GMR
The Original

AI-ENHANCED DATA SCIENCE ARCHITECTURES: A REVIEW OF AUTOMATED LEARNING, ANALYTICS, AND DECISION SUPPORT

Vrushali S. Takate[1], Mrs. Manisha D. Balsaraf[2]

[1]Assistant Professor, Department:Instrumentation and Control Engineering, Pravara Rural Engineering College,Loni., vrushali.takate30@gmail.com, ORCID ID: https://orcid.org/0000-0002-7632-5528?lang=en
[2]Assistant Professor, Automation and Robotics Engineering, Pravara Rural Engineering College, Loni, manisha.balsaraf@pravara.in, ORCID :0000-0003-1839-9227

ABSTRACT
The explosion of genomic, transcriptomic, proteomic, metabolomic and other multi-omics data has resulted in many opportunities to advance understanding of complex biological processes and a significant computational and analytical challenge. To overcome such challenges, the use of AI powered data science architectures is growing, enabling automated learning, scalable data integration, predictive modelling, and decision-support systems. This review explores the place of AI-enabled architectures in genomic and molecular studies focusing on data collection and pre-processing, computational infrastructure, machine learning, deep learning, automated learning and biological interpretation. AI-based multi-omics integration, genomic prediction, biomarker discovery, cancer genomics, pharmacogenomics, molecular diagnostics and precision medicine are discussed. It also features discussions about emerging techniques, such as transformers, graph neural networks, generative AI, multimodal learning, biological foundation models, and federated learning, and how they can enhance molecular representation and clinical translation. Although significant progress has been made, significant limitations remain around model interpretability, model reproducibility, population bias, data heterogeneity, genomic privacy and external validation. There should thus be a focus in future genomic systems using AI for transparency in modelling, standardisation of analysis workflow, biologically meaningful interpretation and clinically responsible deployment. In sum, AI-driven data science architectures hold significant potential to revolutionize the way complex molecular data is analysed to yield meaningful insights into biology and precision-driven decision-making.

KEYWORDS: Artificial intelligence; Genomics; Multi-omics; Machine learning; Precision medicine

1. INTRODUCTION

Advances in high throughput sequencing and computational biology have dramatically changed the field of genetics and molecular studies due to the high amount of complex biological information produced. In this context, the growing use of high-dimensional molecular datasets such as transcriptomics, proteomics, metabolomics, epigenomics, and whole genome sequencing, is increasingly used to characterize biological variation, disease mechanisms, and functional relationships. While these technologies have enhanced the level of detail with which biological systems can be studied, they have also introduced significant challenges on data volume, data heterogeneity, data dimensionality, data interpretation and data integration.

As a result, the use of artificial intelligence (AI) and sophisticated computational techniques able to detect patterns in vast amounts of molecular data have increasingly been adopted in addition to conventional statistical and bioinformatics techniques. Bioinformatics has thus become a pivotal tool in bridging the gap between intricate biological data and actionable scientific and translational insights, accompanied by advancements in AI to facilitate this process.

Incorporation of bioinformatics into AI has therefore become an important tool in translating complex biological data into meaningful scientific and translational insights (Eskandar, 2025). Intelligent computational infrastructures are becoming even more necessary due to advances in molecular life sciences. The next generation of sequencing technologies and multi-omics platforms can provide information across multiple layers of biology in a single experiment and enable scientists to explore relationships between genome variation, gene expression, protein, metabolites, and phenotypes. These are part of the wider shift to data-intensive molecular science where biological experimentation becomes increasingly linked with computational modelling and automated analytical systems. AI-driven approaches can help speed up the analysis of such datasets by detecting nonlinear relationships, uncovering salient molecular features, and enabling integrative analyses that are challenging to achieve with traditional methods (Keenan et al., 2025).

The integration of these technologies, such as sequencing, multi-omics, and AI, is therefore transforming the landscape of genetics and molecular biology. In recent years, machine learning and other forms of AI have gained significance in the field of genomic data analysis due to the high number of variables when compared to the number of observations in genomic datasets. Single datasets can contain millions of genetic variants, thousands of gene-expression measurements or

many molecular interaction features, challenging traditional analytical workflows to compute or statistically analyze such datasets.

AI techniques can aid in Dimensionality Reduction, Classification, Feature Selection, Pattern Recognition, Predictive Modelling, and Functional Interpretation. These are also used more and more in the identification of disease-associated variants, prediction of gene function, molecular diagnostics, discovery of biomarkers, population genomics and personalized medicine. Examples of recent advances in bioinformatics show how AI is not just a tool for analysis but increasingly a core part of the data processing and interpretation workflows for genomic data in modern research and practice (Olawade et al. 2025).

The growing impact of AI in the field of genomics is evident in the variety of algorithms used in processing genetic and molecular data. These machine learning, deep neural networks, representation learning and other sophisticated computational models can handle complex sequence information and uncover associations that might not be apparent from traditional analysis methods. These technologies could be used to enhance genomic prediction, variant interpretation, functional annotation and the modelling of gene–phenotype relationships. Thus, the integration of AI into structured analytical frameworks that can handle various stages of data lifecycle management could prove significant, especially when these systems are embedded in larger, more complex biological knowledge management environments (Shaikh et al., 2026).

In addition to genome analysis, AI models are gaining significance in precision medicine, where molecular data needs to be integrated with medical and phenotypic data to inform personalized forecasts and recommendations for care. With the recent emergence of machine learning models predicting patient outcomes in precision medicine, it is important to develop computational models that combine multiple data sources in a reliable and biologically meaningful way. These models could be used to predict disease risk, patient classification, therapeutic choice and prediction of the drug's effect. But to achieve their effective application, integrated data science architectures are needed, where data is acquired, pre-processed, feature extracted, and then used to build a model, for the subsequent biological interpretation, and then to make decisions (Wang and Tang, 2025). However, there are still significant hurdles in the implementation of AI-integrated genomic and molecular systems such as biological heterogeneity, small sample sizes, population bias, interpretability of models, reproducibility, data privacy, and the integration of different omics platforms. These restrictions highlight the importance of assessing AI as a tool within a wider data science infrastructure that can be used to provide strong analytical foundations with comprehensible and meaningful biological inferences.

