

Genetic Disorders Prediction From Multi-Omics Using Deep Learning Approaches: A Comprehensive Review

Sowmya K¹, Ananth Prabhu G²

¹Research Scholar, Department of CSE, Sahyadri College of Engineering and Management
Visvesvaraya Technological University Belagavi, India

²Research Supervisor, Department of CSE, Sahyadri College of Engineering and Management,
Visvesvaraya Technological University Belagavi, India

Abstract

Genetic disorders pose a substantial challenge to public health due to their many facets and capacity to affect people throughout their lives. The emergence of high-throughput genomic technologies, and the growing availability of multi-omics datasets make the implementation of sophisticated techniques, including deep learning, possible for predicting and identifying genetic disorders at an early stage. This review paper focuses on sophisticated deep learning techniques and their application in the prediction of genetic disorders from multi-omics data. The review encompasses various deep learning designs, including convolutional, recurrent, and generative adversarial networks, as well as their use in genomics and related fields. Moreover, the review analyses how multi-omics data is integrated to improve the accuracy of the predictive models using transfer learning, attention, and graph neural networks. This paper also describes successful cases of utilizing deep learning to predict various genetic disorders such as neurodegenerative disease, cancer predisposition syndromes, cardiovascular disorders, and several rare genetic conditions. Another contributed topic is the results of applying deep learning to general problems of multi-dimensional data integration, and the comparison of the results to approaches based on traditional machine learning. After discussing these deep learning models for genetic disorder prediction, the review demonstrates the necessity of deep evaluation, concerns of model interpretability, and the use and value of these models for aiding health decision making. The objective of this review is to assist researchers and clinicians wishing to personalize predictions and intervene early for patients with genetic disorders through the utilization of deep learning.

Keywords: Genetic disorders, multi-omics, deep learning, convolutional neural networks, recurrent neural networks, transfer learning, interpretability, genetic disease prediction.

1. Introduction

Disorders due to genetics are developed due to an alteration in the structure of the person's Deoxyribonucleic Acid (DNA). These changes, also known as mutations, could range anywhere from a single gene to multiple genes or entire chromosomes. Some genetic disorders are linked with the family pedigree, whereas, others develop on their own during a person's lifetime. The severity of these disorders varies from minor to major and could influence different areas of a person's overall health and development [1]. Genetic disorders are incredibly common worldwide, with each type displaying different features and facing different problems. Some common examples are Down syndrome, cystic fibrosis, and sickle cell anemia. These disorders can spread to any region of the body and may begin to show signs and symptoms at various points in a person's life, whether it be before birth or during adulthood [2]. There are numerous reasons for the importance of comprehending genetic disorders. Most importantly, it assists in providing more accurate diagnoses and developing better treatment methods. Secondly, it helps to understand basic aspects of human physiology and disease. Finally, it helps in proving targeted therapeutic intervention based on the unique genetic composition of the individual [3].

In contemporary medicine, reverse engineering the emergence of genetic problems is exceedingly critical. This affords the patient proper medical attention, which greatly optimizes the Outlook and living standards on their life. In some instances, the symptoms may not only be managed but also completely stifled or the disease progression may be significantly hindered [4]. Take certain metabolic abnormalities as an example: when these are diagnosed early and combined with specific dietary intake, they can avert severe growth disabilities or potentially fatal problems. Much like the proper vigilance tends to the life-threatening cancer predisposing conditions, proper care can also increase the life span [5]. Figure 1 illustrates the dividends of early discovery of disease upon patients. Besides, early forecasts allow for better preparation on the side of families and healthcare practitioners. It permits for advanced forms of genetic counseling which empower families to make sound choices on reproductive issues and health care that arises later on. From the standpoint of medical systems, it can be more focused with allocating scarce resources and dealing with the creation of specific level medicines [6].

