and heterogeneity of these data, whereas AI approaches, particularly machine learning (ML) and deep learning (DL), are uniquely suited to uncover hidden patterns and make clinically relevant predictions. Current applications of AI in genetics include identifying the possible effects of genomic mutations, data base genome mapping, genomic control, and association of different biological data. There has also been some progress in diagnostics of rare diseases, pharmacogenomics, genome-wide association studies (GWAS), and polygenic risk scores analysis. AI has also influenced precision medicine. The use of deep variant, alpha fold, and AI-aided clinical tools are important milestones to note in the arms of genomic medicine. Regardless of progress clinical decision support systems still face challenges like lack interpret interface, reproducibility of data, and equity issues related to privacy. This review aims to describe the dominions in the application AI to human genetics, success tracking, flaws and relief for AI stems from.

KEYWORDS

• Artificial Intelligence • Human Genetics • Machine Learning • Genomics • Precision Medicine

AUTHOR'S AFFILIATION:

[1] MSc Student, Department of Molecular & Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

[2] MSc Student, Department of Molecular & Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

[3] MSc Student, Department of Molecular & Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

[4] Assistant Professor, Department of Molecular & Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

*The first three authors contributed equally to this manuscript

CORRESPONDING AUTHOR:

Rajaneesh Kumar Gupta, Assistant Professor, Department of Molecular & Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India.

E-mail: guptarajaneesh@gmail.com

➤ **Received:** 06-09-2025 ➤ **Accepted:** 07-10-2025

INTRODUCTION

The sequencing of the human genome in 2001 marked a pivotal moment in biomedical research, providing a comprehensive blueprint of the genetic information underlying health and disease. Over the past two decades, advances in next-generation sequencing (NGS), genome-wide association studies (GWAS), and large-scale Bio-banks have greatly expanded the scope of genetic research. Yet, turning this vast amount of information into actionable insights for clinical practice remains a formidable challenge. Complex diseases are influenced not only by single pathogenic mutations but also by thousands of genetic variants with small effects, intricate regulatory networks, epigenetic modifications, and environmental interactions. Traditional bioinformatics and statistical approaches, while essential, often struggle to fully capture the scale, complexity, and nonlinearity of these biological systems.[1] Artificial intelligence (AI), defined as computational systems capable of performing tasks that typically require human intelligence, offers a promising avenue. Machine learning (ML), a subset of AI, allows algorithms to learn from data without explicit programming, while deep learning (DL), a branch of ML using multilayered neural networks, excels at modeling high-dimensional and nonlinear relationships.[2] These techniques have already achieved remarkable success in diverse areas, including image recognition, natural language processing, and drug discovery. Their application to human genetics is now transforming how researchers analyze genomic data, interpret genetic variation, and predict disease risk.[3]

AI is especially appealing in genetics for several reasons. First, modern genomics generates vast, heterogeneous datasets including DNA sequences, transcriptomic profiles, epigenetic modifications, proteomics, metabolomics, and clinical phenotypes that require integrative computational strategies. Second, many genetic challenges involve pattern recognition tasks, such as distinguishing true variants from sequencing errors or predicting the functional impact of mutations, which align naturally with ML approaches. Third, AI can enhance predictive accuracy in polygenic risk scoring and pharmacogenomics, where subtle interactions between genetic variants and environmental factors are difficult to model using linear methods.[4] The field has already reached important milestones. Deep Variant, an AI-based variant caller developed by Google, has set new benchmarks for sequencing accuracy. Alpha Fold, a deep learning system for protein structure prediction, has transformed structural genomics by solving one of biology's most difficult problems. AI-driven tools like Face2Gene have accelerated rare disease diagnostics by linking facial features to genetic syndromes. Meanwhile, large-scale resources such as the UK Biobank, All of Us Research Program, and national genome projects provide the volume of data needed to train robust AI models.[5] Despite its potential, applying AI in genetics comes with challenges. These include the "black-box" nature of many deep learning models, potential biases from underrepresentation of non-European populations in training datasets, and concerns around data privacy and governance. Addressing these issues is critical for translating AI into clinical practice and ensuring its equitable use in genomics.

This review examines how AI is transforming human genetics. We first outline foundational AI principles relevant to genomics, followed by an overview of current applications in variant interpretation, genome annotation, multi-omics integration, and clinical genetics. We then highlight key case studies, challenges, and ethical considerations, concluding with future directions, including explainable AI, integration with genome editing, and predictive modeling of human health.

