



Role of Artificial Intelligence in Genomics

Maham Taqi¹, Sumera Younus¹, Arleen Yousaf^{2}*

¹ Institute of Molecular Biology, Liaquat University of Medical and Health Sciences (LUMHS), Jamshoro 76080, Sindh, Pakistan

¹ Institute of Molecular Biology, Liaquat University of Medical and Health Sciences (LUMHS), Jamshoro 76080, Sindh, Pakistan

² Institute of Biotechnology and Genetic Engineering, University of Sindh, Jamshoro 76080, Sindh, Pakistan

E-mail: arleenyousuf12@gmail.com

ABSTRACT

In biological research and clinical settings, a significant convergence between Artificial Intelligence (AI) and genomics is transforming the landscape. Genomics, which involves the comprehensive study of genomes, has been pivotal in understanding genetic diversity and the foundations of health and disease. The fast-paced progress in genomic sequencing technologies has generated vast datasets, necessitating advanced analytical tools to interpret this information effectively. The narrative starts by providing a foundational overview of genomics, touching on genetic variation, inheritance, and the evolution of sequencing methods. It then shifts to AI, explaining key concepts of machine learning and deep learning—crucial for analyzing high-dimensional genomic data. The discussion illustrates AI's transformative applications in genomics, including variant calling, gene expression profiling, and functional annotation, thus advancing precision medicine by tailoring treatment strategies to individual genetic profiles. It also addresses challenges linked to the integration of AI in genomics, such as data privacy, information security, and the potential biases in AI algorithms that may affect fairness in clinical genomics. Emerging trends, including the integration of multi-omics data, are also examined, underscoring AI's role in synthesizing genomic, proteomic, and metabolomic data for a comprehensive understanding of biological networks. Additionally, the development of explainable AI models is highlighted, promoting transparency in genomic research. This exploration underscores AI's transformative potential in genomics, fostering greater insights into complex biological systems and supporting a precision medicine framework in healthcare.

Key Words: Artificial Intelligence (AI), Genomics, Genetic variation, Inheritance models, Genetic diversity, Sequencing technologies, Genomic datasets, Machine learning, Deep learning, High-dimensional data, Variant calling, Gene expression profiling, Functional annotation, Precision medicine, Data privacy, AI algorithm biases.

1. Introduction to Genomics and Artificial Intelligence

In the realm of biological sciences and medicine, genomics stands as a groundbreaking field dedicated to the exploration of genomes, which encompass the entire collection of DNA within an organism[1]. This discipline has fundamentally altered the way researchers understand genetic structures, functions, evolutionary patterns, and the intricate connections between genes, as well as their impact on health and disease. Genomic research entails collecting extensive DNA sequence data. The advent of next-generation sequencing technologies has dramatically changed the landscape of genomic data acquisition, making it possible to swiftly generate large volumes of sequencing information at significantly reduced costs. Initially, the emphasis of genomic studies was primarily on the act of sequencing DNA[1]. However, as the demand for sequencing genomic data escalated, it became apparent that there was a pressing requirement for the creation of bioinformatics algorithms, methodologies, and software tools essential for storing, managing, and analyzing this wealth of information[2]. Analyzing genomic data is pivotal in translational research, bridging the gap between laboratory-based biological findings and their practical applications in clinical settings[3]. In a world increasingly shaped by technology, artificial intelligence (AI) stands out as a field that closely resembles human thought processes. This innovative technology is proving invaluable in advancing genomic research, opening new avenues for exploration and understanding[4]. As AI makes significant strides into various sectors—ranging from higher education to industry—its influence is unmistakable. In the realm of academic research, the latest large language models (LLMs) and other AI technologies are not only reshaping the way scientists conduct their studies but also how they share their findings with the broader community[5]. The surge in accessible genomic data repositories has created a fertile ground for AI, enabling researchers to extract meaningful insights from complex genomic datasets. At its core, AI includes machine learning (ML) techniques and computational algorithms that analyze existing data, discern underlying patterns, and apply this knowledge to predict outcomes for new datasets[6]. In genomics, DNA sequence data are quantitatively measured and encoded in digital formats, making them suitable for AI application. AI has already made significant advancements in genomics, including but not limited to gene discovery, genome assemblage, genomic annotation, genomic sequence alignment, characterizing genomic variations, and predicting the impact of genomic mutations in Plants and Humans.

