بسم الله الرحمن الرحيم

Omdurman Islamic University Journal of Applied Sciences

(OIUJ)-(ASC)

مجلة جامعة أم درمان الاسلامية للعلوم التطبيقية

https://journal.oiu.edu.sd/index.php/oiuj

https://doi.org/10.52981/oiu v22i1.3490 j

2026 Jan; 22(1):144-170

ISSN: 5361-1858

رجب 1447هـ

Attention-Enhanced Kolmogorov–Arnold Network for Interpretable Subtype Classification and Biomarker Discovery in Lower-Grade Glioma

Eman Mohammed Hamid[1], Murtada K. Elbashir[2*], Mohamed Elhafiz Musa[3]

[1]Department of Computer Science, Faculty of Mathematical and Computer Science, University of Gezira, Wad Madani, Sudan

[2]Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia

[3]Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia

Corresponding Email address: : eman_24@hotmail.com

To cite this article:

, Eman Mohammed Hamid , Murtada K. Elbashir, Mohamed Elhafiz Musa , Attention-Enhanced Kolmogorov–Arnold Network for Interpretable Subtype Classification and Biomarker Discovery in Lower-Grade Glioma, Omdurman Islamic University Journal, Sudan

ISSN: 5361-1858

https://doi.org/10.52981/oiuj. v22i1.3490

Abstract

Precision molecular classification and valid biomedical marker identification are necessary for the optimization of personalized therapy in lower-grade glioma (LGG). In this study, we present a transparent deep learning model built on the Kolmogorov–Arnold Network (KAN) for effective LGG subtype classification and biomarker discovery. The method integrates multiple-omics (mRNA expression, DNA methylation and miRNA profiles) data from the TCGA-LGG sample set for comprehensive modelling of tumour heterogeneity. Before training the model, we performed variance-based feature screening, retaining molecular features and excluding class labels.

The remaining pre-processing, including standard normalisation, SMOTE and LASSO feature selection, was performed only during model training to allow unbiased learning. We introduce a feature-level attention mechanism within an autoencoder-type network architecture to learn low-

dimensional, salient representations. The developed classifier shows excellent performance, as assessed by stratified cross-validation, reaching a classification accuracy of 99.02% across three clinically meaningful subtypes (the IDHmut-codel, the IDHmut-non-codel and the IDHwt), with sensitivity, specificity and precision all averaging above 0.98, suggesting extremely consistent and balanced class-wise performance.

Furthermore, the attention mechanism provided intrinsic interpretability by highlighting important subtype-related biomarkers across multiple omics layers, such as CDH4 (mRNA), MIR196A2 (miRNA), and HumanMethylation450 (DNA methylation), yielding bioinformatically significant results in LGG molecular differentiation. Such a framework therefore offers a scalable and transferable basis for multi-omics integration in precision oncology, with strong potential to accommodate other tumour types.

Keywords:
Lower-grade glioma; Multi-omics integration; Kolmogorov–Arnold Network; Attention mechanism; Molecular subtype classification; Biomarker discovery; Deep learning

1. Introduction

Lower-grade glioma (LGG) represents a heterogeneous group of primary brain tumours characterised by complex molecular abnormalities, variable therapeutic responses, and markedly different clinical outcomes [1]. Accurate molecular subtype classification is therefore essential for reliable prognosis assessment and the development of personalised treatment strategies [2]. However, the pronounced molecular heterogeneity of LGG, spanning transcriptional activity, epigenetic regulation, and genetic alterations, poses significant challenges for conventional computational approaches. Recent advances in high-throughput sequencing technologies and multi-omic profiling have enabled more comprehensive characterisation of tumor biology by integrating complementary molecular layers such as gene expression and DNA methylation[2]. In parallel, deep learning methods have demonstrated strong potential for modelling the high-dimensional and nonlinear relationships inherent in multi-omics data, leading to improved performance in cancer subtype and biomarker discovery tasks [3]. Despite these advances, many existing deep learning frameworks for LGG classification rely on graph-based representations constructed from predefined biological networks, most notably protein–protein interaction (PPI)

databases [4]. While biologically informative, such static networks are often incomplete, context-independent, and may fail to capture subtype-specific or patient-specific molecular interactions, thereby limiting model adaptability and generalizability. To overcome these limitations, recent studies have proposed data-driven graph construction strategies based on correlation or similarity measures derived directly from patient molecular profiles, enabling the modelling of cohort-specific interaction patterns without reliance on external biological knowledge [2]. Although these approaches improve flexibility, they predominantly employ graph neural networks, particularly graph attention networks (GATs), as their core learning paradigm. Despite their effectiveness, GAT-based models may suffer from limited interpretability and suboptimal modelling of nonlinear dependencies across heterogeneous omics layers, especially in high-dimensional and relatively small biomedical datasets [5].