FOUNDATIONS OF ARTIFICIAL INTELLIGENCE IN GENETICS

AI refers to computational approaches designed to perform tasks that typically require human intelligence, such as pattern recognition, classification, and prediction. Within AI, ML allows algorithms to learn from data and improve performance without explicit programming, whereas DL uses multilayered neural networks capable of capturing complex, nonlinear relationships.[1] These strengths make AI particularly well-suited for human genetics, where the size, complexity, and heterogeneity of datasets often exceed the capabilities of traditional statistical methods.

Key Concepts in AI

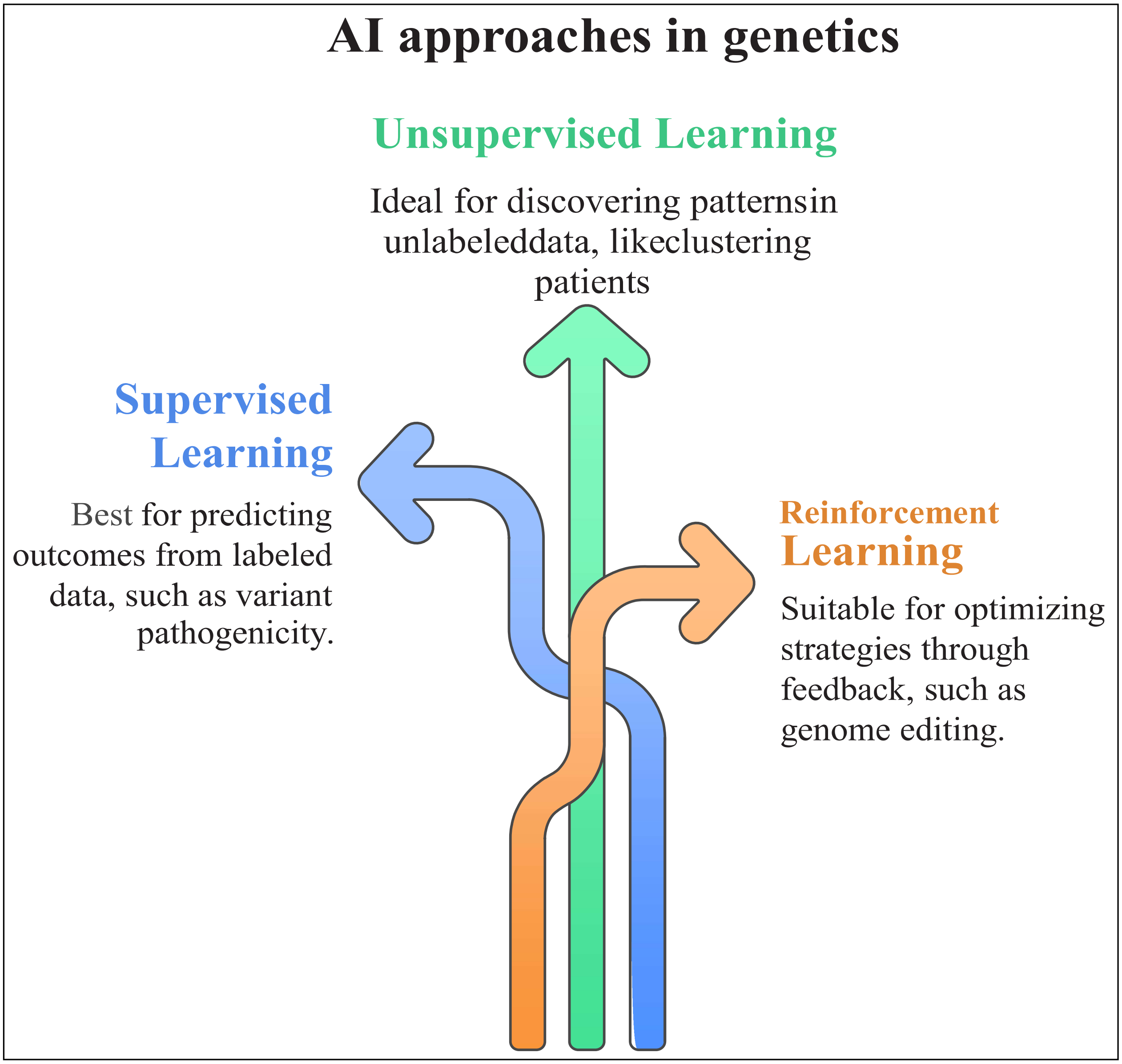

Figure 1: AI approaches in genetics are generally classified (Fig. 1) into:

i. Supervised learning: Models are trained on labeled data, learning to predict outcomes such as variant pathogenicity or disease status from genomic features.[2] Common algorithms include support vector machines (SVMs), random forests, and gradient boosting.

