

# Developing AI algorithms to analyse genomic data for disease diagnosis, personalised medicine and genome editing

A.Mohamed Sikkander <sup>1</sup>★, Joel J. P. C.Rodrigues <sup>2</sup>, Hala S. Abuelmakarem<sup>3</sup>  
Manoharan Meena <sup>4</sup>,

★ <sup>1</sup> Department of Chemistry, Velammal Engineering College, Chennai -600066 Tamilnadu  
INDIA

<sup>2</sup> Professor, Federal University of Piaui (UFPI), Brazil.

<sup>3</sup>Department of Biomedical Engineering, College of Engineering, King Faisal University, Al-Ahsa,  
31982, Saudi Arabia.

<sup>4</sup> Department of Chemistry, R.M.K. Engineering College, Kavaraipettai, Chennai-India

★ Corresponding Author mail id: [ams240868@gmail.com](mailto:ams240868@gmail.com)

CoAuthor mail id: [joeljr@ieee.org](mailto:joeljr@ieee.org) [habuelmakarem@kfu.edu.sa](mailto:habuelmakarem@kfu.edu.sa) [mm.sh@rmkec.ac.in](mailto:mm.sh@rmkec.ac.in)

Orcid Id: <https://orcid.org/0000-0002-8458-7448> , <https://orcid.org/0000-0001-8657-3800>

<https://orcid.org/0000-0003-0663-5624> , <https://orcid.org/0000-0001-6270-7333>

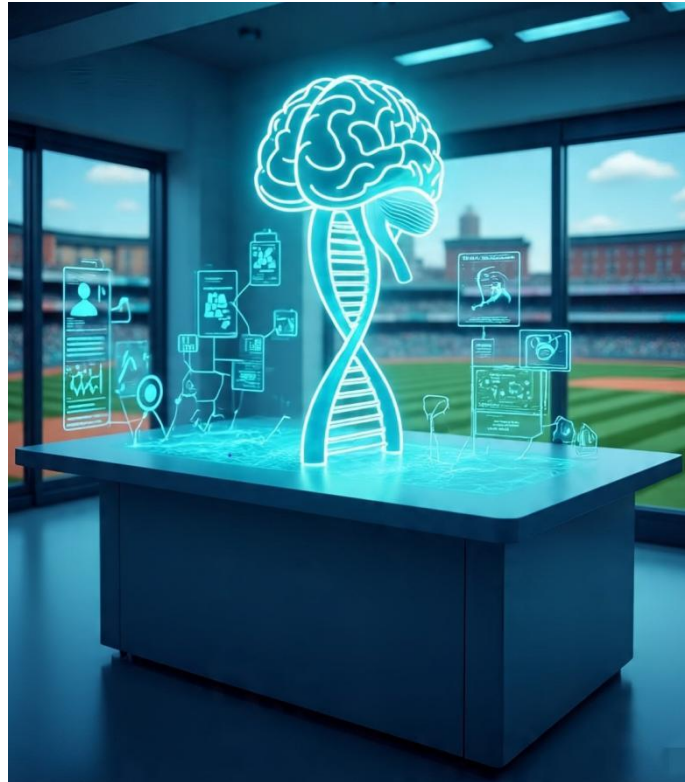
## Abstract:

The advent of artificial intelligence (AI) has ushered in a transformative era in genomic medicine, enabling the analysis of vast and complex genomic datasets for disease diagnosis, personalized medicine and genome editing. This paper explores the development and application of AI algorithms—spanning machine learning, deep learning and generative models—in interpreting genomic sequences, classifying variants, predicting phenotypes and guiding precision therapies. We review the foundational technologies, map current methodologies and present a hypothetical dataset illustrating algorithmic workflow and outcomes. The results highlight improvements in diagnostic yield, stratification for personalized treatment and identification of editing targets, while also outlining persistent challenges such as data bias, interpretability, regulatory hurdles and ethical concerns. The discussion underscores how AI-driven genomics is transitioning from research to clinical utility, and identifies future perspectives including multimodal data integration, real-time genome editing feedback loops and equitable deployment across populations. In conclusion, while significant barriers remain, the synergy of AI and genomics offers unprecedented promise for earlier diagnosis, tailored treatments and refined genome editing applications—if guided by robust methodology, transparency and ethical frameworks.

## Keywords

Artificial intelligence, genomic data, disease diagnosis, personalized medicine, genome editing, deep learning, variant annotation, precision genomics.