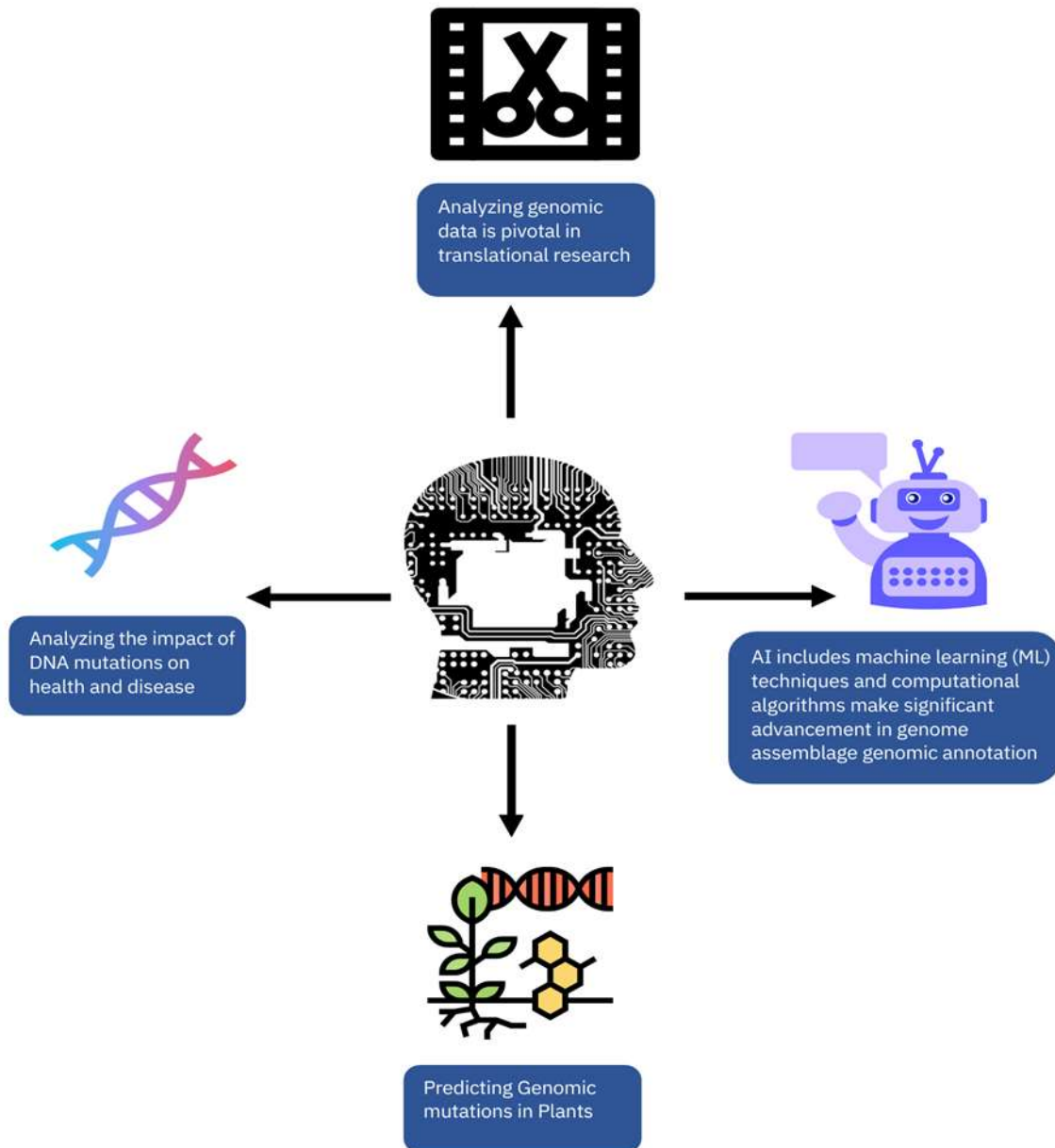


Figure 1 Role of Artificial Intelligence in Genomics

2. Fundamentals of Genomics

Once upon a time, there was a captivating field of science known as genomics, which delved into the intricate world of an organism's entire DNA, or genome. This remarkable genome contained every piece of genetic information necessary for determining the structure and function of the organism. Moreover, it was a treasure trove of knowledge that would be passed on to future generations through the miracle of reproduction[7]. In the realm of genomic structure, the core components include genes, chromosomes, and DNA sequences. Genes are DNA segments that encode proteins essential for biochemical reactions and cellular support. Chromosomes, long DNA strands containing numerous genes, organize these genes with proteins. Different species have distinct chromosome counts; humans have 23 pairs while fruit flies have 4. DNA sequences consist of nucleotides—fundamental DNA units identified by A, T, C, and G[8]. The sequence of these nucleotides encodes vital information for an organism's growth. Variations in DNA sequences lead to alleles, which contribute to the diverse traits in individuals and influence disease susceptibility. Genomic sequencing is a quest to decode the nucleotide order in DNA, revealing genetic information[9]. It began in 1975 with a technique to amplify a DNA fragment chemically, producing a mixture of fragments analyzed through gel electrophoresis. Despite its throughput limitations, this method was crucial for sequencing the first human genome in 2001, which involved approximately 3 billion base pairs. To understand genomic evolution across species, researchers needed extensive genomic data, leading to the development of high-throughput next-generation sequencing (NGS) technologies, which streamlined data generation at a lower cost and faster pace. NGS comprises two stages: library preparation, where DNA is fragmented and adapters added, followed by parallel sequencing of the

fragments. In 2005, 454 pyrosequencing marked a pivotal moment, prompting the rise of various NGS technologies based on different sequencing principles. Currently, over 30 commercial NGS machines exist, with Illumina being the most popular. The field has advanced with long-read sequencing technologies enhancing the reconstruction of complex genomic areas and enabling complete genome assembly. The synergy of NGS technologies and bioinformatics for data analysis has revolutionized genomic research, shedding light on the structures, functions, and evolutionary histories of genomes across diverse species[6].