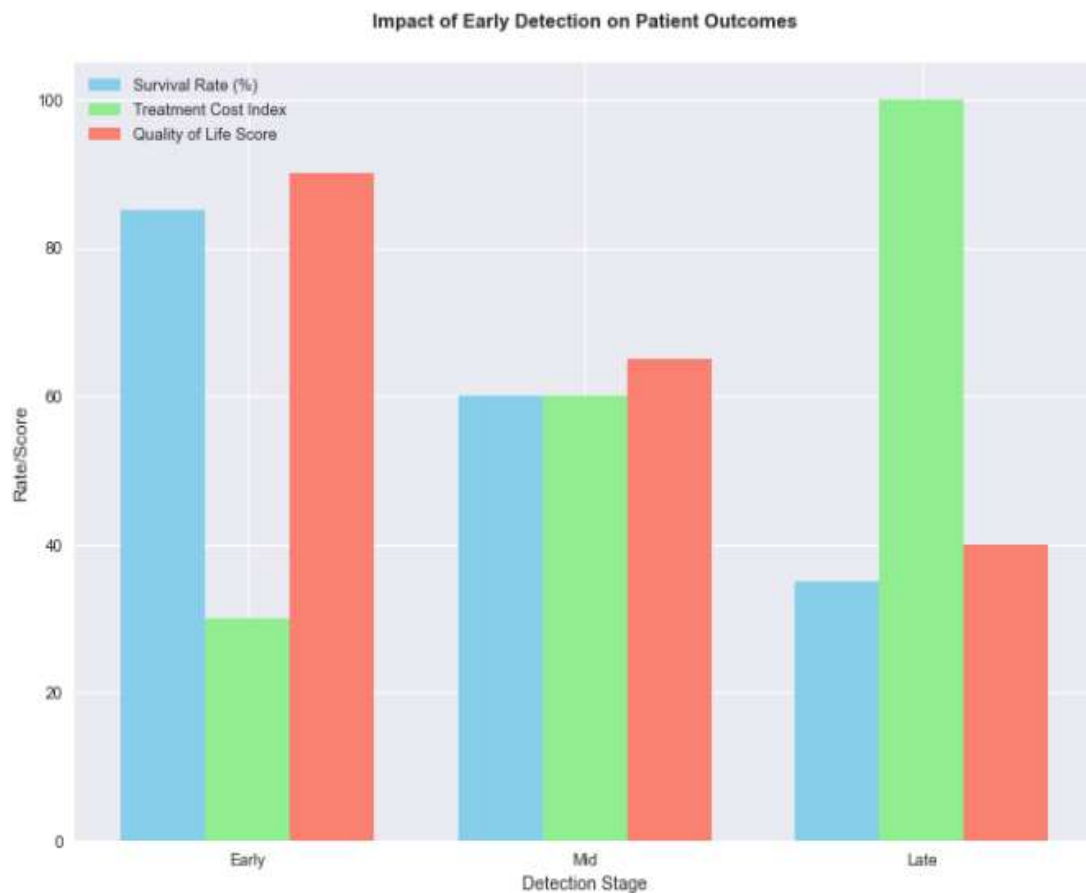


Figure 1: Impact of early detection on patient outcomes

Multi-omics is the combination of two or more biological datasets and has advanced the understanding of living systems. This combination has changed the area of genetics and genomics as it offers a more integrated approach to biological processes [7]. Omics is a term used for the many biological fields that end with -omics such as genomics (study of genes), transcriptomics (study of RNA transcripts), proteomics (study of proteins), and metabolomics (study of metabolites). Each of the stated sciences focuses on a certain level of biological function [8]. By combining the different levels of data, omics provides the possibility of revealing intricate interrelations that would be missed if only one type

of data is considered. For instance, genomic data and matching proteomic data can be studied to explain how genetic differences determine the quantity and activity of proteins produced in an organism, which gives further detail into the understanding of diseases [9]. High-throughput technologies have greatly increased the ability to produce enormous amounts of multi-omics data. Unfortunately, integrating and interpreting these multi-omics data remains a major issue due to the complexity and diversity of the data [10].

With the aid of deep learning, the novel world of IoT systems can now extract information from the genome using artificial neural networks which analyze vast amounts of data through layers. Deep learning algorithms can cross mountains of biological data looking for certain sequences or even predicting certain outcomes. These algorithms can accurately analyze and predict multi-functional genomic applications like gene expression, calling variants, and protein structure mapping [12]. One of the key advantages of deep learning in genomics is automatic feature detection from raw untrained data without any advance preparations [13]. This feature can come in particularly handy when spacing with the proteins or even the DNA in question. Moreover, deep learning models also account for complex structures and sequential relationships contained within biological data, which more often than not get left out by conventional statistical models. This property enables deep learning to be the best choice for predicting variable important functional reasons or associating disease-related genes [14].

1.5. Scope and objectives of the review

This comprehensive review seeks to understand the relationship between deep learning methods, genetic disorders, and multi-omics data. The main goal of the review is to try and explain how deep learning approaches are utilized on multi-omics data to solve and comprehend genetic disorders [15]. The review will start with a description of what multi-omics data is and the problems that come with its integration. We will proceed to explain the different forms of deep learning that have been successfully employed in genomics, particularly those used for multi-omics integration [16]. The review will describe several case studies on the application of these approaches in solving and elucidating various genetic disorders. We will measure the effectiveness of these methods against more conventional approaches and highlight the benefits and shortcomings [17]. Moreover, we will discuss the challenges facing the discipline, such as problems with data validity, model explanation, and other ethical issues [18]. Readers should be able to gauge the comprehensive approaches used for predicting genetic disorders using multi-omics data deep learning framework, alongside gaps for potential research focus in this area by the end of this review [19].

2. Multi-Omics Data in Genetic Disorder Prediction

The advent of high-throughput technologies has enabled the collection of diverse types of omics data, which are crucial for understanding the complex biological underpinnings of genetic disorders. Multi-omics data integration, encompassing genomics, transcriptomics, proteomics, metabolomics, and epigenomics, provides a comprehensive view of biological processes and facilitates the prediction of genetic disorders. This section delves into the various types of omics data, the challenges associated with their integration, and the preprocessing and normalization techniques essential for effective analysis.

2.1. Types of Omics Data

Genomics involves the comprehensive study of an organism's entire genome, including the sequencing and analysis of DNA. It provides insights into genetic variations, such as single nucleotide polymorphisms (SNPs) and copy number variations (CNVs), which are often implicated in genetic disorders.

2.1.1. Genomics

Genomic data serves as the foundational layer for understanding the hereditary aspects of diseases, enabling researchers to identify potential genetic markers associated with specific conditions Chung & Kang [10]. For instance, genomic analyses have been pivotal in identifying mutations linked to various cancers, thereby aiding in risk assessment and personalized treatment strategies [11].

2.1.2. Transcriptomics

The study of transcriptomics centers on the RNA transcript produced by a particular genome in specified conditions or particular cell types. This discipline sheds light onto gene imprints which may be modified in a multitude of genetic aberrations. Researchers can discover altered genes and pathways controlling phenotypes of a disease by examining transcriptomic data [12]. For instance, the differential expression of mRNA and non-coding RNAs provided deep understanding of the mechanisms of schizophrenia and autism spectrum disorders [13]. The combination of transcriptomic and genomic information increases the knowledge of regulation of genes and their role in the pathology of diseases.

2.1.3. Proteomics

Proteomics is the large-scale study of proteins, particularly their functions and structures. Given that proteins are the primary effectors of cellular functions, proteomic data is essential for understanding the biological mechanisms of genetic disorders. Proteomic analyses can identify post-translational modifications and protein-protein interactions that are crucial for cellular signaling pathways [14]. For instance, studies have shown that proteomic profiling can uncover biomarkers for early detection of diseases like Alzheimer's and various cancers, thus facilitating timely interventions [15]. The integration of proteomic data with genomic and transcriptomic information provides a multi-faceted view of the molecular landscape of diseases.

2.1.4. Metabolomics

The metabolomics field focuses on the analysis of metabolites in living organisms within biospecimen samples. It can tell how metabolism works and how it is modified in various diseases. This omics layer is useful for assessing the biochemical changes related to a certain genetic disorder or any disturbance [16]. For instance, metabolites patterns have been defined for diabetes and other cardiovascular disorders which aid in diagnosis and prognosis [17]. The combination of other omics layers with metabolomics fervently aids in appreciating the genetic, transcriptomic, and proteomic alterations concerning particular diseases.

