AI-Driven Genomic Analysis for Prevention of Recessive Disorders, their Emerging Tools, Challenges, and Future Directions: An Overview

ABSTRACT

Recessive genetic disorders, frequently concealed within heterozygous carriers, pose a considerable challenge in the realm of clinical genetics owing to their asymptomatic characteristics in carriers and the profound consequences when transmitted in a bi-allelic manner. The emergence of next-generation sequencing (NGS) has facilitated the accessibility of extensive genomic data; nonetheless, the intricacies of interpretation continue to present a significant impediment. The fields of artificial intelligence (AI) and machine learning (ML) are now transforming the genomic landscape by facilitating comprehensive analyses of genomic variants, amalgamating phenotype data, and forecasting disease risks with enhanced speed and precision. This review examines the contemporary AI-driven methodologies employed in the prevention of recessive disorders through carrier screening, embryo selection, and extensive population analyses. We reference recent advancements, including AI systems such as PhenIX, X rare, Deep Variant, and prioritization frameworks based on GPT-4. Additionally, we address ethical considerations, challenges pertaining to clinical translation, and the potential of generative AI in the context of genetic counseling. By scrutinizing both the technical evolution and translational significance, this review positions AI as an indispensable instrument in predictive and preventive genomic medicine.

Keywords: Artificial Intelligence; Bioinformatics; Carrier Screening; Clinical Decision Support; Genomic Analysis; Machine Learning; Predictive Genomics; Rare Diseases; Recessive Disorder

1. INTRODUCTION

Recessive genetic disorders represent a significant category of inherited conditions arising from mutations in genes located on autosomes or sex chromosomes. These disorders manifest when an individual inherits two copies of the mutated gene, one from each parent, a condition known as being homozygous for the mutation [1].These disorders arise when an individual inherits two copies of the mutated gene, one from each parent, a condition known as being homozygous for the mutation. In contrast to dominant disorders, where a single copy of the mutated gene is sufficient to cause the condition, recessive disorders require the absence of a functional gene product from both alleles for the phenotype to be expressed [2]. The parents, who each carry one copy of the mutated gene and one normal copy, are typically asymptomatic carriers [3]. This carrier status arises because the presence of one functional allele is usually sufficient to produce enough of the required protein to prevent the disorder's manifestation [4]. However, each child of two carrier parents has a 25% chance of inheriting both mutated genes and thus expressing the disorder, a 50% chance of being an asymptomatic carrier like their parents, and a 25% chance of inheriting two normal genes and being neither affected nor a carrier [5].The inheritance pattern of recessive disorders differs based on whether the mutated gene is located on an autosome or a sex chromosome (Table 1).

Table 1: Common Recessive Genetic Disorders

Disorder Name	Type	Affected Gene(s)	System Affected	Symptoms/Impact
Cystic Fibrosis (CF)	Autosomal Recessive	*CFTR*	Respiratory/Digestive	Thick mucus, lung infections, digestive problems, reduced lifespan
Sickle Cell Anemia	Autosomal Recessive	*HBB*	Blood	Anemia, pain crises, organ damage, stroke
Tay-Sachs Disease	Autosomal Recessive	*HEXA*	Nervous System	Neurodegeneration, blindness, death by age 4–5
Phenylketonuria (PKU)	Autosomal Recessive	*PAH*	Metabolic/Brain	Intellectual disability if untreated, requires strict diet
Thalassemia	Autosomal Recessive	*HBA1, HBA2, HBB*	Blood	Severe anemia, fatigue, bone deformities
Albinism (OCA types)	Autosomal Recessive	*TYR, OCA2, etc.*	Skin/Eyes	Lack of pigment, vision problems, sun sensitivity

Disorder Name	Type	Affected Gene(s)	System Affected	Symptoms/Impact
Duchenne Muscular Dystrophy	X-linked Recessive	*DMD*	Muscular	Progressive muscle weakness, wheelchair use, early death
Hemophilia A/B	X-linked Recessive	*F8*, *F9*	Blood Clotting	Excessive bleeding, joint damage, risk of hemorrhage
Color Blindness (Red-Green)	X-linked Recessive	*OPN1LW*, *OPN1MW*	Vision	Difficulty distinguishing red and green colors
Leigh Syndrome (nuclear)	Mitochondrial Recessive	Various (e.g., *SURF1*)	Brain/Nervous System	Developmental delay, respiratory failure, seizures