2.1. Genetic Variation and Inheritance

Over the past five decades, genomic sequencing technologies have revolutionized our understanding of genomes. Various methods have been developed, showcasing significant advancements[10]. This exploration focuses on the evolution of sequencing technologies from a bioanalytical perspective, examining prevalent techniques, their principles, benefits, limitations, and their impact on genomic research[11]. The discussion emphasizes the need for greater accessibility and efficiency in these methods. Sanger sequencing, launched in the 1970s, marked the first generation of sequencing, known for its versatility and accuracy, playing a crucial role in the Human Genome Project. However, its low throughput and high costs hindered smaller labs from participating in genomics. The landscape transformed in 2005 with second-generation sequencing methods, which greatly increased throughput and reduced costs, enabling previously unattainable large-scale projects. Since 2015, third-generation sequencing has emerged, allowing direct sequencing of long DNA strands and enhancing resolution, facilitating the assembly of complex genomes and exploring repetitive areas, structural variations, and haplotypes[12]. Each sequencing technology has unique advantages and drawbacks, influencing the quality of data and subsequent analyses. Researchers in genomic studies and AI integration into genomics benefit from understanding these technologies. The significance of these methods in genomics and metagenomics has led to a rising demand for DNA sequencing applications in biomedical and environmental fields[13]. As sequencing technologies advance and costs decrease, genomic research is becoming standard practice in many laboratories. In public health, agriculture, and biomedicine, genomic studies have made transformative strides, particularly in precision and personalized medicine. Sequencing technologies are vital bioanalytical tools that reveal insights from the nucleotide bases of genomes. The past five decades have seen notable progress in these technologies, driving discussions on their principles, advantages, limitations, and essential considerations for their design and application[13].

2.2. Genomic Sequencing Technologies

Over the last fifty years, genomic sequencing technologies have significantly advanced our understanding of genomes. A variety of sequencing methods have been developed, showcasing substantial progress. This exploration focuses on the evolution of these technologies, particularly from a bioanalytical perspective, discussing the most common sequencing methods, their principles, advantages, disadvantages, and their impact on genomic research[14]. There is a strong emphasis on improving accessibility and efficiency. The 1970s saw the introduction of Sanger sequencing, the first generation of sequencing methods, praised for its versatility and accuracy, crucial to initiatives like the Human Genome Project. However, small labs struggled with the low throughput and high costs, limiting their genomic research participation. In 2005, second-generation sequencing technologies emerged, revolutionizing the field by increasing throughput and reducing costs, enabling large-scale projects. By 2015, third-generation sequencing technologies evolved, allowing direct sequencing of long DNA strands without fragmentation, improving resolution and facilitating the assembly of complex genomes and exploration of repetitive areas and structural variations. Each sequencing technology has unique advantages and drawbacks, influencing data quality and subsequent analysis. Researchers in genomic studies and AI integration in genomics will benefit from understanding these technologies. Genomic methods are crucial to genomics and metagenomics, fostering increasing demand for DNA sequencing applications in biomedical and environmental contexts. As sequencing advancements continue and costs decrease, genomic and metagenomic research is becoming standard in many labs. In public health, agriculture, and biomedicine, genomic studies have transformed these sectors, particularly in precision and personalized medicine. Sequencing technologies are central to these inquiries, unlocking secrets within genome nucleotide bases. The past five decades have witnessed remarkable advancements in genomic sequencing technologies. Exploring bioanalytical methods presents a discussion on the principles, advantages, and limitations of various sequencing approaches, alongside essential considerations for designing and applying these innovative methods[15].

3. Fundamentals of Artificial Intelligence

To understand the role of artificial intelligence in genomics, one must grasp its foundational concepts. "Artificial intelligence" encompasses a variety of technologies, from simple algorithms for efficient routing to complex systems excelling in games like chess. AI systems differ from traditional algorithms by their ability to learn from experience and improve. Traditional programming involves following a specific set of instructions, while AI uses a data-driven approach that enables it to leverage historical data for future decision-making[6]. Every AI system aims to optimize a specific evaluation metric by adjusting its internal parameters, or "weights." Among the branches of AI, "machine learning" is particularly prominent, as it focuses on systems that learn and improve through experience. Recently, "deep learning" has gained attention, utilizing sophisticated, multi-layered neural networks [16]. Core traits of AI systems include data gathering as a basis for predictions or decisions, enhancing accuracy over time. The process of acquiring and quantifying data is critical, with challenges often arising from needing relevant, well-formatted, and abundant data. This requirement can create bottlenecks in AI advancement. Genomic data presents various forms, from entire genomes to single nucleotides. Despite these complexities, AI methodologies offer numerous opportunities to analyze genomic information[17].