ii. Unsupervised learning: These methods detect patterns in unlabeled data, such as clustering patients based on multi-omics profiles or identifying novel cell types in single-cell data.[3] Techniques include k-means clustering, hierarchical clustering, and autoencoders.

iii. Reinforcement learning: Algorithms learn optimal strategies through trial and error, guided by feedback signals. Emerging applications in genetics include genome editing optimization and drug response prediction.[4]

Deep learning architectures have further expanded AI applications. Convolutional neural networks (CNNs) are particularly effective for analyzing DNA sequences and predicting regulatory element activity. Recurrent neural networks (RNNs) and transformers capture sequential dependencies in genomic data, enabling the study of longrange interactions. Graph neural networks (GNNs) are increasingly applied to represent complex biological networks, such as gene–gene and protein–protein interactions.

Genetic Data Types Suitable for AI

Human genetics research produces multiple types of high-dimensional data:

i. **Genomic sequences:** Whole-genome and whole-exome sequencing generate millions of variants per individual.

ii. **Transcriptomics:** RNA-seq and single-cell RNA-seq provide dynamic gene expression profiles across tissues and cell types.

iii. **Epigenomics:** DNA methylation, histone modifications, and chromatin accessibility datasets capture regulatory mechanisms.

iv. **Proteomics and metabolomics:** These reflect the functional outputs of genetic programs and cellular pathways.

v. **Phenotypic and clinical data:** Electronic health records, imaging, and biobank surveys provide crucial context for linking genotype to phenotype.[5]

AI models can integrate these diverse datasets to reveal hidden patterns and generate predictive insights, enabling a systems-level understanding of human biology.

Advantages and Challenges

AI offers several benefits over conventional approaches:

i. **Handling high dimensionality:** AI can analyze millions of features simultaneously.

ii. Capturing nonlinear relationships: Deep learning uncovers complex interactions among variants and across omics layers.

iii. Integration across modalities: AI can combine genomics, transcriptomics, proteomics, and clinical data for more comprehensive predictions.

However, challenges remain:

i. **Data sparsity and noise:** Sequencing errors, batch effects, and missing data can reduce model performance.

ii. **Interpretability:** Deep learning models often act as “black boxes,” limiting their use in clinical decision-making.

iii. **Bias:** Underrepresentation of certain populations in training datasets can lead to inequitable predictions.

iv. **Computational requirements:** Training deep learning models demands substantial computational resources and specialized expertise.[6]