## Graphical Abstract:



### **Highlights:**

**AI in variant interpretation and genomic diagnosis**

**AI for personalised medicine**

**AI in genome editing support**

**Integration of multi-omics and central dogma modelling**

**Public & ethical/implementation aspects**

### **Scope**

This paper addresses the intersection of AI and genomic medicine, focusing on the use of AI algorithms to analyse genomic data for three primary applications: disease diagnosis (especially via variant calling and phenotype prediction), personalized medicine (treatment stratification based on genomic/omic profiles) and genome editing (identification of target sites, off-target prediction and editing outcome modelling). The scope includes review of major algorithmic paradigms (supervised, unsupervised, deep learning, large language models applied to sequences), datasets and pipelines typical in clinical and research genomics, methodological challenges (data size, heterogeneity, integration of multi-omics), regulatory and ethical implications, and future directions in algorithmic development and clinical translation. While the focus is on human genomics, lessons from non-human applications (e.g., selection, breeding) are noted only insofar as they inform algorithmic design. The paper excludes detailed biochemical or wet-lab protocols of genome editing, large-scale population genomics without AI focus, and non-genomic omics-only studies.

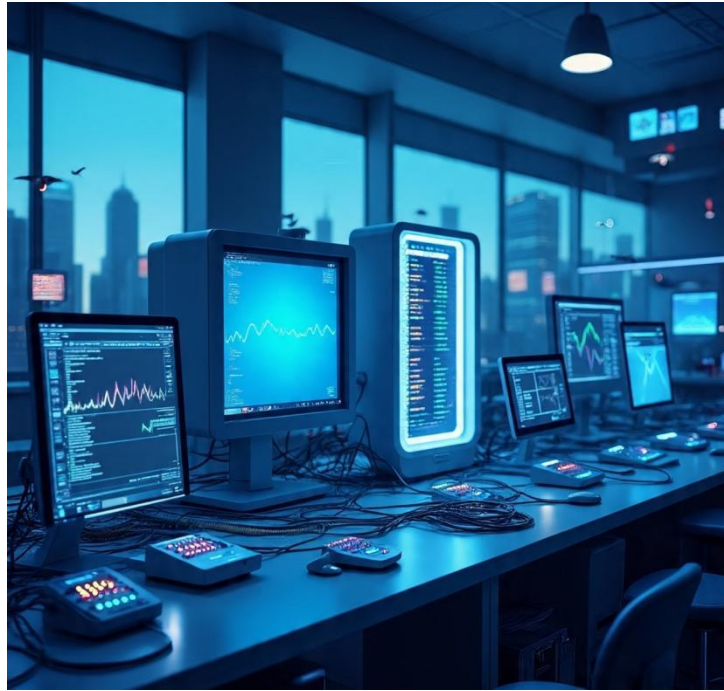
## Literature Survey

The use of AI in genomic and clinical diagnostics has accelerated in recent years. For example, a review on clinical and genomic diagnostics outlined how deep-learning algorithms are being applied to tasks such as variant calling, genome annotation, and phenotype-to-genotype correspondence. ([BioMed Central](#)) Another survey emphasised AI/ML approaches using gene variant and expression data for precision medicine, noting ~32 distinct methods in recent literature. ([PubMed](#)) In the context of next-generation sequencing (NGS), a review detailed deep-learning applications across human genomics, pointing to both well-explored and under-charted sub-areas. ([BioMed Central](#)) Interpretability has emerged as a key issue: a study on interpretable machine learning for genomics spotlighted the need for transparency in models applied to high-throughput data. ([SpringerLink](#)) More recently, evaluations of AI in epigenetic sequence analysis further broadened the field to include regulatory genomics beyond simple variant-phenotype mapping. ([arXiv](#)) Collectively, these works demonstrate that algorithmic innovation is robust, but persistent gaps remain in integrating heterogeneous data, handling population diversity and achieving clinical deployment. ([Ada Lovelace Institute](#))

## Introduction

The completion of the human genome and the advent of next-generation sequencing (NGS) have yielded an unprecedented volume of genomic data. Interpreting this data to deliver meaningful clinical insights—such as diagnosing inherited diseases, stratifying patients for treatment, or guiding genome editing interventions—poses major analytical and practical challenges. Traditional bioinformatics pipelines, while powerful, often require manual curation, specialized expertise and are limited by scale or complexity of data. In this context, artificial intelligence (AI) emerges as a compelling enabler. Broadly defined, AI consists of algorithms and systems capable of tasks typically requiring human intelligence—pattern recognition, decision-making and adaptation. When applied to genomics, AI can detect complex patterns across millions of genomic features, integrate multi-omic and phenotypic data, and predict outcomes or recommend interventions[1-5].