3.1. Machine Learning Basics

In artificial intelligence, machine learning enables machines to learn from data. Unlike traditional computing, which follows set algorithms, machine learning algorithms evolve by learning from data and experience. A prime example is email spam filtering. Initially, an email provider may filter based solely on keywords like "Nigeria," but spammers adapt quickly, necessitating a more sophisticated approach. By training a machine learning algorithm on a dataset of labeled emails, it can identify complex patterns that characterize spam emails, adjusting as new emails are processed. Machine learning algorithms fall into two categories: supervised and unsupervised learning. Supervised learning uses labeled datasets to create predictive models, refining decision boundaries (hyperplanes) during training to distinguish between different classes. Conversely, unsupervised learning explores unlabeled datasets to uncover hidden patterns and groupings, as seen in clustering algorithms that form similar data point clusters. High-quality training datasets are crucial for supervised learning; poorly constructed datasets can lead to ineffective predictive models. For instance, in spam filtering, a dataset with significantly more spam than legitimate emails would present challenges[18]. Machine learning has also made strides in genomics, aiding tasks like variant calling and genotype-phenotype predictions. The rise in available genomic data has driven advancements in precision medicine, which aims to tailor healthcare based on individual genomic profiles. Selecting optimal parameters and understanding the goals of machine learning are essential for effective implementation. Clear objectives help define how to represent data and desired outcomes. However, challenges like overfitting, where a model becomes too tailored to the noise in training data, and data bias can undermine model performance. Overfitting complicates generalization to new data, while biases, such as those arising from facial recognition technologies trained on non-representative samples, can lead to skewed results. In genomics, varying sequencing methods may introduce biases in data accuracy across different genomic regions[16], [18].

3.2. Deep Learning Concepts

In machine learning, deep learning has rapidly evolved over the past fifteen years, particularly in processing extensive datasets. It employs neural networks, structured with layers of interconnected nodes or neurons: an input layer, hidden layers, and an output layer. The input layer receives external data relevant to the specific challenge, while the output layer generates results. Hidden layers perform the bulk of processing, where nodes and weights—tunable parameters—are adjusted to optimize performance[19]. Deep learning enhances large data processing through parallel computation, making it a prime candidate for genomics, an interdisciplinary field merging biology, computer science, statistics, and mathematics. Genomic applications analyze genomes to identify genetic mutations, model species relationships, and discover causative genes for inherited diseases. With next-generation sequencing, genomic data has surged dramatically, producing over 1 petabase in 2021, projected to reach 4.5 petabytes by 2025. The use of deep learning techniques in genomics continues to grow for improved outcomes. Applications include identifying single-nucleotide polymorphisms, detecting indel mutations, structural variations, gene identification, siRNA target prediction, active site prediction, gene expression analysis, and disease classification. Deep learning offers substantial benefits over traditional machine learning, particularly in end-to-end training, which fosters the development of effective feature representations instead of relying on hand-crafted features that demand extensive domain knowledge. Its hierarchical structure allows for grasping features at various resolutions: while lower layers focus on local characteristics, upper layers capture broader traits. This capability enables deep learning to produce accurate results across diverse applications. However, challenges accompany its adoption; extensive training data is usually required, and interpreting its decision-making can be complex. Once, deep learning and genomics were viewed as nascent fields, but their convergence has become clear, promising to transform genomic studies. Emerging conversations about the practical applications of deep learning within genomics point towards a future brimming with potential[7].

4. Applications of AI in Genomics

In the unfolding story of modern science, Artificial Intelligence (AI) is poised to revolutionize genomics and life sciences, especially post-COVID-19. Significant advancements in genomic technologies are flourishing alongside AI's capabilities. In healthcare, AI provides solutions encompassing new drug discovery and improved clinical decision-making. A key area of AI contribution is analyzing vast genomic datasets produced by Next Generation Sequencing (NGS) technologies. Genomics, which explores genes and their roles, is crucial for understanding human health and disease. Researchers employ various computational techniques to extract essential biological insights from genomic data. With the increasing demand for comprehensive genomic analysis, data scientists are engaged in creating innovative solutions. Among these, Machine Learning (ML) stands out for its ability to learn from past experiences and apply this knowledge to new scenarios. When trained on genomic datasets, ML algorithms can identify genetic variants tied to diseases prevalent in specific populations, thus speeding up research initiatives. The advancements in AI have the potential to redefine genomics and healthcare, encouraging researchers and clinicians to enhance existing workflows or embark on new projects to develop innovative applications. As genomics grows in importance within the biomedical field, public and private sectors are increasingly interested in this dynamic study area. Although AI's application in genomic analysis is still nascent, it addresses critical challenges and opens new avenues. The convergence of genomics, data science, and AI stands at a crucial juncture, facilitating collaborative contributions across these fields. Conversations about genomics often focus on data generation processes and analysis methodologies. This includes exploring relevant programming languages and tools that enable these activities. Illustrative case studies showcase AI's transformative impact on healthcare genomics, highlighting successful applications that improve patient outcomes and stimulate further research. Nevertheless, integrating AI into genomic research and product development poses challenges that require careful navigation in this exciting and complex landscape[20].

