REVIEW
Medical Data Mining 2026;9(1):4. https://doi.org/10.53388/MDM202609004

Harnessing deep learning for the discovery of latent patterns in multi-omics medical data

Okechukwu Paul-Chima Ugwu[1*], Fabian C. Ogenyi[1], Chinyere Nkemjika Anyanwu[2], Melvin Nnaemeka Ugwu[3], Esther Ugo Alum[1], Mariam Basajja[4], Joseph Obiezu Chukwujekwu Ezeonwumelu[5], Daniel Ejim Uti[1], Ibe Michael Usman[6], Chukwuebuka Gabriel Eze[7], Simeon Ikechukwu Egba[1]

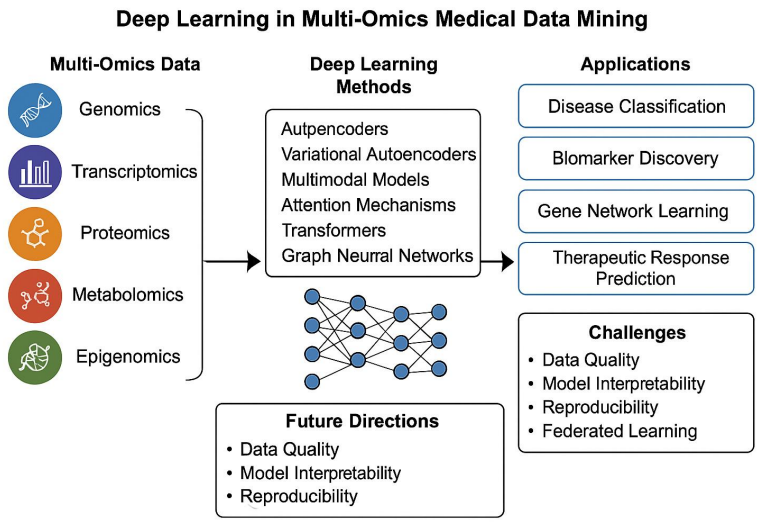
[1]Department of Publication and Extension Kampala International University, Kansanga, Kampala, Box 20000, Uganda.
[2]Department of Microbiology and Immunology Kampala International University, Kansanga, Kampala, Box 20000, Uganda.
[3]Department of Medical Biochemistry, Faculty of Basic Medical Sciences, State University of Medical and Applied Science, Igbo Eno, Enugu Box 412106, Nigeria.
[4]Department of Health Care and Data Management, Leiden University, Leiden, Box 9500 - 2300 RA Leiden, The Netherlands.
[5]Faculty of Pharmacy, Kampala International University, Kampala, Box 20000, Uganda.
[6]Department of Human Anatomy Faculty of Biomedical Science Kampala International University, Kansanga, Kampala, Box 20000, Uganda.
[7]Department of Science Laboratory Technology (Biochemistry Option) Faculty of Physical Sciences University of Nigeria, Nsukka, Enugu Box 410105, Nigeria.

[*]**Correspondence to:** Okechukwu Paul-Chima Ugwu. Department of Publication and Extension, Kampala International University, Box 20000, Ggaba Road, Kansanga, Kampala, Uganda. E-mail: ugwuopc@kiu.ac.ug.

Abstract

The rapid growth of biomedical data, particularly multi-omics data including genomes, transcriptomics, proteomics, metabolomics, and epigenomics, medical research and clinical decision-making confront both new opportunities and obstacles. The huge and diversified nature of these datasets cannot always be managed using traditional data analysis methods. As a consequence, deep learning has emerged as a strong tool for analysing numerous omics data due to its ability to handle complex and non-linear relationships. This paper explores the fundamental concepts of deep learning and how they are used in multi-omics medical data mining. We demonstrate how autoencoders, variational autoencoders, multimodal models, attention mechanisms, transformers, and graph neural networks enable pattern analysis and recognition across all omics data. Deep learning has been found to be effective in illness classification, biomarker identification, gene network learning, and therapeutic efficacy prediction. We also consider critical problems like as data quality, model explainability, whether findings can be repeated, and computational power requirements. We now consider future elements of combining omics with clinical and imaging data, explainable AI, federated learning, and real-time diagnostics. Overall, this study emphasises the need of collaborating across disciplines to advance deep learning-based multi-omics research for precision medicine and comprehending complicated disorders.

Keywords: deep learning; multi-omics integration; biomedical data mining; precision medicine; graph neural networks; autoencoders and transformers



Author contributions

Ugwu OP-C, Ogenyi FC, Anyanwu CN, Ugwu MN, Alum EU, Mariam B, Joseph OCE: Conceptualization, Validation, Visualization, Writing – review & editing. Uti DE, Ibe MU, Eze CG, Egba SI: Resources, Supervision, Writing – original draft.

Competing interests

The authors declare no conflicts of interest.

Acknowledgments

The author(s) reported there is no funding associated with the work featured in this article.

Peer review information

Medical Data Mining thanks all anonymous reviewers for their contribution to the peer review of this paper.

Abbreviations

AI, artificial intelligence; AE, autoencoder; AUC, area under the curve; CNN, convolutional neural network; CNV, copy number variation; CT, computed tomography; DNNs, deep neural networks; DESeq2, differential expression analysis for sequence count data version 2; DL, deep learning; EHR, electronic health records; FAIR, findable, accessible, interoperable, and reusable; FL, federated learning; GCN, graph convolutional network; GEO, Gene Expression Omnibus; GNN, graph neural network; GRNs, gene regulatory networks; HPC, high-performance computing; KEGG, Kyoto Encyclopedia of Genes and Genomes; KNN, k-nearest neighbors; LIME, local interpretable model-agnostic explanations; miRNA, microRNA; ML, machine learning; MRI, magnetic resonance imaging; PCA, principal component analysis; QC, quality control; RNN, recurrent neural network; RNA-seq, RNA sequencing; SHAP, SHapley additive exPlanations; SMOTE, synthetic minority over-sampling technique; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; SVM, support vector machine; TCGA, The Cancer Genome Atlas; TPM, transcripts per million; VAE, variational autoencoder; XAI, explainable artificial intelligence.

Citation

Ugwu OP-C, Ogenyi FC, Anyanwu CN, et al. Harnessing deep learning for the discovery of latent patterns in multi-omics medical data. *Med Data Min*. 2026;9(1):4. doi: 10.53388/MDM202609004.

Executive editor: Ming-Hao Wang.

Received: 17 June 2025; **Revised:** 24 July 2025; **Accepted:** 14 January 2026; **Available online:** 16 January 2026.

© 2026 By Author(s). Published by TMR Publishing Group Limited. This is an open access article under the CC-BY license. (https://creativecommons.org/licenses/by/4.0/)

Introduction

As healthcare becomes more digitized, medical data mining has emerged as a critical subject for extracting meaningful information from big and complicated clinical and biological datasets as shown in Table 1 [1]. In this review, latent patterns refer to hidden structures or connections in complex multi-omics data that aren't easily seen with regular analysis. The patterns are important in medical data mining because they help find new biomarkers, types of diseases, and insights that can enhance diagnosis accuracy and the effectiveness of treatments [1]. Medical data mining is the process of identifying hidden patterns and correlations in large medical datasets, which is critical for diagnosis, prognosis, therapy selection, and individualized treatment planning [2].

Multi-omics approaches extend medical data mining beyond single data types by jointly analysing multiple molecular layers to obtain a systems-level view of cellular activity [3]. Instead of focusing on one omics modality at a time, integrated analysis combines complementary information across genomes, transcriptomes, proteomes, metabolomes, and epigenomes, thereby capturing the complexity of disease mechanisms and treatment responses [3, 4]. A detailed description of individual omics layers, key repositories, and integration challenges is provided in the section "Overview of multi-omics data in medicine".

Because of new high-throughput technologies and improved bioinformatics, there is now an enormous volume and diversity of biological data, which provides both exciting opportunities and difficult issues for analysis. Currently, biomedical research uses data from a variety of sources, including next-generation sequencing, mass spectrometry, imaging, and electronic health records [5]. As the discipline transitions from reductionist to systems biology methods, powerful, adaptable, and intelligent analytic tools are becoming more important. As a result, researchers increasingly recognize that understanding illnesses such as cancer, diabetes, and neurodegenerative disorders necessitates an examination of individual biomarkers as well as the interactions of several biological scales [6]. However, effectively integrating these heterogeneous datasets is challenging because of differences in data types, noise levels, missing values, and batch effects, which can obscure true biological signals and bias downstream models [7]. These multi-omics integration challenges are elaborated in the later section on multi-omics data, where we discuss their implications for biomarker discovery and precision treatment. Conventional statistical approaches cannot address the complex nonlinear and hierarchical relationships seen in biological systems [8]. Because omics analysis is split, it provides only limited insights and precludes the full application of multi-omics techniques. As a result, the biomedical field is currently adopting deep learning, a kind of artificial intelligence (AI), to overcome these challenges [9].

This narrative review examined the role of deep learning methods such as autoencoders, convolutional neural networks (CNNs) and graph neural networks (GNNs) in the discovery of unknown patterns in multi-omics data. Deep learning models use their capability to comprehend complex patterns and correlation to provide a feasible way of integrating various types of input and gleaning information. Case studies and review of recent literature show that these models can help in illness classification, biomarker discovery, and grouping of patients, therefore, promoting precision medicine.

The study is relevant and timely as it shows how deep learning can be useful in making data relevant in therapeutic scenarios. This makes it contribute to the improved understanding of the functioning of illnesses, improvement in the accuracy of medical diagnosis and prognosis, and improvement in the creation of new treatments. In short, multi-omics and deep learning are revolutionizing biomedical research and healthcare to become more comprehensive and data-driven, and more personalized.