As these technologies, such as artificial intelligence, genomics, bioinformatics and multi-omics, converge, they offer a solid foundation for exploring the potential of automated analytical systems for fundamental molecular research and for translation to practical applications. This review, therefore, seeks to critically explore the architectures of AI applications in data science for genomics and molecular research, focusing on the automated learning, integrated molecular analytics, and decision support applications. The review aims to describe the genomic and molecular data ecosystem, analyse the main computational architectures and AI-based analytical strategies for data interpretation, explore their potential in multi-omics data analysis and precision medicine, and outline some future technological trends, methodological challenges and opportunities of creating more explainable, reproducible and clinically useful AI systems for genomic data.

2. Genomic and Molecular Data Ecosystem

Today, a vast high-dimensional data system is being created, yielding multiple types of data on many levels of biological processes. The combination of genomic, transcriptomic, epigenomic, proteomic, metabolomic, single-cell, and spatial data yields a more holistic perspective of the molecular mechanisms underlying biology than any single type of data does on its own. This diversity poses significant challenges with respect to scale, heterogeneity, data harmonization, and computational interpretation, however. The integration of multiple omics has thus emerged as a prime area of focus in modern molecular research, as it allows linkages between genetic variation and downstream molecular and phenotypic effects (Savitha et al., 2025).

2.1 Genomic and Transcriptomic Data

Genomic data give information on variations in DNA sequence, such as single nucleotide variants, insertions and deletions, copy number changes and structural rearrangements. These data are vital for research in disease susceptibility, molecular evolution, functional genetics and precision medicine. In addition to genomics, transcriptomic datasets provide information on gene-expression patterns and regulatory activity in tissues, disease conditions and experimental conditions. Their integration enables researchers to investigate the relationship between genetic changes to transcriptional changes and downstream biological pathways. With the rise of the use of AI in multi-omics studies, the analytical bridge between variation of the genome and variation of the transcriptome has been reinforced, especially in the field of cancer biology. AI-driven integration can enrich the classification of biologically heterogeneous disease subtypes (Hulwan, 2026) and detect molecular signatures that would be hard to identify in the absence of AI-based integration with other datasets. These are especially useful when combined with genomic and transcriptomic data to identify molecular patterns that are relevant to function.

2.2 Epigenomic, Proteomic, and Metabolomic Data

Epigenomic data refer to changes in regulation that affect the activity of the genes without changing the underlying DNA sequence. These include methylation of DNA, modifications of histones, and accessibility of chromatin. Proteomic data provide molecular information on protein abundance, structure, modification and interaction, while metabolomic data reflect molecular products of cellular processes, and therefore provide data on the metabolic function.

This in itself makes metabolomics a growing key element in multi-omics research, with its role in bridging the gap from molecular regulation to physiological and disease-related phenotypes. In the era of integrating metabolomics data with genomic, transcriptomic, and proteomic data, metabolomics-based integration can identify pathway disruptions and molecular signatures linked to complex diseases. The comprehensive analysis of nutritional and metabolic multi-omics data using AI/ML has also shown promising potential for the identification of disease-associated molecular patterns and the establishment of translational frameworks in precision oncology (Alnuaim et al., 2026).

2.3 Single-Cell and Spatial Omics Data

Single cell technologies have broadened the scope of molecular studies and allowed the study of biological variation at the level of the cell. Single-cell studies allow to detect rare cell populations, characterize cellular heterogeneity, and uncover dynamic shifts in gene expression in complex tissues, in contrast to bulk sequencing methods. Spatial omics then builds on this ability by keeping the spatial context of molecular signals, so that molecular profiles can be directly mapped to the anatomy.

In the field of cancer research, the integration of high-resolution and spatial molecular data into AI-driven analytical frameworks is especially relevant, given the complex interplay between the heterogeneity of cancer cells and their microenvironment, which significantly impacts disease progression and patient response to therapy. AI fusion strategies can be implemented to combine multi-omics and medical imaging, creating more comprehensive representations of tumour biology and enhancing molecular interpretation (Marouf et al., 2025).

2.4 Multi-Omics Data Integration Challenges

Multi-omics integration is an approach to understanding biological processes at a systems level, but with significant methodological challenges. The distribution, dimension, scale and missingness of data vary depending on the different omics platforms. Cross-platform integration is further complicated by the fact that there are differences in sample preparation, sequencing technology, normalization procedures, and batch effects. Therefore, computational architectures must, at least, maintain biologically meaningful relationships, and also harmonize heterogeneous datasets. There are several approaches to integrate multiple molecular layers at the feature level, model level, and representation level with the help of AI. But strong preprocessing, suitable dimensionality reduction, data standardization, and biologically interpretable modelling for successful integration. Multi-omics analysis combined with AI is increasingly proposed to extract meaningful molecular relationships from complex data and can be integrated into frameworks for better results. (Kalpar,2025)

Given the increasing complexity of genomics, proteomics, metabolomics and multimodal data, the challenge of scalable and transparent AI-driven architectures that can process and convert diverse molecular information into comprehensive biological insight is more crucial than ever. The genomic and molecular data ecosystem includes different layers of omics data such as genomics, transcriptomics, epigenomics, proteomics, metabolomics, single-cell, and even spatial data, and can be combined in an AI-supported approach to tackle heterogeneity, missing data, batch effects, and complexity of computation (Figure 1). The major types of molecular data (genomics, transcriptomics, proteomics, metabolomics, single cell/spatial omics and integrated multi-omics) are presented in Table 1, along with their main data outputs and their analytical relevance.