4.1. Genomic Data Analysis

Once upon a time, the world of science was transforming due to Artificial Intelligence (AI), which changed how vast data sets were understood in fields like climate change and human brain studies. In genomics, high-throughput technologies produced massive datasets, encouraging researchers to explore algorithms from other disciplines. The combination of AI and genomics opened new opportunities and challenges, especially in clinical genomics, where AI could be harnessed to enhance patient care. The narrative examined genomic data's complexity and how AI could streamline its analysis, emphasizing the importance of precise genomic data for advancing personalized medicine. AI insights could influence treatment patterns, aiding healthcare providers significantly[21]. This exploration included tools and software for data analysis in genomics, highlighting challenges like data integration and computational efficiency, alongside potential solutions. Case studies showcased the effectiveness of AI in improving data analysis outcomes, illustrating the real-world benefits of this integration. As the story progressed, it revealed the intricacies in the development and clinical application of genomics, starting with the diverse datasets from high-throughput technologies, which vary from simple to complex multi-dimensional datasets requiring advanced modeling. The narrative delved into the coding of genomic variants, critical for structural and functional changes affecting proteomic transcripts. Yet, uncertainties remain regarding the actual number of protein-coding genes, their functions, and how coding variants impact both genes and proteins. These ambiguities drive the development of polygenic risk scores (PRS), which further complicate the journey through genomics. Despite hurdles from the vast dimensionality of genomic data and our limited understanding of variant functionalities, the potential for AI remains compelling. However, integrating these technologies into clinical practice poses additional challenges needing resolution as we navigate this complex landscape[22].

4.2. Precision Medicine

In the evolving landscape of healthcare, artificial intelligence (AI) plays a crucial role in precision medicine, which personalizes treatments by matching them to patients' unique genetic profiles. AI algorithms analyze vast genetic data to predict patient responses to therapies, facilitating the creation of more effective treatment plans. Precision medicine can identify new drug targets within the genome, optimize combinatorial therapies, and improve therapeutic strategies. Studies illustrate AI's integration into clinical precision medicine, particularly in cases like breast cancer, glioblastoma, and prostate cancer, showcasing its impact on genomic data analysis and patient outcomes. This highlights technology's potential to revolutionize healthcare through genetic insights[23]. However, challenges remain, including infrastructure issues related to managing datasets and a gap between technology developers and researchers. Ethical concerns regarding sensitive genomic data also arise. Collaboration between AI developers and genomic scientists is essential for successfully merging these fields in healthcare. With insights from AI and genomics, public health could enter a transformative new era (Ahmed, 2020). Overcoming implementation challenges may broaden the scope of this technology in genomic data analysis, ultimately enhancing healthcare worldwide[24].

Table No 1 Applications of Artificial Intelligence in Genomics

Application	Description	Examples	References
Gene Annotation	Identifying and labeling genes within DNA sequences.	Predicting gene functions and regulatory elements.	[25]
Variant Analysis	Detecting and interpreting genetic mutations.	Identifying variants associated with diseases like cancer.	[26]
Drug Discovery	Accelerating the identification of potential drug targets.	Screening compounds for genetic diseases.	[27]
Personalized Medicine	Tailoring treatments based on individual genomic profiles.	Recommending specific therapies for rare genetic disorders.	[28]
Population Genomics	Analyzing genetic data across populations to study evolution and disease.	Understanding genetic diversity and susceptibility to diseases.	[29]
Epigenomics	Studying changes in gene expression not caused by DNA sequence changes.	Identifying epigenetic markers linked to environmental exposure.	[30]

5. Challenges and Limitations

The promise of AI in genomics is vast, yet there are significant barriers. Quality concerns in genomic data hinder integration, as effective clinical applications rely on high-quality data collection. Mass data lacks standards, leading to issues like sequencing artifacts that affect AI accuracy. Junior AI researchers face limited access to clinical genomic data, complicating model training. Ideally, resources for AI model creation—like time and data—