Table 1 Comprehensive table on multi-omics and deep learning

Concept	Description	Implications
Medical data mining	Extraction of meaningful patterns and correlations from complex clinical and biological datasets for diagnostic and therapeutic insights.	Enhances diagnosis, prognosis, and personalized treatment planning.
Multi-omics data	Analysis of genomes, transcriptomics, proteomics, metabolomics, and epigenomics to provide a holistic view of cellular functions.	Enables comprehensive understanding of human health and disease networks.
High-throughput technologies	Advances in sequencing and data acquisition technologies generating large, diverse datasets for research.	Offers vast opportunities for discovery but also complex data analysis challenges.
Biomedical data sources	Sources include next-generation sequencing, mass spectrometry, imaging, and electronic health records.	Provides multifaceted datasets essential for integrated biomedical research.
Analytical transition	Shift from reductionist to systems biology necessitating powerful and intelligent analytical tools.	Encourages development of adaptable, AI-driven methods for holistic analysis.
Systems biology relevance	Understanding disease requires analyzing individual biomarkers and their interactions across biological scales.	Facilitates identification of novel therapeutic targets and biomarker interactions.
Integration challenges	Integration is hampered by data heterogeneity, noise, missing values, and batch effects across omics layers.	Necessitates advanced computational methods to combine and interpret data effectively.
Limitations of conventional methods	Statistical tools fall short in modeling complex, nonlinear, and hierarchical biological relationships.	Restricts insights when used alone; limits full potential of multi-omics studies.
Deep learning applications	Deep learning models like CNNs, GNNs, and autoencoders offer robust pattern recognition across omics layers, aiding in precision medicine.	Transforms healthcare by enabling accurate prediction, classification, and therapy development.

Adapted with permission. Olson K. A Comprehensive Review on Healthcare Data Analytic. *Biomed Sustainable Healthcare Appl.* 2023;3(2):95–105. Copyright 2023, AnaPub Publications. CNNs, convolutional neural networks; GNNs, graph neural networks.

Multi-omics data preprocessing and integration

To overcome the heterogeneity of the genomic, transcriptomic, proteomic, metabolomic, and epigenomic data, multi-omics data integration involves a systematic preprocessing. The technical steps followed are as following pipeline.

Harmonization and data acquisition

Data: The data used were obtained in public repositories like TCGA, Gene Expression Omnibus (GEO), and European Molecular Biology Laboratory-European Bioinformatics Institute on the profiles of RNA-seq, DNA methylation, CNV, proteomics, and miRNA.

Sample matching: Incomplete multi-omics profiles in any of the modalities were discarded due to the consistency of downstream analysis.

Quality control

RNA-seq: Genes with low expression (e.g., TPM < 1 in 80 per cent of samples) were discarded.

Methylation: The probes with a high detection *P*-value (> 0.01), cross-reactive sites and those that were on sex chromosomes were eliminated.

Proteomics/Metabolomics: Mahalanobis distance or principal component analysis (PCA) to identify outliers, then manual curation.

Normalization

Transcriptomic data: Log2 (TPM + 1) transformation and z-score normalisation among samples.

Proteomics: Quantile normalization and median centering.

Metabolomics: Pareto scaling and log-transformation.

Methylation: Beta-values were transformed into M-values to make them statistically sound.

Feature selection/dimensionality reduction

Univariate filtering: Variance thresholding (e.g. top 10 percent most variable features per omics layer).

Statistical test: ANOVA, t-tests or DESeq2 to select differentially expressed/methylated features.

Biological priors: Inclusion of pathway information (e.g., KEGG, Reactome) to use in a network-constrained selection.

Dimensionality reduction: Representation of each omic layer was created by employing autoencoders or PCA to come up with compact representations.

Techniques of data integration

The various fusion strategies were implemented depending on the deep learning model:

Early integration (concatenation): All the omics features are concatenated to a single matrix as the input of CNNs or fully connected networks.

Intermediate integration: autoencoders or variational autoencoders (VAEs) trained on an omics-by-omics basis; the latent representations were integrated through shared layers.

Late integration: The outputs of modality-specific models (e.g., GNN on transcriptomics, CNN on imaging, recurrent neural network (RNN) on time-series proteomics) were combined through attention mechanisms or consensus-based voting systems.

Graph-based integration: Omics data as nodes in biological interaction networks (e.g., STRING), with edges representing co-expression or co-regulation; graph convolutional networks (GCNs) pass signals between layers.

Correction of batch effect

Use of Applied ComBat or Limma-based empirical Bayes to match datasets that were measured across labs or platforms.

Imputation and data augmentation

Imputed omics features obtained with KNN, missForest, or deep generative models (e.g. denoising autoencoders).

SMOTE or variational autoencoder based synthetic data generation in class imbalanced situations.

To enhance transparency and trust in complex models, explainable AI (XAI) approaches were employed. SHapley Additive exPlanations (SHAP) values were used to quantify the contribution of each omic feature to the model's prediction, while attention-based layers in multimodal architectures provided direct interpretability of modality relevance. Local interpretable model-agnostic explanations (LIME) was applied in black-box scenarios for model-agnostic explanations. Furthermore, in scenarios involving data from multiple institutions, federated learning (FL) was considered. FL techniques enabled the collaborative training of models across distributed omics datasets without sharing raw data, thereby preserving data privacy. Advanced methods such as secure multiparty computation and homomorphic encryption were integrated to ensure confidentiality during model updates. These techniques address the key challenges of privacy, interpretability, and reproducibility in large-scale omics AI workflows.

Overview of multi-omics data in medicine

Using multi-omics data has become critical in contemporary biomedical research, providing fresh insights into the complexities of health and illness as illustrated in Figure 1 [10]. Multi-omics entails investigating all "omics" levels simultaneously in order to comprehend the complicated ways in which cells are controlled [11–13]. These omics dimensions consist of genomics, which looks at the entire DNA sequence and the genetic variations that impact phenotypes and disease chances; transcriptomics, which analyses RNA expression to reveal gene activity and regulation; proteomics, which measures proteins that carry out cell functions; metabolomics, which detects small molecules that reflect biochemical activities and cellular conditions; and epigenomics, which focusses on changes in DNA and histones. Each omics layer focuses on a specific but related aspect of biology, providing further insights into cell function, illness development, and how the body reacts to treatment [14]. Genomics provides the fundamental design, while transcriptomics demonstrates how genes respond to various factors [15]. Proteomics demonstrates how the response is carried out, while metabolomics exposes the outcomes of biochemical events, which combined characterise the body's physiology [16]. Epigenomics, on the other hand, aids in the regulation of gene activity in diverse tissues and the body's response to its environment [17]. This technique allows researchers to explore beyond basic studies and into how complicated illnesses like cancer, heart disease, and brain problems function.

With so many public repositories accessible, multi-omics research has been able to access vast amounts of credible data. TCGA is currently a valuable resource for cancer research, including extensive data from a variety of omics for a wide range of tumour types [18]. The Genotype-Tissue Expression (GTEx) project provides data on gene expression and regulation in a variety of non-diseased human tissues, which is very valuable for comparing normal and aberrant states [19]. GEO collects a vast volume of transcriptomics and functional genomics data from researchers throughout the globe [20]. In the same manner, the Encyclopaedia of DNA Elements (ENCODE) project aims to identify all active portions of the human genome and give critical information on epigenomics and transcription factors [21]. These repositories, along with other initiatives such as the METABOLOMICS Workbench and the Human Proteome Project, contribute to the feasibility of multi-omics investigations and guarantee that studies are repeatable, can be analysed collectively, and can be compared across various research [22–24]. Even still, using such different datasets is challenging due to the significant computational and analytical hurdles they present. One of the most significant issues is that data from different omics research do not match because of changes in how experiments are conducted, materials are processed, and findings are quantified [25]. Genomic data typically indicate if a mutation exists or not, but transcriptomics

and proteomics provide continuous expression levels, and epigenomics may incorporate both binary and quantitative measures [26]. Because of this diversity, it is difficult to normalise, scale, or model several elements together.

Because there are more characteristics (genes, transcripts, proteins, and metabolites) than samples, the danger of overfitting increases, while the model's capacity to handle additional data reduces [27]. Because high-dimensional data contains a large number of variables, sophisticated approaches must be used to select key characteristics and reduce noise [27, 28]. Missing data is widespread in multi-omics datasets due to technological difficulties, sampling mistakes, or insufficient profiling of distinct omics [29]. If data is lacking, it may distort subsequent studies, necessitating the use of specialised procedures to fill in the gaps [29]. Furthermore, changes in labs, equipment, or time points during data creation might produce batch effects, which may obscure actual biological signals and lead to incorrect results [30]. They should be adjusted using statistical methods to ensure that the interpretations are valid. Handling these problems is critical for successfully integrating multi-omics data, which might lead to the discovery of novel biomarkers, a better understanding of disease mechanisms, and improved precision treatment. While multi-omics provides a deep look into human biology, its full potential can only be realised if key integration challenges are addressed. Using modern computational methods, particularly deep learning, is a potential way to address these challenges and promote overall interpretations, which may significantly improve biological research and treatment.