Table 1. Major Molecular Data Types and Their Analytical Relevance

Data type	Primary information generated	Major analytical relevance	Supporting reference
Genomics	DNA sequence and genetic variation	Variant analysis, disease susceptibility, functional genetics	(Hulwan, 2026)
Transcriptomics	Gene-expression profiles	Regulatory analysis and molecular classification	(Savitha et al., 2025)
Proteomics	Protein abundance and interactions	Functional pathway and biomarker analysis	(Savitha et al., 2025)
Metabolomics	Small-molecule and metabolic profiles	Phenotyping and pathway characterization	(Wang et al., 2026)
Single cell/spatial omics	Cell-specific and spatial molecular profiles	Cellular heterogeneity and tissue-context analysis	(Marouf et al., 2025)
Integrated multi-omics	Combined molecular layers	Systems-level interpretation and precision applications	(Alnuaim et al., 2026)

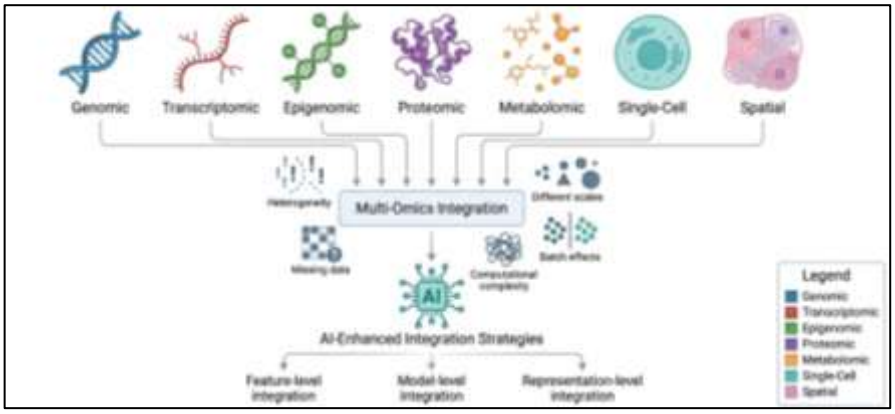

Figure 1. Genomic and Molecular Data Ecosystem: Multi-Omics Generation, Integration, and AI-Driven Analysis

3. AI-Enhanced Data Science Architectures for Genomic Research

AI-powered architectures of data science enable the computation necessary to convert large, diverse genomic data into biological knowledge. Such architectures usually include data acquisition, data pre-processing, data storage, machine learning, and data interpretation as coherent analysis pipelines. In the field of genomic research, these systems are becoming more crucial, as the data generated by sequencing, imaging, and multi-omics platforms can vary widely in terms of their format, dimensionality, and biological context. Artificial intelligence in next-generation sequencing has been making it easier to deal with intricate genomic data and identify significant molecular patterns, as well as to aid in downstream biological interpretation (Athanasopoulou et al., 2025).

3.1 Data Acquisition, Preprocessing, and Quality Control

The quality of the genomic analysis with AI relies heavily on the reliability of the data it is fed. The data that can be acquired ranges from next-generation sequencing, transcriptomic profiling, multi-omics assays, clinical records, imaging systems and experimental phenotypic measurements. Typically, genomic data needs to be subjected to multiple preprocessing steps before they can be used for model development, such as assessment of data quality, alignment, normalization, filtering, imputation, feature selection, and dimensionality reduction.

Increasingly, AI is being integrated into sequencing pipelines to increase the accuracy of base calling, variant detection, sequence interpretation and quality control. These advances can help to minimize manual steps and facilitate the automated processing of large-scale molecular data (Athanasopoulou et al., 2025). Furthermore, preprocessing should also address the issue of harmonizing genomics, medical imaging, and Ehr data with differences in structure and scale in multimodal precision-medicine systems (Khan et al., 2025).

3.2 Data Storage, Management, and Computational Infrastructure

Computational infrastructures that can store, process and retrieve large amounts of data efficiently are needed for large-scale study of genomics and multi-omics. Baseline computing environments might not be enough for analyses of whole genomes, high resolution images, longitudinal clinical data or multiple omics layers. As a result, scalable architectures are more and more embracing cloud computing, distributed storage, high-performance computing and parallel processing. In the field of cancer research, AI-driven big-data frameworks are especially pertinent as they allow for the seamless incorporation of genomic data with clinical and treatment factors.

These systems can facilitate fast analysis and allow researchers to perform data-heavy systems throughout the investigative process (Kumarachari et al., 2026). The need for similar architectures is also increasing in the field of plant genomics, where AI-based prediction models for crop improvement and genomic selection of large-scale molecular and phenotypic data are becoming common (Mukherjee et al., 2026).

3.3 Machine Learning, Deep Learning, and Automated Learning Layers

The analytical intelligence layer is the heart of genomic architectures that integrate AI. Machine-learning models can be used for classification, regression, clustering, feature prioritization, and pattern recognition; and deep-learning models can learn hierarchical representations directly from high-dimensional data in biology. Automated learning algorithms also minimize the need for manual feature, algorithm, and hyperparameter selection. Multimodal AI systems are particularly useful due to their ability to integrate genomic data with imaging and clinical information to create more holistic patient profiles and enhance diagnostic or predictive capabilities (Khan et al., 2025).

Personalized medicine also benefits from a multimodal integration of the heterogeneous information sources within an integrated predictive framework through AI (Manivannan et al., 2025). The application of deep learning methods to multi-omics pipelines in drug discovery has been gaining traction, with the potential to uncover intricate relationships between molecular attributes and therapeutic targets (Vora et al., 2025).

3.4 Biological Interpretation and Knowledge Integration

Predictive performance is not enough for genomic research without connecting the predictions of the model to genes, pathways, regulatory mechanisms, clinically relevant molecular processes, etc. Biological interpretation is thus a key

architectural component that links computational predictions with mechanistic understanding. This can include functional annotation, pathway analysis, molecular interaction networks, gene-set enrichment or incorporation with other knowledge bases. AI-driven data science architectures can augment this process by creating a connection from predictive signals to interpretable molecular features. These methods can enable genomic prediction in plants to correlate genomic variation to complex agronomic traits and inform plant breeding with biological approaches (Mukherjee et al., 2026).

