GMR The Original

A Scalable Big Data–Enabled Deep Learning Architecture for Integrated Analysis of DNA and RNA Sequencing Data in Cancer Genomics

Sathasivam Sivamalar[1], Jeyaseelan R[2], Anish Kumar A[3], Dinesh Kumar R[4], Dr. Janani Balachndran [5],Rajashri CK [6]

[1]Scientist, Department of Research, Meenakshi Academy of Higher Education and Research. malarss@maher.ac.in, ORCID: 0000-0002-8774-2773 Scientist, Central Research Laboratory, Meenakshi Medical College Hospital & Research Institute, [2]Meenakshi Academy of Higher Education and Research. Suresh@maher.ac.in, ORCID: 0000-0001-6247-1156
Assisant Professor, Department of Oral Pathology, Meenakshi Ammal Dental College and Hospital, Meenakshi Academy of Higher [3]Education and Research. jeyas@maher.ac.in, ORCID: 0000-0003-2722-3549
Lecturer, Meenakshi College of Pharmacy, Meenakshi Academy of Higher Education and Research. anishka@maher.ac.in, ORCID: 0009-0008-2162-4919
[4]Associate Professor, Arulmigu Meenakshi College of Nursing, Meenakshi Academy of Higher Education and Research. dineshkr@maher.ac.in, ORCID: 0009-0007-5288-5305
[5]Conservative Dentistry and Endodontics , Reader , Sree Balaji Dental College and Hospital, (Affiliated to Bharath Institute of Higher Education and Research), Pallikaranai, chennai – 600100, Email id: Janani.balachandran@gmail.com, Orcid ID: 0000-0001-9432-9672
[6]Assistant Professor,Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research , Chennai, Tamil Nadu, India- 600087 E-mail: rajashrick@maher.ac.in

ABSTRACT

Cancer is a very heterogeneous condition provoked by complicated interplay between the changes on a genetic level and transcriptional regulation, and it is difficult to properly characterise and predict without the involvement of both single-omics data sets. The high dimensionality, nonlinearity and scale of these new sequencing data can be difficult to say the least with traditional statistical and machine learning methods. In order to overcome these shortcomings, this paper suggests a scalable deep learning solution based on big data analysis to combine DNA and RNA sequencing to analyse cancer genomics. In the given framework, a multi-omics fusion strategy, i.e., the combination of DNA- and RNA-based features, is learned via a deep neural network (DNN): in this framework, a set of encoding branches is used to learn features based on DNA and RNA sequences, and a learned combination is created to create a unified predictor. The effectiveness of the proposed approach is tested using publicly available cancer genomics data, such as those of The Cancer Genome Atlas (TCGA). The experimental evidence shows that the integrated DNA RNA model is always better than the single-omics models, as it provides a better accuracy, F1-score, and area under the ROC curve (AUC) in various evaluation environments. These results demonstrate the usefulness of deep learning-based multi-omics integration to provide complementary information on the molecular level and best predictive outcomes. The suggested architecture in general offers a highly scalable, and extended architecture on integrative cancer genomics analysis, which may find potential applications in translational oncology, disease stratification, and precision clinical decision making, supported by data.

Key words: *Cancer genomics; Multi-omics integration; DNA sequencing; RNA sequencing; deep neural networks; big data analytics*

INTRODUCTION

Cancer is a complex and highly diversified condition that is associated with widespread genetic changes and transcription reprogramming that are dynamic over various patients, different tumours, and various levels of clinical progression. This heterogeneity results in a considerable difference in the ensuing diseased behaviour, therapy response, and patient outcome making it incredibly difficult to easily characterise and predict the disease. The recent developments in high-throughput sequencing technologies have facilitated profiling of cancer genomes and transcriptomics across the board, producing high volumes of DNA and RNA sequencing

data that complement each other in normally explaining tumour biology. DNA-based data reflects somatic mutations, copy number changes and structural changes whereas the RNA expression data reflect downstream functional activities and regulatory states in cancer cells (Ballard et al., 2024; Fan et al., 2023; Wekesa and Kimwele, 2023). These heterogeneous molecular layers when appropriately incorporated are important in cancer genomics in predicting tumour progression as well in predictive modelling.

Though there is an increased access to multi-omics data, most cancer studies are still based on single-omics analyses that are inadequate to reveal the multiple layers of interaction between genomic changes and transcriptional regulation. Traditional statistical methods and simple machine learning models are ineffective with high-dimensional sequencing data because the methods have low scalability, are unable to characterise nonlinear relationships, and are not favourable when used to predict (Wekesa and Kimwele, 2023). It has been recently discovered that cancer subtype classification, survival forecast, and prognostic stratification could be greatly enhanced with the use of integrated DNA-RNA methods in comparison with single-omics methodologies (Fan et al., 2023; Ji et al., 2023; Shi et al., 2024; Sun et al., 2023). Nevertheless, multi- omics integration is not easily facilitated due to the differences in the level of data, noise and the distribution of features in omics layers.