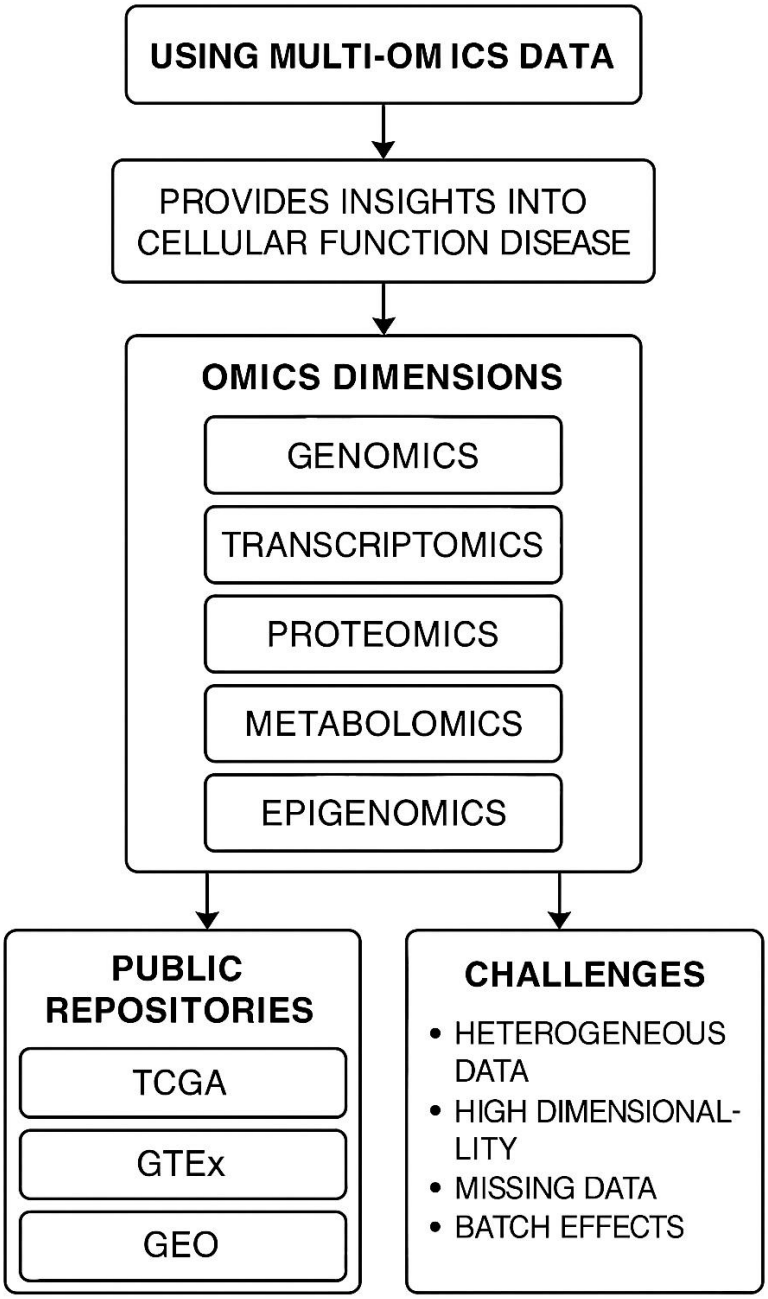


Figure 1 Multi-omics data integration and challenges.
Adapted with permission. Babu M, Snyder M. Multi-Omics Profiling for Health. *Mol Cell Proteomics.* 2023;22(6):100561. Copyright 2023, Elsevier.
GEO, Gene Expression Omnibus; Genotype-Tissue Expression (GTEx); TCGA, The Cancer Genome Atlas.

Fundamentals of deep learning

Deep learning, which is based on how the human brain functions, has gained popularity in data-driven research, particularly for analysing complicated and high-dimensional biological data as shown in Table 2 [31]. An artificial neural network (ANN) is the primary structure, consisting of many connected nodes that analyse and change data via a series of mathematical processes [32, 33]. Deep neural networks (DNNs) are defined as networks with several hidden layers between the input and output [34]. In these designs, the lower layers detect basic characteristics while the higher levels acquire more complex ideas.

CNNs, RNNs, and autoencoders are three of the most prominent deep learning designs, each designed to handle a certain kind of input [35]. CNNs excel at detecting local correlations in data by using image-specific filters, which are increasingly being employed in biology for sequencing, expression tables, and omics heatmaps [36]. This expertise enables them to model data in DNA, histone, and gene expression patterns [37]. RNNs are intended to cope with sequential data by including memory units that store information for a set duration [38]. This capability is required for building models of gene regulation, clinical records across time, and changes in different biological profiles as the illness progresses or medication is administered.

Autoencoders are a common deep learning model that is particularly helpful for merging multi-omics data [7]. They are designed to learn shorter, concealed forms of data by reducing input data to a smaller area and recreating it with the fewest potential errors [39]. Autoencoders clean up input data, remove batch effects, and expose underlying structures in huge omics datasets [7]. VAEs, a probabilistic extension, enhance capabilities by learning the distribution of the latent space, allowing them to produce and impute data, which is critical for multi-omics investigations with missing values [7, 40].

Deep learning stands out in biomedical applications mostly because of its ability to learn characteristics autonomously. Deep learning models, unlike classical machine learning, do not need extensive human feature building since they can learn from raw data on their own [41]. This skill is critical for analysing omics data since the relationships between genes, proteins, and metabolites are often unclear, nonlinear, and occur in high-dimensional spaces. Deep learning aids in the discovery of hidden molecular fingerprints that are difficult to detect using typical statistical approaches due to the many complicated connections between biological factors [27]. Furthermore, deep learning models can manage vast volumes of data, making them suitable for biological research. Deep learning can manage the rising volume and variety of omics data effectively since it is parallelizable and works with high-performance computers [42]. The capacity to handle big data volumes is very beneficial in multi-omics investigations because it enables researchers to connect and analyse data from several biological levels [27].

Furthermore, deep learning excels at detecting patterns that are not straight lines, which is significant since biological systems are nonlinear [43]. Biology's processes are often nonlinear, including feedback loops, random interactions, and situation-specific restrictions [44]. Deep neural networks can express these connections over several layers, enabling them to handle epistatic interactions, gene-environment interactions, and nonlinear dose-response effects. Because of this power, deep learning is critical for combining genomes, transcriptomics, proteomics, and other omics layers to reveal hidden biological activities and therapy targets.

Taken together, these properties of hierarchical representation learning, scalability to high-dimensional data, and the ability to model complex non-linear dependencies make deep learning an attractive foundation on which more specialised architectures for multi-omics integration can be built, as detailed in the subsequent section on deep learning architectures.

Table 2 Deep learning architectures in multi-omics

Deep learning architecture	Key features	Applications in multi-omics and biomedicine	Advantages
Artificial neural networks	Basic structure with nodes and connections that process data through mathematical operations.	Foundation for all deep learning models used in biomedical data mining.	Mimics brain-like processing; flexible and adaptive.
Deep neural networks	Multiple hidden layers enabling hierarchical feature learning—basic to complex patterns.	Facilitates modeling of complex biological systems and hierarchical data structures.	Improves model depth and prediction capability.
Convolutional neural networks	Uses filters to detect local patterns; effective for imaging and sequence data.	Applied to DNA sequences, gene expression profiles, omics heatmaps, and image-based datasets.	Captures spatial dependencies; high performance in biological image analysis.
Recurrent neural networks	Includes memory units; effective for sequential or temporal biological data.	Used in modeling temporal gene expression, clinical trajectories, and disease progression.	Captures temporal relationships in biological sequences and time-series data.
Autoencoders	Learns compressed data representations; helps clean and integrate multi-omics datasets.	Reduces noise, aligns omics datasets, identifies hidden biological structures.	Enables data dimensionality reduction and feature extraction.
Variational autoencoders	Probabilistic extension of autoencoders; learns latent space distribution for data generation/imputation.	Imputes missing data, augments datasets, and supports integration of multi-omics layers with uncertainty modeling.	Enhances robustness and completeness in data analysis; useful for sparse datasets.

Adapted with permission. Sharifani K, Amini M. Machine Learning and Deep Learning for Big Data Analytics: a Review of Methods and Applications. *World Inf Technol Eng J.* 2023;10(7):3897–3904. Copyright 2024, Elsevier.

Applications of deep learning in multi-omics medical data mining

Biomedical research and personalised healthcare may now reveal hidden patterns, valuable traits, and processes that were previously impossible to uncover using normal approaches as shown in Table 3 [27]. Using deep neural networks to study genomes, transcriptomics, proteomics, metabolomics, and epigenomics data simultaneously, researchers have been able to address critical issues in disease knowledge and clinical decision-making [45]. Disease categorisation and subtyping, biomarker discovery, gene regulatory network inference, and medication response prediction are some of the most

common applications of deep learning in this field, all of which contribute to the aims of precision medicine [46].

Deep learning is extremely beneficial in multi-omics analysis for sorting and categorising illnesses, particularly cancer and neurological disorders [7]. Traditional approaches often fail to detect the minor changes in molecular markers that help distinguish distinct types of illnesses, while deep learning (DL) models, particularly convolutional and autoencoder networks, do a better job of doing so [47]. In cancer research, scientists have begun to use multi-omics DL models to classify tumours into significant subtypes by combining mutations, gene expression, DNA methylation, and protein analysis [48]. As a consequence, scientists discovered novel subgroups of glioblastoma, breast, and colorectal cancers with distinct genomic characteristics and therapeutic responses. Similarly, deep learning models have assisted researchers in Alzheimer's and Parkinson's disease by merging several forms of data to distinguish between early-stage and late-stage illness, as well as identify subgroups with varying rates of progression [49].

Deep learning is also highly promising in biomarker development because of its capacity to detect precise traits and the fact that it is understandable to biologists [50]. Most conventional biomarkers are discovered using statistical approaches, which may not provide a complete picture of how many variables interact [51]. Unlike other models, DL models can discover relationships between disease states or treatment outcomes and several biological layers [7, 27, 47]. For example, using attention mechanisms and approaches like integrated gradients or SHAP, researchers may determine which genes, pathways, or epigenetic markers are most relevant in the model's predictions [52]. This skill allows us to better understand illnesses and leverage our discoveries to develop biomarkers for clinical application in diagnosis, prognosis, and monitoring.

Deep learning enables scientists to rebuild the complex relationships between genes that control biological processes [53]. Gene regulatory networks (GRNs) are often shown as networks with nodes representing genes or regulatory elements and edges indicating whether one gene activates or represses another [53]. Traditional GRN models struggle to manage big and complicated multi-omics datasets, while RNNs and GNNs are intended to handle both time- and network-based relationships [54]. They may use information regarding gene expression, epigenetic marks, transcription factors, and chromatin accessibility to anticipate how genes are regulated in various circumstances. For example, variational autoencoder models have assisted researchers in identifying key regulators in cancer and the process of stem cell development [55]. They are critical for understanding how cells pick their destiny, what causes illnesses, and why certain therapies fail.

Deep learning can currently anticipate how medications will behave in specific patients, allowing for more personalised treatment regimens based on a wide range of data [27, 56]. DL frameworks use genetic mutations, transcriptome markers, proteomic alterations, and pharmacogenomic data to predict how a patient will respond to certain medications [57]. They can detect both the effects of individual medications and the potential consequences of mixing many drugs. Deep learning for drug repositioning (DeepDR) for example, integrates deep learning with data on gene expression, methylation, and mutations to assist doctors in determining the optimum therapy for each cancer patient [58]. Similarly, integrated deep learning is currently being used to predict how medications will function in autoimmune illnesses and mental health issues, where complicated pathways make it difficult to choose the best therapy [59].

Overall, deep learning is transforming biomedical research and clinical practice. Deep learning aids researchers and doctors in illness classification, biomarker discovery, gene network reconstruction, and therapy prediction, enabling precision medicine. The following section builds on these broad application categories by examining specific case studies and benchmarking efforts that demonstrate how different architectures and integration strategies perform in real-world multi-omics settings.