Notwithstanding the existence of genomic sequencing technologies, the early identification and prevention of these disorders are still constrained by interpretive intricacies and the integration within clinical workflows [6]. Artificial intelligence (AI) presents a transformative opportunity to mitigate this issue by automating variant interpretation, augmenting data integration, and enhancing diagnostic efficacy. Over the past decade, AI instruments have progressed from rudimentary rule-based systems to sophisticated deep learning networks capable of analyzing intricate genetic and phenotypic data [7-8] . Recent advancements encompass the incorporation of natural language processing (NLP), large language models (LLMs), and reinforcement learning into genomic workflows. These instruments are currently being employed not solely for diagnostic purposes but also for preventive strategies such as preconception carrier screening and IVF embryo selection [9]. This review offers a comprehensive examination of how AI facilitates genomic analysis with a concentration on the prevention of recessive disorders. Authors conducted a critical analysis of its application in variant prioritization, risk prediction, and phenotype-genotype mapping, while also addressing the challenges posed by data scarcity, ethical considerations, and clinical validation. Furthermore, we emphasize emerging models and case studies from contemporary literature to delineate a roadmap for prospective advancements in this interdisciplinary field [10-11]. As AI technologies continue to evolve, their incorporation into genomic workflows will require sustained collaboration between professionals in computer science and healthcare to ensure an effective and ethical implementation [12]. This collaboration will be crucial in addressing the complexities inherent to data interpretation and in maintaining ethical standards in the application of AI within the realm of genomic medicine. The successful integration of AI into genomic workflows could markedly enhance the precision of diagnosing recessive genetic disorders and ultimately lead to improved patient outcomes [13-14]. In addition, the use of AI-driven platforms can significantly aid in the identification of genetic variants linked to recessive disorders, thereby informing targeted interventions and preventive measures. This approach not only enhances diagnostic accuracy but also supports the development of personalized medicine strategies for individuals at risk of inheriting these disorders [15]. Moreover, as AI technologies advance, they may also facilitate the integration of genetic information with environmental factors, enhancing the precision of risk assessments for recessive disorders.

2. FOUNDATIONS OF GENPMIC AANALYSIS IN RECESSIVE DISORDERS

2.1 Genetic Architecture of Recessive Inheritance

Recessive disorders manifest when both alleles of a particular gene contain pathogenic variants, which typically leads to the loss or modification of the gene's functional capacity. These disorders can be classified as either autosomal or X-linked and may present a diverse array of clinical phenotypes, ranging from mild metabolic disturbances to severe neurodevelopmental disorders [16].The challenge in managing these conditions arises from the fact that heterozygous carriers are generally asymptomatic and, as a result, remain oblivious to their risk of transmitting the disorder. The advent of next-generation sequences (NGS) technologies, such as whole-exome sequencing (WES) and whole-genome sequencing (WGS), has enabled the identification of single-nucleotide variants (SNVs), insertions/deletions (indels), and structural variants with exceptional precision. Nevertheless, the clinical obstacle is not found in the sequencing process itself but rather in the interpretation of variants — specifically, the determination of which among the millions of variants are genuinely pathogenic and causative [17-18].

2.2 Genomic Pipelines and Carrier Screening

A typical genomic analysis pipeline includes:

(1) Variant Calling (GATK, DeepVariant)

The first step in genomic analysis involves identifying variants such as single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) from raw sequencing data. Two commonly used tools for this task are **GATK** (Genome Analysis Toolkit) and **DeepVariant**.

GATK, developed by the Broad Institute, is widely regarded as the gold standard for variant calling in clinical and research settings. Its **Haplotype Caller** module reconstructs haplotypes through local re-assembly and accurately identifies variants [19-20]. Alternatively, **Deep Variant**, developed by Google, uses deep learning to improve the accuracy of variant calls by transforming aligned sequencing reads into images and applying a convolutional neural network [21]. Both tools are capable of handling whole-exome and whole-genome data, offering high sensitivity and specificity.