would be ample, but reality often results in a trade-off between precision and scope due to restricted access to quality datasets. Clinical data are costly to obtain, and using public or in-house data as substitutes can lead to inaccuracies. Ethical concerns abound, particularly regarding genomic data privacy and ownership, as well as potential misuse leading to discrimination. In genomic medicine, trust in AI model predictions requires interpretability, yet current deep learning architectures tend to be opaque. While some black-box models may be acceptable in low-risk scenarios, the stakes in genomics are high, necessitating transparency. Developers must prioritize clear communication about algorithmic decisions to build trust. Complicating matters, malicious attacks could compromise patient privacy and genomic data, while generative modeling might lead to sensitive health data breaches[21]. The potential for bias in AI systems is another concern, as bias can infiltrate genomic data and models at various stages. Factors such as rigid inclusion/exclusion criteria in clinical studies may inadvertently bias results. Drug trials involving underrepresented populations could worsen health disparities. Regularization is crucial for managing bias-variance issues in high-dimensional genomic data, but overfitting is a risk when features outnumber observations. Selection bias also affects genomic data, as models trained on one population may not generalize to others, exacerbating inequity. Recognizing these modeling limitations is essential. AltLabs is working towards responsible genomic AI, aiming to address these challenges while unlocking AI's potential in genomics[20].

5.1. Data Quality and Quantity

As AI advances, its application in genomics is increasingly important. Researchers must carefully consider the challenges associated with using AI in genomic medicine, where genomic data is utilized for healthcare. Significant progress has been made in managing genomic information through advanced algorithms, spurred by the Human Genome Project's data influx. The evolution of open-source and commercial genomic analysis tools highlights the importance of high-performance computing and efficient algorithms. Unique challenges arise in genomics since genomic data differ from traditional big data sources. The discussions aim to establish a genetically-informed framework addressing concerns and potential shifts in adopting AI technologies in genomics, particularly regarding data integrity[31]. Ensuring high-quality, precise, and accurate data for AI model training is essential, as errors in data lead to unreliable analyses. Genomic data's distinct nature necessitates careful scrutiny when applying AI. While AI and machine learning (ML) methods unveil complex patterns in various research fields, genomics faces particular challenges with dataset comprehensiveness for robust AI model development. Lack of diverse genomic representation can undermine the effectiveness and credibility of genomic AI models. Additionally, AI's performance hinges on the volume of available data; without sufficient data, AI's ability to address complexities is compromised. Presently, genomics restricts AI implementation to simpler approaches, posing risks of overfitting and limited generalizability due to complexity demands. Lastly, data quality and integrity are pivotal for refining AI models. AI algorithms tend to settle on locally optimal solutions, necessitating significant adjustments and configurations based on domain insights. In genomics, the integration of domain knowledge is crucial for successful implementations[32].

5.2. Interpretability and Ethical Concerns

Despite the potential of artificial intelligence (AI) to enhance genomics, its adoption is limited. Key factors include the need for interpretability and transparency, as well as ethical issues linked to AI algorithms. The challenging complexity of AI results complicates their understanding for healthcare professionals and patients, particularly in clinical applications affecting patient outcomes. As genomics data integrates into electronic health records, AI is poised to generate new insights [33]. However, the intricate nature of AI results demands interpretable solutions for the various stakeholders involved. Ethical implications arise with diverse genomic data used to train AI, raising concerns about privacy, consent, and ownership. Moreover, biases from training data can unintentionally manifest in algorithms, mirroring issues seen in traditional genomics. Although AI's ethical discussions are prevalent in other fields, healthcare and genomics confront unique challenges requiring focused efforts on interpretability and ethics. As AI integration expands in genomics, examining these aspects is vital. This overview emphasizes the importance of these considerations for genomics researchers employing AI and encourages dialogue on developing interpretable and ethical AI solutions. Regulations are essential to ensure AI algorithms are both interpretable and ethical prior to their use in genomics. Researchers need to weave interpretability and ethics into their AI systems to prevent a lack of trust in AI solutions. Promoting discussions about algorithm interpretability and ethical data usage is critical, as the scientific community must identify how genomic data could compromise privacy and ownership rights. Concurrently, the genomics field needs to evaluate biases from alternative non-genomic data sources in AI algorithms, recognizing their potential to impede research and clinical practices. As AI becomes integral to these processes, biases in training data might skew outcomes, impacting diagnoses and treatments. While biases in AI are acknowledged within genomics, it's also crucial to view alternative data sources as potentially biased, fostering collaboration to tackle interpretability and ethical challenges in AI applications in genomics[34].

6. Future Directions and Emerging Trends

This review explores the intersection of artificial intelligence (AI) and genomics, focusing on recent advancements and future prospects. AI methods are evolving quickly, offering significant potential for genomic research and clinical applications. Genomic data serves as a vital resource for AI tools, allowing for the discovery of intricate biological insights (A Walton et al., 2023). As AI continues to advance, its integration with cloud computing for genomic data management and blockchain for secure sharing can be expected. These innovations will help safeguard genomic data as it is distributed. AI analysis is anticipated to improve personalized medicine, customizing treatments based on individual genetic profiles.