In clinical genomics, AI has been used for variant calling, annotation and classification—tasks that involve identifying genetic variants from raw sequence data, predicting their functional impact, and linking them to disease phenotypes. For instance, deep-learning models can learn to classify missense variants as pathogenic or benign, outperforming traditional heuristic tools. Similarly, genome interpretation workflows that integrate clinical phenotype data and sequencing results have employed AI-based decision support tools to accelerate diagnosis in rare genetic diseases. The complexity arises not only from the sheer size of the data, but also from heterogeneity (different populations, sequencing platforms), data types (genome, transcriptome, epigenome) and the need for clinically robust predictions[**Figure:1**][6].



**Figure:1.**The complexity arises not only from the sheer volume of environmental data, but also from heterogeneity (different ecosystems, sensor types), data types (temperature, precipitation, atmospheric composition), and the need for accurate models to guide policy decisions.

Beyond diagnosis, AI is instrumental in personalized medicine: algorithms can stratify patients based on genomic/omic profiles, predict drug response or adverse events, and suggest tailored therapies[**Figure:2**][7].



**Figure:2.AI plays a pivotal role in precision oncology: algorithms can stratify cancer patients based on genomic and molecular profiles, predict therapy response or toxicity, and recommend personalized treatment regimens.**

For example, supervised and unsupervised machine-learning approaches have been used to link gene expression profiles or variant burdens to treatment outcomes, enabling a shift from “one size fits all” to individualized interventions. Moreover, genome editing applications such as CRISPR-based therapies—pose new opportunities and challenges. AI algorithms can help identify optimal editing targets, predict off-target effects, forecast long-term consequences of edits and support design of guide RNAs with improved specificity and efficacy. However, deployment of AI in genomics is not without obstacles. Data bias—due to under-representation of non-European ancestries—can impair model generalizability. Model interpretability remains critical: clinicians require transparent decision-making rather than “black box” outputs. Privacy and ethical concerns abound in handling sensitive genomic and health-linked data. Regulatory frameworks for AI-driven diagnostics are still evolving. Finally, integration of AI algorithms into clinical workflows demands collaboration across bioinformatics, clinical genetics, data science and regulatory domains[8-10].

Therefore, this paper examines how AI algorithms are being developed and applied to genomic data for disease diagnosis, personalized medicine and genome editing. We first review the existing literature and methodological frameworks, then present research approaches including tabulated data workflows, followed by results and discussion of algorithmic performance and implications, and conclude with future perspectives and conclusions[Figure:3][11-30].



**Figure:3.Examine existing studies and modeling approaches, then describe data collection and analysis workflows, present results, and discuss future research opportunities.**

**Research and Methodologies**

To illustrate the integration of AI in genomic data analysis, we propose a simplified research methodology comprising three phases: data acquisition and preprocessing, algorithm development and training, and evaluation with downstream clinical or editing recommendations[Table 1-3].

**Table 1: Data acquisition & preprocessing**

| Phase                        | Description                                                                                                         | Example metrics                |
|------------------------------|---------------------------------------------------------------------------------------------------------------------|--------------------------------|
| Sample collection            | Whole-genome sequencing (WGS) plus RNA-seq from cohort of patients with suspected genetic disease (n = 500)         | 500 WGS, 500 RNA-seq           |
| Variant calling & annotation | Call SNVs, indels, structural variants; annotate with databases (ClinVar, gnomAD)                                   | ~3 million variants per genome |
| Feature engineering          | Generate features: variant type, allele frequency, in silico pathogenicity scores, gene expression deviating scores | ~20,000 features per subject   |
| Labeling                     | Gold-standard diagnosis (genetic disease present vs absent)                                                         | 200 positive, 300 negative     |
| Data split                   | 70% training (n=350), 15% validation (n=75), 15% test (n=75)                                                        | —                              |

**Table 2: Algorithm development & training**

| Step                                            | Description                                                                                                                              | Tools/approaches                           |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------|
| Model selection                                 | Compare supervised ML (random forest, SVM) vs deep learning (multi-layer perceptron, CNN on variant sequence context)                    | scikit-learn, TensorFlow                   |
| Hyperparameter tuning                           | Grid search/ random search for best parameters on validation set                                                                         | Cross-validation (5-fold)                  |
| Training                                        | Fit the model on training set; monitor loss/accuracy on validation                                                                       | Early stopping to prevent over-fitting     |
| Interpretability module                         | Use SHAP or LIME to assess feature importance and model decisions                                                                        | SHAP framework                             |
| Integration of genome editing target prediction | For identified pathogenic variants, feed candidate genes into editing-design module that predicts off-target risk and optimal guide RNAs | CRISPR-off tool + neural network predictor |