(2) Annotation (ANNOVAR, VEP)

Once variants are identified, the next step is to annotate them with biological and clinical information. **ANNOVAR** and **VEP (Variant Effect Predictor)** are two powerful tools used for this purpose. ANNOVAR is a command-line tool that provides extensive annotation capabilities using multiple databases, such as RefSeq, ClinVar, dbSNP, and gnomAD [22]. It supports gene-based, region-based, and filter-based annotations and can help interpret the biological significance of variants [23]. On the other hand, **VEP**, developed by Ensembl, annotates variants with information about their potential effects on genes and proteins. It also integrates data from Ensembl's comprehensive database to predict variant consequences, including synonymous, missense, and nonsense mutations [24]. Both tools help researchers prioritize variants that may be functionally or clinically relevant.

(3) Pathogenicity Scoring (SIFT, PolyPhen-2)

After annotation, variants are often assessed for their likely pathogenicity using computational prediction tools. Two of the most established tools are **SIFT** and **PolyPhen-2**. **SIFT (Sorting Intolerant From Tolerant)** predicts whether an amino acid substitution affects protein function based on sequence homology and the degree of conservation of amino acid residues across species [25]. Variants are classified as "Tolerated" or "Deleterious [26]. **PolyPhen-2 (Polymorphism Phenotyping v2)**, on the other hand, evaluates the potential impact of a mutation on the structure and function of a protein using features such as sequence conservation and structural modeling. Its output includes scores that classify variants as "Benign,""Possibly Damaging," or "Probably Damaging" [27]. These tools are essential for prioritizing variants likely to be involved in disease.

To narrow down potential disease-causing variants, population frequency filtering is commonly performed using databases like **gnomAD (Genome Aggregation Database)**. gnomAD aggregates exome and genome sequencing data from over 140,000 individuals and provides allele frequency data across diverse populations [28-29]. Variants that are too common in the general population are unlikely to be highly penetrant causes of rare diseases and are typically filtered out. This step is critical to reduce false positives and focus on rare, potentially pathogenic variants that warrant further investigation.

(4) Phenotype Correlation (HPO, OMIM)

The final step in the pipeline often involves correlating the patient's phenotype with potential genetic findings. Tools like the **Human Phenotype Ontology (HPO)** and **Online Mendelian Inheritance in Man (OMIM)** play a central role in this phase . **HPO** provides a standardized vocabulary for phenotypic abnormalities, which helps in computationally matching patient symptoms to known genetic diseases [30]. **OMIM** is a curated database of human genes and genetic phenotypes, providing comprehensive information on gene-disease relationships [31]. Together, these tools facilitate the identification of variants that are not only biologically plausible but also clinically relevant, making them invaluable in diagnostic genomics.For carrier screening, especially in the context of pre-marital, pre-conception, or population-wide programs, the process must be efficient, scalable, and accurate. Traditional bioinformatics tools often generate hundreds of variants of uncertain significance (VUS), making manual review burdensome. This is where AI steps in as a transformative layer, automating variant classification, integrating phenotype data, and even learning novel genotype–phenotype relationships from heterogeneous datasets [32].

3. ROLE OF ARTIFICIAL INHERITENCE IN GENOMIC ANALYSIS

3.1 Machine Learning and Deep Learning Architectures

AI in genomics relies predominantly on: (i) Supervised learning, where algorithms are trained on labeled genomic data to classify or predict variant pathogenicity; (ii) Unsupervised learning, which identifies hidden patterns or clusters (e.g., subtypes of disorders or novel associations); (iii) Deep learning (DL), a powerful subset of ML, excels in high-dimensional data such as genomic sequences and clinical images. For instance, deep neural networks have been used to distinguish pathogenic from benign variants by learning complex sequence features that traditional tools overlook. Models like DeepVariant leverage convolutional neural networks (CNNs) to improve variant calling accuracy over standard pipelines [34]

3.2 AI Tools for Variant Prioritization

Several tools have been developed specifically to assist in variant analysis as: (i) PhenIX combines phenotypic descriptors from the Human Phenotype Ontology (HPO) with variant data to prioritize Mendelian disease candidates [35]; (ii) Xrare enhances this by integrating clinical similarity scoring and population frequency data, making it effective for recessive condition diagnosis [36]; (iii) Dx29, an open-access AI tool, uses both genotypic and phenotypic input to produce ranked differential diagnoses. It has demonstrated >90% success in placing the correct diagnosis within the top 10 [37]. Recently, generative AI models and large language models (LLMs) like GPT-4 have revealed surprising capabilities in gene prioritization when provided with phenotypic descriptions. However, current accuracy is still below traditional bioinformatics tools, necessitating hybrid AI–human-in-the-loop systems