Collaborations between AI and genomic researchers are expected to increase, addressing key genomic questions through AI analysis (Greatbatch et al., 2019). Understanding the complexity of genomic data is essential for uncovering insights beyond AI's current and future capabilities. Ethical and data

privacy issues must be managed as AI-driven genomic analysis grows. It's important to consider the impact on public health, especially in developing countries with limited access to advanced technologies. This overview examines the state of AI in genomic data analysis and future trends.

References:

- [1] H. H. Smith, "EFFECTS OF GENOME BALANCE, POLYPLOIDY, AND SINGLE EXTRA CHROMOSOMES ON SIZE IN NICOTIANA," *Genetics*, vol. 28, no. 3, pp. 227–236, May 1943, doi: 10.1093/genetics/28.3.227.
- [2] J. P. F. C J Rawlings 1, "Artificial intelligence in molecular biology: a review and assessment".
- [3] N. A. Walton et al., "Enabling the clinical application of artificial intelligence in genomics: a perspective of the AMIA Genomics and Translational Bioinformatics Workgroup," *Journal of the American Medical Informatics Association*, vol. 31, no. 2, pp. 536–541, Jan. 2024, doi: 10.1093/jamia/ocad211.
- [4] C. J. RAWLINGS, "Databases, artificial intelligence and knowledge-based systems for molecular biology," *Biochem Soc Trans*, vol. 17, no. 5, pp. 851–855, Oct. 1989, doi: 10.1042/bst0170851.
- [5] J. Zou, H. Li, and S. Plevritis, "Editorial: Artificial Intelligence, machine learning and the changing landscape of molecular biology," *J Mol Biol*, vol. 434, no. 15, p. 167712, Aug. 2022, doi: 10.1016/j.jmb.2022.167712.
- [6] O. Greatbatch, A. Garrett, and K. Snape, "The impact of artificial intelligence on the current and future practice of clinical cancer genomics," *Genet Res (Camb)*, vol. 101, p. e9, Oct. 2019, doi: 10.1017/S0016672319000089.
- [7] W. S. Alharbi and M. Rashid, "A review of deep learning applications in human genomics using next-generation sequencing data," *Hum Genomics*, vol. 16, no. 1, p. 26, Jul. 2022, doi: 10.1186/s40246-022-00396-x.
- [8] P. L. Reichertz, "Preparing for change: concepts and education in medical informatics," *Comput Methods Programs Biomed*, vol. 25, no. 2, pp. 89–101, Sep. 1987, doi: 10.1016/0169-2607(87)90044-7.
- [9] J. A. Bossinger, "The Annotator's Assistant: an expert system for direct submission of genetic sequence data," *Bioinformatics*, vol. 4, no. 1, pp. 197–202, 1988, doi: 10.1093/bioinformatics/4.1.197.
- [10] J. H. Tjio and T. T. Puck, "GENETICS OF SOMATIC MAMMALIAN CELLS," *J Exp Med*, vol. 108, no. 2, pp. 259–268, Aug. 1958, doi: 10.1084/jem.108.2.259.
- [11] K. C. Atwood and S. L. Scheinberg, "Somatic variation in human erythrocyte antigens," *J Cell Comp Physiol*, vol. 52, no. S1, pp. 97–123, Dec. 1958, doi: 10.1002/jcp.1030520408.
- [12] A. G. Bearn and E. C. Franklin, "COMPARATIVE STUDIES ON THE PHYSICAL CHARACTERISTICS OF THE HERITABLE HAPTOGLOBIN GROUPS OF HUMAN SERUM," *J Exp Med*, vol. 109, no. 1, pp. 55–68, Jan. 1959, doi: 10.1084/jem.109.1.55.
- [13] H. Cleve and A. G. Bearn, "INHERITED VARIATIONS IN HUMAN SERUM PROTEINS: STUDIES ON THE GROUP - SPECIFIC COMPONENT*," *Ann N Y Acad Sci*, vol. 94, no. 1, pp. 218 – 224, Aug. 1961, doi: 10.1111/j.1749-6632.1961.tb35543.x.
- [14] B. T. J. M. S Pospíšilová 1, "Human genome sequencing--next generation technology or will the routine sequencing of human genome be possible".
- [15] B. M. Paegel, C. A. Emrich, G. J. Wedemayer, J. R. Scherer, and R. A. Mathies, "High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor," *Proceedings of the National Academy of Sciences*, vol. 99, no. 2, pp. 574–579, Jan. 2002, doi: 10.1073/pnas.012608699.
- [16] Y. Park, D. Heider, and A.-C. Hauschild, "Integrative Analysis of Next-Generation Sequencing for Next-Generation Cancer Research toward Artificial Intelligence," *Cancers (Basel)*, vol. 13, no. 13, p. 3148, Jun. 2021, doi: 10.3390/cancers13133148.
- [17] R. Esmailyfard, N. Esmaeeli, and M. Paknahad, "An artificial intelligence mechanism for detecting cystic lesions on CBCT images using deep learning," *J Stomatol Oral Maxillofac Surg*, vol. 126, no. 6, p. 102152, Dec. 2025, doi: 10.1016/j.jormas.2024.102152.
- [18] C. R. Cunha, A. Moreira, S. Coelho, V. Mendonça, and J. P. Gomes, "Converging extended reality and Machine Learning to improve the lecturing of geometry in basic education," *Journal of Engineering Research*, Nov. 2024, doi: 10.1016/j.jer.2024.10.016.
- [19] Q. van der Linden, E. Vanmassenhove, F. Zamberlan, and N. Rankovic, "Simple integrated circuit reverse-engineering with deep learning: A proof of concept for automating die-polygon-capturing," *Expert Syst Appl*, vol. 261, p. 125479, Feb. 2025, doi: 10.1016/j.eswa.2024.125479.
- [20] J. D. Lantos, "Neonatal bioethics, AI, and genomics," *Early Hum Dev*, vol. 198, p. 106130, Nov. 2024, doi: 10.1016/j.earlhumdev.2024.106130.
- [21] P. Guillod et al., "A systematic approach to prioritize diagnostically useful findings for inclusion in electronic health records as discrete data to improve clinical AI tools and genomic research," *Clin Oncol*, p. 103737, Dec. 2024, doi: 10.1016/j.clon.2024.103737.