Deep learning is a potent paradigm model to fit high-dimensional biological data, to provide the power to automatically learn hierarchical representations and learn non-linear relationships amongst multi-data modalities. Auto encoders, attention mechanisms, and graph-based models based on deep neural networks have been demonstrated to have high prospects of integration and predictive analysis of multi-omics in cancer genomics (Benkirane et al., 2025; Dou and Mirzaei, 2025; Li and Nabavi, 2024; Li et al., 2022; Wu et al., 2024; Zhang et al., 2025). The methods allow the study to learn jointly the representations using both DNA and RNA data, which allows modelling of tumour molecular state with greater precision and with stronger biological predictions. However, the vast amount of models in deep learning is restricted in one way or another by scalability, model complexity, or lack of validation on large and heterogeneous cancer data (Ballard et al., 2024; Li et al., 2025; Wang and O'Connell, 2025).

With these issues driving the study, a scalable deep learning framework of integrated DNA and RNA sequencing analysis of cancer genomics is proposed. The contributions made by the work are three. It, first, proposes a DNN-based multi-omics integration framework, where a dedicated DNA and RNA encoding branch is used, and then an effective fusion plan is followed to learn shared molecular code. Second, the proposed architecture is developed to scale analysis of large cancer genomics datasets, which will make it possible to learn through high-dimensional sequencing data. Third, comprehensive experimental verification of publicly available cancer datasets proves the claim that the integrated DNA RNA model is always of better predictive performance, in terms of accuracy, F1-score, and AUC, when compared to single omics models (Ballard et al., 2024; Benkirane et al., 2025; Fan et al., 2023; Wu et al., 2024; Zhang et al., 2025).

Related Work

The early cancer genomics studies mainly depended on single-omics analysis and traditional statistical results to examine heterogeneity of tumours, subtype analysis, and tumour prognosis. DNA sequencing observations were typically compared with RNA sequencing observation in order to determine somatic mutations and copy number changes, but the latter were studied independently to analyse the differences in gene expression and pathway activity. These methods were useful in giving great biological clues, but they had shortcomings in their skills to reflect the intricate association between genetic modifications and transcription control. Since the development of cancer is regulated by complex multicellular interactions among the molecular processes across the levels of different omics, the study of DNA or RNA data independently tends to provide incomplete or biassed information on tumour behaviour and clinical outcomes (Shi et al., 2024; Wekesa and Kimwele, 2023).

As the data of high-throughput sequencing were growing exponentially, machine learning approaches grew in prominence in cancer genomics to enable better predictive power and automated extract feature extractions. Conventional support vector machine, random forest and shallow neural network are traditional classifiers that have been used in identifying subtypes of cancers and predicting survival with the use of DNA or RNA features. Although these approaches achieved some experiments by being better than purely statistical models, they were often weak when dealing with high dimensionality, noise, and nonlinear relationships of sequencing data. Moreover, most machine learning models need a significant amount of manual feature engineering and do not scale to be viable when using large cohorts of cancer with genomics research (Ballard et al., 2024; Wekesa and Kimwele, 2023).

The most recent developments in the field of deep learning have also changed the sphere of cancer genomics research greatly, especially in the field of the multi-omics integration. The models based on deep neural networks, such as autoencoders, attention mechanisms, and graph neural networks have also shown high potential of jointly modelling DNA and RNA data. It has been demonstrated that integrated multi-omics representations can significantly enhance the subtype classification of cancer, prognosis, and molecular stratification by incorporating complementary information in the various biological layers. CustOmics, MO-GCAN, MoGCN, SADLN, DeepMoIC, and MOFNet are examples of such representative methods (Benkirane et al., 2025; Dou and Mirzaei, 2025; Li et al., 2022; Sun et al., 2023; Wu et al., 2024; Zhang et al., 2025). The review studies also confirm that the multi-omics integration by deep learning will always be superior over the individual-omics (Ballard et al., 2024; Wekesa and Kimwele, 2023; Zhang et al., 2025).