3.3 Data Challenges and AI Solutions

One major limitation in AI for rare or recessive diseases is data scarcity. With limited training data, especially for ultra-rare variants, model generalization becomes difficult. Several workarounds have been proposed as (a) **Transfer learning:** Pre-training on large genomic datasets and fine-tuning on smaller RD cohorts [38]; (b) **Data augmentation:** Generating synthetic variants or simulated genomes to train models on diverse data [39]; (c) **Bayesian networks**: Useful in recognizing comorbidity patterns and subtypes from sparse datasets [40]. Table 2 reveals the comparison of major AI tools (PhenIX, Xrare, DeepVariant, Dx29, GPT-4) based on inputs, learning approach, output, accuracy, and limitations.

Table 2: Comparison of Major AI Tools (PhenIX, Xrare, DeepVariant, Dx29, GPT-4).

Tool	Inputs	Learning Approach	Output	Accuracy/Performance	Limitations
PhenIX	Phenotypic terms (HPO),	Semantic similarity + ML	Ranked list of candidate genes	High accuracy in Mendelian disorders	Limited by phenotype specificity

Tool	Inputs	Learning Approach	Output	Accuracy/Performance	Limitations
	genetic data				
Xrare	Genomic variants, phenotype (HPO terms)	Bayesian + Deep Learning	Disease-causing variant prioritization	Strong accuracy in rare diseases	Requires curated phenotype input
DeepVariant	Raw sequencing reads (FASTQ)	Deep Neural Networks (CNN)	Highly accurate variant calls (VCF)	State-of-the-art SNP/indel detection	Computationally intensive, not diagnostic
Dx29	Phenotypic data, clinical notes	NLP + ML (proprietary)	Probable diagnoses, gene suggestions	Good with diverse case data	Dependent on EHR data quality
GPT-4	Free-text queries, structured data	Transformer-based LLM	General clinical insights, summarization, suggestions	Flexible, supports multi-domain tasks	Not purpose-built for genomics; hallucination risk

4. RECENT TRENDS IN AI TOOLS FOR VARIENT DETECTION AND PRIORITIZATION

4.1 DeepVariant and Next-Generation Variant Calling

Developed by Google, DeepVariant is a deep learning-based variant caller that transforms NGS read data into images and uses convolutional neural networks (CNNs) to call variants. Unlike traditional heuristic-based methods, DeepVariant learns the underlying distribution of sequencing errors and biological variants directly from the data [41]. It achieves remarkable accuracy, especially in calling indels and complex substitutions, making it a valuable first step in AI-enabled genomic pipelines.

4.2 PhenIX and Xrare: Phenotype-Aware Tools

1. PhenIX (Phenotypic Interpretation of eXomes) maps HPO terms with variant-level data to prioritize genes in Mendelian diseases. It was among the first to incorporate structured phenotypic information in a computationally meaningful way [42].

2. Xrare, a more recent advancement, goes further by combining clinical phenotype similarity scores with variant annotations, population frequency databases, and ACMG-based guidelines to filter and rank likely pathogenic variants for recessive disorders [43].

4.3 Dx29 and LLM-based Systems

Dx29 is a diagnostic platform that takes both genomic and clinical inputs to produce a ranked list of differential diagnoses. In a 2023–24 evaluation, it ranked the correct diagnosis in the top 10 in over 92% of rare disease cases, including several autosomal recessive conditions [44]. It is designed to assist clinicians rather than replace them, offering real-time interpretability. Additionally, large language models (LLMs) like GPT-4 have been tested for gene prioritization. While still in development, studies show that when provided with HPO-based phenotype inputs, LLMs can suggest relevant disease genes, although their precision remains inferior to established tools [45].

4.4 Explainability and Trust in Clinical Contexts

AI explainability remains a key barrier to clinical adoption. Black-box models, while powerful, lack transparency needed in medical genetics. Tools like SHAP (SHapley Additive exPlanations) are increasingly integrated into AI workflows to make predictions interpretable — critical when decisions may lead to reproductive or prenatal interventions[46].

5. APPLICATIONS IN CARRIER SCREEING AND PREVENTIVE GENOMICS

5.1 AI-Driven Pre-Conception Screening

AI enables scalable, rapid analysis of whole-exome or whole-genome data from healthy individuals to detect heterozygous carriers. A study by Khan et al. (2024) demonstrated how AI-assisted pipelines successfully analyzed a healthy couple's exome data to predict possible recessive inheritance risks. This approach significantly reduces variant review burden and improves detection of actionable findings.