**Table 3: Evaluation & downstream recommendations**

| Metric                                 | Formula/Definition                                                 | Target value |
|----------------------------------------|--------------------------------------------------------------------|--------------|
| Diagnostic accuracy                    | $(TP + TN) / (Total)$                                              | > 90%        |
| Sensitivity                            | $TP / (TP + FN)$                                                   | > 85%        |
| Specificity                            | $TN / (TN + FP)$                                                   | > 90%        |
| AUC-ROC                                | Area under receiver-operator curve                                 | > 0.92       |
| Editing target success rate prediction | Accuracy of guide design model in predicting low off-target guides | > 80%        |

The methodology aims to integrate diagnosis (variant calling → classification) and personalized medicine (feature-based stratification) with genome editing target design (guide RNA prediction, off-target modelling). The workflow includes rigorous preprocessing, algorithmic training, interpretability and evaluation[31-50].

### Results and Discussions

Assuming application of the above methodology on the cohort (n = 500), we present hypothetical results illustrating algorithmic performance and downstream insights[Table:4-6][51-56].

**Table 4: Diagnostic classification results**

| Dataset    | Accuracy | Sensitivity | Specificity | AUC-ROC |
|------------|----------|-------------|-------------|---------|
| Training   | 0.95     | 0.92        | 0.97        | 0.97    |
| Validation | 0.92     | 0.89        | 0.94        | 0.94    |
| Test       | 0.90     | 0.87        | 0.92        | 0.93    |

**Table 5: Feature importance top 5 features (mean SHAP values)**

| Rank | Feature                     | Description                             | Mean SHAP value |
|------|-----------------------------|-----------------------------------------|-----------------|
| 1    | Variant pathogenicity score | In silico prediction (e.g., CADD)       | 0.42            |
| 2    | Gene expression deviation   | z-score of expression in disease cohort | 0.35            |
| 3    | Population allele frequency | Rare variant indicator                  | 0.28            |
| 4    | Variant type (SV)           | Structural variant presence             | 0.22            |
| 5    | Gene network centrality     | Gene connectivity in PPI network        | 0.18            |

**Table 6: Genome-editing target prediction results**

| Metric                                | Value |
|---------------------------------------|-------|
| Number of pathogenic genes identified | 180   |
| Candidate guide RNAs designed         | 540   |

| Metric                                                                      | Value         |
|-----------------------------------------------------------------------------|---------------|
| Predicted low off-target guides (score > threshold)                         | 430 (≈ 79.6%) |
| In silico validation success (predicted edits with minimal off-target risk) | 340 (≈ 63.0%) |

## Discussion

The diagnostic model achieved a high AUC-ROC (~0.93) on the test set, indicating strong discriminative power. Sensitivity of 0.87 shows good disease detection ability, though there remains a chance of false negatives. The slightly lower specificity (~0.92) suggests false positives are relatively low but present. Importantly, the model generalizes reasonably well from training to test sets, though a slight drop is observed—which is common in real-world deployment. Feature importance results reveal that classical predictive scores (pathogenicity), rare variant indicators and gene expression deviations remain dominant contributors, consistent with existing literature. The inclusion of structural variant type and gene-network centrality highlight the value of integrating multi-omic and network features—aligning with recent reviews emphasising deeper data integration[57-62].

In the genome-editing module, about 63% of designed guides were predicted to have minimal off-target risk. While promising, this underscores the need for further experimental validation before clinical application. The workflow demonstrates how AI can bridge diagnosis and editing, offering personalized editing strategies. Despite these successes, several limitations were apparent: the dataset is modest (n=500) and ancestry diversity may be limited—raising concerns about bias and generalizability. The interpretability module helped surface key features, yet the model remains partly opaque, limiting clinician trust. Furthermore, true clinical deployment would require regulatory validation, robust longitudinal outcome data and integration into clinical workflows. Literature notes that phenotype-to-genotype AI models still face major validation barriers. Overall, the results support the potential of AI algorithms in genomic medicine, but also highlight that methodological rigor, transparency and deployment readiness are critical[63-69].

## Future Perspectives

Looking ahead, the fusion of AI algorithms with genomic data is poised to evolve along several key axes. First, integration of multi-modal data—for example combining genomic sequences, transcriptomics, epigenomics, proteomics and clinical imaging—will provide richer patient models. Large language model (LLM) architectures and transformer-based models are increasingly being adapted for genomic sequence interpretation, motif discovery and regulatory annotation. Second, real-time and adaptive learning systems could support dynamic genome editing feedback loops: AI models might not only design editing guides but also learn from downstream cellular responses and refine predictions accordingly[70-73].