These developments notwithstanding, current multi-omics deep learning methods do have a number of weaknesses. Other models are computationally taxing, or are not scalable to large sequencing data and have complex architectures which make them reproducible and deployable. Other methods are focused on a particular type of cancer or a small dataset, making their use less applicable to heterogeneous populations of patients. Moreover, the types of feature fusion strategies are diverse, and poor integration also can result in the low predictive performance (Fan et al., 2023; Li et al., 2025). To address such issues, a scaled DNN-based multi-omics fusion architecture is suggested in the given work that can be effectively used to combine DNA and RNA sequencing data, provide the opportunity to analyse cancer genomics on a large scale, and provide a more consistent performance increase than similar single-omics approaches.

MATERIALS AND DATA SOURCES

Cancer Genomics Datasets

The current research paper is based on publicly available cancer genomics datasets with the overall emphasis on those received in the Cancer Genome Atlas (TCGA), which comprises detailed molecular characterizations of various cancer subtypes. Sequencing data was demonstrated to capture genomic changes such as somatic mutations, copy numbers variants (CNVs) and other structural changes that may have an important role in tumour initiation and progression. These genomic signatures are consistent cancer genome alterations that give information on oncogenic drivers and tumour suppressor perturbations that are applicable in characterising and prognosticating the disease. Simultaneously, gene expression comparative data of RNA sequencing (RNA-Seq) were used to describe the nature of transcriptional activity and functional downstream outcomes of genomic changes. The RNA-seq profiles can be used to measure gene expression in thousands of genes and they indicate dynamic regulation and cell state in tumours. Using the combination of the DNA and RNA modalities the dataset will record complementary information in both genetic variation and transcriptional regulation, a crucial part in the bright analysis of cancer genomics. The proposed framework is tested in a wide range of cancer types, allowing to consider the model generalizability in quite heterogeneous tumour groups. The sample size within the cancer classes and categories against class labels is represented in (Figure 1) that gives an overview of the data composition and the class distribution.

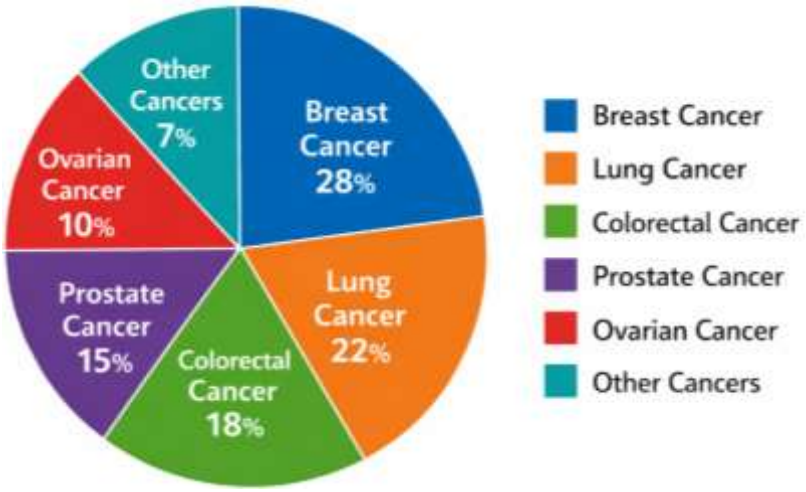


Figure 1. Proportional distribution of cancer types included in the study datasets.

Data Preprocessing

The quality of data and the analytical strength of the models were ensured by applying extensive preprocessing to the DNA and RNA data before the models were trained. Quality control was originally done in order to eliminate the samples and features of low quality and those having high noise or missing values. The expression data obtained by RNA-seq was normalised to ensure that it accounted for both the depth of sequencing and technical variation and the DNA-derived phenomena e.g., mutation counts and CNVs were standardised to allow straightforward integration into the deep learning framework. The steps will enable to minimize systematic bias and enhance the stability of the models. After the normalisation, feature selection and filtering were performed to alleviate the curse of dimensionality as well-known with high-throughput sequencing information. The low-variance and low-frequency features were filtered out to obtain informative genomic and transcriptomics cues of use in cancer prediction tasks. Any missing values that could be a result of platform constraints or sparsity of data were dealt with in an adequate imputation scheme so as to maintain sample integrity without necessarily added pattern. Lastly, a stratified splitting strategy was applied to split the processed data into a training, validation, and testing set to ensure consistency in the distribution of classes. This data splitting gives an impartial evaluation of the model and sound judgement of prediction performance in various experimental conditions.