5.2 Clinical Use Case:

An AI tool filtered through 30,000 variants per individual, narrowed potential pathogenic matches to <10 candidate pairs within 3 minutes, showing its potential in real-time reproductive counseling[47].

5.3 Embryo Selection in IVF

AI can also be applied to preimplantation genetic testing (PGT) by ranking embryos based on their carrier status for known recessive conditions. Machine learning models trained on parental and embryonic sequence data can detect compound heterozygosity and filter embryos at risk before implantation [48].

5.4 Population-Wide Genomic Screening Programs

Pilot programs in countries like the UAE, Qatar, and Israel have integrated AI to run large-scale population screening campaigns. AI-assisted platforms help manage bioinformatics workflows, report turnaround, and prioritize follow-up cases — essential features for public health scale screening for disorders like beta-thalassemia, Tay-Sachs, and cystic fibrosis.

6. INTEGRATION OF AI WITH CLINICAL DECISION SUPPORT SYSTEMS (CDSS)

6.1 Decision Support for Genomic Counseling

Clinical Decision Support Systems (CDSS) powered by AI are of prime significance in converting genomic information into actionable information, particularly for non-genetics professionals like physicians. CDSS make use of variant interpretation algorithms, databases like ClinVar, and phenotype model tools to produce ranked diagnoses or clinical suggestions. Tools like Dx29 and Phenoxome not only rank genes but also suggest potential diagnoses consonant with clinical presentation, thus aiding diagnosis as well as prevention [49]. CDSS are capable of identifying pathogenic variant pairs of partners and predicting offspring genotypes in carrier screening based on Mendelian inheritance models. Software is also being created to offer genetic compatibility reports with decision trees that lead couples to reproductive choices, like in vitro fertilization with preimplantation genetic diagnosis or adoption.

6.2 Genetic Counseling Improved with Generative AI

In 2025, a study examined the use of generative AI models like GPT-4 for patient education on an automatic scale in genetic counseling [50].These AI models are not a replacement for a trained genetic counselor but can provide explainable descriptions, carrier risk summaries, and answer patient queries in an interactive manner. This is especially useful for regions where there is a shortage of genetics specialists [50].

6.3 Real-Time Monitoring and Adaptive Risk Prediction

Artificial Intelligence also enables longitudinal patient care by using dynamic models of risk. For example, Bayesian networks can be trained on follow-up data and update recurrence risk estimates based on changing clinical data. This adaptive learning is important for conditions with variable expressivity or incomplete penetrance, which are characteristically seen in recessive conditions [51].

7. ETHICAL, LEGAL AND SOCIAL Issues (ELSI)

7.1 Data Privacy and Security

AI algorithms in genomics handle sensitive individual health information. Genomic privacy issues, especially in population-level screening, are of prime importance. Laws like GDPR in the EU and HIPAA in the United States offer guidelines, but the nature of AI models complicates complete anonymization and auditability.Secure multiparty computation (SMPC) and federated learning are becoming new hope technologies for safeguarding individual-level information while allowing for collaborative AI model training across institutions [51]

7.2 Algorithmic Equity and Bias

These machine learning models are typically trained on Western-predominant genomic data, such as gnomAD, and thus perform suboptimally on underrepresented populations. This may lead to pathogenic misclassification or incorrect carrier risk prediction in African, Asian, or Indigenous populations [51]. Attempts to diversify genomic data and fairness-aware machine learning models are underway but need to be accelerated to avoid exacerbating health disparities.

7.3 Informed Consent and Clinical Transparency

A particular challenge to AI-based genomics is the transparency of algorithms. Black-box models make it difficult for patients to understand how their risk calculations were derived, thus increasing ethical issues regarding informed consent. While methods like SHAP, LIME, and rule-based hybrid models hold the promise to increase transparency, they have not yet become integral parts of clinical reporting pipelines [49-51]. Table 3 summarizes ELSI concerns in AI-genomic screening.

Table 3: ELSI Concerns in AI-genomic Screening.