Consequently, there remains a need for an interpretable and scalable multi-omics learning framework that can capture complex nonlinear relationships across heterogeneous molecular layers without dependence on predefined or graph-based interaction structures, while simultaneously enabling transparent biomarker discovery. To address this gap, this study proposes a novel LGG classification framework based on the KAN, leveraging its strong representational capacity for modellng nonlinear functions and its inherent interpretability. By integrating multi-mics features within a KAN-based architecture, the proposed approach enables accurate LGG molecular subtyping and robust biomarker identification, offering a compelling alternative to existingg graph-based deep learning models.

2. Related Work

Deep learning has been widely adopted for cancer subtype classification and biomarker discovery due to its ability to model complex, high-dimensional biological data [7 ,6]. Early studies primarily focused on single-omics data, particularly gene expression, using architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs)[8, 9] . While these approaches demonstrate reasonable predictive performance, their limited molecular scope restricts their ability to capture tumour heterogeneity[10].

Subsequent studies shifted toward multi-omics integration to provide a more comprehensive molecular characterisation of tumours. Song et al. (2021)[11] introduced a computational framework for cancer subtype identification by jointly modelling driver mutation profiles and gene expression data, highlighting the benefit of integrating heterogeneous molecular features to capture tumour heterogeneity. Singh et al. (2019)[12] extended the DIABLO framework for supervised multi-omics integration, improving subtype discrimination and biomarker interpretability, although its linear latent-variable formulation limited modelling of complex nonlinear interactions. Sun et al. (2023)[13] introduced SADLN, an attention-based framework that adaptively weights different omics modalities and achieves improved predictive performance compared to early fusion strategies.

More recent approaches have leveraged graph-based deep learning models to explicitly capture molecular relationships across omics layers. Wang et al. (2021)[14] introduced GraphOmics, an interactive platform for integrating transcriptomics, proteomics, and metabolomics data by mapping biological entities to shared biochemical reactions and pathways. More recently, Wu et al. (2024) [15] introduced DeepMoIC, a supervised framework designed to capture cross-omics dependencies, while Munquad et al. (2023)[16] proposed DeepAutoGlioma for LGG subtype classification using gene expression and DNA methylation data. Ryan et al. (2024)[17] subsequently introduced MOGDx, demonstrating the feasibility of heterogeneous multi-omics integration, although performance was affected by high dimensionality and limited cohort size. Elbashir et al. (2025)[18] introduced a PPI-informed graph attention–based multi-omics framework for LGG subtype classification, achieving high accuracy and identifying biologically relevant subtype-specific biomarkers.

2.1 Research Gap and Motivation

Despite significant progress in deep learning–based cancer subtype classification, most existing frameworks rely on either vector-based multi-omics integration or graph-based representations that depend on predefined or correlation-derived interaction structures. These approaches often struggle to balance predictive performance with feature-level interpretability, particularly under conditions of high dimensionality and class imbalance.

Recently, KANs have emerged as an alternative deep learning paradigm capable of representing complex multivariate nonlinear functions through compositions of learnable univariate functions.

Owing to their spline-parameterised formulation, KANs offer enhanced functional interpretability and strong representational capacity compared to conventional neural networks. While KAN-ba sed models have demonstrated promising results in biological applications, their application to LGG molecular subtype classification in a fully data-driven multi-omics setting remains largely unexplored.

3. Methodology

This article proposes an interpretable deep learning model framework utilising the KAN for the subtyping of LGG patients' molecules (IDHmut-codel subtype (n = 169), IDHmut-non-codel subtype (n = 245), and IDH wildtype (IDHwt subtype; n = 95)) based on the integration of multi-omics information to achieve the identification of biologically significant markers within the TCGA-LGG dataset.

As opposed to graph-based deep learning models which require the construction of interaction networks or correlation graphs as a prerequisite, the proposed method in this study directly works upon the tabular multi-omics feature space and hence is free from the burden of network construction and allows the network to explicitly learn subtype discriminative patterns solely based upon data-driven feature interactions and remains transparent at the feature level.

Each omics layer was preprocessed and filtered using independent statistical criteria to ensure robustness and biological relevance: normalization, differential analysis, and refinement of features based on their variance to discard those features with low informativeness that usually appear noisy. These omics-specific features were then integrated into one unified multi-omics matrix to serve as the backbone for downstream machine learning.