Proposed Deep Learning Architecture

Overall Framework

The presented framework is to be a scalable deep-learning architecture performing the combined analysis of RNA and DNA sequencing data on cancer genomics. The architecture is multi-branch as illustrated in (Figure 2), the architecture consists of two parallel feature encoding modules, one that takes up DNA-derived features and the other feature is RNA-derived gene expression profiles. A fusion layer is then used to merge these modality-specific representations, to form a single molecular representation on which downstream prediction is performed. This scalable architecture facilitates successful learning of complementing information with heterogeneous sources of omics and allows scalability and flexibility in applications of large-scale cancer genomics tasks. The framework is suitable in high-dimensional and large-volume sequencing datasets because it achieves scalability by using batch-based training, performing efficient matrix operations and simple architecture.

DNA Sequencing Feature Encoder

The encoder DNA sequencing characterises the data of genomic changes, such as somatic mutation indicators and copy number variation characteristics, represented as numbers vectors of each sample. Since DNA specifications are sparsely and high-dimensional, the encoder uses a sequence of fully connected layers to encode raw genomic inputs into the small latent representations. Regularization strategies such as dropout are used to minimize over fitting and improve on generalization, whereas nonlinear activation functions are used to learn intricate relationships between genomic alterations. This is a branch of encoding that dwells on acquiring the patterns, which remain stable through mutation and this pattern displays long-term genomic features regarding the development and progression of cancer.

RNA Sequencing Feature Encoder

The RNA sequencing characteristic encoder is created to process big dimensional gene expression information produced by the RNA-Sequencing experiment. Before inputting the model, normalised levels of expressed genes should be used to depict the transcriptional activity on samples. The encoder uses a series of fully connected layers that more and more implement successive dimensionality reductions to the data whilst allowing biologically meaningful expression patterns through the encoder. The deep neural network architecture makes the learning of features hierarchical and the model can translate the global trends of expression with the fine regulation cues. This branch focuses on dynamic transcriptional data that compliments the relatively stable genomic properties that are quantified by the DNA encoder.

Multi-Omics Fusion Strategy

After the encoding of the DNA genome and RNA genome in a modality-specific manner, the latent representations of the DNA and RNA encoders are fused together in a feature-based fusion mechanism as depicted in (Figure 2). Fusion layer combines the encoded features to create a combined representation which reflects the cross-omics interactions among the genomic changes and the patterns of expression of genes. This combined output is then fed to an ultimate prediction layer which generates task specific output including cancer subtype classification or prognostic prediction. The fusion strategy allows mutual DNA plus RNA learning, which increases predictive ability and strength against uncommercial models. In general, the suggested

architecture is effective and scalable in handling the integration of deep learning-based multi-omics in cancer genomics.

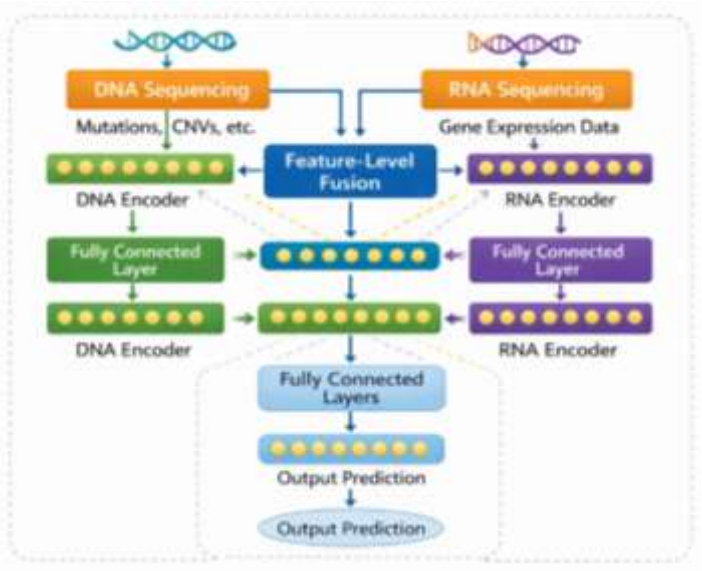


Figure 2. Schematic overview of the proposed deep learning architecture for integrated DNA–RNA analysis.