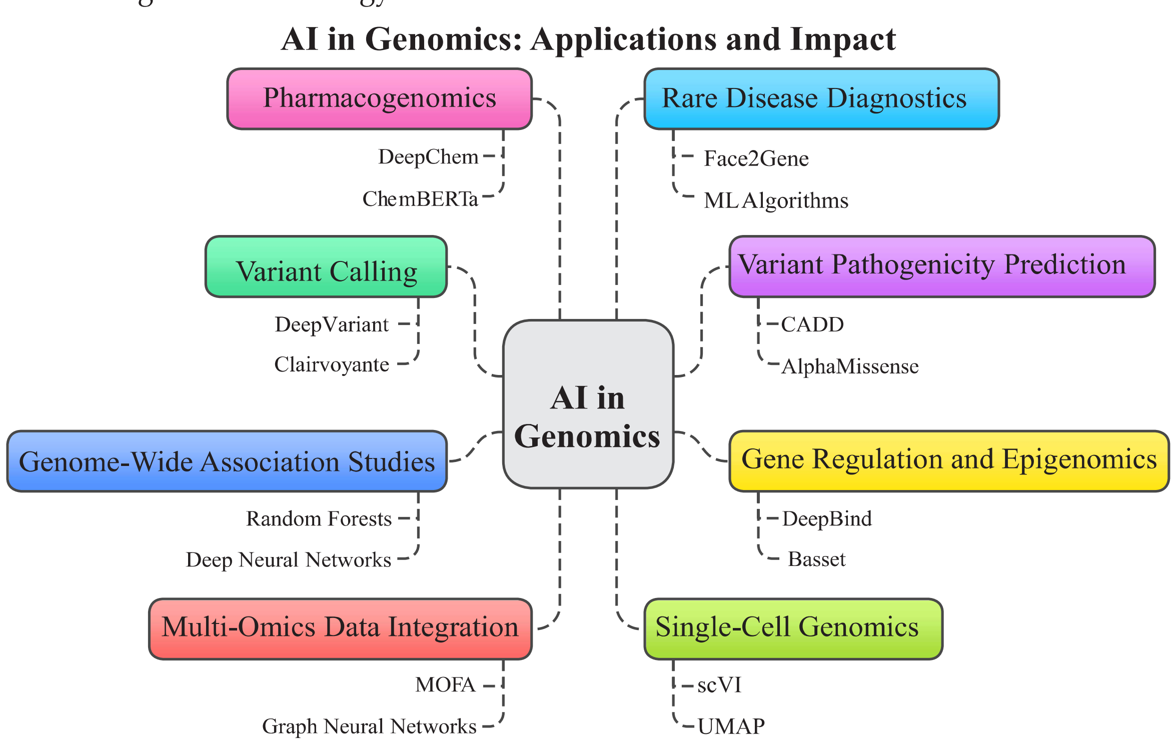

Figure 2:

Current efforts focus on explainable AI (XAI), model regularization, and incorporating biological priors to improve interpretability, reliability, and generalizability. By tackling these challenges, AI has the potential to become a transformative tool for both research and clinical applications.

Current Applications of AI in Human Genetics

Artificial intelligence has rapidly transformed human genetics by enabling the analysis, interpretation, and integration of large-scale genomic data (Figure 2). Its applications span from fundamental research to clinical translation, providing insights that were previously difficult or impossible to achieve with conventional bioinformatics and statistical methods.

Variant Calling and Genome Annotation

Accurate identification of genetic variants is a cornerstone of genomics. Traditional variant calling methods rely on probabilistic models, which can be error-prone in low-coverage or complex genomic regions. AI approaches, especially deep learning, have substantially improved both sensitivity and specificity. For instance, DeepVariant, developed by Google, frames variant calling as an image classification task, converting aligned sequencing reads into images that are analyzed using convolutional neural networks.[1] DeepVariant consistently outperforms traditional pipelines in detecting single nucleotide variants (SNVs) and small insertions/deletions (indels). Other tools, such as Clairvoyante and DeepTrio, extend deep learning to long-read sequencing and trio-based analyses, improving the detection of complex structural variants.[2] Neural network-based genome annotation has also advanced, enabling more accurate identification of coding regions, regulatory elements, and noncoding functional sequences.[3]

Variant Pathogenicity Prediction

A major challenge in genetics is interpreting variants of uncertain significance (VUS). AI models can predict the functional impact and pathogenicity of both coding and noncoding variants. Tools such as CADD (Combined Annotation-Dependent Depletion), PolyPhen-2, and SIFT employ supervised machine learning to integrate evolutionary conservation, protein structure, and functional annotations.[4] More recent deep learning models, including DeepSEA and PrimateAI, use convolutional neural networks trained on large functional genomic datasets to predict regulatory effects and disease relevance.[5] AlphaMissense, which incorporates protein structure information from AlphaFold, demonstrates high accuracy in classifying missense variants and prioritizing candidate pathogenic mutations.[6] These models are increasingly incorporated into clinical pipelines, helping to close the diagnostic gap for rare and complex diseases.

Gene Regulation and Epigenomics

AI has proven highly effective for modeling gene regulation. Noncoding regions, which make up over 98% of the human genome, play essential roles in transcriptional control and disease susceptibility. Models such as DeepBind, Basset, and Basenji use CNNs to predict transcription factor binding, chromatin accessibility, and enhancer activity from DNA sequences.[7] Autoencoders and variational autoencoders (VAEs) are applied to high-dimensional epigenomic datasets, reducing noise and capturing latent regulatory patterns.[8] These approaches help identify functional noncoding variants associated with conditions like cancer, autoimmune disorders, and neurodegeneration.