In each cross-validation fold, LASSO-based embedded feature selection was applied to further reduce the dimensionality and retain the most informative and non-redundant features in a leakage-free manner. Selected features were standardized and then used as input to the KAN model. To alleviate class imbalance across LGG subtypes, the SMOTE oversampling technique was exclusively performed on the training data, ensuring that learning is well-balanced without contaminating any of the validation samples[19].

The proposed KAN model combines a learnable sigmoid function-based feature attention mechanism, which helps the model adaptively weigh the features while training the model. Using attention-based weighting, it helps to enhance predictive performance as well as directly interpret the weights of subtype-specific biomarkers. A reliable training process with early stopping and adaptive adjustment of the learning rate was used for training the model. A complete process pipeline of the proposed methodology is shown in Figure1 below.



Figure 1. Overview of the proposed Kan-based workflow for low-grade glioma (LGG) classification and biomarker discovery.

3.1 Omics Data Sources

Profiling data of multi-omics in patients with LGG were obtained from the Cancer Genome Atlas (TCGA-LGG) database through the TCGAbiolinks Bioconductor package in R software[20]. The multi-omics profiling information used in this study had three dimensions of measurements, which were messenger RNA (mRNA) expression, microRNA (miRNA) expression, and DNA methylation measured in tumor samples. These samples were pre-processed by the GDCquery and GDCprepare functions in R software.

3.1.1 mRNA Expression Processing

The raw count RNA-seq data were normalized by the use of DESeq2 (v1.44.0) version 1.44.0[21]. Low expression genes were filtered out based on count filtering thresholds in order for the data to be free from noise. This resulted in a new set of informative genes for analysis based on differential expression.

3.1.2 miRNA Expression Processing

Expression data of the miRNAs was analyzed using the limma approach (version 3.60.6) including voom transformation. [22, 23]. This reduced the data to only those miRNAs in which the counts per million (CPM) value is above 1. This brought down the numbers to 771 of the original 1,881 to only the top differentially expressed in which 588 were found to be significantly differentially expressed (adjusted p-value of <0.05).

3.1.3 DNA Methylation Processing

For DNA methylation data, this entailed the removal of probes associated with sex chromosomes, cross-reactive probes, and probes containing missing values. After these quality-control steps, 298,588 CpG sites remained. Differential methylation analysis by limma identified 2,838 CpG loci showing significant methylation differences [adjusted $p < 0.05$].

3.1.4 Omics-Specific Filtering

The series of filtering and preprocessing steps performed on each omics dataset independently prior to integration is encapsulated in Table1. These processes significantly decreased dimensionality while retaining biologically important features.

3.1.5 Variance-Based Refinement and Data Integration

After the omics-specific preprocessing, a further variance-based unsupervised filtering step was introduced to further remove uninformative features and reduce redundancy. Since this type of filtering is based only on the variability of the features and does not require knowledge of class labels, it does not introduce information related to the subtype. The number of features remaining at each level of the refinement process is expressed in Table2.

The resulting mRNA, miRNA, and DNA methylation features were then integrated at the feature level to build the multi-omics matrix. After variance filtration, the filtered dataset was composed of 7,581 high-variance features in the 509 LGG samples, which was the input to the machine learning process.

3.1.6 Considerations on Data Leakage

Although variance filtering is label agnostic, there still appears to be a trivial dependency between the training and validation datasets from a strict machine learning point of view if it is carried out before splitting the data. In order to be maximally transparent about our approach, we choose to point this out explicitly. More importantly, both feature scaling and LASSO-based feature selection, as well as oversampling by SMOTE and training models, were each carried out independently in each of the training folds.

Table 1. Feature Filtering per Omics Data Type (Pre-integration)

Step	mRNA	miRNA	DNA methylation
Original features	60,660	1,881	330,974
Removal of sex chromosomes	57,670	–	324,232
Differential analysis	19,193	771	256,584
Limma-selected features	–	588	238,245
Global LASSO	3,395	161	90

Table 2. Final Feature Refinement Used for Model Training

Step	mRNA	miRNA	DNA methylation
Filtered features (input)	25,315	588	238,245
Low-variance filtering	10,500	310	7,800
Final selected features	6,328	180	1,073

3.2 Model Architecture

The proposed model follows the KAN paradigm and operates directly on a tabular multi-omics feature matrix without relying on predefined interaction graphs. Given an input feature vector

$$\mathrm{x} \in \mathbb{R}^{d}, \qquad (1)$$

where d denotes the number of selected features per fold, the network transforms the input through four sequential components: (1) feature-level attention, (2) encoding, (3) reconstruction, and (4) classification.