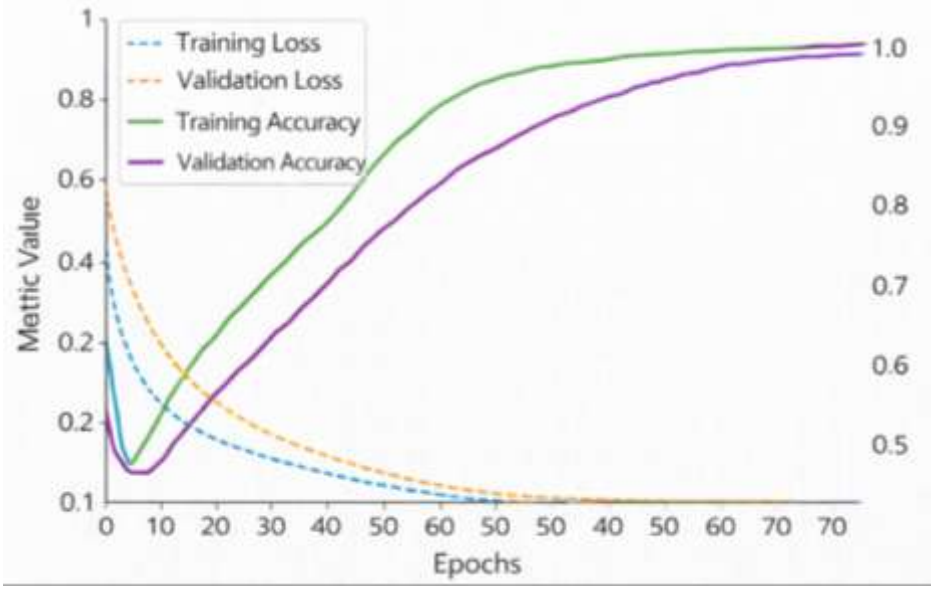
Model Training and Implementation

The proposed multi-omics deep learning model was trained with a supervised learning approach; in which, the goal was to minimise predictive accuracy in cancer-related classification or prediction problems. They used a suitable loss function based on the task formulation where categorical cross-entropy was applied in a multi-class classification task and binary cross-entropy was applied in binary prediction tasks. The adaptive gradient-based optimization algorithm was used to perform model optimization and this allows high-dimensional feature space convergence common to DNA and RNA sequencing data. This optimization scheme provides stable learning of multicomponent nonlinear correlation in combined modalities of omics.

In order to obtain good performance, major training parameters and hyper parameters were distinctly chosen and optimised. These were the learning rate, hidden layers and hidden layer neurons, dropout rate, and batch size. The tuning of the hyper parameters was performed using the validation subset to avoid the over fitting and also determining the best model settings that can provide the right balance between model complexity and generalisation. Early stopping was used on the validation loss in order to prevent unnecessary iterations of training and overtraining. Training convergence behaviour of the model such as loss levelling curve and improvement in performance over the number of epochs is as depicted in (Figure 3).

The experiments have all been done in a standardised computational setting to help with the large-scale genomics data analysis. The deep learning networks were run with popular machine learning libraries, making it possible to make efficient use of matrices and parallel calculations. The system used in the training and evaluation had adequate memory and computation power to handle high-dimensional DNA and RNA sets of features. This implementation environment can guarantee reproducibility and aid viable application of the suggested framework in the real cancer genomics studies scenarios.

A major factor in the design and implementation of the model was scalability. The structure encourages the concept of batch-based training, which makes it possible to process the massive amounts of data with the limitations of the memory in an effective manner. Incremental loading and preprocessing of the data were done to support large sequencing cohorts and therefore, the model could efficiently scale with the size of the sample. The modular architecture also allows its expansion to other omics layers or more massive datasets without redesigning the architecture in a significant way. The scalability properties indicate that the suggested deep learning architecture can be applied in cancer genomics process sourced by big data, where the volume of sequencing data should continuously increase.



Evaluation Metrics

Predictive / Learning Performance Metrics

In order to fully evaluate performance of the suggested deep learning framework, various predictive and learning performance indicators were used. To measure various model performance aspects, especially cancer genomics, whereby, the class imbalance and nonhomogeneous molecular patterns are ordinary, these metrics have been chosen. A threshold dependent plus a threshold independent measure is used in order to provide a reliable assessment of predictive ability in various cancer classes and datasets. The measure of accuracy was employed as a general measure of related predictions which is the proportion of samples that would be assigned accurately by the model. Although accuracy is useful to give a general performance advert, it is also sensitive to class imbalance as present in most cancer genomics datasets. In order to overcome this drawback, other metrics that concentrated on performance of classes were added. These complementary measures allow the more subtle evaluation of model behaviour to be more than just correct.

The reliability and completeness of positive predictions using precision and recall were used to assess the accuracy of the prediction. Precision is used to measure the percentage of the correct positive samples out of all the positive samples predicted and recall is used to measure the ability of the model to identify all the relevant positive samples. The harmonic mean between the concepts of precision and recall was used to weight these two measures and obtain one, strong measure of classification performance and especially when dealing with imbalanced data-sets, as is the case with cancer genomics research. Along with these measurements, the area under the receiver operating characteristic curve (AUC–ROC) was applied in order to test the discriminative capacity of the model with various classification levels. AUC–ROC is a threshold free measure of performance and represents the trade-off between sensitivity and specificity. The metric can be particularly useful to make comparisons between models that are used in clinical and translational research as model decision thresholds can be different. Combined, these evaluation metrics can provide a more detailed framework of the predictive performance of the suggested multi-omics deep learning structure.