Table 3 Applications of deep learning in multi-omics and precision medicine

Application area	Deep learning models/techniques used	Biological data types involved	Key contributions to precision medicine
Disease classification & subtyping	Convolutional neural networks (CNNs), autoencoders	Genomics, transcriptomics, DNA methylation, proteomics	Improved classification of cancer (e.g., glioblastoma, breast cancer) and neurodegenerative diseases (e.g., Alzheimer's, Parkinson's); identification of subgroups with distinct molecular profiles.
Biomarker discovery	Attention mechanisms, integrated gradients, SHAP	Gene expression, epigenetic modifications, pathways	Uncovered complex interactions between molecular traits; facilitated clinical translation of disease-specific diagnostic, prognostic, and monitoring biomarkers.
Gene regulatory network inference	Recurrent neural networks, graph neural networks, variational autoencoders	Gene expression, transcription factors, chromatin accessibility, epigenetics	Enabled reconstruction of complex gene regulation networks, identification of master regulators, and understanding of disease mechanisms and therapeutic resistance.
Drug response prediction	Integrated DL frameworks (e.g., DeepDR), CNNs, multi-omics neural models	Genetic mutations, transcriptomic and proteomic markers, pharmacogenomics	Personalized therapy predictions based on multi-omics inputs; optimization of drug choice and combination strategies in cancer, autoimmune, and psychiatric diseases.

Adapted with permission. Mukherjee A, Abraham S, Singh A, Balaji S, Mukunthan KS. From Data to Cure: A Comprehensive Exploration of Multi-omics Data Analysis for Targeted Therapies. *Mol Biotechnol.* 2024;67(4):1269–1289. Copyright 2024, Springer Nature. CNNs, convolutional neural networks; DeepDR, deep learning for drug repositioning; SHAP, SHapley additive exPlanations.

Deep learning architectures for multi-omics integration

Because multi-omics data is complicated and diverse, it necessitates sophisticated computational approaches to extract useful information as illustrated in Figure 2 [27, 40, 45]. Building on the general deep learning principles described earlier, specialised architectures have been developed to encode each omics modality, fuse complementary information, and exploit known biological structure. Autoencoders, VAEs, multimodal deep learning, attention mechanisms such as transformer networks, and GNNs are among the finest DL designs for dealing with different kinds of input [7, 40, 59]. Every architecture contributes something unique to the processing of omics data, making it simpler to model illnesses, find biomarkers, and apply analytics to precision medicine.

Many researchers employ AEs and VAEs to incorporate various

kinds of omics data because they can do unsupervised learning [7, 40]. AEs reduce omics data from several dimensions to fewer, preserving the fundamental biological changes while removing any unnecessary noise [60]. Following that, the latent representation may be used to perform tasks such as grouping, class identification, and data visualisation. VAEs include a probabilistic layer into the architecture, allowing them to manage data uncertainty and generate fresh fake samples [55, 61]. When it comes to multi-omics, VAEs excel at merging several omics layers by learning representations that emphasise biological commonalities. For example, utilising VAEs, researchers were able to connect transcriptomic and methylomic data to discover novel cancer types with distinct predictive outcomes [62]. VAEs may also help fill in data gaps, which is a common challenge in multi-omics research [7, 40, 62, 63].

Another method for integrating multi-omics data is to use multimodal deep learning models, which are designed to handle data from many kinds of experiments simultaneously [5, 7, 40, 56, 58]. Each sort of omics data is processed by its own branch, and all of the branches' information is combined in a final common layer for interpretation. It manages each data type's specific statistics while also allowing for detailed modelling of how multiple forms of data interact with one another. A multimodal neural network might use convolutional layers for genomic pictures, dense layers for expression tables, and recurrent layers for epigenomic sequences [64]. The next stage is to combine the signals and create a full biological picture that may aid in activities such as grouping patients or predicting their survival [5, 56]. This approach improves the model's performance while also making it easy to identify how each omics category influences the final output.

Multi-omics analysis now includes attention processes and transformer structures to improve inter-omics modelling. Attention processes allow the model to prioritise the most relevant biological signals for each task [5, 40, 47, 65]. For example, when predicting the effects of a medicine, an attention-based model might prioritise drug-targeted gene pathways above extraneous information [66]. Transformers were originally used in natural language processing, but they now utilise self-attention to identify links between all inputs [67]. In multi-omics, transformers may reflect complex, far-reaching linkages between genes, proteins, and epigenetic markers at all omics levels [68]. It is particularly important when attempting to comprehend various kinds of illnesses and the processes behind them, since the impacts of biology may be observed in several levels and do not necessarily follow a linear sequence.

GNNs take a novel approach to multi-omics integration, presenting biological data as graphs rather than simply vectors or sequences [69]. Nodes in these graphs may be genes, proteins, metabolites, or patients, while edges represent the different ways genes or proteins interact, are co-expressed, or belong to the same pathway [70]. The structure enables GNNs to update the characteristics of each node while taking into account information from its neighbours, allowing them to learn both local and global topological data. This is especially useful in multi-omics investigations, where established biological networks (such as protein-protein interaction maps and transcription factor binding networks) may be utilised to guide the process. To demonstrate, GNNs have merged proteomic and transcriptome information in cancer research by using signalling pathway information, resulting in more accurate predictions and more relevant findings [69, 71, 72]. Furthermore, GNNs are inherently simple to grasp since researchers can track the information as it flows through the network and identify key regulatory regions or groups [72].

In practice, the choice of architecture depends on the scientific question and data structure: autoencoders and VAEs are preferred for unsupervised representation learning and data completion, multimodal networks for flexible fusion of heterogeneous inputs, attention and transformers for capturing long-range dependencies, and GNNs when prior biological networks are available. Rather than re-stating the general advantages of deep learning, this section emphasises how different architectures operationalise those advantages for multi-omics integration.

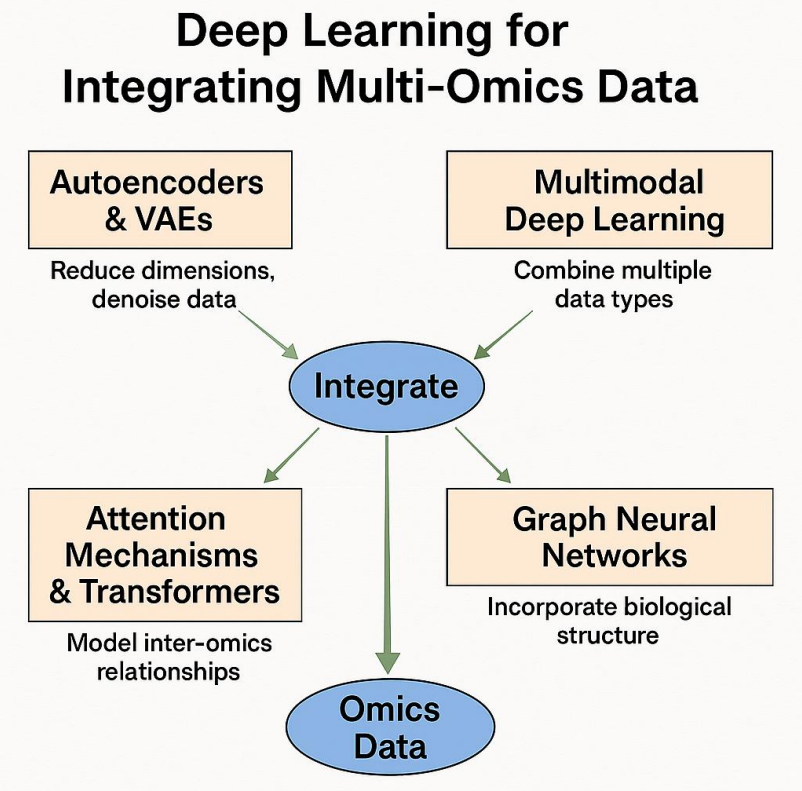


Figure 2 Deep learning for integrating multi-omics data.
Adapted with permission. Ballard J, Wang Z, Li W, et al. Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min.* 2024;17(1):38. Copyright 2024, Springer Nature. VAE, variational autoencoder.

Case studies and benchmarking

To complement the conceptual application areas described above, this section summarises representative case studies and benchmarking efforts that apply DL to multi-omics data in concrete disease settings as shown in Table 4 [26, 27, 57, 73–75]. By examining specific models, datasets, and evaluation metrics, these examples illustrate how DL-based approaches behave in practice and how they compare with classical machine learning (ML) methods.

Lee employed a deep autoencoder to incorporate transcriptome, epigenomic, and miRNA data from TCGA to predict survival subgroups in hepatocellular carcinoma [58, 76]. The DL model was able to classify patients into more accurate risk groups than classic clustering methods like k-means and PCA [77]. It is worth mentioning that the autoencoder's latent features caught the complex interactions between diverse omics data, resulting in better survival prediction and subgroup separation than standard techniques [5, 56, 78]. It highlights how deep learning can transform challenging and complex omics data into clinically relevant tools. Another major study, Zitnik et al. (2019), developed a GCNs model that integrates data from genomic, transcriptomic, proteomic, and chemical structure networks to discover new therapeutic applications [73]. GCN outperformed random forest and Support Vector Machines (SVM), in predicting drug-disease connections, demonstrating that graph-based deep learning is useful for such applications [79]. Modelling the relationships between genes, proteins, and pathways could lead to novel strategies to treat diseases, demonstrating the potential of deep learning in drug discovery and repositioning. This method of learning shifted from focussing on individual aspects to examining the whole structure, providing a biologically sound technique to combine various omics data [80].

Another notable example is the Deep Learning for Multi-Omics Integration (DeepMO) stands for model (Lin et al. 2020), which employed a deep neural network to combine genomes, transcriptomics, and epigenomics to classify various forms of cancer [74]. DeepMO outperformed logistic regression, naïve Bayes, and random forests for classification, particularly with noisy and incomplete data [81]. It employed subnetworks for each type of input and attention-based fusion, allowing the model to select the most significant omics layer for each prediction. Because DL can handle data in a variety of formats, it is more useful than classical ML, which typically assumes that the value of features remains constant and does not operate well with missing data.