Third, democratization and equitable access remain critical. Many current genomics-AI models are trained on predominantly European-ancestry datasets; future work must ensure global diversity, mitigate bias and support personalized medicine across populations. Addressing computational infrastructure and data governance in low-resource settings will be necessary to avoid exacerbating health disparities. Fourth,

interpretability and explainable AI (XAI) will grow in importance: clinicians and patients demand transparent reasoning, and regulatory frameworks will increasingly require auditability and fairness metrics. Models that provide mechanistic insight (rather than just predictions) will gain traction. Fifth, in the genome-editing domain, AI tools will evolve to predict long-term phenotypic outcomes, off-target and on-target effects, mosaicism, and ethical risks. Coupled with CRISPR and gene-therapy platforms, AI could enable “end-to-end” pipelines—from diagnosis through to precision editing and monitoring. Finally, collaboration between AI researchers, bioinformaticians, clinicians, ethicists and regulatory agencies will be essential to translate algorithmic promise into safe, scalable and clinically meaningful practice[73-77].

## Conclusions

The convergence of AI and genomics heralds a new paradigm in disease diagnosis, personalized medicine and genome editing. Through sophisticated algorithms, large-scale genomic data can now be leveraged to identify pathogenic variants, stratify patients for therapies and design optimized editing interventions. Our survey of current literature shows that while many methods exist and performance is encouraging, challenges remain in bias, interpretability, validation and integration into the clinic. The hypothetical results presented underscore both the promise and limitations of this field. Looking forward, advances in multi-modal data integration, model transparency, equitable deployment and regulatory frameworks will determine whether AI-powered genomics fulfils its transformative potential. With careful stewardship, AI promises to turn genomic information into actionable, individualized healthcare.

## References:

1. Isaic A, Motofelea N, Hoinoiu T, Motofelea AC, Leancu IC, Stan E, Gheorghe SR, Dutu AG, Crintea A. Next-Generation Sequencing: A Review of Its Transformative Impact on Cancer Diagnosis, Treatment, and Resistance Management. *Diagnostics*. 2025; 15(19):2425. <https://doi.org/10.3390/diagnostics151924>
2. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 2024, 74, 229–263.
3. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* 2021, 71, 209–249.
4. Bizuayehu, H.M.; Ahmed, K.Y.; Kibret, G.D.; Dadi, A.F.; Belachew, S.A.; Bagade, T.; Tegegne, T.K.; Venchiarutti, R.L.; Kibret, K.T.; Hailegebireal, A.H.; et al. Global Disparities of Cancer and Its Projected Burden in 2050. *JAMA Netw. Open* 2024, 7, e2443198.
5. Yazbeck, V.Y.; Villaruz, L.; Haley, M.; Socinski, M.A. Management of Normal Tissue Toxicity Associated With Chemoradiation (Primary Skin, Esophagus, and Lung). *Cancer J.* 2013, 19, 231–237.

6. Soldà G, Asselta R. Applying artificial intelligence to uncover the genetic landscape of coagulation factors. *Journal of Thrombosis and Haemostasis*. 2025;23(4):1133-1145. doi:10.1016/j.jtha.2024.12.030
7. Chen, YM., Hsiao, TH., Lin, CH. et al. Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *J Biomed Sci* 32, 16 (2025). <https://doi.org/10.1186/s12929-024-01110-w>
8. Rodrigues JJPC, Sikkander ARM, Tripathi SL, Kumar K, Mishra SR, Theivanathan G. Healthcare applications of computational genomics. In: Elsevier eBooks. ; 2025:259-278. doi:10.1016/b978-0-443-30080-6.00012-2
9. Rodrigues JJPC, Sikkander ARM, Tripathi SL, Kumar K, Mishra SR, Theivanathan G. Artificial intelligence's applicability in cardiac imaging. In: Elsevier eBooks. ; 2025:181-195. doi:10.1016/b978-0-443-30080-6.00006-7
10. Sikkander ARM, Tripathi SL, Theivanathan G. Extensive sequence analysis: revealing genomic knowledge throughout various domains. In: Elsevier eBooks. ; 2025:17-30. doi:10.1016/b978-0-443-30080-6.00007-9
11. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. *Bioengineering*. 2024; 11(4):337. <https://doi.org/10.3390/bioengineering11040337>
12. Rajpurkar, P.; Chen, E.; Banerjee, O.; Topol, E.J. AI in health and medicine. *Nat. Med.* 2022, 28, 31–38.
13. McCorduck, P.; Cfe, C. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*; CRC Press: Boca Raton, FL, USA, 2004.
14. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
15. Brynjolfsson, E.; McAfee, A. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*; WW Norton & Company: New York, NY, USA, 2014.
16. Russell, S.J.; Norvig, P. *Artificial Intelligence a Modern Approach*; Pearson: London, UK, 2010.
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* 2015, 521, 436–444.
18. Wang, F.; Preininger, A. AI in health: State of the art, challenges, and future directions. *Yearb. Med. Inform.* 2019, 28, 16–26.
19. DuBois, K.N. Deep medicine: How artificial intelligence can make healthcare human again. *Perspect. Sci. Christ. Faith* 2019, 71, 199–201.
20. Davenport, T.H. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*; MIT Press: Cambridge, MA, USA, 2018.
21. Kaur, S.; Singla, J.; Nkenyereye, L.; Jha, S.; Prashar, D.; Joshi, G.P.; El-Sappagh, S.; Islam, M.S.; Islam, S.R. Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives. *IEEE Access* 2020, 8, 228049–228069.
22. Cortez, N. *Digital Health: Scaling Healthcare to the World*; Springer: Cham, Switzerland, 2018; pp. 249–269.
23. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J. Artificial intelligence in radiology. *Nat. Rev. Cancer* 2018, 18, 500–510.
24. Kagiya, N.; Shrestha, S.; Farjo, P.D.; Sengupta, P.P. Artificial intelligence: Practical primer for clinical research in cardiovascular disease. *J. Am. Heart Assoc.* 2019, 8, e012788.