In the context of drug discovery, the fusion of multi-omics data with deep-learning algorithms can facilitate the linking of molecular signatures to potential therapeutic targets, pathways, and opportunities (Vora et al., 2025). Thus, the most successful genomic architectures are ones that are scalable, but that can also be understood in terms of biological knowledge, rather than AI as a black box predictor. For an AI-enabled genomic data science architecture, the data is acquired, pre-processed, processed through a computational infrastructure, analysed with AI intelligence and then interpreted biologically, illustrated in Figure 2, can be organized into a series of interconnected layers to convert heterogeneous genomic data into biologically meaningful insights. The key components of architectures of AI enhanced genomic data science include the acquisition of data, integration of multimodal data, computational infrastructure, analytical intelligence, genomic prediction and deep learning based multi-omics analysis as shown in Table 2.

Table 2. Core Components of AI-Enhanced Genomic Data Science Architectures

Architectural component	Principal function	Typical genomic application	Supporting reference
Data acquisition and preprocessing	Data cleaning, alignment, normalization, and quality control	NGS analysis and variant processing	(Athanasopoulou et al., 2025)
Multimodal data integration	Harmonization of heterogeneous data types	Genomics, imaging, and EHR integration	(Khan et al., 2025)
Scalable computational infrastructure	Storage and large-scale processing	Cancer genomic big-data analysis	(Kumarachari et al., 2026)
AI analytical layer	Classification, prediction, representation learning	Personalized medicine applications	(Manivannan et al., 2025)
Genomic prediction systems	Learning genotype–phenotype relationships	Plant genomic prediction	(Mukherjee et al., 2026)
Deep-learning multi-omics analysis	Molecular pattern and target discovery	AI-assisted drug discovery	(Vora et al., 2025)

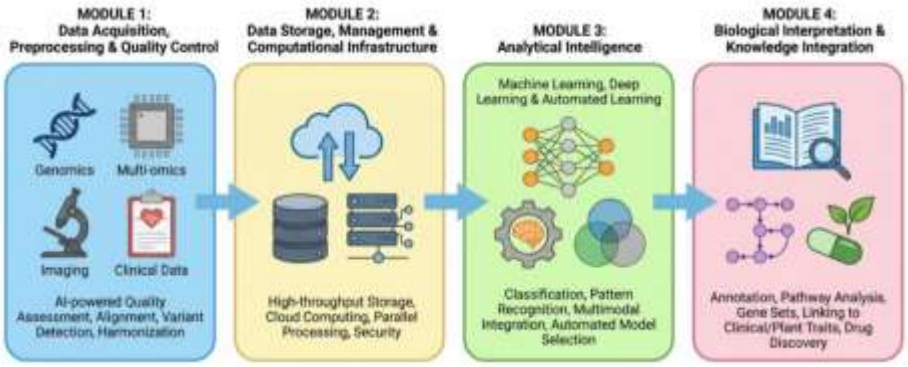

Figure 2. AI-Enhanced Data Science Architecture for Genomic Research: From Data Acquisition to Biological Interpretation

4. Automated Learning and AI-Driven Molecular Analytics

AI has significantly boosted the analytical power of genomic and molecular research, facilitating automated pattern recognition, feature selection, predictive modelling, and the integration of high-dimensional data. Such methods are particularly useful when the data set contains a high dimensionality of variables, nonlinear interaction, and appreciable biological variability. Machine learning and deep learning can be used to discover key genomic signatures, identify disease states, prioritize molecular features, and aid in multi-omics interpretation. The increasing use of these technologies in fields such as cancer genomics, spatial omics, immunogenomics and plant genomics highlights the wide range of relevance of AI-powered molecular analytics to current genetics research.

4.1 Machine Learning and Deep Learning for Genomic Prediction

Genomic data is being used increasingly to predict outcomes that are biologically and clinically relevant using machine-learning models. Supervised algorithms can be used to classify samples by molecular subtype, disease status, therapeutic response or genomic risk, while unsupervised learning can identify novel patterns within complex and heterogeneous molecular datasets.

Deep-learning models have been developed to enhance these functions by learning complex representations automatically from high-dimensional genomic and omics inputs. Cancer Genomics: AI models have been used to identify genes involved in multidrug resistance, offering a computational way of recognizing molecular determinants that can affect treatment failure (Anand, 2026). Machine learning and artificial intelligence more generally are increasingly being used in precision oncology to synthesise complex molecular signals and to aid in predicting disease behaviour and treatment outcomes (Moar et al., 2026).

4.2 Automated Machine Learning, Transfer Learning, and Self-Supervised Learning

The goal of automated learning techniques is to minimize the need for manually building models, designing features, selecting algorithms, tuning hyperparameters, and evaluating the model. These are especially valuable in the case of thousands of genomic variables in molecular datasets, and when the optimal configurations for prediction are hard to develop through traditional experimentation. Insert also the topics of transfer learning and self-supervised learning: many genomic studies have small numbers of labelled samples.

Models learned from larger biological datasets can be applied to smaller disease-specific datasets; similarly, self-supervised approaches can be used to learn useful molecular representations without extensive annotation. Multi-omics systems using neural networks illustrate how automated representation learning can be applied in the context of diagnostic screening and biomarker discovery for complex diseases like breast cancer (Kumari et al., 2026). These methods can enhance the analytical scalability and lessen manual work in feature building.

4.3 AI Applications in Variant, Gene Expression, and Regulatory Analysis

Molecular analytics, which are based on artificial intelligence, are used more and more to detect genetic mutations, gene expression signatures and regulatory elements linked to a disease phenotype and to biological processes. Computational models can select potentially functional genes, identify molecular patterns associated with drug resistance and separate clinically relevant genomic profiles from background variation. The use of AI-based genomic analysis has been particularly promising in the field of cancer, where predictive models can help identify the genetic markers associated with therapeutic resistance.

In the context of cancer, AI-driven genomic analysis has also been used to identify multidrug-resistance genes, with predictive models to aid in the recognition of genetic determinants of therapeutic resistance (Anand, 2026). Furthermore, spatial multi-omics can integrate the molecular profile with the tissue context, enhancing the comprehension of tumour heterogeneity, organization, and regulatory patterns specific to the disease (Shao et al., 2024). Beyond human disease, genomic and multi-omics data are increasingly being combined with crop improvement strategies to uncover the traits of environmental adaptation and climate resilience (Syeda, 2025).