The same results are obtained when benchmarking studies are conducted on diverse datasets. Nguyen et al. (2021) compared the performance of deep learning models such as autoencoders, CNNs, and variational models to random forests, decision trees, and SVMs for breast cancer subtype classification and survival analysis [75]. It was regularly discovered that DL models performed better in Area under the Curve (AUC), recall, and error reduction, particularly when more than two omics modalities were utilised. Another advantage of these models was their ability to deal with noise and missing data, making them more effective in a variety of datasets and validation groups. Stories of effective DL integration in multi-omics research are now being used in therapeutic contexts. For example, deep learning is applied in non-small cell lung cancer (NSCLC), and a DL model trained on various forms of data was able to accurately predict treatment response [82]. This strategy enabled clinicians to select the most successful therapies while avoiding ineffective ones, reducing both patient troubles and costs. Similarly, utilising DL models that incorporate transcriptome and proteomic data has enabled the early detection of Alzheimer's disease, even before clinical symptoms appear [83].

Recent studies on other diseases have found that DL models can capture complex multi-omics patterns in both cancer and non-cancer contexts, reinforcing the application themes described earlier disease subtyping, biomarker discovery, regulatory network inference, and treatment response prediction while grounding them in concrete datasets and benchmarking protocols [7, 11, 27, 40, 56, 72, 74, 84]. Instead of restating these applications in general terms, the case studies emphasise how specific architecture choices and integration strategies translate into measurable gains in predictive and clinical utility.

Table 4 Case studies on deep learning applications in multi-omics

Study	DL model studied	Omics data integrated	Biomedical application	Performance compared to ML
Mukherjee et al. (2024) [27]	Deep autoencoder	Transcriptomics, Epigenomics, miRNA	Hepatocellular carcinoma survival subgroup prediction	Outperformed k-means and PCA in survival prediction and risk stratification
Zitnik et al. (2019) [73]	Graph convolutional network	Genomics, Transcriptomics, Proteomics, Chemical Structure Networks	Drug discovery and repositioning	Outperformed random forest and SVM in drug-disease association prediction
Lin et al. (2020) [74]	Deep neural network with attention-based fusion	Genomics, Transcriptomics, Epigenomics	Cancer subtype classification	Outperformed multiple ML integration approaches, elastic net, random forest, ensemble methods, multiple kernel learning
Nguyen et al. (2021) [75]	Autoencoders, CNNs, VAEs	Genomics, Transcriptomics, Proteomics	Breast cancer subtype classification and survival prediction	Higher AUC, recall, and reduced error than random forest, decision trees, SVM
NSCLC DL Model Davri et al. (2023) [26]	Multi-input DL model	Genomics, imaging, clinical data	NSCLC treatment response prediction	Enabled accurate therapy predictions, better patient outcomes
Alzheimer's early detection model Ogunjobi et al. (2024) [57]	Transcriptome + Proteome DL Integration	Transcriptomics, Proteomics	Early Alzheimer's detection	Detected disease prior to clinical symptoms, improving early intervention

AUC, area under the curve; CNNs, convolutional neural networks; ML, machine learning; PCA, principal component analysis; SVM, support vector machine; VAEs, variational autoencoders.

Challenges and limitations

Even though DL has enormous potential for medical data mining in multi-omics, various barriers prevent it from being properly utilised in research and clinical settings as shown in Table 5 [7, 84]. Such barriers affect technological, methodological, and infrastructure

elements, therefore designers and users of DL models should proceed with caution and critical thought. Recognising these limitations is critical for improving the discipline both scientifically and clinically. This section addresses five major issues: data quality and annotation, model explanation, overfitting and generalisation, repeatability and standardisation, and computational resources. The main and long-term issue is that inadequate data and annotations can significantly impair the usefulness and accuracy of deep learning models [85, 86]. Most multi-omics datasets have issues with noise, missing data, incorrect labelling, and insufficient clinical data [5]. Technical variability, batch effects, and measurement mistakes in high-throughput technologies such as genomics, transcriptomics, proteomics, and metabolomics can all diminish data accuracy [87]. In addition, many public datasets lack clear and standard annotations, particularly for clinical outcomes and phenotypes, making it difficult to train and evaluate models. As a result of these flaws, DL models may be inaccurate and may not perform well in many contexts, indicating that we need to enhance data cleaning, imputation, and the usage of well-annotated data sets.

Another critical issue is that DL models must be intelligible and transparent, which is still a major worry when utilising DL to make biological decisions [65]. While classic statistical models are simple to explain, deep neural networks operate in a complex manner [88]. Failure to be upfront about the approach might raise serious ethical and medical concerns, particularly in cases involving cancer diagnosis and treatment response prediction. Even though techniques like SHAP, attention mechanisms, and layer-wise relevance propagation have been developed to improve model interpretability, they are continually evolving and lack a consistent set of best practices [89]. It is critical to strike a balance between how complicated a model is and how simple it is to understand, so that both doctors and regulators believe it.

Using deep learning models in multi-omics presents additional challenges due to the potential of overfitting and poor generalisation [7, 56, 84]. Because there are numerous features but few instances in omics data, DL models are more likely to recall the training data than to discover patterns that can be applied elsewhere. When models are evaluated on new data, their performance typically reduces dramatically. Overfitting can be reduced by dropout regularisation, early halting, data augmentation, and transfer learning, but all of these require careful adjustment and domain understanding [90]. Furthermore, the intrinsic heterogeneity and diversity observed in patients, diseases, and diverse omics platforms make the task more complex [91].

It is also challenging to achieve consistent results while adhering to established methods. It is still difficult to replicate DL investigations in biological research because there are no standard methods for handling data, designing models, adjusting hyperparameters, and reporting results [92]. If many research organisations use the same deep learning algorithm with different settings, the findings may differ. Furthermore, many published models lack the code or data required for independent verification and reuse. Because of this lack of transparency, the field is not moving forward as swiftly as it should. Following Findable, Accessible, Interoperable, and Reusable (FAIR) data principles and developing standardised ML pipelines are required to address these issues and ensure that results are reproducible [93]. The demand for computing resources might be difficult to meet, particularly in areas with limited infrastructure. Deep neural networks, particularly those that combine many omics layers, require a significant amount of memory, storage, and computational capacity to train, hence high-performance computing (HPC), systems or cloud-based GPUs are typically required [94]. Such requirements may preclude small research groups or universities in resource-limited areas from participating in research. Because energy-intensive computations hurt the environment, there is a growing interest in discovering ways to improve deep learning algorithms [95]. People are looking into model pruning, quantisation, and federated learning to reduce the amount of compute required while preserving model performance [96]. Overall, deep learning excels at handling complicated biological data, but its application in multi-omics medical data mining presents numerous challenges. To maximise the potential of deep learning in precision medicine, we must address concerns such as data quality, interpretability, overfitting, reproducibility, and computing intensity. To address these concerns, experts in bioinformatics, computer science, clinical medicine, and regulatory policy must collaborate to create robust frameworks that ensure accuracy, ethical behaviour, and equitable access to AI-based healthcare services.

Table 5 Challenges and solutions in deep learning for multi-omics integration

Challenge area	Description	Proposed solutions	Implications for precision medicine
Data quality and annotation	Multi-omics data often suffer from noise, missing values, batch effects, and poor clinical annotation, impairing DL model accuracy and reliability.	Enhance data preprocessing, annotation, cleaning, and imputation; adopt standardized, high-quality clinical datasets.	Improved data quality will enhance predictive accuracy and clinical relevance of DL models in multi-omics.
Model explanation and interpretability	DL models are complex and difficult to interpret compared to traditional statistical models; lack of transparency may raise ethical and clinical concerns.	Use SHAP, attention mechanisms, and relevance propagation; develop best practices for explainable AI in omics.	Interpretable models foster clinician trust and regulatory acceptance, enabling real-world application of DL.
Overfitting and generalisation	High-dimensional omics data with limited samples can lead to overfitting; models may fail to generalize to new datasets without proper regularization.	Apply dropout, early stopping, data augmentation, and transfer learning; require domain expertise for tuning.	Better generalisation will allow models to be robust across diverse patient populations and datasets.
Repeatability and standardisation	Absence of standardized DL protocols leads to difficulty in reproducing results; inconsistencies in code, hyperparameters, and data handling hinder validation.	Follow FAIR principles; create standardized ML pipelines and open access code repositories for reproducibility.	Reproducible workflows improve confidence in DL applications, facilitating clinical translation and policy adoption.
Computational resource demands	Training DL models requires significant computing resources (e.g., GPUs, HPC); limits participation from low-resource settings and increases environmental costs.	Adopt model compression, pruning, quantization, and federated learning to reduce compute needs; expand cloud-based research support.	Resource-efficient DL strategies promote equitable access to AI-driven healthcare innovations globally.

Adapted with permission. Mohammed MA, Abdulkareem KH, Dinar AM, Zapirain BG. Rise of Deep Learning Clinical Applications and Challenges in Omics Data: A Systematic Review. *Diagnostics.* 2023;13(4):664. Copyright 2023, MDPI. FAIR, Findable, Accessible, Interoperable, and Reusable.

Future perspectives

Multi-omics medical data mining is a fast-growing area, facilitated by the development of DL and AI technology [74, 78]. Nevertheless, there are major challenges in translating these research advances into clinical practice. To address this issue, new approaches need to focus on key technical and practical problems, such as making data consistent, dealing with different types of data, and aligning clinical,

imaging, and molecular data over time and space [83, 85–87]. We discuss some of the future-oriented views below, addressing certain unresolved issues to fully realise their potential.

Integrated multi-modal data fusion combining multi-omics data with clinical data and medical images could create detailed models that better show the differences in diseases and individual patients. Although omics data are rich in molecular data, they can be deficient in the spatial and anatomical contextualisation provided by imaging modalities, such as MRI, CT, and PET, or in temporal dynamics, such as electronic health records (EHRs) [5]. In theory, multimodal transformers and graph neural networks, which are types of deep learning models, can combine all these different types of data to improve how we predict and plan treatments for diseases [96–99]. Key challenge: An important challenge is that there is no standard for data formats and ontologies across these domains, which hinders successful fusion and interpretation. Additionally, it is very difficult to match data that have different levels of detail in space and time, such as gene expression profiles and imaging scans. With stricter privacy laws, like HIPAA and GDPR, FL presents an attractive decentralised model for training AI models collectively without sharing data directly [98–100]. This strategy will allow institutions to develop strong models with distributed multi-omics data without infringing on patient confidentiality. The main challenge is that FL faces real-world issues related to differences in data types, maintaining model stability, and managing communication costs, despite its enormous potential. The sequencing depths, population biases, and EHR structures may differ across different institutions, which makes it difficult to harmonise the models. Also, FL should be applied to hospitals, where scalable and secure infrastructures are still rudimentary.