Strategy of Comparative Evaluation

In order to measure the value of multi-omics integration, a comparative evaluation plan was put down. The 3 model configurations were tested in the same experimental conditions: the DNA-only model, RNA-only model and integrated DNA-RNA model. The DNA-only model involves the use of genomic characteristics like mutations and variations of the copy number, but the RNA-only takes into account only the information of the gene expression. Such single-omics models are used as control groups to determine the contribution made by each type of data. The integrated DNA RNA model fuses both the information of the genomic and the transcriptomics based on the proposed deep learning fusion strategy. The inference of the use of multi-omics fusion can be quantitatively measured by the direct comparisons of predictive performance of the combined model over that of the individual omics baselines across all measures of evaluation. This comparative study indicates the value-added advantage compared to the joint representation learning technique and how supplementary molecular information can result in a higher predictive accuracy, robustness, and generalisation in cancer genomics analysis.

Results and Discussion

Predictive Performance of the Proposed Model

The choice of predictive performance was analysed with the metrics introduced in Section 6 to assess the proposed integrated deep neural network. Numerical findings indicate that the DNARNA integrated model is better than the single-omics baselines in all the evaluation metrics. According to the results in (Table 1) the integrated model has better accuracy, F1-score and AUC than the DNA-only and RNA only configurations. These advances imply that the integration of genomic and transcriptional-activity accommodates complementary information of the model that cannot be obtained via the independent layers of omics. The tendency of the comparative performances is graphically represented in (Figure 4) which shows the evident superiority of multi-omics fusion compared to individual models.

Impact of Multi-Omics Integration

In order to examine the contribution of respective modalities of data further, an ablation-based comparison was carried out between the DNA-only, RNA-only, and the integrated setup. Findings reveal that RNA-only models are highly favoured over DNA-only models in terms of able to capture expression-level predictions, however, the joint DNA-RNA model is the most robust option, as discussed in (Table 1). This attests the fact that DNA characteristics that include mutations and copy number variations consent to complementary and stable genomic context that amplifies transcriptomics signal interpretation. It has its biological explanation because joint modelling indicates the inherent pathology of cancer development whereby genomic changes determine transcriptional dysregulations thus enhancing predictive accuracy when the two modalities are interrelated.

Comparative Analysis with the Existing Methods

It was also compared to the traditional machine learning models such as support vector machines and random forest classifiers trained under the same conditions of the experiment. As (Figure 4) illustrates, the DNN-based multi-omics model has the highest performance in all the metrics, especially with regards to the F1-score and AUC. Even though traditional models are effective on lower-dimensional data, they have a limitation on their ability to model nonlinear interactions and high-dimensional dependencies that characterise multi-omics sequencing data. Even though deep learning models can be computationally complicated, their capacity to independently learn hierarchical representations is an enormous benefit compared to traditional methods, especially in built-in cancer genomics analysis.

Strength, Generalisation and Scaling Assumptions

Cross-validation and testing of robustness and generalisation across various types of cancer and partitions of their datasets were conducted. The model consisted of an integrated scheme demonstrates the presence of low variance between folds, which means the high ability of generalisation. The consistency in the performance across the varied cancer data also indicate the versatility of the suggested architecture to the diverse tumour profiles. Moreover, scalability experiments prove that efficiency of training does not decrease with the size of the dataset and is sustained by both processing batch wise and structuring a network to be modular. The summary of these results in (Table 1) and the visualisation of the results in (Figure 4) suggest that the suggested framework is not just predictive, but also practically viable to large-scale studies of cancer genomics, where data volume and complexity are only increasing.

Table 1. Performance comparison of DNA-only, RNA-only, and integrated DNA–RNA models

Model	Accuracy	Precision	Recall	F1-score	AUC
DNA-only model	0.78	0.76	0.74	0.75	0.82
RNA-only model	0.84	0.83	0.81	0.82	0.88
Integrated DNA–RNA model	0.90	0.89	0.88	0.88	0.94