4.4 AI-Based Multi-Omics and Systems Biology Analytics

One of the biggest applications of AI is multi-omics analysis, which examines the interactions between multiple layers of the molecular landscape and is vital for understanding complex phenotypes, which are usually driven by multiple types of molecular changes. AI-based systems can combine all of these different types of information together, including genomic, transcriptomic, proteomic, metabolomic, and spatial information, to create a more comprehensive model of the biological system and to identify coordinated molecular signatures. Multi-omics integration with neural networks in breast cancer has shown promise for diagnostic screening and biomarker discovery, leveraging a variety of molecular features in a unified predictive model (Kumari et al., 2026).

AI-driven multi-omics models are also contributing to the prediction of response and resistance to immune checkpoint inhibitors, demonstrating their potential in understanding the molecular mechanisms of treatment and aiding in precision treatment strategies (Wang et al., 2026). Likewise, incorporating AI, ML, and multi-omics into precision oncology is seen as a key factor in next-generation precision medicine due to the ability to interpret the complex, heterogeneous nature of biological data on a systems level (Moar et al., 2026). Table 3 shows the significant applications of AI-based molecular analytics such as genomic prediction, resistance-gene analysis, multi-omics neural networks, spatial multi-omics, agricultural genomics, and immunogenomic prediction.

Table 3. Major AI-Driven Molecular Analytics Applications

Analytical area	Principal AI function	Representative application	Supporting reference
Genomic prediction	Classification and risk modelling	Cancer and precision oncology	(Moar et al., 2026)
Resistance-gene analysis	Feature prioritization and prediction	Multidrug-resistance gene identification	(Anand, 2026)
Multi-omics neural networks	Representation learning	Breast cancer screening and biomarker discovery	(Kumari et al., 2026)
Spatial multi-omics	Context-aware molecular integration	Lymphoma precision medicine	(Shao et al., 2024)
Agricultural genomics	Genotype–phenotype learning	Climate-resilient crop improvement	(Syeda, 2025)
Immunogenomic prediction	Multi-omics response modelling	Immune checkpoint inhibitor response	(Wang et al., 2026)

5. AI-Enabled Decision Support in Genetics and Precision Medicine

With the advent of AI, genomic and molecular information is being translated to clinical decisions at a greater rate. Precision medicine approaches include decision-support systems that combine molecular markers, genomic data, patient characteristics, and treatment-related factors to aid in diagnosis, risk stratification, prognosis, and therapeutic choice. With the increasing volume of multi-omics data available, these systems have been reinforced to give a more holistic picture of the biology of the disease. AI-driven platforms have the potential to uncover intricate relationships between molecular aspects and clinical outcomes, enabling personalized healthcare approaches and genomic translation.

5.1 Molecular Diagnostics and Biomarker Discovery

AI has proven to be a key technology in the identification of molecular biomarkers, which have diagnostic, prognostic and predictive value. Machine-learning models can identify subtle molecular signatures, such as disease subtypes and/or treatment response, by analysing high-dimensional genomic and multi-omics datasets. AI-driven genomic biomarker identification can enhance the ability to identify clinically relevant variants and molecular patterns that influence tumour behaviour, which is a crucial aspect in the field of precision oncology (Judijanto & Nurhayati, 2026).

Biomarker discovery is further bolstered by the integration of genomic, transcriptomic, proteomic and other molecular data into AI-based analytical models through the lens of translational bioinformatics. These can help to identify strong molecular markers which can be used to improve diagnosis and patient stratification (Kara & Gunel, 2026). Smart biosensor platforms and AI-driven multi-omics analysis are also the opportunities for real-time or minimally invasive detection of biomarkers in precision oncology (Akpe & Cock, 2026).

5.2 Cancer Genomics and Disease Risk Prediction

The genetic and molecular diversity of tumours makes cancer genomics one of the most mature domains of the AI field of decision support. These AI models can process genomic mutations, gene-expression patterns, and biomarker profiles to distinguish between different types of tumours, assess disease prognosis, and pinpoint patients for specific treatments. Genomic data can be used to identify potential biomarkers that can be used to optimize treatments, as they can be associated with a specific molecular profile and predicted therapeutic response (Kumar & Metta, 2024).

In addition to oncology, AI decision support systems are now being used for complex immunologic diseases. These systems have the potential to combine the power of omics data, clinical variables, and extensive biomedical data to enhance disease classification, risk prediction, and tailored therapeutic strategies for individual patients (Jariwala, 2025). These applications are just some examples of how AI can translate complex molecular data into clinically relevant information.

5.3 Pharmacogenomics and Personalized Medicine

Pharmacogenomics is the study of the influence of genetic variation on drug response, efficacy and toxicity. AI can be used to improve pharmacogenomic analysis, by finding genotype–drug relationships, predicting response to a drug, and recommending drugs based on a patient's molecular profile. However, the use of AI in Pharmacogenomics is growing in significance as it enables the utilization of genomic information to make personalized treatment choices (Jakka et al., 2025).

The benefits of AI-based personalized medicine are not constrained to pharmacogenomics alone; it also involves incorporating molecular, clinical, and phenotypic data into predictive models. These models can help identify patient subgroups, to estimate treatment benefit, and to minimize the time and expense of trial-and-error prescribing. With the ongoing growth of genomic data, AI-driven decision algorithms could facilitate more targeted treatment choices and tracking.

5.4 Clinical and Translational Genomic Decision Support

The true power of AI-powered genomic analysis lies in its ability to convert molecular information into effective clinical decisions. The use of clinical decision-support systems can integrate genomic, multi-omics, laboratory, imaging, and patient-level data to support diagnosis, prognosis, choice of treatment, and monitoring of a patient's condition. The path from molecular data analysis to clinical decision-making represents a wider trend towards the adoption of computational sciences in precision medicine where AI acts as a mediator between discovery and treatment (Primorac & Südhof, 2026). However, the clinical usability of effective translation requires clinically validated models, easy to interpret predictions and standardized workflows, as well as careful integration with professional judgement. AI should thus be used as an aid to analysis, not as a clinical replacement.