2.1.5. Epigenomics

Epigenomics investigate the added chemical modifications on the DNA strands as well as the histone proteins that regulate the gene expression without changing the base DNA structure. For example, dye modification and acetylation of the histones serve as controls for many genes and can be affected by the environment [18]. Epigenomic alterations

were shown to be associated with several heritable conditions including malignancies and neurodevelopmental disorders. By combining epigenomic information with genomic and transcriptomic data, scientists can understand more about the regulation of the disease and its progression [19].

2.2. Integration Challenges of Multi-Omics Data

Despite the potential benefits of integrating multi-omics data, several challenges hinder effective analysis. One significant challenge is the heterogeneity of data types, which often exhibit different distributions, scales, and measurement units. This variability complicates the integration process, as combining disparate data sources can lead to biased or misleading results [20]. Additionally, missing data is a common issue in multi-omics studies, as not all samples may have complete information across all omics layers. This incompleteness can limit the statistical power of analyses and complicate the interpretation of results [21]. Furthermore, the computational complexity associated with analyzing large-scale multi-omics datasets necessitates the development of robust algorithms and frameworks capable of handling such intricacies [22].

2.3. Preprocessing and Normalization Techniques

To address the challenges of multi-omics data integration, preprocessing and normalization techniques are essential. Preprocessing involves cleaning the data to remove noise and artifacts, ensuring that the datasets are suitable for analysis. Common preprocessing steps include quality control, filtering out low-quality samples, and correcting batch effects that may arise from variations in sample processing [23]. Normalization techniques are crucial for adjusting the data to account for systematic biases and differences in measurement scales across omics layers. Methods such as quantile normalization, log transformation, and variance stabilization are frequently employed to achieve comparability among datasets [24]. By applying these preprocessing and normalization techniques, researchers can enhance the reliability and interpretability of multi-omics analyses, ultimately leading to more accurate predictions of genetic disorders. By leveraging the strengths of genomics, transcriptomics, proteomics, metabolomics, and epigenomics, researchers can gain a comprehensive understanding of the molecular mechanisms underlying diseases. However, addressing the challenges associated with data integration and employing robust preprocessing techniques are critical for maximizing the potential of multi-omics studies in genetic disorder prediction.

3. Deep Learning Approaches in Genomics

Convolutional Neural Networks (CNNs) represent a specific deep learning architecture that has found extensive application in the field of genomics [25]. CNNs excel in identifying local patterns and features within sequential data, including DNA sequences [26]. Convolutional and pooling layers are employed to extract hierarchical features from the input data, rendering them effective for tasks such as predicting transcription factor binding, regulatory elements, and various genomic features from DNA sequences [27]. Figure 2 illustrates the publications related to deep learning methodologies in genomics throughout the years.

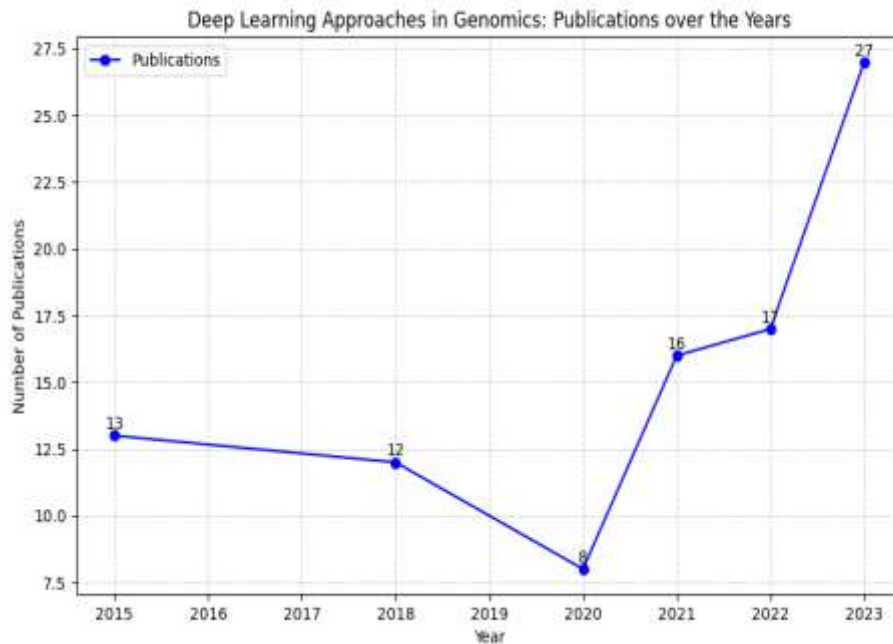


Figure 2: Publication over the Years

3.1.2. Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNN's) are a different type of model that fall under deep learning, capable of handling sequential information like DNA and RNA sequences [28]. They can incorporate long-term dependencies in sequential data by keeping a hidden state that is updated on each time step [29]. Hence, they are suitable for performing DNA modification prediction, gene annotation, and metagenomics classification [30].

3.1.3. Long Short-Term Memory (LSTM) networks

Long Short-Term Memory (LSTM) networks are a specific type of RNN that are designed to better handle long-term dependencies in sequential data [31]. LSTM networks introduce gating mechanisms that allow the model to selectively remember and forget information, improving their ability to capture long-range contextual information in genomic sequences [32].

3.1.4. Autoencoders

Autoencoders are a type of neural network that can learn efficient data representations in an unsupervised manner. In genomics, autoencoders have been used to learn low-dimensional embeddings of DNA sequences that capture higher-order sequence features, which can then be used as input to other deep learning models [33].

3.1.5. Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a deep learning framework that can be used to generate synthetic data, such as DNA sequences. In genomics, GANs have been explored for tasks like predicting 3D genome folding from DNA sequence [34].

3.2. Transfer learning in genomics

The adopted approach of transfer learning, which consists of implementing a model trained on one task to enhance the results of a similar task, has proven rather effective within genomics. By employing models that were pre-trained on large genomic datasets, researchers were able to reach the best performance for many genomic prediction tasks, even when the available training data was scant [35].