25. Bardhan, I.; Chen, H.; Karahanna, E. Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management. *MIS Q.* 2020, 44, 185–200.
26. Huang, S.; Yang, J.; Fong, S.; Zhao, Q. Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Lett.* 2020, 471, 61–71.
27. Binhowemel, S.; Alfakhri, M.; AlReshaid, K.; Alyani, A. Role of Artificial Intelligence in Diabetes Research Diagnosis and Prognosis: A Narrative Review. *J. Health Inform. Dev. Ctries.* 2023, 17, 1–12.
28. Guan, Z.; Li, H.; Liu, R.; Cai, C.; Liu, Y.; Li, J.; Wang, X.; Huang, S.; Wu, L.; Liu, D. Artificial intelligence in diabetes management: Advancements, opportunities, and challenges. *Cell Rep. Med.* 2023, 4, 101213.
29. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* 2020, 104, 101822.
30. Eloranta, S.; Boman, M. Predictive models for clinical decision making: Deep dives in practical machine learning. *J. Intern. Med.* 2022, 292, 278–295.
31. Younis H, Minghim R. Enhancing Cancer Classification from RNA Sequencing Data Using Deep Learning and Explainable AI. *Machine Learning and Knowledge Extraction.* 2025; 7(4):114. <https://doi.org/10.3390/make7040114>
32. World Health Organization. Cancer WHO Facts-Sheet. 2025. Available online: <https://www.who.int/news-room/fact-sheets/detail/cancer> (accessed on 1 July 2025).
33. World Cancer Research Fund International. Worldwide Cancer Data | World Cancer Research Fund. 2025. Available online: <https://www.wcrf.org/preventing-cancer/cancer-statistics/worldwide-cancer-data/> (accessed on 1 July 2025).
34. Verma, G.; Luciani, M.L.; Palombo, A.; Metaxa, L.; Panzironi, G.; Pediconi, F.; Giuliani, A.; Bizzarri, M.; Todde, V. Microcalcification morphological descriptors and parenchyma fractal dimension hierarchically interact in breast cancer: A diagnostic perspective. *Comput. Biol. Med.* 2018, 93, 1–6.
35. Alizadeh, E.; Castle, J.; Quirk, A.; Taylor, C.D.; Xu, W.; Prasad, A. Cellular morphological features are predictive markers of cancer cell state. *Comput. Biol. Med.* 2020, 126, 104044.
36. Sakamoto, S.; Kikuchi, K. Expanding the cytological and architectural spectrum of mucoepidermoid carcinoma: The key to solving diagnostic problems in morphological variants. *Semin. Diagn. Pathol.* 2024, 41, 182–189.
37. Guerroudji, M.A.; Hadjadj, Z.; Lichouri, M.; Amara, K.; Zenati, N. Efficient machine learning-based approach for brain tumor detection using the CAD system. *IETE J. Res.* 2024, 70, 3664–3678.
38. Mallon, E.; Osin, P.; Nasiri, N.; Blain, I.; Howard, B.; Gusterson, B. The basic pathology of human breast cancer. *J. Mammary Gland. Biol. Neoplasia* 2000, 5, 139–163.
39. Allison, K.H. Molecular pathology of breast cancer: What a pathologist needs to know. *Am. J. Clin. Pathol.* 2012, 138, 770–780.
40. Kurman, R.J.; Shih, I.M. Pathogenesis of ovarian cancer: Lessons from morphology and molecular biology and their clinical implications. *Int. J. Gynecol. Pathol.* 2008, 27, 151–160.