Genome-Wide Association Studies and Polygenic Risk Scores

Genome-wide association studies (GWAS) have identified thousands of loci linked to complex traits. AI enhances GWAS by modeling nonlinear relationships between variants and phenotypes, improving predictive accuracy. Machine learning methods such as random forests, gradient boosting, and deep neural networks can integrate thousands of variants to compute polygenic risk scores (PRS).[9] These AI-derived PRS predict disease susceptibility—from cardiovascular disease to diabetes—with higher accuracy than traditional linear models.[10] Integrating genetic data with clinical and lifestyle information through AI further strengthens risk stratification and personalized medicine approaches.

Multi-Omics Data Integration

Human diseases often result from interactions across multiple molecular layers. AI provides a framework to integrate genomics, transcriptomics, proteomics, metabolomics, and epigenomics, enabling a holistic view of disease mechanisms. Deep learning-based integrative models, such as MOFA (Multi-Omics Factor Analysis) and graph neural networks, uncover coordinated patterns across datasets.[11] In cancer genomics, AI models that combine somatic mutation data, gene expression, and protein alterations have successfully predicted tumor progression and therapy response.[12] Transfer learning allows models trained on one omics type to enhance predictions in others, which is particularly valuable for rare diseases with limited datasets.

Single-Cell Genomics

Single-cell technologies provide high-resolution maps of cellular heterogeneity but produce sparse and noisy data. AI is crucial for processing, denoising, and interpreting these datasets. Unsupervised clustering with ML algorithms, combined with dimensionality reduction techniques like t-SNE and UMAP, reveals novel cell populations. Deep generative models, such as scVI (single-cell variational inference), probabilistically model gene expression distributions, enabling trajectory inference and lineage prediction.[13] These approaches have illuminated developmental pathways, immune cell differentiation, and tumor evolution at single-cell resolution.

Pharmacogenomics and Drug Response Prediction

AI accelerates pharmacogenomics by predicting how genetic variation affects drug efficacy and toxicity. ML models integrating genomic data with chemical properties of drugs can identify gene–drug interactions and potential adverse reactions.[14] Deep learning frameworks, including DeepChem and ChemBERTa, have been applied to predict patient-specific therapy responses, particularly in oncology.[15] By combining tumor genomics with transcriptomic and proteomic profiles, AI supports precision oncology approaches that optimize treatment selection and improve outcomes.

RARE DISEASE DIAGNOSTICS

Rare genetic disorders often remain undiagnosed due to the vast number of potential variants. AI-assisted diagnostic tools speed up variant prioritization and clinical interpretation. For example, Face2Gene uses computer vision to analyze craniofacial features and suggest candidate syndromes, complementing exome or genome sequencing.[16] ML algorithms applied to NICU sequencing data enable rapid triage of potentially pathogenic variants, reducing diagnostic delays and improving clinical outcomes.[17]

Practical applications and translational relevance

Several landmark applications illustrate the transformative impact of AI in human genetics, showcasing both research and clinical utility.

DeepVariant in Genome Sequencing

Google's DeepVariant highlights how AI can enhance genome analysis. By treating variant calling as an image classification problem, DeepVariant uses convolutional neural networks to detect single nucleotide variants and small insertions/deletions with higher accuracy than traditional approaches.[18] Its deployment across large-scale sequencing projects has streamlined variant detection, reduced false positives, and enabled more reliable downstream analyses, particularly in rare disease diagnostics.[19]

AlphaFold in Protein Structure Prediction

AlphaFold, developed by DeepMind, represents a major breakthrough in structural genomics.[20] Using deep learning, AlphaFold predicts three-dimensional protein structures directly from amino acid sequences with remarkable accuracy, often approaching experimental resolution. This advance has profound implications for understanding the functional impact of genetic variants, guiding drug development, and interpreting pathogenic mutations in clinical genetics.[21]

Face2Gene in Rare Disease Diagnostics

AI has also made inroads into clinical genetics through diagnostic tools like Face2Gene, which applies deep learning to patient facial photographs to suggest potential syndromes.[22] When combined with exome or genome sequencing, this method accelerates the identification of rare genetic disorders, reducing diagnostic delays and supporting patient management.[23]

AI in Cancer Genomics

In oncology, AI models that integrate

genomic, transcriptomic, and proteomic data have been applied to predict tumor progression, identify driver mutations, and guide therapy selection.[24] For instance, ML-based polygenic risk scores, combined with clinical features, have improved risk stratification for breast, prostate, and colorectal cancers.[25] These models enhance prognostic accuracy and facilitate precision medicine by tailoring interventions to individual genetic profiles.