3.3. Interpretability and explainability in deep learning models

With the growing implementation of deep learning models in the field of genomics, the issue of making these models interpretable and explainable is critical in providing biological insights. Some attempts such as attention, saliency maps, and layer visualization have been employed to decipher workings of deep learning models in genomic applications [36]. Figure 3 illustrates how machine learning methods perform accuracy predictions compared to ordinary ones with greater success.

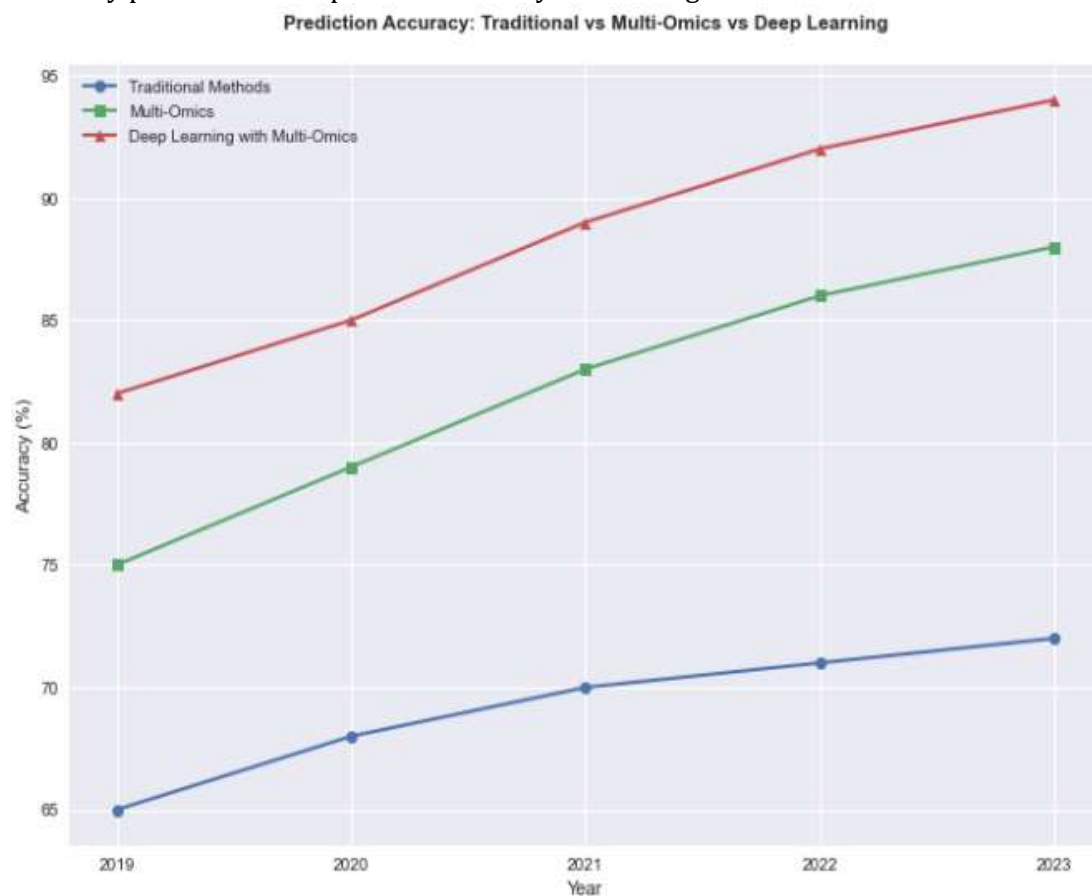


Figure 3: Prediction Accuracy, Traditional vs MultiOmics vs Deep Learning

4. Deep Learning Models for Multi-Omics Integration

4.1. Multi-modal deep learning architectures

Deep learning models have shown great potential in integrating multi-omics data by leveraging their ability to capture complex non-linear relationships. Approaches like autoencoders and graph neural networks (GNNs) have been particularly effective at fusing multi-omics data and learning comprehensive representations of biological systems [37].

4.2. Attention mechanisms for multi-omics data

Deep learning models are trained with an attention mechanism to capture relationships from different data types across various omics domains. During multi-omics integration tasks, attention-based deep learning models focus on the most pertinent features and relationships, leading to superior performance of the model [38].

4.3. Graph neural networks for biological network integration

Graph neural networks (GNNs) have emerged as a powerful tool for integrating multi-omics data by modeling the complex relationships between different molecular entities as a graph structure [14], [15]. GNNs can effectively capture the topological information of biological networks and leverage it to make more accurate predictions [39].

4.4. Ensemble methods combining multiple deep learning models

Ensemble methods that combine the predictions of multiple deep learning models have been shown to improve the performance of multi-omics integration tasks. By leveraging the complementary strengths of different architectures and learning strategies, ensemble approaches can achieve more robust and accurate results compared to single models [40].

5. Applications in Genetic Disorder Prediction

5.1. Case studies of successful predictions

5.1.1. Neurodegenerative disorders

Models based on deep learning techniques are emerging as a powerful option for predicting and detecting various neurodegenerative disorders like Alzheimer's and Parkinson's diseases [41], [42], [43]. They have been able to use neuroimaging, genetics, and other clinical data to achieve high accuracy not only in differentiation of patients from healthy controls but also between various subtypes of neurodegenerative diseases [44], [45].

5.1.2. Cancer predisposition syndromes

The technique of deep learning has been utilized in the forecasting of the genetic cancer predisposition syndromes, particularly focus on the breast and ovarian cancer (HBOC) syndrome [46]. These models have proven that, through genomic analysis, they can pinpoint individuals at increased likelihood for certain cancer types and facilitate early intervention and more individualized prevention methods [47].

5.1.3. Cardiovascular genetic disorders

Deep learning approaches have been employed in the prediction of genetic cardiovascular disorders, including familial hypercholesterolemia and cardiomyopathies. These models have shown the potential to improve the accuracy of genetic risk assessment and facilitate the implementation of preventive measures for individuals with a genetic predisposition to these conditions [48].