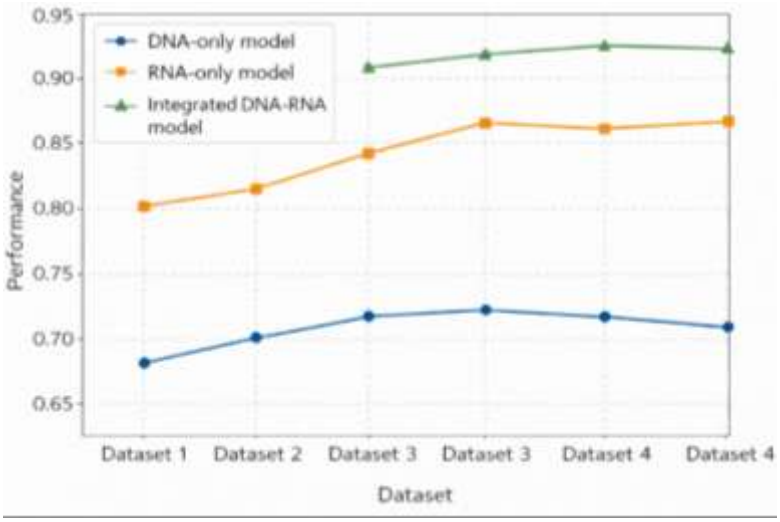


Figure 4. Comparative performance of single-omics and multi-omics models across datasets.

Limitations and Future Directions

Although the suggested deep learning scheme is promising in performance, it has a number of limitations associated with this paper. Firstly, the experimental analysis is based mainly on publicly available cancer genomics data, including those of big consortium projects. Although these datasets are exhaustive and have wide application, their ability to represent the heterogeneity of clinics, population, or sequencing platform used in practise might be limited. Moreover, inconsistent quality of data, batch bias and inconsistent annotations across the public repositories may condition model performance and generalizability. The other significant weakness is associated with the interpretability of the deep learning models. Even though the predictive capability of the proposed architecture is very high, internal representations acquired by deep neural networks are not always physically interpretable. This is a limitation to direct translation on model predictions into mechanistic information or clinical decision-making. The interpretation of the underlying genomic change or transcriptional signature prediction of predictions is a vital aspect in cancer genomics to create trust and promote hypothesis samples, and thus better interpretability solutions are required.

The future research will be devoted to the extension of the offered framework to include other layers of omics, including DNA methylation, proteomics, and epigenomics profiles. Combining such complementary data modalities can further improve predictive performance and give the tumour biology in a more detailed picture. More detailed characterization of cancer progression and treatment response would also be supported with multi-omics expansion since it would also allow exploring regulatory interactions between molecular layers. Lastly, the implementation of explainable artificial intelligence (XAI) methods is also a valuable research direction in the future. Attention mechanisms, feature attribution techniques or post-hoc explanation models would be useful in determining the important molecular characteristics behind predictions of the model. These strategies would enhance the biological explanability of the outputs in deep learning and help them to accept the results in translational and clinical research settings. Combined, these extensions in the future are planned to enhance the strength, openness, and effective influence of deep learning-based multi-omics analysis in cancer genomics.

CONCLUSION

This paper provided a scalable deep learning architecture to the combined analysis of DNA and RNA sequencing data in cancer genomics to resolve major challenges related to the high-dimensional and non-homogeneous molecular data. With the proposed architecture, by using a DNN-based multi-omics fusion strategy, the architecture effectively integrates complementary genomic and transcriptomics information by which predictive performance consistently improves in comparison to single-omics schemes. The experimental evidence shows that joint representation learning results in the improved accuracy, robustness, and generalizability in various cancer datasets, which is a good indicator of how deep learning can be utilised in integrated molecular data analysis. All in all, the article introduces a practically applicable and generalizable computational framework elucidating more detailed characterizations of cancer biology. In the future, it is anticipated that the further combination of other layers of omics and explainable AI methods will more

optimally improve deep learning-based cancer studies and enable translating multi-omics findings into precision oncology practises**.**

REFERENCES

[1] AlAfeshat, N. (2024). *Data Point Attention and Multi-Omics Fusion for Prostate Cancer" Diagnosis Age" Prediction* (Master's thesis, Princess Sumaya University for Technology (Jordan)).

[2] Ballard, J. L., Wang, Z., Li, W., Shen, L., & Long, Q. (2024). Deep learning-based approaches for multi-omics data integration and analysis. *BioData Mining*, *17*(1), 38.

[3] Benkirane, H., Vakalopoulou, M., Planchard, D., Adam, J., Olaussen, K., Michiels, S., & Cournède, P. H. (2025). Multimodal CustOmics: A unified and interpretable multi-task deep learning framework for multimodal integrative data analysis in oncology. *PLOS Computational Biology*, *21*(6), e1013012.