Multi-Omics Integration for Complex Traits

Integrative AI models, such as MOFA and graph neural networks, have successfully connected genomic, epigenomic, and transcriptomic layers to uncover disease mechanisms.[26] In neurodegenerative disorders, these approaches have revealed novel regulatory interactions and candidate therapeutic targets, demonstrating AI's ability to translate multidimensional genetic data into actionable biological insights.[27] Collectively, these examples show that AI is not only accelerating basic research but also reshaping clinical practice, enabling more precise, data-driven approaches to diagnosis, prognosis, and treatment.

CHALLENGES AND LIMITATIONS

Despite the transformative potential of AI in human genetics, several challenges and limitations must be addressed to ensure reliability, fairness, and clinical applicability.

Data Quality and Heterogeneity

AI models depend on high-quality, well-annotated datasets. However, genomic and multi-omics data often contain sequencing errors, batch effects, missing values, and uneven coverage.[28] Noisy or incomplete datasets can compromise model accuracy and lead to erroneous predictions. Additionally, integrating heterogeneous datasets from different platforms, laboratories, or populations remains challenging, requiring advanced normalization and harmonization strategies.[29]

Model Interpretability

Deep learning models, particularly convolutional and recurrent neural networks, are frequently criticized as "black boxes".[30] Although they often achieve high predictive performance, their lack of transparency can limit trust and adoption in clinical genetics. Clinicians and researchers need to understand how predictions are generated, especially when they influence patient care. Current efforts in explainable AI (XAI), attention mechanisms, and model visualization aim to improve model interpretability and transparency.[31]

Bias and Population Representation

Many AI models are trained on datasets dominated by populations of European ancestry.[32] This can introduce bias, reducing predictive accuracy for underrepresented groups and potentially worsening health disparities. Ensuring diverse representation in training datasets and evaluating model performance across populations are critical steps for equitable application.[33]

Computational Resources and Expertise

Training and deploying AI models, particularly deep learning architectures, requires substantial computational resources, specialized hardware (e.g., GPUs or TPUs), and expertise in both genomics and AI.[34] Smaller research labs or clinical centers may face barriers to adoption due to these resource constraints.

Ethical and Privacy Concerns

Genomic data are inherently sensitive. AI models often need large-scale datasets, raising concerns about privacy and informed consent.[34] Potential risks include unauthorized data sharing, re-identification of individuals, and misuse of predictive information. Regulatory frameworks and secure data-sharing platforms are necessary to protect patient information while enabling AI research.[35] Addressing these limitations is essential to fully realize AI's potential in genetics. Strategies include developing interpretable models, improving data quality and diversity, implementing ethical guidelines, and fostering collaborations between AI experts and geneticists. By overcoming these challenges, AI can move from a research tool to a robust, clinically impactful technology.

FUTURE DIRECTIONS AND CONCLUSION

Artificial intelligence is set to transform human genetics by moving beyond data analysis toward predictive, interpretable, and actionable tools. A key future direction is the development of explainable and interpretable AI, which clarifies the reasoning behind complex predictions. Techniques such as attention mechanisms, feature importance mapping, and interpretable neural architectures can enhance clinician trust, support regulatory approval, and facilitate adoption in diagnostic workflows. Integrating AI with genome editing technologies, including CRISPR-Cas systems, represents another promising avenue. Predictive models can optimize guide RNA design, reduce off-target effects, and forecast the functional outcomes of edited sequences,[38] accelerating therapeutic development for monogenic disorders and targeted gene therapies.

The concept of predictive digital twins—computational representations of an individual's genetic, molecular, and clinical profile—offers the potential for truly personalized medicine.[39] AI can simulate disease progression, treatment response, and lifestyle interventions, enabling proactive, patient-specific healthcare decisions. Furthermore, AI models will increasingly integrate multi-omics and longitudinal datasets, combining genomics, transcriptomics, proteomics, metabolomics, and imaging over time.[40] These integrative approaches can capture dynamic biological processes, reveal causal relationships, and enhance understanding of disease onset, progression, and treatment response, with advanced architectures such as graph neural networks and transformers enabling complex analyses. Ensuring population diversity in genomic datasets is another critical focus. Including diverse ancestries improves model generalizability, reduces bias, and supports equitable application of AI in healthcare.[41] AI can also identify population-specific variants and risk factors, contributing to global genomics initiatives and helping to mitigate health disparities. Ethical and regulatory considerations remain central, requiring transparent guidelines, robust data governance, and frameworks that balance innovation with privacy, consent, and fairness.[42] These measures are essential for the responsible translation of AI into research and clinical practice.