First, **(1) a feature-level attention module** is applied to explicitly model the contribution of each feature. The attention mechanism is implemented as two fully connected layers with dimensions $d \to 128 \to d$, and is defined as

$$\mathrm{a} = \sigma(\mathrm{W}_2\,\mathrm{ReLU}(\mathrm{W}_1\mathrm{x} + \mathrm{b}_1) + \mathrm{b}_2), \qquad (2)$$

where $\mathrm{a} \in [0,1]^{d}$ represents feature-wise attention weights and $\sigma(\cdot)$ denotes the sigmoid activation. The attended input is then obtained by element-wise scaling:

$$\mathrm{x}' = \mathrm{x} \odot \mathrm{a}. \qquad (3)$$

Second, **(2) the encoder network** processes the zattended features $\mathbf{x}'$ through three fully connected layers with decreasing dimensionality $d \to 128 \to 64 \to 32$. Each layer employs a ReLU activation and dropout regularization with a rate of 0.5. The encoder produces a compact latent representation

$$z = f_{enc}(x'), \qquad z \in \mathbb{R}^{32}. \tag{4}$$

Third, **(3) a symmetric decoder network** reconstructs the original input from the latent space using fully connected layers with dimensions $32 \to 64 \to 128 \to d$:

$$\hat{x} = f_{dec}(z), \tag{5}$$

where ReLU activations are applied in the hidden layers and a sigmoid activation is used at the output layer to constrain reconstructed feature values.

Finally, **(4) a linear classification head** maps the latent representation **z** to the output space through

$$\hat{y} = \text{softmax}(W_c z + b_c), \tag{6}$$

producing class probability estimates for the three LGG molecular subtypes: IDHmut-codel, IDHmut-non-codel, and IDHwt.

This parameterized architecture jointly optimizes attention-guided feature weighting, latent encoding, and reconstruction, thereby ensuring stable learning in high-dimensional settings. Importantly, the learned attention weights a provide a direct quantitative measure of feature importance, facilitating interpretable subtype classification and biologically meaningful biomarker discovery.

3.3 Model Training

After data preprocessing and feature selection, the multi-omics data was utilized as training data for the proposed KAN model. The training of models was done with a stratified 10-fold cross-validation setting that allowed effective estimation of model performance and maintenance of equal class probability as per original class distribution in all folds. The entire training and modeling activity was limited within the training set of each fold, with no leakage of validation information that could compromise model performance estimation on respective validation set instances.

3.3.1 Cross-Validation Strategy

The training and testing of models, as well as evaluation, were done using stratified 10-fold cross-validation, with the objective of ensuring that there is representation of the class distribution of

LGG molecular types. In this case, after creating folds, another split would occur to generate independent testing sets.

In each of the training folds, the process of applying feature standardization began with the application of z-score normalization. This involved the standardization of the features by shifting them along the z-score scale with a mean of 0 and a variance of 1, without utilizing the normalization on any data other than the training dataset statistics.

Accordingly, L1-regularized multinomial logistic regression, also known as LASSO [24, 25], was run on the standardized training data to perform an embedded feature selection and dimensionality reduction, in order to retain the most informative and non-redundant features. Finally, to handle the problem of class imbalance, SMOTE-based oversampling was only applied on the training data following the feature selection.

The resulting standardized, balanced, and reduced feature set was used to train the proposed KAN model. Validation data never participated in normalization, feature selection, or oversampling and were used only for performance evaluation. This fold-wise design ensures a leakage-free evaluation and provides a reliable estimate of model generalization across LGG subtypes.

3.3.2 Loss Function and Optimisation

The KAN model was trained with a multi-objective loss function that allowed the classification accuracy and representation to be optimised jointly. This was achieved through the overall training goal being defined as a weighted sum of a categorical cross-entropy loss function for subtype classification tasks and a reconstruction loss function obtained from the autoencoder component, as follows:

$$\mathcal{L} = \alpha\, \mathcal{L}_{cls} + (1 - \alpha)\, \mathcal{L}_{rec}, \qquad (7)$$

where $\mathcal{L}_{cls}$ denotes the cross-entropy loss computed on the predicted subtype probabilities, and $\mathcal{L}_{rec}$ represents the mean squared error (MSE) between the reconstructed and original input features. The weighting parameter α was set to 0.4 to balance discriminative learning with faithful feature reconstruction.

To account for residual class imbalance after oversampling, class-weighted cross-entropy was employed, with class weights computed from the training data of each fold. Model parameters were optimized using the Adam optimizer, with an initial learning rate of 5×10^{-4} and L2 regularization implemented via weight decay (5×10^{-4}) to reduce overfitting.