Ethical Issue	Clinical Risk	Proposed Mitigation
Data Privacy	Unauthorized access or misuse of sensitive genomic information	Use secure data encryption, strict access controls, and transparent data governance
Bias	Inaccurate results due to underrepresentation of certain populations in training data	Diversify genomic datasets and regularly audit algorithms for demographic fairness
Explainability	Difficulty understanding AI-generated results may hinder clinical decision-making	Develop interpretable AI models and provide clinician-friendly summaries
Informed Consent	Patients may not fully grasp the implications of AI use in genomic screening	Create clear, accessible consent materials and involve genetic counselors in the process

8. CONCLUSION AND FUTURE PERSPECTIVES

Artificial intelligence is reshaping the landscape of genomic medicine, particularly in the early identification and prevention of recessive genetic disorders. From advanced variant callers like DeepVariant to phenotype-integrated platforms like Xrare and Dx29, AI systems are enabling faster, more accurate interpretation of genomic data and facilitating personalized reproductive decisions. With the incorporation of large language models and dynamic Bayesian networks, AI is expanding from static analysis tools to interactive clinical assistants. However, challenges remain. The scarcity of rare disease data, algorithmic biases, privacy concerns, and a lack of clinical explain ability continue to slow full-scale adoption. The ethical landscape, particularly regarding AI transparency and equity, must evolve in tandem with technical innovation. Regulatory standards and rigorous validation pipelines are needed to translate promising AI tools into clinical-grade platforms. Looking forward, the integration of multi-omics data, generative models, and causal AI frameworks promises a new era of precision prevention. AI will not merely diagnose genetic disorders but help preempt them, empowering patients and clinicians with foresight rather than hindsight. Taken together the earlier findings and relevant hypotheses [52-54], authors emphatically state, "As the field matures, interdisciplinary collaboration across genetics, computer science, and bioethics will be the key to building AI systems that are not just smart, but accountable and comprehensive".

References

1. Duong D, Solomon BD. Artificial intelligence in clinical genetics. European Journal of Human Genetics 2025; 33(3): 281–288.
2. van Dijk PJ, Ellis TH. The full breadth of Mendel's genetics. Genetics 2016; 204(4):1327-1336.
3. Dias R, Torkamani A. (2019). Artificial intelligence in clinical and genomic diagnostics. Genome Medicine 2019; 11(1): 70.
4. Khan SS, Norouzi M, Tahir M. (2025). Artificial intelligence and deep learning algorithms for epigenetic sequence analysis: A review for epigeneticists and AI experts. Computers in Biology and Medicine 2025; 183: 109302.
5. Loomans JI, Lock J, Peters M, Leebeek FW, Cnossen MH, Fijnvandraat K. Hemofilie [Haemophilia]. Ned Tijdschr Geneeskd. 2014;158: A7357.
6. Park J, Render PharmD KP, Cates PharmD DW. Daridorexant: Comprehensive review of a new oral agent for the treatment of insomnia. Ann Pharmacother. 2023; 57(9):1076-1087.
7. Dauer W, Przedborski S. Parkinson's disease: Mechanisms and models. Neuron. 2003; 39(6):889-909.
8. Onciul R, Tataru C-I, Dumitru AV, Crivoi C, Serban M, Covache-Busuioc R-A, Radoi MP, Toader C. Artificial intelligence and neuroscience: Transformative synergies in brain Research and clinical applications. J. Clin. Med. 2025; 14: 550.
9. Walsh RR, Krismer F, Galpern WR, Wenning GK, Low PA, Halliday G, Koroshetz WJ, Holton J, Quinn NP, Rascol O, Shaw LM, Eidelberg D, Bower P, Cummings JL, Abler V, Biedenharn J, Bitan G, Brooks DJ, Brundin P, Fernandez H, Fortier P, Freeman R, Gasser T, Hewitt A, Höglinger GU, Huentelman MJ, Jensen PH, Jeromin A, Kang UJ, Kaufmann H, Kellerman L, Khurana V, Klockgether T, Kim WS, Langer C, LeWitt P, Masliah E, Meissner W, Melki R, Ostrowitzki S, Piantadosi S, Poewe W, Robertson D, Roemer C, Schenk D, Schlossmacher M, Schmahmann JD, Seppi K, Shih L, Siderowf A, Stebbins GT, Stefanova N, Tsuji S, Sutton S, Zhang J. Recommendations of the global multiple system atrophy research roadmap meeting. Neurology 2018; 90(2):74-82.