XAI in Omics Research: Complex DL models lack transparency, which reduces clinical trust and acceptance by regulators [101]. Explainable AI tries to uncover the biological reasons for the model's predictions using tools like saliency maps, attention mechanisms, or counterfactual reasoning. Major limitation: Existing XAI approaches do not necessarily yield biologically interpretable explanations, and they are known to be inadequate in high-dimensional omics data. One of them is the lack of incorporation of domain-specific priors, like pathway databases, gene ontologies, or known molecular interactions, into the learning process. Specifying standardised explanation measures for use in a medical setting is also crucial.

Real-time multi-omics diagnostics and decision support the development of biosensors, edge computing, and high-throughput sequencing enabling the possibility of real-time diagnostics using multi-omics data [102]. By using real-time information from wearable devices, lab results, and clinical notes, deep learning models can help keep track of patients, spot diseases that are getting worse, and provide immediate personalised treatment [103, 104]. Key challenge: To achieve this vision, it will require the creation of stream-compatible machine learning models capable of processing temporal-dependent streams of varying data. In addition to that, real-time deployment should focus on regulatory compliance, low-latency computing infrastructure, and integration into clinical workflows, which remain in their infancy. The combination of deep learning and multi-omics data integration can bring transformative power to precision medicine. Nevertheless, to convert existing advances into clinically applicable instruments, a number of strategic areas should be covered. The main disadvantage of existing deep learning models is the fact that they are black-boxed. Future studies should focus on creating inherently interpretable models, e.g., attention-based mechanisms or concept bottleneck models, which can give clinicians explanations of their decisions.

Solving data quality and model interpretability in deep learning in multi-omics

As outlined in the "Challenges and limitations" section, issues such as batch effects, missing values, and limited interpretability remain major barriers to deploying DL models in practice. Future work, therefore, needs to move from merely describing these problems to systematically implementing and evaluating concrete solutions, some of which are summarised below.

Assuring data quality in multi-omics integration

Deep learning can be a very important tool in multi-omics analysis, but only to the extent that the input data is of high quality, complete, and consistent. Promising directions for improving data quality include.

Normalization and batch effect correction: The data normalization and the correction of the batch effect can be achieved with the help of such tools as Combat-Seq or surrogate variable analysis.

Missing data imputation: Denoising autoencoders or generative adversarial imputation nets are deep generative models that provide effective imputation of missing values in transcriptomic and proteomic data.

Noise reduction: Wavelet transforms, low-rank matrix approximations or PCA may be used to denoise high-throughput omics datasets without sacrificing important biological signal.

Quality control pipelines: The inclusion of pipelines like FastQC (in genomics) and MSstats (in proteomics) would make sure that the low-quality or low-coverage dataset is discarded prior to entering a model.

Enhancing model interpretability

Deep learning models cannot be used clinically without interpretability. Rather than re-listing interpretability as a challenge, future research should prioritise the systematic use and benchmarking of the following technical strategies.

SHAP: Provides feature-wise attribution values, so it is possible to identify molecular drivers in a prediction with precision.

Integrated gradients: It highlights the most significant features that contribute to the output of a neural network in various layers, particularly in CNNs and transformers.

Attention mechanisms: In multi-modal models, in particular, attention-based architectures (e.g., transformers) are capable of automatically assigning different weights to different omics layers or individual genes when making predictions.

Concept bottleneck models: In these models, intermediate representations are constrained to correspond to human-interpretable concepts (e.g. pathway activity), which allows the decision process to be transparent and better validated.

Model disentanglement: Disentangled latent space: VAEs enable the decomposition of individual biological factors that can be examined independently.

Confirmation and credibility

In order to make sure that interpretability is not deceptive, it is necessary to:

Confirm explanations with biological priors (e.g. KEGG, GO annotations).

Use stability tests (e.g., explainability consistency with data perturbation).

Cross-validate the results of various explanatory methods using model-agnostic interpretability toolkits (such as LIME, ELI5 and Captum).

By explicitly linking these solution-oriented strategies to the earlier discussion of challenges, the future-perspectives section avoids repeating the same lists of limitations and instead emphasises how the field can move towards robust, trustworthy, and clinically actionable deep learning models for multi-omics data.

Conclusion

The combination of deep learning and multi-omics data mining is a potentially life-changing transition in biomedical research that presents immense potential to advance our knowledge of the biological complexity of human health and disease. This review looks

at how deep learning is currently being used to analyse multi-omics medical data, emphasising its growing importance in new precision medicine efforts. The results indicate that deep learning models, due to their hierarchical representation and scalability, may be the best fit to manage the high-dimensional, heterogeneous, and non-linear nature of multi-omics data. Neural network architectures, including DNNs, CNNs, RNNs, autoencoders, and GNNs, have shown potential in disease classification, biomarker identification, regulatory network inference, and personalised therapeutic prediction. One of the key strengths of deep learning is that it can discover latent, possibly physiologically meaningful patterns that may not be detected by traditional statistical and traditional machine learning methods. Unlike traditional methods, deep learning models can automatically learn important features from the raw input data without needing manual adjustments, which can be quicker and help the models work better in different situations. Some of the applications discussed in this section demonstrate that deep learning models have the potential to have more success than conventional methods in addressing challenging biomedical problems, including tumour subtyping, gene interaction prediction, and individualised drug response prediction. These developments, which are still being developed, can possibly have significant implications for clinical practice, as they bring us closer to a personalised, data-driven healthcare system that allows early disease identification, individualised treatment planning, and dynamic monitoring. Recently, new types of deep learning models, like variational autoencoders, transformers, and graph-based models, have made deep learning even more helpful in combining different types of biological data with clinical or imaging information. These integrative models can offer a better picture of the health of patients, contributing to more accurate diagnostics and systems-level understanding of the disease process. Although the above developments are encouraging, there are still several issues. Concerns about the data quality, interpretability, reproducibility, and scalability of computations remain a major hindrance to widespread clinical adoption. However, recent developments like XAI, federated learning to train models in a privacy-preserving manner, and real-time diagnostic algorithms point to positive future research and practice. The technical advancements alone may not be sufficient to make deep learning a successful long-term approach to multi-omics data mining, but rather will require long-term interdisciplinary collaboration. Collaboration between computational scientists, bioinformaticians, clinicians, biologists, ethicists, and policymakers will be necessary to ensure that the models developed will be not only scientifically sound but also ethically responsible, clinically actionable, and of societal good. To conclude, the synergy between deep learning and multi-omics data analysis promises to be a game-changer in the field of disease comprehension, diagnosis, and treatment, but achieving this promise will necessitate optimism, rigorous testing, and teamwork. In this age of data-driven medicine, the responsible use of this powerful intersection can play a key role in the future of medicine with greater precision, a more profound understanding, and more compassion.

References

1. Olson K. A Comprehensive Review on Healthcare Data Analytics. *J Biomed Sustainable Healthcare Appl.* 2023;3(2):95–105. Available at: https://doi.org/10.53759/0088/JBSHA202303010
2. Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res.* 2021;8(1):44. Available at: https://doi.org/10.1186/s40779-021-00338-z
3. Vahabi N, Michailidis G. Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review. *Front Genet.* 2022;13:854752. Available at: https://doi.org/10.3389/fgene.2022.854752
4. Mohr AE, Ortega-Santos CP, Whisner CM, Klein-Seetharaman J, Jasbi P. Navigating Challenges and Opportunities in Multi-Omics Integration for Personalized Healthcare. *Biomedicines.* 2024;12(7):1496. Available at: https://doi.org/10.3390/biomedicines12071496
5. Duan J, Xiong J, Li Y, Ding W. Deep learning based multimodal biomedical data fusion: An overview and comparative review. *Inf Fusion.* 2024;112:102536. Available at: https://doi.org/10.1016/j.inffus.2024.102536
6. Chasman D, Fotuhi Siahpirani A, Roy S. Network-based approaches for analysis of complex biological systems. *Curr Opin Biotechnol.* 2016;39:157–166. Available at: https://doi.org/10.1016/j.copbio.2016.04.007
7. Ballard JL, Wang Z, Li W, Shen L, Long Q. Deep learning-based approaches for multi-omics data integration and analysis. *BioData Min.* 2024;17(1):38. Available at: https://doi.org/10.1186/s13040-024-00391-z
8. Tavassoly I, Goldfarb J, Iyengar R. Systems biology primer: the basic methods and approaches. *Essays Biochem.* 2018;62(4):487–500. Available at: https://doi.org/10.1042/EBC20180003
9. Naskar S, Sharma S, Kuotsu K, et al. The biomedical applications of artificial intelligence: an overview of decades of research. *J Drug Target.* 2025;33(5):717–748. Available at: https://doi.org/10.1080/1061186X.2024.2448711
10. Babu M, Snyder M. Multi-Omics Profiling for Health. *Mol Cell Proteomics.* 2023;22(6):100561. Available at: https://doi.org/10.1016/j.mcpro.2023.100561
11. Shahrajabian MH, Sun W. Survey on Multi-omics, and Multi-omics Data Analysis, Integration and Application. *Curr Pharm Anal.* 2023;19(4):267–281. Available at: https://doi.org/10.2174/1573412919666230406100948
12. Srivastava U, Kanchan S, Kesheri M, Gupta MK, Singh S. Types of omics data. *Integr Omics.* 2024:13–34. Available at: https://doi.org/10.1016/B978-0-443-16092-9.00002-3
13. Chen C, Wang J, Pan D, et al. Applications of multi-omics analysis in human diseases. *MedComm* (2020). 2023;4(4):e315. Available at: https://doi.org/10.1002/mco2.315
14. Mudadla T, Sharma G, Mishra A, Gola S. Multifaceted Landscape of Omics Data. *Bio-Algorithms Med-Syst.* 2024;20(1):22–36. Available at: https://doi.org/10.5604/01.3001.0054.8093
15. D Magar N, Shah P, Harish K, et al. Gene Expression and Transcriptome Sequencing: Basics, Analysis, Advances. *Gene Expression.* 2022. Available at: https://doi.org/10.5772/intechopen.105929
16. Wishart DS. Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiol Rev.* 2019;99(4):1819–1875. Available at: https://doi.org/10.1152/physrev.00035.2018
17. Marsit CJ. Influence of environmental exposure on human epigenetic regulation. *J Exp Biol.* 2015;218(1):71–79. Available at: https://doi.org/10.1242/jeb.106971
18. Chakraborty S, Hosen MdI, Ahmed M, Shekhar HU. Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *Biomed Res Int.* 2018;2018:9836256. Available at: https://doi.org/10.1155/2018/9836256
19. Keen J, Moore H. The Genotype-Tissue Expression (GTEx) Project: Linking Clinical Data with Molecular Analysis to Advance Personalized Medicine. *J Pers Med.* 2015;5(1):22–29. Available at: https://doi.org/10.3390/jpm5010022
20. Clough E, Barrett T. The Gene Expression Omnibus Database. *Methods Mol Biol.* 2016:93–110. Available at: https://doi.org/10.1007/978-1-4939-3578-9_5
21. Luo Y, Hitz BC, Gabdank I, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* 2019;48(D1):D882–D889. Available at: https://doi.org/10.1093/nar/gkz1062