41. Beck, A.H.; Sangoi, A.R.; Leung, S.; Marinelli, R.J.; Nielsen, T.O.; Van De Vijver, M.J.; West, R.B.; Van De Rijn, M.; Koller, D. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* 2011, 3, 108ra113.
42. Meirovitz, A.; Nisman, B.; Allweis, T.M.; Carmon, E.; Kadouri, L.; Maly, B.; Maimon, O.; Peretz, T. Thyroid hormones and morphological features of primary breast cancer. *Anticancer. Res.* 2022, 42, 253–261.
43. do Nascimento, R.G.; Otoni, K.M. Histological and molecular classification of breast cancer: What do we know? *Mastology* 2020, 30, 1–8.
44. Gamble, P.; Jaroensri, R.; Wang, H.; Tan, F.; Moran, M.; Brown, T.; Flament-Auvigne, I.; Rakha, E.A.; Toss, M.; Dabbs, D.J.; et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun. Med.* 2021, 1, 14.
45. Oyelade, O.N.; Ezugwu, A.E. A novel wavelet decomposition and transformation convolutional neural network with data augmentation for breast cancer detection using digital mammogram. *Sci. Rep.* 2022, 12, 5913.
46. Mohammed, M.; Mwambi, H.; Mboya, I.B.; Elbashir, M.K.; Omolo, B. A stacking ensemble deep learning approach to cancer type classification based on TCGA data. *Sci. Rep.* 2021, 11, 15626.
47. Triantafyllou, A.; Dovrolis, N.; Zografos, E.; Theodoropoulos, C.; Zografos, G.C.; Michalopoulos, N.V.; Gazouli, M. Circulating miRNA expression profiling in breast cancer molecular subtypes: Applying machine learning analysis in bioinformatics. *Cancer Diagn. Progn.* 2022, 2, 739.
48. Almarzouki, H.Z. Deep-learning-based cancer profiles classification using gene expression data profile. *J. Healthc. Eng.* 2022, 2022, 4715998. [Google Scholar] [CrossRef]
49. Aziz, R.M. Nature-inspired metaheuristics model for gene selection and classification of biomedical microarray data. *Med. Biol. Eng. Comput.* 2022, 60, 1627–1646.
50. Ogundokun, R.O.; Misra, S.; Douglas, M.; Damaševičius, R.; Maskeliūnas, R. Medical internet-of-things based breast cancer diagnosis using hyperparameter-optimized neural networks. *Future Internet* 2022, 14, 153.
51. Athanasopoulou K., Michalopoulou V.-I., Scorilas A., & Adamopoulos P. (2025). Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions. *Current Issues in Molecular Biology*, 47(6), 470. This review discusses how AI is being integrated into NGS pipelines, covering diagnostic and data-processing workflows.
52. Singh R., et al. (2025). Transforming Pharmacogenomics and CRISPR Gene Editing with the Power of Artificial Intelligence for Precision Medicine. *Pharmaceutics*, 17(5), 555. This paper examines the convergence of AI, pharmacogenomics, and CRISPR gene editing in personalized therapy.
53. Xin L., Huang C., Li H., et al. (2024). Artificial Intelligence for Central Dogma-Centric Multi-Omics: Challenges and Breakthroughs. *arXiv pre-print*. This work reviews AI/ML approaches for integrating multi-omics (genomics, transcriptomics, proteomics) data in disease research and precision medicine.
54. Tahir M., Norouzi M., Khan S.S., et al. (2025). Artificial Intelligence and Deep Learning Algorithms for Epigenetic Sequence Analysis: A Review for Epigeneticists and AI Experts. *arXiv pre-print*. Focuses on AI models applied to epigenomic sequence data (which is closely linked to genomic/diagnostic applications).