5.1.4. Rare genetic diseases

For rare genetic diseases, deep learning models have been utilized to assist in the diagnosis and management of these conditions. By leveraging genomic data, such as gene expression profiles and sequence information, these models have been able to identify

disease-associated patterns and facilitate the development of personalized treatment strategies [49].

5.2. Comparison with traditional machine learning approaches

While traditional machine learning methods have been widely used in genetic disorder prediction, deep learning models have demonstrated superior performance in many cases. Deep learning's ability to automatically extract and learn complex features from raw data, as well as its capacity to handle large and diverse datasets, have contributed to its advantages over traditional approaches.

5.3. Performance metrics and evaluation strategies

The efficacy of deep learning models on predicting genetic disorders was evaluated using parameters such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (ROC) [50]. Cross-validation, independent test sets, and benchmark datasets were strategies used to make the models to ensure that they are robust and generalizable [51]. Figure 4 depicts the distribution of different types of practices in genetic disorder prediction. Applicants to cancer genetics made up approximately 35 percent of the total, compared to 25 percent for applicants pursuing genetics of neurological disorders. Other genetic disorders like cardiovascular and metabolic diseases account for 20 percent and 15 percent respectively.

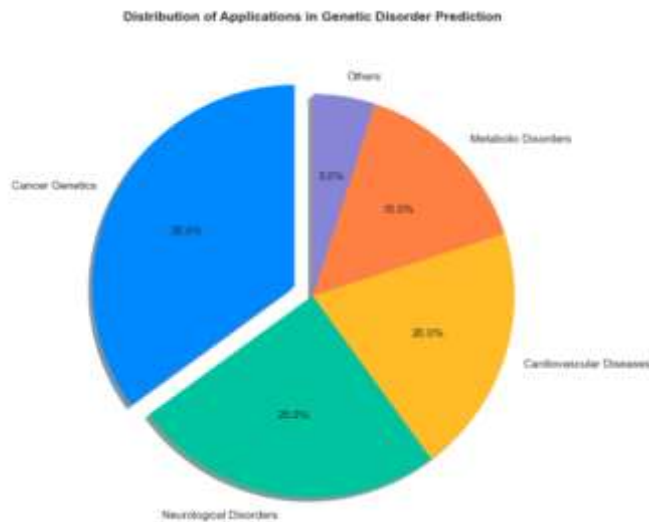


Figure 4: Applications in Genetic Disorder Prediction

6. Challenges and Limitations

6.1. Data quality and standardization issues

One of the key challenges in applying deep learning to genomics is the quality and standardization of the available data. Genomic datasets can be highly heterogeneous, with variations in data formats, experimental protocols, and annotation standards across different studies and sources. This lack of data standardization can hinder the effective integration and analysis of multi-omics data using deep learning models [52].

6.2. Computational resources and scalability

The training and deployment of deep learning models for genomic applications can be computationally intensive, requiring significant hardware resources, such as high-performance GPUs and large memory capacities. As the size and complexity of genomic

datasets continue to grow, the scalability of deep learning approaches becomes a critical concern, particularly for real-time or large-scale applications [53].

6.3. Overfitting and generalization

When there's a lack of training information, deep learning models are likely to overfit. In this situation, the model is able to successfully predict information from the training data, but is unable to make predictions on new data. Addressing overfitting and enhancing the generalization capabilities of deep learning models, particularly in genomics, is an active research area. Regularization, data augmentation, and ensemble methods are some of the techniques being studied [54].

6.4. Ethical considerations and privacy concerns

The application of deep learning in genomics poses significant ethical and privacy challenges, especially on how to manage sensitive individual genetic data. Researchers have to make sure that appropriate measures are taken to protect the data, and that the creation and implementation of deep learning models observe ethical standards and legal frameworks.

6.5. Interpretability of complex models

Interpreting and understanding deep learning models is complex due to their sophisticated design and high parameter count. This lack of interpretability poses a challenge for the deployment of deep learning in genomics, particularly for clinical decision-making and biological reasoning where explanation is fundamental. There are attempts to create deep learning models that are more interpretable and techniques that can provide better explanations of their predictions. Table 1 summarizes the challenges and limitation of the existing approaches for genetic disorder prediction using deep learning techniques.

Table 1: summary of the challenges and limitations

Challenge/ Limitation	Description	References
Data Heterogeneity	Multi-omics data (e.g., genomic, transcriptomic, epigenomic) have different data types, scales, and distributions, making integration and joint analysis challenging.	[54]
Small Sample Size	Many genetic disorder studies have limited sample sizes, which can lead to overfitting and poor generalization of deep learning models.	[55]
Interpretability	Deep learning models are often considered "black boxes", making it difficult to understand the underlying biological mechanisms driving the predictions.	[56]
Genetic Overlap	Genetic disorders often exhibit substantial overlap and pleiotropy, which can complicate the development of disorder-specific prediction models.	[57]
Missing Heritability	The "missing heritability" problem, where the observed genetic variants explain only a fraction of the estimated heritability, limits the predictive power of genetic models.	[58]

Validation and Evaluation	Robust validation and evaluation strategies, such as the use of independent test sets and standardized performance metrics, are crucial but often lacking in genetic disorder prediction studies.	[59]
Translational Challenges	Translating deep learning-based genetic disorder prediction models into clinical practice faces challenges related to interpretability, cost-effectiveness, and ethical considerations.	[60]

7. Future Directions

7.1. Emerging deep learning architectures for multi-omics

As the field of deep learning in genomics continues to evolve, researchers are exploring novel neural network architectures that can more effectively integrate and model the complex relationships within multi-omics data. Heo et al. approaches like hierarchical integration, graph neural networks, and multi-view factorization autoencoders have shown promise in capturing the intricate connections between different data modalities [59].

7.2. Integration of clinical and environmental data

In a bid to improve the accuracy and clinical utility of deep learning models in genomics, there is increasing focus on combining clinical and environmental data such as age, gender, medical history, and even lifestyle together with multi-omics data. This approach has the potential to refine, and make more accurate, the predictions made regarding the risk of diseases, prognosis, and response to treatment [60].