10. Coghlan S, Gyngell C, Vears DF. Ethics of artificial intelligence in prenatal and pediatric genomic medicine. J Community Genet. 2024; 15(1):13-24.
11. Sharma J, Goel P. The use of AI for phenotype-genotype mapping. Methods Mol Biol. 2025; 2952:369-410.
12. Ray S. Human evolutionary genetics. Yale J Biol Med. 2014; 87(4):603.
13. Carini C, Seyhan AA. Tribulations and future opportunities for artificial intelligence in precision medicine. J Transl Med. 2024; 22(1):411.
14. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. Med Oncol. 2022; 39(8):120.
15. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, Zhao J, Snowdon JL. Precision medicine, AI, and the future of personalized health care. Clin Transl Sci. 2021; 14(1):86-93.
16. Isrie M, Froyen G, Devriendt K, de Ravel T, Fryns JP, Vermeesch JR, Van Esch H. Sporadic male patients with intellectual disability: contribution of X-chromosome copy number variants. Eur J Med Genet. 2012; 55(11):577-585.
17. Azimi M, Schmaus K, Greger V, Neitzel D, Rochelle R, Dinh T. Carrier screening by next-generation sequencing: health benefits and cost effectiveness. Mol Genet Genomic Med. 2016; 4(3):292-302.
18. Yang M, Kim JW. Principles of genetic counseling in the era of next-generation sequencing. Ann Lab Med. 2018; 38(4):291-295.
19. Zhou Y, Kathiresan N, Yu Z, Rivera LF, Yang Y, Thimma M, Manickam K, Chebotarov D, Mauleon R, Chougule K, Wei S, Gao T, Green CD, Zuccolo A, Xie W, Ware D, Zhang J, McNally KL, Wing RA. A high-performance computational workflow to accelerate GATK SNP detection across a 25-genome dataset. BMC Biol. 2024; 22(1):13.
20. Nussbaum RL, McInnes RR, Willard HF. (2021). Thompson & Thompson Genetics in Medicine 2021; 9th ed., Elsevier.
21. Turnpenny PD, Ellard S. (2016). Emery's Elements of Medical Genetics 2-16; 15th ed., Elsevier.
22. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015; 10(10):1556-66.
23. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012; 40 (Web Server issue):W452-457.
24. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. (2023). Artificial intelligence in rare disease diagnosis and treatment. Clinical and Translational Science 2023; 16(10), 2106–2111.
25. Tahir M, Norouzi M, Khan SS. Evaluating the use of generative artificial intelligence to support genetic counseling for rare diseases. Diagnostics 2025; 15(6), 672.

26. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M. Rehm HL, MacArthur DG, O'Donnell-Luria A. Genome aggregation database consortium; Variant interpretation using population databases. Hum Mutat. 2022; 43(8):1012-1030.
27. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, Danis D, Balagura G, Baynam G, Brower AM, Callahan TJ, Chute CG, Est JL, Galer PD, Ganesan S, Griese M, Haimel M, Pazmandi J, Hanauer M, Harris NL, Hartnett MJ, Hastreiter M, Hauck F, He Y, Jeske T, Kearney H, Kindle G, Klein C, Knoflach K, Krause R, Lagorce D, McMurry JA, Miller JA, Munoz-Torres MC, Peters RL, Rapp CK, Rath AM, Rind SA, Rosenberg AZ, Segal MM, Smedley D, Talmy T, Thomas Y, Wiafe SA, Xian J, Yüksel Z, Helbig I, Mungall CJ, Haendel MA, Robinson PN. The human phenotype ontology in 2021. Nucleic Acids Res. 2021; 49(D1):D1207-D1217.
28. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, Gross SS, Dorfman L, McLean CY, DePristo MA. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018; 36(10):983-987.
29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. Genome Biol. 2016; 17(1):122.
30. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31(13):3812-3814.
31. Chen Q, Quan L, Cao L, Zhang B, Zhang Z, Peng L, Wang J, Jiang Y, Nie L, Li G, Wu T, Lyu Q. DS-MVP: identifying disease-specific pathogenicity of missense variants by pre-training representation. Brief Bioinform. 2025; 26(2):bbaf119.
32. Kelly C, Szabo A, Pontikos N, Arno G, Robinson PN, Jacobsen JOB, Smedley D, Cipriani V. Phenotype-aware prioritisation of rare Mendelian disease variants. Trends Genet. 2022; 38(12):1271-1283.
33. Li Q, Zhao K, Bustamante CD, Ma X, Wong WH. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. Genet Med. 2019; 21(9):2126-2134.
34. Reiley J, Botas P, Miller CE, Zhao J, Malone Jenkins S, Best H, Grubb PH, Mao R, Isla J, Brunelli L. Open-source artificial intelligence system supports diagnosis of Mendelian diseases in acutely Ill infants. Children (Basel). 2023; 10(6):991.
35. Tejani AS, Ng YS, Xi Y, Fielding JR, Browning TG, Rayan JC. Performance of multiple pretrained BERT models to automate and accelerate data annotation for large datasets. Radiol Artif Intell. 2022; 4(4):e220007.
36. Mendes JM, Barbar A, Refaie M. Synthetic data generation: a privacy-preserving approach to accelerate rare disease research. Front Digit Health. 2025; 7:1563991.
37. Tsang G, Zhou SM, Xie X. Modeling large sparse data for feature selection: Hospital admission predictions of the dementia patients using primary care electronic health records. IEEE J Transl Eng Health Med. 2020; 9:3000113.