In an ideal genomic data science system, these systems can enhance the quality and personalization of molecularly informed decision making in health care. Table 4 illustrates the key areas where AI-driven decision support is currently being applied in genetics and precision medicine, highlighting the discovery of biomarkers, the use of translational bioinformatics, cancer genomics, pharmacogenomics, and clinical decision-making. AI-based decision support is presented in Figure 3 where genomic and multi-omics data, clinical variables, imaging, biosensors and laboratory results are combined to facilitate the diagnosis, risk stratification, prognosis, monitoring, and personalized therapy.

Table 4. Major Applications of AI-Enabled Decision Support in Genetics and Precision Medicine

Application area	AI-supported function	Translational relevance	Supporting reference

Molecular biomarker discovery	Identification of diagnostic and predictive molecular signatures	Precision oncology	(Judijanto & Nurhayati, 2026)
Translational bioinformatics	Multi-omics biomarker integration	Disease classification and stratification	(Kara & Gunel, 2026)
Smart biosensors	AI-assisted molecular detection	Precision oncology monitoring	(Akpe & Cock, 2026)
Cancer genomics	Predictive biomarker modelling	Personalized treatment optimization	(Kumar & Metta, 2024)
Immunological disorders	Omics-based decision support	Risk assessment and therapeutic planning	(Jariwala, 2025)
Pharmacogenomics	Prediction of genotype-specific drug response	Personalized medicine	(Jakka et al., 2025)
Clinical genomics	Molecular-to-clinical decision translation	Diagnosis and treatment selection	(Primorac & Südhof, 2026)

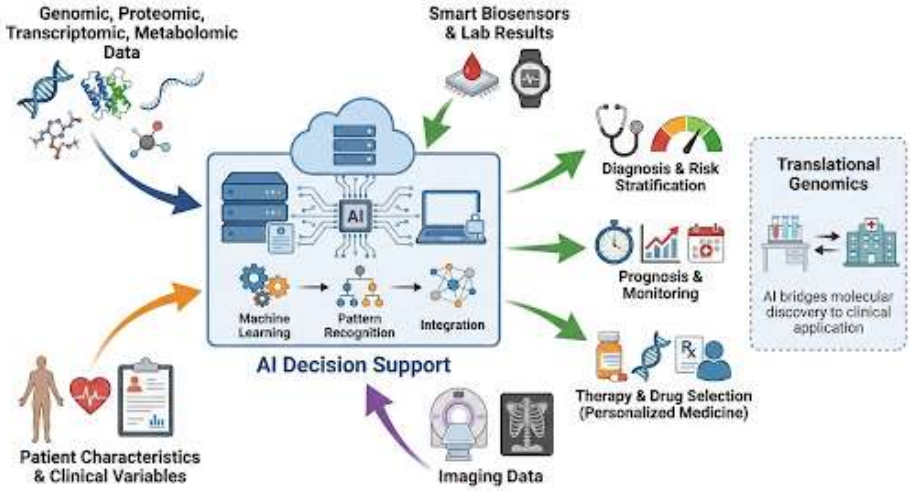

Figure 3. AI-Enabled Decision-Support Framework Integrating Multi-Omics, Clinical, Imaging, and Biosensor Data for Precision Medicine6. Emerging Architectures, Challenges, and Future Directions

The field of genomics and molecular research is undergoing significant transformation thanks to the swift progress in Artificial Intelligence. The integration of multi-omics imaging, deep learning, generative modelling, and precision medicine into single computing architectures has grown common across new architectures.

A unified computing system is emerging that brings together multi-omics integration, deep learning, generative modelling and precision medicine. Such methods go beyond the traditional prediction task and allow for representation learning, cross modal data fusion, target identification, and clinically relevant decision-making support. Their future implications will rely on the capacity to handle heterogeneous bio data, ensuring reproducibility, interpretability, privacy and translational validity.

6.1 Transformers, Graph Neural Networks, and Biological Foundation Models

As a result of this sequential, relational and networked nature of biological information, transformer-based models and graph neural networks are increasingly relevant to molecular research. Transformers are capable of capturing long-range dependencies in genomic and molecular sequences, while graph neural networks can encode the structure of genes, proteins, pathways and molecular interactions as graphs. These architectures have the potential to provide benefits in the domains of functional annotation, molecular representation learning, pathway analysis and therapeutic target identification.

AI's integration with multi-omics is also increasingly driving precision drug discovery by facilitating the development of models that uncover complex patterns between molecular profiles, disease processes, and prospects for new therapeutics (Liu et al., 2026). The use of sophisticated AI and machine-learning systems, which could combine patient-specific clinical data with genomic features to further enhance the accuracy of prediction, may further benefit personalized medicine (Paul, 2026). Thus, biological foundation models will be increasingly crucial for transferable genomic and biomedical applications, built from large scale molecular data.

6.2 Generative AI, Multimodal Learning, and Federated Genomic Learning

New opportunities with Generative AI for molecular design, drug repurposing, generation of synthetic data and biological hypothesis generation. Using generative architectures, molecular distributions can be learned and used to suggest new molecular structures, alternative drug targets and representations that can be leveraged in downstream prediction. The combination of AI in drug discovery with multi-omics data has been shown to have potential applications in the identification and repurposing of therapeutic candidates for complex diseases (Sabry et al., 2025).

Multimodal learning is also crucial since future precision-medicine systems will have to integrate genomics, transcriptomics, proteomics, metabolomics, imaging, and clinical data. Combining such information sources can produce more comprehensive patient representations than any one of these on its own. Another emerging direction for federated learning is that it enables models to be trained across geographically distributed datasets without the need to centrally share sensitive genomic data. These architectures can help to facilitate collaborative genomic studies and alleviate certain privacy and data sharing concerns.