In conclusion, AI has already made substantial contributions in variant calling, pathogenicity prediction, gene regulation, multi-omics integration, pharmacogenomics, and rare disease diagnostics. The combination of AI with expanding genomic datasets, advanced modeling techniques, and clinical insights promises a new era of precision genomics, enabling individualized risk prediction, targeted therapies, and improved patient outcomes. While challenges such as data quality, interpretability, bias, computational demands, and ethical considerations persist, ongoing methodological innovations and collaborative efforts are likely to overcome these barriers. By integrating predictive AI tools with genome editing, longitudinal multi-omics, and personalized digital twins, the field is moving toward a future in which AI not only interprets genetic data but actively informs decision-making, accelerates discovery, and transforms human healthcare.

REFERENCES

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
3. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet*. 2015;16:321–332.
4. Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878.
5. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589.
6. Jaganathan K, *et al.* Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176:535–548.e24.
7. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat Methods*. 2015;12:931–934.
8. Argelaguet R, *et al*. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14:e8124.
9. Kourou K, *et al.* Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
10. Mavaddat N, et al. Polygenic risk scores for prediction of breast cancer and its subtypes. *Am J Hum Genet*. 2019;104:21–34.
11. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multi-omics data

integration methods. *Front Genet*. 2017;8:84.

12. Zitnik M, *et al*. Machine learning for integrating biological networks. *Curr Opin Biotechnol*. 2019;58:17–23.

13. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods*. 2018;15:1053–1058.

14. Lee SI, Celik S, Logsdon BA, *et al.* A machine learning approach to integrate big data for precision medicine in cancer. *Nat Rev Genet*. 2018;19:298–310.

15. Ramsundar B, Kearnes S, Riley P, Webster D, Konerding D, Pande V. Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*. 2015.

16. Gurovich Y, Hanani Y, Bar O, *et al*. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med*. 2019;25:60–64.

17. Clark MM, *et al*. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping. *Sci Transl Med*. 2019;11:eaat6177.

18. Poplin R, *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36:983–987.

19. Luo R, Sedlazeck FJ, Lam TW, Schatz MC. Clairvoyante: a multi-task convolutional deep neural network for variant calling in single molecule sequencing. *Nat Commun*. 2019;10:998.

20. Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589.

21. Tunyasuvunakool K, *et al.* Accurate prediction of protein structures for the human proteome. *Nature*. 2021;596:590–596.

22. Gurovich Y, *et al.* Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med*. 2019;25:60–64.

23. Loos R, *et al.* AI-assisted diagnosis in rare diseases. *NPJ Genom Med*. 2020;5:53.

24. Kourou K, *et al.* Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.

25. Mavaddat N, *et al.* Polygenic risk scores for prediction of breast cancer and its subtypes. *Am J Hum Genet*. 2019;104:21–34.

26. Argelaguet R, *et al.* Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14:e8124.

27. Zitnik M, *et al.* Machine learning for integrating biological networks. *Curr Opin Biotechnol*. 2019;58:17–23.

28. Chen X, *et al.* The challenges of AI in genomics: data quality and heterogeneity. *Nat Rev Genet*. 2020;21:445–458.

29. Huang S, *et al.* More is better: recent progress in multi-omics data integration. *Front Genet*. 2017;8:84.

30. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206–215.

31. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–4774.

32. Martin AR, *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584–591.

33. Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell*. 2019;177:26–31.

34. Angermueller C, *et al.* Deep learning for computational biology. *Mol Syst Biol*. 2016;12:878.

35. Gymrek M, *et al.* Identifying personal genomes by surname inference. *Science*. 2013;339: 321–324.

36. Shabani M, Borry P. Rules for processing genetic data: ethical considerations. *Genome Med*. 2018;10:72.

37. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*. 2017.

38. Chuai G, *et al.* DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol*. 2018;19:80.

39. Schork NJ, Goetz LH. Personalized medicine: from genome to digital twin. *Nat Rev Genet*. 2017;18:291–300.

40. Yuan H, *et al.* Multi-omics data integration for precision medicine. *Front Genet*. 2020;11:131.

41. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161–164.

42. Vayena E, *et al.* Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018;15:e1002689.