38. Ramachandran A, Lumetta SS, Klee EW, Chen D. HELLO: improved neural network architectures and methodologies for small variant calling. BMC Bioinformatics. 2021; 22(1):404.
39. Wu C, Devkota B, Evans P, Zhao X, Baker SW, Niazi R, Cao K, Gonzalez MA, Jayaraman P, Conlin LK, Krock BL, Deardorff MA, Spinner NB, Krantz ID, Santani AB, Tayoun ANA, Sarmady M. Rapid and accurate interpretation of clinical exomes using Phenoxome: a computational phenotype-driven approach. Eur J Hum Genet. 2019; 27(4):612-620.
40. Harrison SM, Biesecker LG, Rehm HL. Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. Curr Protoc Hum Genet. 2019; 103(1):e93.
41. Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: A systematic review. PLOS Digit Health 2025; 4(6):e0000877.
42. Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. Front Robot AI. 2024; 11:1444763.
43. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248-249.
44. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME; Genome Aggregation Database Consortium; Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581(7809):434-443.
45. Elhaddad M, Hamam S. AI-driven clinical decision support systems: An ongoing pursuit of potential. Cureus. 2024; 16(4):e57728.
46. Jeon S, Lee SA, Chung HS, Yun JY, Park EA, So MK, Huh J. Evaluating the use of generative artificial intelligence to support genetic counseling for rare diseases. Diagnostics (Basel). 2025; 15(6):672.
47. Coen E, Del Fiol G, Kaphingst KA, Borsato E, Shannon J, Smith H, Masino A, Allen CG. Chatbot for the return of positive genetic screening results for hereditary cancer syndromes: Prompt Engineering Project. JMIR Cancer 2025; 11:e65848.
48. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Humaniz Comput. 2023;14(7):8459-8486.

49. Maleki VS, Forouzanfar M. The role of AI in hospitals and clinics: Transforming healthcare in the 21st century. Bioengineering (Basel). 2024; 11(4):337.
50. Bonet D, Levin M, Montserrat DM, Ioannidis AG. Machine learning strategies for improved phenotype prediction in underrepresented populations. Pac Symp Biocomput. 2024; 29:404-418.
51. Boudi AL, Boudi M, Chan C, Boudi FB. Ethical challenges of artificial intelligence in medicine. Cureus. 2024; 16(11):e74495.
52. Bajpai D, Mishra MK, Srivastava S, Tiwari AM, Mishra MK, Gupta P, Singh B, Mishra S. (2025). Advancements and challenges of artificial intelligence in healthcare and medicine: An overview. Asian Journal of Current Research 10(1)**:**145-160**.**
53. Chaurasia N, Yadav D, Bajpai D, Saxena R, Mishra MK, Singh B, Gupta P, Mishra, S. Integrating Artificial Intelligence in Medical Robotics: Current Applications, Challenges, and Future Vistas. Asian Journal of Current Research 2025; 10 (3):108-123.
54. Bajpai D, Mishra MK, Singh B, Siddiqui KM, Singh P, Ahmad M, Gupta P, Singh S, Mishra S. AI driven biotechnology for climate resilient agriculture, healthcare and food system. i-manager's Journal of Life Sciences 2025; 4(3): 38-49.