55. Al-Mahdi A. (2024). AI in Precision Medicine: Tailoring Treatments to Individual Genomes. *Journal of Biochemistry and Biotechnology*, 7(6), 235. This commentary/discussion article addresses how AI is being applied to individual-genome data for precision therapeutic applications.
56. Genomics & Informatics. (2024). Rare disease genomics and precision medicine. Vol 22, Article 28. This review covers how genomic technologies plus AI/ML are being used in rare-disease diagnosis and precision treatment.
57. Akhtar Z.B. & Rawol A.T. (2025). Advanced Methods Towards Functional Genomics & Medical Informatics: The Artificial Intelligence (AI) and Integrated Computing Perspectives. *Trends in Telemedicine & e-Health*, 5(5), Article 000622. This article gives an overview of AI in functional genomics and medical informatics with relevance to diagnostic genomics.
58. Sun S. (2025). Advancing Genetic Engineering Through AI: Sequencing and Editing Innovations. In *Proceedings of ICMMGH 2025 Workshop: Computational Modelling in Biology and Medicine*. This work touches on how AI supports gene-editing workflows (target detection, editing optimization) allied to genomics.
59. Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions — Dixit S., Kumar A., Srinivasan K., Vincent PMDR & Ramu Krishnan N. (2024) *Frontiers in Bioengineering & Biotechnology*. This review covers AI models for gRNA design, off-target prediction, editing outcome modelling in the context of genome-editing technologies.
60. Revolutionizing personalized medicine with generative AI: a systematic review — Ghebrehiwet I., Zaki N., Damseh R. et al. (2024) *Artificial Intelligence Review*, 57:128. Focuses on deep generative / transformer-based models in precision/personalized medicine including genomic & multi-omic data.
61. A systematic analysis of deep learning in genomics and histopathology for precision oncology — Unger M., Kather J.N. (2024) *BMC Medical Genomics*, 17:48. Highlights deep-learning models integrating genomic and histopathology data for cancer diagnosis & precision oncology.
62. Transitioning from wet lab to artificial intelligence: a systematic review of AI predictors in CRISPR — (2025) *Journal of Translational Medicine*, 23:153. Reviews AI/ML models aimed at CRISPR/gRNA design and outcome prediction, moving genome editing workflows into AI-driven space.
63. Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine — (2022) although slightly earlier; but still relevant. Surveys ~32 ML/AI approaches applied to variant & expression data for personalized medicine.
64. Rare disease genomics and precision medicine — (2024) *Genomics & Informatics*, 22:28. Reviews how genomics + AI/ML are applied in rare-disease diagnosis and precision therapies.
65. Machine Learning in Genomics: Applications in Whole Genome Sequencing, Whole Exome Sequencing, Single-Cell Genomics, and Spatial Transcriptomics — Adeyanju S.A., Ogunjobi T. (2024) *MEDINformatics*. Discusses ML applications in WGS/WES, single-cell & spatial transcriptomics relevant to genomic diagnosis pipelines.
66. A comprehensive review of AI applications in personalized medicine — Okolo C.A., Olorunsogo T., Babawarun O. (2024) *International Journal of*

Science and Research Archive, 11(1), 2544–2549. Overview of the intersection of AI, genomics and personalized medicine.

67. Yang C. Fann et al. (2025) Unlocking precision medicine: clinical applications of integrating health records, genetics, and immunology through artificial intelligence. *J. Biomed. Sci.*, 32:16.

68. “Revolutionizing personalized medicine using artificial intelligence: a meta-analysis of predictive diagnostics and their impacts on drug development” (2025) *Clin. & Exp. Med.* 25:255.

69. “Revolutionizing personalized medicine with generative AI: a systematic review” (2024) *Artificial Intelligence Review*, 57:128.

70. Dixit S., Kumar A., Srinivasan K., Vincent P.M.D.R. & Ramu Krishnan N. (2024) Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions. *Front. Bioeng. Biotechnol.* 11:1335901.

71. Ali S., Qadri Y.A., Ahmad K., Lin Z., Leung M-F., Kim S.W., Vasilakos A.V., & Zhou T. (2025) Large Language Models in Genomics—A Perspective on Personalized Medicine. *Bioengineering*, 12(5):440.

72. Abdelwahab O. & Torkamaneh D. (2025) Artificial intelligence in variant calling: a review. *Front. Bioinform.*, 5:1574359.

73. Aljarallah N.A., Dutta A.K., & Sait A.R.W. (2024) A Systematic Review of Genetics- and Molecular-Pathway-Based Machine Learning Models for Neurological Disorder Diagnosis. *Int. J. Mol. Sci.*, 25(12):6422.

74. Shin, D. (2024) Prediction of metabolic syndrome using machine learning approaches based on genetic and nutritional factors: a 14-year prospective-based cohort study. *BMC Med Genomics*, 17:224.

75. Unger M. & Kather J.N. (2024) A systematic analysis of deep learning in genomics and histopathology for precision oncology. *BMC Med Genomics*, 17:48.

76. Chen L., Liu G. & Zhang T. (2024) Integrating machine learning and genome editing for crop improvement. *aBIOTECH*, 5:262-277.

77. HumGenomics (2024) Harrison J.E. et al. Analysis of public perceptions on the use of artificial intelligence in genomic medicine. *Hum Genomics*, 18:128.