7.3. Federated learning for collaborative research

The adoption of federated learning, a distributed machine learning technique that allows for collaborative model training without the need to share sensitive data, could enable the development of deep learning models for genomics while addressing privacy concerns. This approach can facilitate the pooling of data and resources across multiple institutions, leading to more robust and generalizable models.

7.4. Real-time prediction and monitoring systems

As deep learning models become more accurate and efficient, there is an opportunity to develop real-time prediction and monitoring systems for genetic disorders and other health conditions. These systems could leverage streaming data from wearable devices, electronic health records, and continuous genomic monitoring to provide timely insights and personalized interventions.

7.5. Personalized medicine and treatment optimization

The application of deep learning in genomics could enable revolutionary advances in personalized medicine through strategies that treat the patient based on their genetic makeup, characteristics of disease, and response to therapy. Deep learning models have the capability of discovering new biomarkers, predicting the results of treatment, and optimizing dosages and combinations of drugs for better results in patients.

8. Conclusion

This review has explored the different ways that deep learning has been used in genomics. It has also provided an overview of some of the state-of-the-art architectures in deep learning, which include CNNs, RNNs, and LSTM networks, billing their application in genomic data as their ability to extract intricate patterns and features from the data. The review has noted the successful use of deep models for prediction of DNA changes, gene annotation, and metagenomics classification. Finally, the review has discussed the possibilities of transfer learning and multi-omics data fusion using attention mechanisms and graph neural networks. These innovations have shown how deep learning can utilize multiple data sources to give impressive results in almost all tasks that involve the prediction of genetic disorders. The ability to implement deep learning successfully in genomics on the other hand brings hope to the entire field of genetic disorders investigations and clinical applications. More importantly, the accurate predictions, early diagnosis, and enhance targeted treatment approaches that these models will provide will greatly change the healthcare outcomes of the patients and the methods used for treating them. The analysis mentions examples that concern the use of deep learning for spotting neurodegenerative disorders, cancer predisposition syndromes, cardiovascular genetic disorders, and some rare genetic diseases. These innovations stand to markedly improve the accuracy and effectiveness of diagnosing, managing, and treating deep-seated genetic disorders, providing more effective solutions for the patients.

References

- [1] J. Zhou, C. Park, C. Theesfeld, A. Wong, Y. Yuan, C. Scheckelet al., "Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk", *Nature Genetics*, vol. 51, no. 6, p. 973-980, 2019.
<https://doi.org/10.1038/s41588-019-0420-0>
- [2] A. Rahman, M. Nasir, M. Gollapalli, M. Zubair, M. Saleem, S. Mehmoodet al., "Advance genome disorder prediction model empowered with deep learning", *Ieee Access*, vol. 10, p. 70317-70328, 2022.
<https://doi.org/10.1109/access.2022.3186998>
- [3] K. Kendler, S. Lönn, J. Sundquist, & K. Sundquist, "The actions and interactions of family genetic risk scores for alcohol use disorder and major depression on the risk for these two disorders", *American Journal of Medical Genetics Part B Neuropsychiatric Genetics*, vol. 189, no. 5, p. 128-138, 2022.
<https://doi.org/10.1002/ajmg.b.32909>
- [4] R. Maier, G. Moser, G. Chen, S. Ripke, W. Coryell, J. Potashet al., "Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder", *The American Journal of Human Genetics*, vol. 96, no. 2, p. 283-294, 2015.
<https://doi.org/10.1016/j.ajhg.2014.12.006>
- [5] A. Hilten, S. Kushner, M. Kayser, M. Ikram, H. Adams, C. Klaveret al., "Gennet framework: interpretable deep learning for predicting phenotypes from genetic data", *Communications Biology*, vol. 4, no. 1, 2021.
<https://doi.org/10.1038/s42003-021-02622-z>
- [6] Y. Liu, H. Qu, F. Mentch, J. Qu, X. Chang, K. Nguyenet al., "Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in african american patients", *Molecular Psychiatry*, vol. 27, no. 3, p. 1469-1478, 2022.
<https://doi.org/10.1038/s41380-021-01418-1>

- [7] A. Dingemans, B. Vries, L. Vissers, M. Gerven, & M. Hinne, "Comparing facial feature extraction methods in the diagnosis of rare genetic syndromes", 2022. <https://doi.org/10.1101/2022.08.26.22279217>
- [8] K. Chaudhary, O. Poirion, & L. Lu, "Deep learning-based multi-omics integration robustly predicts survival in liver cancer", *Clinical Cancer Research*, vol. 24, no. 6, p. 1248-1259, 2018. <https://doi.org/10.1158/1078-0432.ccr-17-0853>
- [9] Z. Liu, "Deep multiple-instance learning accurately predicts gene haploinsufficiency and deletion pathogenicity", 2023. <https://doi.org/10.1101/2023.08.29.555384>.
- [10] R. Chung and C. Kang, "A multi-omics data simulator for complex disease studies and its application to evaluate multi-omics data analysis methods for disease classification", *Gigascience*, vol. 8, no. 5, 2019. <https://doi.org/10.1093/gigascience/giz045>
- [11] K. Chaudhary, O. Poirion, & L. Lu, "Deep learning-based multi-omics integration robustly predicts survival in liver cancer", *Clinical Cancer Research*, vol. 24, no. 6, p. 1248-1259, 2018. <https://doi.org/10.1158/1078-0432.ccr-17-0853>
- [12] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. Marioniet al., "Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets", *Molecular Systems Biology*, vol. 14, no. 6, 2018. <https://doi.org/10.15252/msb.20178124>
- [13] S. Tarazona, L. Balzano-Nogueira, A. Schmidt, A. Imhof, T. Hankemeier, J. Tegn  ret al., "Harmonization of quality metrics and power calculation in multi-omic studies", *Nature Communications*, vol. 11, no. 1, 2020. <https://doi.org/10.1038/s41467-020-16937-8>
- [14] M. Bersanelli, E. Mosca, D. Remondini, E. Giampieri, C. Sala, G. Castellani et al., "Methods for the integration of multi-omics data: mathematical aspects", *BMC Bioinformatics*, vol. 17, no. S2, 2016. <https://doi.org/10.1186/s12859-015-0857-9>
- [15] K. Zhang, "Snapatac2: a fast, scalable and versatile tool for analysis of single-cell omics data", 2023. <https://doi.org/10.1101/2023.09.11.557221>
- [16] A. Paganelli, V. Righi, E. Tarentini, & C. Magnoni, "Current knowledge in skin metabolomics: updates from literature review", *International Journal of Molecular Sciences*, vol. 23, no. 15, p. 8776, 2022. <https://doi.org/10.3390/ijms23158776>
- [17] H. Luo, H. Liang, M. Jia, Y. Li, X. Yao, & S. Cong, "Gremi: an explainable multi-omics integration framework for enhanced disease prediction and module identification", 2023. <https://doi.org/10.1101/2023.03.19.533326>
- [18] T. Wang, W. Shao, Z. Huang, H. Tang, J. Zhang, Z. Ding et al., "Mogonet integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification", *Nature Communications*, vol. 12, no. 1, 2021. <https://doi.org/10.1038/s41467-021-23774-w>
- [19] N. Rappoport, R. Safran, & R. Shamir, "Monet: multi-omic module discovery by omic selection", *Plos Computational Biology*, vol. 16, no. 9, p. e1008182, 2020. <https://doi.org/10.1371/journal.pcbi.1008182>
- [20] M. Osseni, P. Tossou, F. Laviolette, & J. Corbeil, "Mot: a multi-omics transformer for multiclass classification tumour types predictions", 2022. <https://doi.org/10.21203/rs.3.rs-1348696/v1>