6.3 Explainability, Reproducibility, Privacy, and Ethical Challenges

While the applications of AI-driven genomic systems are promising, there are ethical and methodological considerations to address. Complex deep-learning models can make predictive accuracy and only offer limited insight into the molecular features that underlie the predictions. Lack of interpretability can undermine biological validity and cause problems in the application of AI's recommendations to the clinical care pathway. For AI-driven personalized cancer care, it is crucial to have transparent models that connect predictive results to clinically relevant genomic markers and therapeutic targets (Sheikh et al., 2025).

Another significant hurdle is reproducibility, as AI models could yield varying results in different populations, sequencing platforms, institutions, and clinical settings. If there is bias in the training data, it can result in unequal prediction performance, especially when certain populations are not well-represented in the genomic data. One aspect of genomic privacy is the identifiability of the genetic information, and their potential to contain information about biological relatives. Thus, secure data governance, privacy-preserving computation, transparent model validation and external replication are to be integrated into future architectures.

6.4 Future Directions for AI-Integrated Genomic and Molecular Research

Next-generation genomic data science architectures will be more multimodal, automated, and clinically integrated. More molecular layers of integration can help integrate models from correlation-based disease prediction to systems level characterization of disease mechanisms. The recent developments of multi-omics integration and AI-based precision medicine demonstrate the increasing focus on the integration of heterogeneous biological information for enhanced diagnoses, prognosis, and precision treatment (Zhang, 2025). Additional studies are also needed to focus on the development of biologically interpretable AI, a standardized benchmarking, population-diverse data sets, and prospective clinical validation.

The convergence of AI-driven molecular predictions and multi-omics data will facilitate the identification of new targets and prioritize them for therapeutic action, which is likely to enhance precision drug discovery (Liu et al., 2026). Another facet of personalized medicine could involve the use of computational systems that will continually assimilate genomic, molecular, and clinical information to guide individual treatment decisions (Paul, 2026). Finally, the most useful architectures will be those that are both biologically valid and sophisticated in computation, are easy for the reader to follow, and are clinically appropriate.

7. CONCLUSION

To make sense of high dimensional, heterogeneous and fast-growing biological data, AI is increasingly embedded in genomic and molecular research. Data science architectures that integrate genomic, transcriptomic, epigenomic, proteomic, metabolomic, single-cell, spatial, imaging, and clinical data into coherent data analysis pipelines are being facilitated by the power of AI. These architectures can be used to enhance molecular pattern recognition, biomarker discovery, disease classification, genomic prediction, and therapeutic decision support through machine learning, deep learning, automated learning, and multimodal strategies. The increasing advancement of multi-omics-based, AI-assisted approaches has also enhanced the shift from descriptive molecular analysis to predictive and precision applications. AI-powered systems are particularly helping in molecular diagnostics, cancer genomics, pharmacogenomics, personalized medicine and prediction of treatment responses. New models like transformers, graph neural networks, generative AI, biological foundation models, and federated learning will further enhance the analytical power of all genomic studies with their advancements in representation learning, cross-modal integration, scalability, and privacy-preserving collaboration. Nevertheless, there are still significant concerns about the interpretability of the models, lack of reproducibility, population bias, data standardization, genomic privacy, and clinical validation. Future development should thus focus on biologically meaningful models, multi- and representative data sets, uniform benchmarking, secure management of data and a robust evaluation of translation. The future of AI-driven genomic data science will not just be about computational power but also the capacity to produce clinically actionable, repeatable, and biologically relevant insights. The well-designed AI architectures can thus serve as a crucial link in the chain of complex molecular data, mechanistic understanding and precision decision making.

REFERENCES

1. Akpe, V., & Cock, I. E. (2026). Artificial intelligence meets biosensors: smart platforms for precision oncology and multi-omics integration. Digital Discovery.
2. Alnuaim, L., AlGarni, A., Ishaqui, A. A., AlQahtani, N. M., Alshuaibi, M., Almansour, E. A., ... & Iqbal, Z. (2026). AI/ML-Integrated Nutritional and Metabolic Multi-Omics in Cancer: A Systematic Review, Meta-Analysis, and Proposed Translational Framework.