22. Fuloria NK, Malviya R, Verma S, Balusamy B. *Big Data in Oncology: Impact, Challenges, and Risk Assessment*. *River Publishers*. 2023. Available at: https://doi.org/10.1201/9781003442639
23. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 2015;44(D1):D463–D470. Available at: https://doi.org/10.1093/nar/gkv1042
24. Smith LM, Agar JN, Chamot-Rooke J, et al. The Human Proteoform Project: Defining the human proteome. *Sci Adv.* 2021;7(46):eabk0734. Available at: https://doi.org/10.1126/sciadv.abk0734
25. Brooks TG, Lahens NF, Mrčela A, Grant GR. Challenges and best practices in omics benchmarking. *Nat Rev Genet.* 2024;25(5):326–339. Available at: https://doi.org/10.1038/s41576-023-00679-6
26. Davri A, Birbas E, Kanavos T, et al. Deep Learning for Lung Cancer Diagnosis, Prognosis and Prediction Using Histological and Cytological Images: A Systematic Review. *Cancers (Basel)*. 2023;15(15):3981. Available at: https://doi.org/10.3390/cancers15153981
27. Mukherjee A, Abraham S, Singh A, Balaji S, Mukunthan KS. From Data to Cure: A Comprehensive Exploration of Multi-omics Data Analysis for Targeted Therapies. *Mol Biotechnol.* 2024;67(4):1269–1289. Available at: https://doi.org/10.1007/s12033-024-01133-6
28. Leclercq M, Vittrant B, Martin-Magniette ML, et al. Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICs Data. *Front Genet.* 2019;10:452. Available at: https://doi.org/10.3389/fgene.2019.00452
29. Schumann Y, Gocke A, Neumann JE. Computational Methods for Data Integration and Imputation of Missing Values in Omics Datasets. *Proteomics.* 2024;25(1–2):e202400100. Available at: https://doi.org/10.1002/pmic.202400100
30. Micheletti S, Schlauch D, Quackenbush J, Ben Guebila M. Higher-order correction of persistent batch effects in correlation networks. *Bioinformatics.* 2024;40(9):btae531. Available at: https://doi.org/10.1093/bioinformatics/btae531
31. Sharifani K, Amini M. Machine Learning and Deep Learning for Big Data Analytics: a Review of Methods and Applications. *World Inf Technol Eng J.* 2023;10(7):3897–3904. Available at: https://ssrn.com/abstract=4458723
32. Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon.* 2018;4(11):e00938. Available at: https://doi.org/10.1016/j.heliyon.2018.e00938
33. Christobel YA, Suji RJ. A Comprehensive Review on Neural Network Architectures. *JoCAAA*. 2024;33(07):1443–1448. Available at: https://eudoxuspress.com/index.php/pub/article/view/1329
34. Nehra N, Sangwan P, Kumar D. Artificial Neural Networks. *Handbook of Machine Learning for Computational Optimization.* 2021:203–227. Available at: https://doi.org/10.1201/9781003138020-11
35. Hossain Y, Ferdous Z, Wahid T, et al. Enhancing intrusion detection systems: Innovative deep learning approaches using CNN, RNN, DBN and autoencoders for robust network security. *Appl Comput Sci*. 2025;21(1):111–125. Available at: https://doi.org/10.35784/acs_6667
36. Godinez WJ, Hossain I, Lazic SE, Davies JW, Zhang X. A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics.* 2017;33(13):2010–2019. Available at: https://doi.org/10.1093/bioinformatics/btx069
37. Pimpalkar A, Gandhewar N, Shelke N, Patil S, Chhabria S. An Efficient Deep Convolutional Neural Networks Model for Genomic Sequence Classification. *Genomics at the Nexus of AI, Computer Vision, and Machine Learning.* 2024:345–375. Available at: https://doi.org/10.1002/9781394268832.ch16
38. Mienye ID, Swart TG, Obaido G. Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications. *Information.* 2024;15(9):517. Available at: https://doi.org/10.3390/info15090517
39. Li P, Pei Y, Li J. A comprehensive survey on design and application of autoencoder in deep learning. *Applied Soft Computing.* 2023;138:110176. Available at: https://doi.org/10.1016/j.asoc.2023.110176
40. Wani SA, Khan SA, Quadri S. Application of Deep Learning for Single Cell Multi-Omics: A State-of-the-Art Review. *Arch Computat Methods Eng.* 2025;32(5):2987–3029. Available at: https://doi.org/10.1007/s11831-025-10230-x
41. Taye MM. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers.* 2023;12(5):91. Available at: https://doi.org/10.3390/computers12050091
42. Zhang J, Che Y, Liu R, Wang Z, Liu W. Deep learning-driven multi-omics analysis: enhancing cancer diagnostics and therapeutics. *Brief Bioinform*. 2025;26(4):bbaf440. Available at: https://doi.org/10.1093/bib/bbaf440
43. Wang Y, Liu L, Wang C. Trends in using deep learning algorithms in biomedical prediction systems. *Front Neurosci.* 2023;17:1256351. Available at: https://doi.org/10.3389/fnins.2023.1256351
44. Bressloff PC. Stochastic gene expression and regulatory networks. *Inderdiscip Appl Math.* 2021:233–355. Available at: https://doi.org/10.1007/978-3-030-72515-0_5
45. Savitha S, Keerthana R, Logeswaran K, Keerthika P, Sharmila V, Sangeetha M. Integration of Multi-Omics Data. *Advances in Computational Intelligence and Robotics.* 2025:149–184. Available at: https://doi.org/10.4018/979-8-3693-9521-9.ch006
46. Wang RC, Wang Z. Precision Medicine: Disease Subtyping and Tailored Treatment. *Cancers.* 2023;15(15):3837. Available at: https://doi.org/10.3390/cancers15153837
47. Zafar I, Anwar S, kanwal F, et al. Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine. *Biomed Signal Process Control.* 2023;86:105263. Available at: https://doi.org/10.1016/j.bspc.2023.105263
48. Menyhárt O, Győrffy B. Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Comput Struct Biotechnol J.* 2021;19:949–960. Available at: https://doi.org/10.1016/j.csbj.2021.01.009
49. Akan T, Akan S, Alp S, et al. Deep Learning in neuroimaging for neurodegenerative diseases: State-of-the art, Challenges, and Opportunities. *J Neurol Sci*. 2025;478:123735. Available at: https://doi.org/10.1016/j.jns.2025.123735
50. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. *NPJ Breast Cancer.* 2023;9(1):21. Available at: https://doi.org/10.1038/s41523-023-00518-1
51. Dai Y, Wu D, Carroll I, Zou F, Zou B. High-dimensional biomarker identification for interpretable disease prediction via machine learning models. *Bioinformatics.* 2025;41(5):btaf266. Available at: https://doi.org/10.1093/bioinformatics/btaf266
52. Miao K, Houssou Hounye A, Su L, et al. Exploring explainable machine learning and Shapley additive exPlanations (SHAP) technique to uncover key factors of HNSC cancer: An analysis of the best practices. *Biomed Signal Process Control.* 2024;89:105752. Available at:

https://doi.org/10.1016/j.bspc.2023.105752
53. Ali SIM, Alrashid SZ. A review of methods for gene regulatory networks reconstruction and analysis. *Artif Intell Rev*. 2025;58(8):256. Available at: https://doi.org/10.1007/s10462-025-11257-z
54. Alharbi F, Vakanski A. Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review. *Bioengineering*. 2023;10(2):173. Available at: https://doi.org/10.3390/bioengineering10020173
55. Simidjievski N, Bodnar C, Tariq I, et al. Variational Autoencoders for Cancer Data Integration: Design Principles and Computational Practice. *Front Genet*. 2019;10:1205. Available at: https://doi.org/10.3389/fgene.2019.01205
56. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. *Brief Bioinform*. 2021;23(1):bbab454. Available at: https://doi.org/10.1093/bib/bbab454
57. Ogunjobi TT, Ohaeri PN, Akintola OT, et al. Bioinformatics Applications in Chronic Diseases: A Comprehensive Review of Genomic, Transcriptomics, Proteomic, Metabolomics, and Machine Learning Approaches. *Medinformatics*. 2024. Available at: https://doi.org/10.47852/bonviewMEDIN42022335
58. Mathema VB, Sen P, Lamichhane S, Orešič M, Khoomrung S. Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine. *Comput Struct Biotechnol J*. 2023;21:1372–1382. Available at: https://doi.org/10.1016/j.csbj.2023.01.043
59. Hassan Ali. Artificial intelligence in multi-omics data integration: Advancing precision medicine, biomarker discovery and genomic-driven disease interventions. *Int J Sci Res Arch*. 2023;8(1):1012–1030. Available at: https://doi.org/10.30574/ijsra.2023.8.1.0189
60. Neloy AA, Turgeon M. A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs. *Mach Learn Appl*. 2024;17:100572. Available at: https://doi.org/10.1016/j.mlwa.2024.100572
61. Ghosh P, Sajjadi MSM, Vergari A, Black M, Schölkopf B. From Variational to Deterministic Autoencoders. *ArXiv*. 2019. Available at: https://doi.org/10.48550/arXiv.1903.12436
62. Rahmanian M, Mansoori EG. MoVAE: Multi-Omics Variational Auto-Encoder for Cancer Subtype Detection. *IEEE Access*. 2024;12:133617–13331. Available at: https://doi.org/10.1109/ACCESS.2024.3462543
63. Erfanian N, Heydari AA, Feriz AM, et al. Deep learning applications in single-cell genomics and transcriptomics data analysis. *Biomed Pharmacother*. 2023;165:115077. Available at: https://doi.org/10.1016/j.biopha.2023.115077
64. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. *Brief Bioinform*. 2022;23(2):bbab569. Available at: https://doi.org/10.1093/bib/bbab569
65. Alqudah AM, Moussavi Z. A Review of Deep Learning for Biomedical Signals: Current Applications, Advancements, Future Prospects, Interpretation, and Challenges. *Comput Mater Contin*. 2025;83(3):3753–3841. Available at: https://doi.org/10.32604/cmc.2025.063643
66. Mobasher G. WeLT: Weighted Loss Trainer for Biomedical Joint Entity and Relation Extraction. *Heidelberg University*. 2025. Available at: https://archiv.ub.uni-heidelberg.de/volltextserver/35978/
67. Palanisamy B, Hassija V, Chatterjee A, et al. Transformers for Vision: A Survey on Innovative Methods for Computer Vision. *IEEE Access*. 2025;13:95496–95523. Available at: https://doi.org/10.1109/ACCESS.2025.3571735
68. Hegde A, Nguyen T, Cheng J. Machine learning methods for gene regulatory network inference. *Brief Bioinform*. 2025;26(5):bbaf470. Available at: https://doi.org/10.1093/bib/bbaf470
69. Saxena A, Nixon B, Boyd A, Evans J, Faraone SV. A Systematic Review of the Application of Graph Neural Networks to Extract Candidate Genes and Biological Associations. *Am J Med Genet B Neuropsychiatr Genet*. 2025;198(6):3–18. Available at: https://doi.org/10.1002/ajmg.b.33031
70. Zhang X, Liu Q. A graph neural network approach for hierarchical mapping of breast cancer protein communities. *BMC Bioinformatics*. 2025;26(1):23. Available at: https://doi.org/10.1186/s12859-024-06015-x
71. Ponzi V, Napoli C. Graph Neural Networks: Architectures, Applications, and Future Directions. *IEEE Access*. 2025;13:62870–62891. Available at: https://doi.org/10.1109/ACCESS.2025.3558752
72. Jiang W, Ye W, Tan X, Bao YJ. Network-based multi-omics integrative analysis methods in drug discovery: a systematic review. *BioData Min*. 2025;18(1):27. Available at: https://doi.org/10.1186/s13040-025-00442-z
73. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Inf Fusion*. 2019;50:71–91. Available at: https://doi.org/10.1016/j.inffus.2018.09.012
74. Lin Y, Zhang W, Cao H, Li G, Du W. Classifying Breast Cancer Subtypes Using Deep Neural Networks Based on Multi-Omics Data. *Genes (Basel)*. 2020;11(8):888. Available at: https://doi.org/10.3390/genes11080888
75. Nguyen TM, Kim N, Kim DH, et al. Deep Learning for Human Disease Detection, Subtype Classification, and Treatment Response Prediction Using Epigenomic Data. *Biomedicines*. 2021;9(11):1733. Available at: https://doi.org/10.3390/biomedicines9111733
76. Lee M. Deep Learning Techniques with Genomic Data in Cancer Prognosis: A Comprehensive Review of the 2021 – 2023 Literature. *Biology*. 2023;12(7):893. Available at: https://doi.org/10.3390/biology12070893
77. Hegde MG, Shenoy PD, Venugopal KR, Canchi A. A deep learning framework for chronic kidney disease stage classification. *Healthcare Anal*. 2025;7:100398. Available at: https://doi.org/10.1016/j.health.2025.100398
78. Sartori F, Codicè F, Caranzano I, et al. A Comprehensive Review of Deep Learning Applications with Multi-Omics Data in Cancer Research. *Genes*. 2025;16(6):648. Available at: https://doi.org/10.3390/genes16060648
79. Shi W, Yang H, Xie L, Yin XX, Zhang Y. A review of machine learning-based methods for predicting drug–target interactions. *Health Inf Sci Syst*. 2024;12(1):30. Available at: https://doi.org/10.1007/s13755-024-00287-6
80. Vefghi A, Rahmati Z, Akbari M. Drug-target interaction/affinity prediction: Deep learning models and advances review. *Comput Biol Med*. 2025;196:110438. Available at: https://doi.org/10.1016/j.compbiomed.2025.110438
81. Ali AM, Mohammed MA. A Comprehensive Review of Artificial Intelligence Approaches in Omics Data Processing: Evaluating Progress and Challenges. *IJMSCS*. 2023;2:114–167. Available at: https://doi.org/10.59543/ijmscs.v2i.8703
82. Huang D, Li Z, Jiang T, Yang C, Li N. Artificial intelligence in lung cancer: current applications, future perspectives, and challenges. *Front Oncol*. 2024;14:1486310. Available at: https://doi.org/10.3389/fonc.2024.1486310
83. Mishra A, Sahu SK, Mistry S. Role of computational biology in the diagnosis of neurodegenerative disorders. *Computational Intelligence for Genomics Data (Academic Press)*. 2025:167–179. Available at: https://doi.org/10.1016/B978-0-443-30080-6.00004-3
84. Mohammed MA, Abdulkareem KH, Dinar AM, Zapirain BG. Rise of Deep Learning Clinical Applications and Challenges in Omics Data: A Systematic Review. *Diagnostics*. 2023;13(4):664. Available at:

https://doi.org/10.3390/diagnostics13040664
85. Sapoval N, Aghazadeh A, Nute MG, et al. Current progress and open challenges for applying deep learning across the biosciences. *Nat Commun*. 2022;13(1):1728. Available at: https://doi.org/10.1038/s41467-022-29268-7
86. Ahmed SF, Alam MSB, Hassan M, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif Intell Rev*. 2023;56(11):13521–13617. Available at: https://doi.org/10.1007/s10462-023-10466-8
87. Yu Y, Mai Y, Zheng Y, Shi L. Assessing and mitigating batch effects in large-scale omics studies. *Genome Biol.* 2024;25(1):254. Available at: https://doi.org/10.1186/s13059-024-03401-9
88. Hassija V, Chamola V, Mahapatra A, et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn Comput*. 2023;16(1):45–74. Available at: https://doi.org/10.1007/s12559-023-10179-8
89. Mersha M, Lam K, Wood J, AlShami AK, Kalita J. Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing*. 2024;599:128111. Available at: https://doi.org/10.1016/j.neucom.2024.128111
90. Fayaz S, Ahmad Shah SZ, ud din NM, Gul N, Assad A. Advancements in Data Augmentation and Transfer Learning: A Comprehensive Survey to Address Data Scarcity Challenges. *Recent Adv Comput Sci Commun*. 2024;17(8):e100124225452. Available at: https://doi.org/10.2174/0126662558286875231215054324
91. Braconi D, Nadwa H, Bernardini G, Santucci A. Omics and rare diseases: challenges, applications, and future perspectives. *Expert Rev Proteomics*. 2025;22(3):107–122. Available at: https://doi.org/10.1080/14789450.2025.2468300
92. Han H. Challenges of reproducible AI in biomedical data science. *BMC Med Genomics*. 2025;18(Suppl 1):8. Available at: https://doi.org/10.1186/s12920-024-02072-6
93. Semmelrock H, Ross - Hellauer T, Kopeinik S, et al. Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*. 2025;46(2):e70002. Available at: https://doi.org/10.1002/aaai.70002
94. Murugaiah M. Application of Machine Learning and Deep Learning in High Performance Computing. *High Performance Computing in Biomimetics*. 2024:271–286. Available at: https://doi.org/10.1007/978-981-97-1017-1_14
95. Reddy R. Sustainable Computing: A Comprehensive Review of Energy-Efficient Algorithms and Systems. *TechRxiv*. 2024. Available at: https://doi.org/10.36227/techrxiv.172831425.52143965/v1
96. Simon BD, Ozyoruk KB, Gelikman DG, Harmon SA, Türkbey B. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. *Diagn Interv Radiol*. 2025;31(4):303–312. Available at: https://doi.org/10.4274/dir.2024.242631
97. Xu X, Li J, Zhu Z, et al. A Comprehensive Review on Synergy of Multi-Modal Data and AI Technologies in Medical Diagnosis. *Bioengineering (Basel)*. 2024;11(3):219. Available at: https://doi.org/10.3390/bioengineering11030219
98. Waqas A, Tripathi A, Ramachandran RP, Stewart PA, Rasool G. Multimodal data integration for oncology in the era of deep neural networks: a review. *Front Artif Intell*. 2024;7:1408843. Available at: https://doi.org/10.3389/frai.2024.1408843
99. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf Fusion*. 2022,77:29–52. Available at: https://doi.org/10.1016/j.inffus.2021.07.016
100. Nguyen DC, Ding M, Pathirana PN, Seneviratne A, Li J, Vincent Poor H. Federated Learning for Internet of Things: A Comprehensive Survey. *IEEE Commun Surv Tutorials*. 2021;23(3):1622–1658. Available at: https://doi.org/10.1109/COMST.2021.3075439
101. Tuan DA. Bridging the Gap Between Black Box AI and Clinical Practice: Advancing Explainable AI for Trust, Ethics, and Personalized Healthcare Diagnostics. *Preprints*. 2024. Available at: https://doi.org/10.20944/preprints202409.1974.v1
102. Alobaidi S. Emerging Biomarkers and Advanced Diagnostics in Chronic Kidney Disease: Early Detection Through Multi-Omics and AI. *Diagnostics (Basel)*. 2025;15(10):1225. Available at: https://doi.org/10.3390/diagnostics15101225
103. Komolafe TE, Monkam P, Komolafe BF, Wang N. *Modern Technologies in Healthcare: AI, Computer Vision*. CRC Press. 2025. Available at: https://doi.org/10.1201/9781003481959
104. Li F, Wang S, Gao Z, et al. Harnessing artificial intelligence in sepsis care: advances in early detection, personalized treatment, and real-time monitoring. *Front Med*. 2025;11:1510792. Available at: https://doi.org/10.3389/fmed.2024.1510792