- [21] T. Eicher, G. Kinnebrew, A. Patt, K. Spencer, K. Ying, Q. Maet al., "Metabolomics and multi-omics integration: a survey of computational methods and resources", *Metabolites*, vol. 10, no. 5, p. 202, 2020. <https://doi.org/10.3390/metabo10050202>
- [22] P. Geng, X. Tong, & Q. Lu, "An integrative u method for joint analysis of multi-level omic data", *BMC Genetics*, vol. 20, no. 1, 2019. <https://doi.org/10.1186/s12863-019-0742-z>
- [23] A. Vázquez, Y. Veturi, M. Behring, S. Shrestha, M. Kirst, M. Resendeet al., "Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles", *Genetics*, vol. 203, no. 3, p. 1425-1438, 2016. <https://doi.org/10.1534/genetics.115.185181>
- [24] J. Hawe, F. Theis, & M. Heinig, "Inferring interaction networks from multi-omics data", *Frontiers in Genetics*, vol. 10, 2019. <https://doi.org/10.3389/fgene.2019.00535>
- [25] Q. Liu, F. Li, G. Yu, D. Wang, C. Xiao, & K. Wang, "Detection of dna base modifications by deep recurrent neural network on oxford nanopore sequencing data", *Nature Communications*, vol. 10, no. 1, 2019. <https://doi.org/10.1038/s41467-019-10168-2>
- [26] R. Schweßinger, M. Gosden, D. Downes, R. Brown, A. Oudelaar, J. Teleniuset al., "Deepc: predicting 3d genome folding using megabase-scale transfer learning", *Nature Chemical Biology*, vol. 17, no. 11, p. 1118-1124, 2020. <https://doi.org/10.1038/s41592-020-0960-3>
- [27] D. Kelley, Y. Reshef, M. Bileschi, D. Belanger, C. McLean, & J. Snoek, "Sequential regulatory activity prediction across chromosomes with convolutional neural networks", *Genome Research*, vol. 28, no. 5, p. 739-750, 2018. <https://doi.org/10.1101/gr.227819.117>
- [28] D. Hindarto, "Comparison of rnn architectures and non-rnn architectures in sentiment analysis", *Sinkron*, vol. 8, no. 4, p. 2537-2546, 2023. <https://doi.org/10.33395/sinkron.v8i4.13048>
- [29] D. Quang and X. Xie, "Danq: a hybrid convolutional and recurrent deep neural network for quantifying the function of dna sequences", 2015. <https://doi.org/10.1101/032821>
- [30] N. Gyawali, "Using recurrent neural networks to detect supernumerary chromosomes in fungal strains causing blast diseases", 2023. <https://doi.org/10.1101/2023.09.17.558148>
- [31] J. Clauwaert, G. Menschaert, & W. Waegeman, "Deepribo: a neural network for precise gene annotation of prokaryotes by combining ribosome profiling signal and binding site patterns", *Nucleic Acids Research*, vol. 47, no. 6, p. e36-e36, 2019. <https://doi.org/10.1093/nar/gkz061>
- [32] Q. Liang, P. Bible, Y. Liu, B. Zou, & L. Wei, "Deepmicrobes: taxonomic classification for metagenomics with deep learning", 2019. <https://doi.org/10.1101/694851>
- [33] W. Kopp, R. Monti, A. Tamburrini, U. Ohler, & A. Akalin, "Janggu - deep learning for genomics", 2019. <https://doi.org/10.1101/700450>
- [34] G. Fudenberg, D. Kelley, & K. Pollard, "Predicting 3d genome folding from dna sequence with akita", *Nature Chemical Biology*, vol. 17, no. 11, p. 1111-1117, 2020. <https://doi.org/10.1038/s41592-020-0958-x>

