

Machine Learning for Variant Interpretation in Clinical Genomics: A New Era of Precision

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: August 30, 2025 | AI Diagnostics

DOI: [10.5281/zenodo.17996575](https://doi.org/10.5281/zenodo.17996575)

Abstract

The rapid advancement of Next-Generation Sequencing (NGS) has led to a deluge of genomic data, creating a significant challenge: interpreting the millions of g...

The rapid advancement of Next-Generation Sequencing (NGS) has led to a deluge of genomic data, creating a significant challenge: interpreting the millions of genetic variants discovered in each individual. The crucial step in clinical genomics—determining which variants are pathogenic (disease-causing) and which are benign—is often a time-consuming and complex process. **Machine Learning (ML)** is emerging as a transformative force, streamlining and enhancing the accuracy of **Variant Interpretation in Clinical Genomics** [1].

The Challenge of Variant Interpretation

In a clinical setting, the interpretation of genomic variants traditionally relies on a manual, evidence-based framework, most notably the guidelines established by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP). This process involves integrating multiple lines of evidence, including population frequency, computational prediction, functional studies, and segregation data. The sheer volume of variants of Unknown Significance (VUS) often overwhelms clinical laboratories, leading to diagnostic delays and uncertainty [2].

Machine Learning: Augmenting the ACMG/AMP Framework

Machine learning models are uniquely suited to tackle this data-intensive challenge. By training on vast, curated datasets of previously classified variants, ML algorithms can learn complex patterns and relationships that are difficult for human experts to discern.

Key Applications of ML in Variant Interpretation:

- Pathogenicity Prediction:** ML models, including deep learning architectures, are now routinely used to predict the likelihood of a variant

being pathogenic. Tools like **DeepVariant** and **Clair** have demonstrated high accuracy in variant calling, which is the initial step of identifying variants from raw sequencing data [3]. More advanced models, such as those that integrate the ACMG/AMP criteria directly into their feature set, provide a probabilistic score of pathogenicity, moving beyond simple binary classification [4]. 2. **Variant Prioritization:** In whole-exome or whole-genome sequencing, ML algorithms help prioritize the most clinically relevant variants from a long list of candidates. Platforms like **3ASC** use explainable algorithms to not only prioritize variants but also annotate the evidence used, increasing transparency and trust in the clinical workflow [5]. 3. **Phenotype-Genotype Correlation:** ML is being used to connect specific genetic variants with clinical phenotypes, particularly in rare diseases. By analyzing electronic health records (EHRs) alongside genomic data, ML can uncover subtle correlations, improving diagnostic yield and facilitating the identification of novel disease genes [6]. Recent work has even introduced ML-derived measurements of **penetrance**—the probability of a gene variant being expressed—offering a more nuanced view of disease risk [7].

Deep Learning: The Cutting Edge

Deep learning (DL), a subset of ML, has proven particularly powerful in genomics. DL models can process raw data directly, such as images of sequencing reads or complex genomic features, to make highly accurate predictions. DL models are used for identifying genetic variants and for predicting the functional impact of non-coding variants, which are notoriously difficult to interpret [8].

Challenges and the Future Outlook

Despite the promise, the integration of ML into clinical genomics faces several hurdles:

Data Quality and Bias: *The performance of ML models is heavily dependent on the quality and representativeness of the training data. Biases in existing databases can lead to models that perform poorly on underrepresented populations.* **Interpretability (Explainable AI - XAI):** Clinicians require transparency. A "black box" model that provides a score without explaining *why* a variant is classified as pathogenic is unlikely to be adopted in critical diagnostic decisions. The development of Explainable AI (XAI) tools is a major focus to ensure clinical utility [5]. **Standardization:** *A lack of standardized data formats and model validation protocols across different clinical labs hinders the widespread adoption of these tools.*

The future of variant interpretation is undeniably intertwined with machine learning. As models become more sophisticated, interpretable, and integrated into clinical workflows, they will transform the VUS problem, accelerate diagnoses for rare diseases, and usher in a new era of truly personalized and precise medicine.

*

References

[1] O. Abdelwahab, "Artificial intelligence in variant calling: a review," PMC, 2025. [URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12055765/>] [2] S. Vadapalli et al., "Artificial intelligence and machine learning approaches using clinical genomics data," PMC, 2022. [URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10233311/>] [3] T. Jo et al., "Deep learning-based identification of genetic variants," Briefings in Bioinformatics, 2022. [URL: <https://academic.oup.com/bib/article/23/2/bbac022/6532536>] [4] G. Nicora et al., "A machine learning approach based on ACMG/AMP guidelines for germline variant interpretation," Scientific Reports, 2022. [URL: <https://www.nature.com/articles/s41598-022-06547-3>] [5] H. H. Kim et al., "Explicable prioritization of genetic variants by integration of multi-omics data," PMC, 2024. [URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10956189/>] [6] J. A. Diao et al., "Biomedical informatics and machine learning for clinical genomics," PMC, 2018. [URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5946905/>] [7] Science Editorial, "Machine learning-based penetrance of genetic variants," Science, 2025. [URL: <https://www.science.org/doi/10.1126/science.adm7066>] [8] Academic OUP, "Advancing genome-based precision medicine: a review on machine learning for genomic variant interpretation," Briefings in Bioinformatics*, 2025. [URL: <https://academic.oup.com/bib/article/doi/10.1093/bib/bbaf329/8203342?rss=1>]