[4] Dahlan Abdullah, & Robbi Rahim. (2025). Multimodal Integration of EEG and fMRI Data for Understanding Visual Attention Networks. *Advances in Cognitive and Neural Studies*, *2*(1), 27-33. https://aasrresearch.com/index.php/ACNS/article/view/459

[5] Dou, Y., & Mirzaei, G. (2025). MO-GCAN: multi-omics integration based on graph convolutional and attention networks. *Bioinformatics*, *41*(8), btaf405.

[6] Fan, Z., Jiang, Z., Liang, H., & Han, C. (2023). Pancancer survival prediction using a deep learning architecture with multimodal representation and integration. *Bioinformatics Advances*, *3*(1), vbad006.

[7] Huang, Z., Li, Y., Bucci, V., & Haran, J. P. (2025). Decoding cancer prognosis with deep learning: the ASD-cancer framework for tumor microenvironment analysis. *MSystems*, *10*(5), e01455-24.

[8] Ji, Y., Dutta, P., & Davuluri, R. (2023). Deep multi-omics integration by learning correlation-maximizing representation identifies prognostically stratified cancer subtypes. *Bioinformatics Advances*, *3*(1), vbad075.

[9] Li, B., & Nabavi, S. (2024). A multimodal graph neural network framework for cancer molecular subtype classification. *BMC bioinformatics*, *25*(1), 27.

[10] Li, M., Jin, M., Lou, M., Deng, S., Wang, L., & Rao, H. (2025). Multiview-cooperated graph neural network enables novel multi-omics cancer subtype classification. *Computational Biology and Chemistry*, 108665.

[11] Li, X., Ma, J., Leng, L., Han, M., Li, M., He, F., & Zhu, Y. (2022). MoGCN: a multi-omics integration method based on graph convolutional network for cancer subtype analysis. *Frontiers in Genetics*, *13*, 806842.

[12] Q. Hugh, & Freddy Soria. (2025). Advances in Cognitive and Neural Studies: Bridging Cognitive Science, Neuroscience, and Brain-Inspired Technology. *Advances in Cognitive and Neural Studies*, *1*(3), 29-36. https://aasrresearch.com/index.php/ACNS/article/view/457

[13] Rajan.C, & S. Praveen Kumar. (2025). Meta-Learning for Personalized Brain-Computer Interfaces: A Computational Neuroscience Approach. *Advances in Cognitive and Neural Studies*, *2*(2), 29-36. https://aasrresearch.com/index.php/ACNS/article/view/462

[14] Shi, T., Ye, X., Huang, D., & Sakurai, T. (2024). Cancer subtype identification by multi-omics clustering based on interpretable feature and latent subspace learning. *Methods*, *231*, 144-153.

[15] Sun, Q., Cheng, L., Meng, A., Ge, S., Chen, J., Zhang, L., & Gong, P. (2023). SADLN: Self-attention based deep learning network of integrating multi-omics data for cancer subtype recognition. *Frontiers in Genetics*, *13*, 1032768.

[16] Wang, C., & O'Connell, M. J. (2025). Autoencoders with shared and specific embeddings for multi-omics data integration. *BMC bioinformatics*, *26*(1), 214.

[17] Wekesa, J. S., & Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. *Frontiers in Genetics*, *14*, 1199087.

[18] Wu, J., Chen, Z., Xiao, S., Liu, G., Wu, W., & Wang, S. (2024). DeepMoIC: multi-omics data integration via deep graph convolutional networks for cancer subtype classification. *BMC genomics*, *25*(1), 1209.

[19] Zhang, G., Zhang, C., Li, P., Sun, D., Yang, Z., & Liu, Z. P. (2025). MOFNet: a deep learning framework for multi-omics data fusion in cancer subtype classification. *Molecular Omics*, *21*(6), 690-705.

[20] Zhang, J., Che, Y., Liu, R., Wang, Z., & Liu, W. (2025). Deep learning–driven multi-omics analysis: Enhancing cancer diagnostics and therapeutics. Briefings in bioinformatics, 26(4), bbaf440.

[21] S. Sindhu. (2024). AI-Driven Precision Livestock Farming: A Smart Framework for Sustainable Animal Health, Welfare, and Environmental Efficiency. National Journal of Animal Health and Sustainable Livestock , 2(1), 26-32.

[22] H. T. Rai. (2026). Smart Vision-Based Sorting and Grading of Fruits Using Deep Learning and Real-Time Conveyor Analytics. Journal of Environmental Sustainability, Climate Resilience, and Agro-Ecosystems, 1–7.

[23] Shaik Sadulla, & K P Uvarajan. (2025). Environmental DNA and AI-Driven Drug Discovery: Unlocking Nature's Genomic Reservoirs for Novel Therapeutics. Frontiers in Life Sciences Research, 8–13.