- [35] E. Lin, C. Lin, & H. Lane, "Deep learning with neuroimaging and genomics in alzheimer's disease", *International Journal of Molecular Sciences*, vol. 22, no. 15, p. 7911, 2021. <https://doi.org/10.3390/ijms22157911>
- [36] Ž. Avsec, M. Weilert, A. Shrikumar, S. Krueger, A. Alexandari, K. Dalalet al., "Base-resolution models of transcription factor binding reveal soft motif syntax", 2019. <https://doi.org/10.1101/737981>
- [37] Q. Liu, F. Li, G. Yu, D. Wang, C. Xiao, & K. Wang, "Detection of dna base modifications by deep recurrent neural network on oxford nanopore sequencing data", *Nature Communications*, vol. 10, no. 1, 2019. <https://doi.org/10.1038/s41467-019-10168-2>.
- [38] D. Jeong, "Goat: gene-level biomarker discovery from multi-omics data using graph attention neural network for eosinophilic asthma subtype", *Bioinformatics*, vol. 39, no. 10, 2023. <https://doi.org/10.1093/bioinformatics/btad582>
- [39] X. Li, J. Ma, L. Leng, M. Han, M. Li, F. Heet al., "Mogcn: a multi-omics integration method based on graph convolutional network for cancer subtype analysis", *Frontiers in Genetics*, vol. 13, 2022. <https://doi.org/10.3389/fgene.2022.806842>
- [40] R. Faurina, A. Wijanarko, A. Heryuanti, S. Ishak, & I. Agustian, "Comparative study of ensemble deep learning models to determine the classification of turtle species", *Computer Science and Information Technologies*, vol. 4, no. 1, p. 24-32, 2023. <https://doi.org/10.11591/cs.it.v4i1.p24-32>
- [41] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo et al., "Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks", *Neuroimage Clinical*, vol. 21, p. 101645, 2019. <https://doi.org/10.1016/j.nicl.2018.101645>
- [42] S. Al-Shoukry, T. Rassem, & N. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: a mini-review", *Ieee Access*, vol. 8, p. 77131-77141, 2020. <https://doi.org/10.1109/access.2020.2989396>
- [43] D. Payares-Garcia, J. Mateu, & J. Krukar, "Spatially informed bayesian neural network for neurodegenerative diseases classification", *Statistics in Medicine*, vol. 42, no. 2, p. 105-121, 2022. <https://doi.org/10.1002/sim.9604>
- [44] N. Chintalapudi, G. Battineni, M. Hossain, & F. Amenta, "Cascaded deep learning frameworks in contribution to the detection of parkinson's disease", *Bioengineering*, vol. 9, no. 3, p. 116, 2022. <https://doi.org/10.3390/bioengineering9030116>
- [45] I. Kollia, A. Stafylopatis, & S. Kollias, "Predicting parkinson's disease using latent information extracted from deep neural networks", 2019. <https://doi.org/10.1109/ijcnn.2019.8851995>
- [46] H. Xiong, P. Lin, J. Yu, J. Ye, L. Xiao, T. Yuan et al., "Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images", *Ebiomedicine*, vol. 48, p. 92-99, 2019. <https://doi.org/10.1016/j.ebiom.2019.08.075>
- [47] A. Rahman, M. Nasir, M. Gollapalli, M. Zubair, M. Saleem, S. Mehmood et al., "Advance genome disorder prediction model empowered with deep learning", *IEEE Access*, vol. 10, p. 70317-70328, 2022. <https://doi.org/10.1109/access.2022.3186998>

- [48] M. Abdelwahab, "Deep learning-based prediction of alzheimer's disease using microarray gene expression data", *Biomedicines*, vol. 11, no. 12, p. 3304, 2023. <https://doi.org/10.3390/biomedicines11123304>
- [49] L. Zhang, M. Wang, M. Liu, & D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis", 2020. <https://doi.org/10.48550/arxiv.2005.04573>
- [50] A. Tagaris, D. Kollias, A. Stafylopatis, G. Tagaris, & S. Kollias, "Machine learning for neurodegenerative disorder diagnosis — survey of practices and launch of benchmark dataset", *International Journal on Artificial Intelligence Tools*, vol. 27, no. 03, p. 1850011, 2018. <https://doi.org/10.1142/s0218213018500112>
- [51] A. Grotzinger, M. Rhemtulla, R. Vlaming, S. Ritchie, T. Mallard, W. Hillet al., "Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits", *Nature Human Behaviour*, vol. 3, no. 5, p. 513-525, 2019. <https://doi.org/10.1038/s41562-019-0566-x>
- [52] A. Montanari and A. Rakhlin, "Deep learning: a statistical viewpoint", 2021. <https://doi.org/10.48550/arxiv.2103.09177>
- [53] C. Xu, P. Coen-Pirani, & X. Jiang, "Empirical study of overfitting in deep learning for predicting breast cancer metastasis", *Cancers*, vol. 15, no. 7, p. 1969, 2023. <https://doi.org/10.3390/cancers15071969>
- [54] A. Grotzinger, M. Rhemtulla, R. Vlaming, S. Ritchie, T. Mallard, W. Hillet al., "Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits", *Nature Human Behaviour*, vol. 3, no. 5, p. 513-525, 2019. <https://doi.org/10.1038/s41562-019-0566-x>
- [55] L. Zhang, M. Wang, M. Liu, & D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis", 2020. <https://doi.org/10.48550/arxiv.2005.04573>
- [56] I. Waldman, H. Poore, J. Lunningham, & J. Yang, "Testing structural models of psychopathology at the genomic level", 2018. <https://doi.org/10.1101/502039>
- [57] Y. Chai, "Deep learning in data science: theoretical foundations, practical applications, and comparative analysis", *Applied and Computational Engineering*, vol. 69, no. 1, p. 1-6, 2024. <https://doi.org/10.54254/2755-2721/69/20241419>
- [58] S. Han, H. Mao, & W. Dally, "Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding", 2015. <https://doi.org/10.48550/arxiv.1510.00149>
- [59] Y. Heo, C. Hwa, G. Lee, J. Park, & J. An, "Integrative multi-omics approaches in cancer research: from biological networks to clinical subtypes", *Molecules and Cells*, vol. 44, no. 7, p. 433-443, 2021. <https://doi.org/10.14348/molcells.2021.0042>
- [60] J. Xu, P. Wu, Y. Chen, Q. Meng, H. Dawood, & H. Dawood, "A hierarchical integration deep flexible neural forest framework for cancer subtype classification by integrating multi-omics data", *BMC Bioinformatics*, vol. 20, no. 1, 2019. <https://doi.org/10.1186/s12859-019-3116-7>
- [61] Wang, Qingyong, Minfan He, Longyi Guo, and Hua Chai. "AFEI: adaptive optimized vertical federated learning for heterogeneous multi-omics data integration." *Briefings in Bioinformatics* 24, no. 5, 2023.