- [22] D. B. Melo, R. G. S. Bruno, R. S. Bisinotto, and F. S. Lima, "Association of Genomic Predictions for Fertility Traits with Reproductive Outcomes in Dairy Cows Under Two Estrus Detection and Timed AI Programs.," *J Dairy Sci*, Dec. 2024, doi: 10.3168/jds.2024-25499.
- [23] S. Rezayi, S. R. Niakan Kalhari, and S. Saeedi, "Effectiveness of Artificial Intelligence for Personalized Medicine in Neoplasms: A Systematic Review," *Biomed Res Int*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/7842566.
- [24] N. Hangel, A. Buyx, and M.-C. Fritzsche, "The interrelation of scientific, ethical, and translational challenges for precision medicine with multimodal biomarkers – A qualitative expert interview study in dermatology research," *Heliyon*, vol. 10, no. 13, p. e31723, Jul. 2024, doi: 10.1016/j.heliyon.2024.e31723.
- [25] A. Ghelfi and S. Isobe, "Hayai-Annotation: A functional gene prediction tool that integrates orthologs and gene ontology for network analysis in plant species," *Comput Struct Biotechnol J*, vol. 27, pp. 117–126, 2025, doi: 10.1016/j.csbj.2024.12.011.
- [26] F. P. Idris et al., "Functional analysis of SRY variants in individuals with 46,XY differences of sex development," *Mol Cell Endocrinol*, vol. 598, p. 112458, Mar. 2025, doi: 10.1016/j.mce.2025.112458.
- [27] M. K. Yadav et al., "Unleashing the future: The revolutionary role of machine learning and artificial intelligence in drug discovery," *Eur J Pharmacol*, vol. 985, p. 177103, Dec. 2024, doi: 10.1016/j.ejphar.2024.177103.
- [28] H. Taherdoost and A. Ghofrani, "AI's role in revolutionizing personalized medicine by reshaping pharmacogenomics and drug therapy," *Intelligent Pharmacy*, vol. 2, no. 5, pp. 643–650, Oct. 2024, doi: 10.1016/j.ipha.2024.08.005.
- [29] X. Wei et al., "Genomic selection based on random regression test-day model in dairy cattle with respect to different reference populations," *Animal - Open Space*, vol. 4, p. 100087, Dec. 2025, doi: 10.1016/j.anopes.2024.100087.
- [30] P. Sa, M. Gòdia, N. Lewis, Y. Lian, and A. Clop, "Genomic, transcriptomic and epigenomic analysis towards the understanding of porcine semen quality traits. Past, current and future trends," *Anim Reprod Sci*, vol. 269, p. 107543, Oct. 2024, doi: 10.1016/j.anireprosci.2024.107543.
- [31] C. Krittanawong et al., "Artificial Intelligence and Cardiovascular Genetics," *Life*, vol. 12, no. 2, p. 279, Feb. 2022, doi: 10.3390/life12020279.
- [32] Q. Lin, P. K.-H. Tam, and C. S.-M. Tang, "Artificial intelligence-based approaches for the detection and prioritization of genomic mutations in congenital surgical diseases," *Front Pediatr*, vol. 11, Aug. 2023, doi: 10.3389/fped.2023.1203289.
- [33] S. K. Sen, E. D. Green, C. M. Hutter, M. Craven, T. Ideker, and V. Di Francesco, "Opportunities for basic, clinical, and bioethics research at the intersection of machine learning and genomics," *Cell Genomics*, vol. 4, no. 1, p. 100466, Jan. 2024, doi: 10.1016/j.xgen.2023.100466.
- [34] Z. Ahmed, "Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis," *Hum Genomics*, vol. 14, no. 1, p. 35, Dec. 2020, doi: 10.1186/s40246-020-00287-z.