AI-Powered Tools for Ancestry and Population Genomics in Healthcare

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: August 21, 2025 | Digital Therapeutics

DOI: 10.5281/zenodo.17996587

Abstract

The convergence of Artificial Intelligence AI and genomics is rapidly transforming precision medicine. The application of AI-powered tools to ancestry and po...

The convergence of Artificial Intelligence (AI) and genomics is rapidly transforming precision medicine. The application of AI-powered tools to **ancestry and population genomics** is unlocking new capabilities in disease risk prediction, drug development, and personalized healthcare strategies. For digital health and AI professionals, understanding this synergy is crucial, as it represents a fundamental shift toward more equitable and accurate genomic medicine [1].

The Role of AI in Genomic Data Analysis

Genomic data is characterized by its immense scale and complexity. Population-level studies aggregate and analyze data from thousands to millions of individuals. Traditional statistical methods often struggle to discern the subtle, non-linear patterns within this data that are indicative of disease risk or drug response.

This is where AI, particularly **Machine Learning (ML)** and **Deep Learning (DL)**, excels. AI algorithms can process vast, high-dimensional genomic datasets—including Single Nucleotide Polymorphisms (SNPs), transcriptomic data, and clinical records—to identify complex genotype-phenotype relationships [2] [3]. Key applications include:

Variant Interpretation: *Identifying and classifying genetic variants of uncertain significance (VUS) by predicting their pathogenicity [4].* **Disease Risk Prediction:** Developing sophisticated models, such as **Polygenic Risk Scores (PRS)**, that integrate thousands of genetic markers to estimate an individual's lifetime risk for complex diseases like cancer, diabetes, and cardiovascular conditions [5].

Ancestry and the Challenge of Bias in Genomic Data

A critical challenge in population genomics is the historical

overrepresentation of individuals of European ancestry in genomic studies. This bias has led to PRS and other predictive models that perform poorly when applied to non-European populations, exacerbating health disparities [6]. AI is now being deployed to actively address this issue.

AI-powered tools are essential for accurately inferring **genetic ancestry** at a super-population level, which is vital for contextualizing genetic findings [7]. Furthermore, advanced ML techniques are being developed to mitigate the inherent ancestral bias in training data. By recognizing and correcting for these imbalances, AI can help ensure that genomic medicine tools are equitable and effective across diverse global populations [8].

| AI Application | Impact on Ancestry/Population Genomics | Key Benefit | | :--- | :--- | :--- | | **Ancestry Inference** | Accurate identification of genetic background for contextualizing variants. | Improved clinical relevance of genetic findings. | | **Bias Mitigation** | Correcting for the overrepresentation of European ancestry data in training sets. | Enhanced equity and accuracy of predictive models across all populations. | | **Polygenic Risk Scores (PRS)** | Integrating population-specific genetic markers for more precise risk estimation. | Personalized and proactive disease prevention strategies. |

Advancing Precision Medicine with Population-Scale AI

The ultimate goal of integrating AI into ancestry and population genomics is to enable true precision medicine. Leveraging population-scale data, AI can move beyond individual risk assessment to inform public health strategies and drug development.

For instance, AI can analyze population-specific genetic variations to identify novel drug targets or predict how different populations will respond to existing therapies. The ability to accurately map genotype to phenotype, even for rare or complex genetic variations, is significantly enhanced by the pattern-recognition capabilities of deep learning models [9].

The development of tools like `ntRoot`, a computationally lightweight method for inferring human super-population-level ancestry, demonstrates the ongoing innovation [10]. Such tools are crucial for researchers and clinicians who need rapid, accurate ancestry information to guide their analysis and clinical decisions.

Ethical Considerations and the Future Outlook

While the potential is immense, the integration of AI in this field is not without ethical and technical challenges. Issues surrounding data privacy, the potential for misuse of ancestry information, and the need for transparency in AI model development remain paramount [11]. The scientific community must ensure the application of genetic ancestry information is conducted within a robust ethical framework to prevent misappropriation and ensure patient trust.

Looking ahead, the future of AI in ancestry and population genomics will be defined by the continued push for more diverse and larger datasets. As AI models become more sophisticated, they will not only predict disease risk but

also guide preventative interventions and personalize treatment pathways with unprecedented accuracy, ultimately transforming healthcare from reactive to proactive.

***

References

[1] [Artificial Intelligence in Genetics - PMC - PubMed Central - NIH] (https://pmc.ncbi.nlm.nih.gov/articles/PMC10856672/) [2] [The transformative role of Artificial Intelligence in genomics] (https://www.sciencedirect.com/science/article/abs/pii/S2452014425001876) [3] [A Systematic Review on the Generative AI Applications in ...] (https://arxiv.org/html/2508.20275v1) [4] [Artificial intelligence in clinical genetics | European Journal ...](https://www.nature.com/articles/s41431-024-01782-w) [5] [AI-powered precision medicine: utilizing genetic risk factor ...] (https://pmc.ncbi.nlm.nih.gov/articles/PMC12051108/) [6] [Evaluating the promise of inclusion of African ancestry populations in genomics] (https://www.nature.com/articles/s41525-019-0111-x) [7] [Inferring genetic ancestry: opportunities, challenges, and implications] (https://www.cell.com/ajhg/fulltext/S0002-9297(10)00155-2?cc=y%3Fc) [8] [Equitable machine learning counteracts ancestral bias in ...] (https://pmc.ncbi.nlm.nih.gov/articles/PMC10402189/) [9] [Enabling the clinical application of artificial intelligence in ...] (https://academic.oup.com/jamia/article/31/2/536/7456434) [10] [ntRoot: computational inference of human ancestry at scale ...] (https://academic.oup.com/bioinformaticsadvances/advance-article/doi/10.1093/bioadv/vbaf287/8317431) [11] [An ethical framework for research using genetic ancestry] (https://muse.jhu.edu/pub/1/article/894253/summary)