3. Athanasopoulou, K., Michalopoulou, V. I., Scorilas, A., & Adamopoulos, P. G. (2025). Integrating artificial intelligence in next-generation sequencing: advances, challenges, and future directions. Current Issues in Molecular Biology, 47(6), 470.
4. Eskandar, K. (2025). Bioinformatics meets AI: transforming big data into breakthroughs in biotechnology. Annales Universitatis Paedagogicae Cracoviensis Studia Naturae, 167-190.
5. G Anand, L. (2026). Artificial Intelligence–Driven Identification of Multidrug Resistance Genes in Cancer Genomics. ScienceOpen Preprints.
6. Hulwan, A. B. (2026). Integrating AI and Multiomics for Breakthroughs in Cancer Studies. Psychopathology Early Prediction and Classification: A Deep Generative AI Approach, 349-367.
7. Jakka, A. L., Chacko, R. M., Vasam, M., Alagarsamy, S., Chandragiri, S. S., Gavini, S. D., & Mathew, M. J. (2025). From genomics to clinic: the transformative impact of AI in pharmacogenomics and personalized medicine. Pharmacogenomics, 26(13-14), 573-585.
8. Jariwala, M. (2025). Ai-driven decision support systems for immunological disorders: Bridging big data, omics, and precision medicine. In AI-Assisted Computational Approaches for Immunological Disorders (pp. 353-392). IGI Global Scientific Publishing.
9. Judijanto, L., & Nurhayati, S. (2026, April). AI-Augmented Genomic Biomarker Identification for Precision Oncology: A Review. In 2026 IEEE 2nd International Conference on Robotics and Technologies for Industrial Automation (ROBOTHIA) (pp. 1-6). IEEE.
10. KALPAR, R. I. INTEGRATION OF MULTI-OMICS AND ARTIFICIAL INTELLIGENCE (2025).
11. Kara, K., & Gunel, T. (2026). Translational Bioinformatics and Multi-Omics Integration for AI-Driven Biomarker Discovery in Precision Oncology. Acta Infologica, 10(1), 1-20.
12. Keenan, L., Lorna Younger, R., Race, P. R., & Bawn, M. (2025). Molecular life sciences in the era of the Fourth Industrial Revolution: sequencing, multi-omics and artificial intelligence. Emerging Topics in Life Sciences, 9(2), 35-45.
13. Khan, S. N., Danishuddin, Khan, M. W. A., Guarnera, L., & Akhtar, S. M. F. (2025). Multi-modal AI in precision medicine: integrating genomics, imaging, and EHR data for clinical insights. Frontiers in Artificial Intelligence, 8, 1743921.
14. Kumar, A., & Metta, D. (2024). AI-driven precision oncology: predictive biomarker discovery and personalized treatment optimization using genomic data. Int J Adv Res Publ Rev, 1(3), 21-38.
15. Kumarachari, R. K., Tulasi, K. S. K., Rajasekhar, E. S. K., Neerugatti, D., Sowjanya, P., & Chandrakala, A. (2026). AI-Enhanced Big Data Analysis in Cancer Research and Treatment. In AI and Precision Medicine in Cancer Management (pp. 74-104). CRC Press.
16. Kumari, V., Tiwari, H., Singh, S., Kailashiya, V., Hoai, N. T. T., Sachan, S., ... & Gautam, V. (2026). AI-Driven Multi-Omics Integrated Applications Using Diverse Neural Networks for Breast Cancer Diagnostic Screening and Biomarker Discovery. Medicinal Research Reviews.
17. Liu, Y., Zhu, K., Peng, W., Liu, Z., & Mao, X. (2026). Multi-omics and artificial intelligence for precision drug discovery and potential clinical applications. Signal Transduction and Targeted Therapy, 11(1), 210.
18. Manivannan, E., Bhuvaneswari, P., Yamini Priya, V., Sasikala, R., & Prabhu, R. (2025, June). A review on AI-driven multi-modal data integration in personalized medicine: Advancements in diagnosis, prognosis, and treatment optimization. In 2025 International Conference on Emerging Technologies in Engineering Applications (ICETEA) (pp. 1-6). IEEE.
19. Marouf, A. A., Rokne, J. G., & Alhajj, R. (2025). Integrating multi-omics and medical imaging in artificial intelligence-based cancer research: an umbrella review of fusion strategies and applications. Cancers, 17(22), 3638.
20. Moar, K., Pant, A., & Maurya, P. K. (2026). Artificial Intelligence, Machine Learning, and Multi-omics: The Future of Cancer and Precision Oncology. Indian Journal of Clinical Biochemistry, 1-18.
21. Mukherjee, R., Paul, I., Patel, D., Majgaonkar, A., Roy, S., Kali, S., ... & Sinha, D. (2026). Artificial Intelligence Approaches in Plant Data Science. In Plant Genomic Prediction: AI-Enhanced Models for Plant Genomics and Crop Improvement (pp. 315-334). GB: CABI.
22. Olawade, D. B., Kade, A., Egbon, E., Usman, S. O., Fapohunda, O., Ijiwade, J., & Ogbonna, C. E. (2025). Bioinformatics and artificial intelligence in genomic data analysis: current advances and future directions. Molecular Genetics and Genomics, 300(1), 111.
23. Paul, I. (2026). Artificial Intelligence (AI) and Machine Learning (ML) in Personalized Medicine. In Personalized Cancer Therapy: Advances in Precision Medicine for Targeted Cancer Treatment (pp. 437-462). Singapore: Springer Nature Singapore.
24. Primorac, D., & Südhof, T. C. (2026). Artificial intelligence in medicine: from molecular data analysis to clinical decision-making. Croatian medical journal, 67(3), 135.
25. Sabry, Z., Arora, H. S., Chandrasekaran, S., & Wang, Z. (2025). AI-driven drug discovery and repurposing using multi-omics for myocardial infarction and heart failure. Exploration of medicine, 6, 1001340.
26. Savitha, S., Keerthana, R., Logeswaran, K., Keerthika, P., Sharmila, V., & Sangeetha, M. (2025). Integration of multi-omics data: Genomics, proteomics, metabolomics. In Harnessing AI and Machine Learning for Precision Wellness (pp. 149-184). IGI Global Scientific Publishing.
27. Shaikh, K., Thanki, R., & Shah, S. Artificial Intelligence in Genomics (2026).
28. Shao, Y., Lv, X., Ying, S., & Guo, Q. (2024). Artificial intelligence-driven precision medicine: Multi-omics and spatial multi-omics approaches in diffuse large B-cell lymphoma (DLBCL). Frontiers in Bioscience-Landmark, 29(12).

29. Sheikh, R., Dean, C. J., Akhlaq, M., Naveed, S. A., ur Rehman, A., Hamza, A., ... & Iram, M. (2025). AI-Driven Personalized Cancer Treatment: Integrating Genomic Data for Targeted Therapy Decisions. Pak-Euro Journal of Medical and Life Sciences, 8(3), 609-620.
30. Syeda, A. (2025). Harnessing multi-omics and genome-editing technologies for climate-resilient agriculture: bridging AI-driven insights with sustainable crop improvement. Plant Molecular Biology, 115(6), 120.
31. Vora, N., Shah, S., Patel, P., & Shah, M. (2025). Artificial intelligence and multi-omics in drug discovery: A deep learning-powered revolution. Cure & Care, 100011.
32. Wang, X., He, J., Ding, G., Tang, Y., & Wang, Q. (2026). Artificial intelligence–enabled multi-omics for predicting immune checkpoint inhibitor response and resistance. Journal of Multidisciplinary Healthcare, 572089.
33. Wang, Z., & Tang, H. (2025). A Comprehensive Survey of AI Models in Precision Medicine.
34. Wang, Z., Deng, S., Wang, X., Chang, X., Guo, N., Chen, P., & Yang, H. (2026). Progress of metabolomics-centric multi-omics research in medicine. iMetaOmics, e70096.
35. Zhang, S. (2025, October). Advances in Multi-Omics Integration and AI-Driven Precision Medicine for Diabetes. In Proceedings of the 2025 6th International Symposium on Artificial Intelligence for Medical Sciences (pp. 96-102).