This encourages learning of compact and informative latent variables, which are simultaneously predictive of LGG subtypes as well as informative at the feature level

3.3.4 Stable Training and Convergence Control

Throughout the training process, multiple strategies for convergence control were evaluated in order to improve the model's overall performance and reduce the likelihood that results could potentially be influenced by overfitting. The first strategy applied early stopping strategy based on model validation loss; as part of this process, the training process would automatically stop after 20 epochs, where no improvements had been realised in the model's performance after this timeframe. It ought to be noted that stopping the training process once the model reaches a plateau, minimizes the unnecessary time spent training and also eliminates the potential for the implementation of overfitting within the context of high-dimensional datasets.

The second strategy that was implemented to improve the model's overall convergence process was the application of gradient clipping as part of the back propagation process of the model's training cycle to enhance the models numeric stability. By using gradient clipping, the maximum norm value of the gradient is restricted to 1.0. This prevents the occurrence of unbounded or exploding gradients during the learning process and permits smoother convergence of the model.

The final plan of action to aid in improving the performance characteristics of the model was the inclusion of adaptive learning rate behaviour as part of the INL model's capabilities; via the implementation of this feature, if the model's validation loss did not improve over the course of three epochs, then the learning rate would be halved in size until it reaches its end point, which is 1×10^{-6}.

Finally, within this work, a new methodology was introduced for performing more equitable and interpretable comparisons of both the training and validation curves for loss. Rather than summing the training loss incurred for each measure of the mini-batch, the training loss for a given epoch will be computed from the complete training set, while using the model in inference mode. The

pre-emptive use of this evaluation modality means that the random or stochastic nature of dropout and mini-batches will no longer introduce noise into aggregate training loss values. The application of this technique will yield loss plots that have a much more homogeneous appearance and will closely resemble the validation loss curves. Additionally, note that the evaluation phases are not used in updates to the parameter set during the training process. Collectively, these mechanisms contribute to robust convergence, improved generalisation, and reliable interpretation of learning dynamics across cross-validation folds.

3.4 Evaluation Metrics

Cross-validated accuracy, precision, recall, and F1-score were used to evaluate the proposed KAN-based framework's overall and class-wise classification performance across LGG molecular subtypes. Misclassification patterns between classes were examined using confusion matrices. Additionally, to offer a threshold-independent assessment of discriminative performance, one-vs-rest ROC curves and corresponding AUC values, including micro-averaged AUC, were calculated. To determine the model's capacity for generalization, all metrics were computed on validation folds and combined.

3.5 Implementation Details

The proposed model was trained and evaluated using the Kaggle cloud computing environment, which offers approximately 13 GB of RAM and optional access to an NVIDIA Tesla T4 GPU. Prior to model training, the multi-omics datasets were downloaded and preprocessed on a local workstation equipped with an Intel® Core™ i7-9700 CPU (3.00 GHz), 32 GB RAM, and an NVIDIA GeForce GTX 1080 GPU (8 GB).

Model evaluation was carried out using stratified 10-fold cross-validation, with each fold requiring approximately 4–6 minutes of training time. The total runtime for completing all folds was therefore around 45 minutes. This computational setup enabled efficient experimentation while ensuring reproducibility.

4. Results

This section presents the experimental evaluation of the proposed KAN-based framework on the TCGA-LGG dataset. Model performance was assessed using stratified 10-fold cross-validation to

ensure a reliable and unbiased evaluation. The proposed model achieved an overall classification accuracy of 99.02%, with a macro-averaged precision, recall, and F1-score of 0.99 across the three LGG molecular subtypes.

In addition to quantitative classification performance, the results include comparative analyses, visual performance assessments, and attention-based interpretability findings. These results confirmed the ability of the proposed framework to accurately distinguish LGG molecular subtypes while simultaneously identifying biologically meaningful subtype-associated biomarkers.

4.1 Classification Performance

The proposed KAN-based framework was evaluated on the TCGA-LGG cohort comprising 509 samples using stratified 10-fold cross-validation. The model demonstrated strong generalisation capability and highly consistent performance across all folds, achieving a mean classification accuracy of 99.02% ± 1.32%.

When aggregating predictions from all validation folds, the model attained an overall accuracy of 99.02%. The corresponding confusion matrix (Figure 2) indicates near-perfect discrimination among the three LGG molecular subtypes. Specifically, all IDHmut-codel samples were correctly classified, while only four IDHmut-non-codel samples and one IDHwt sample were misclassified, resulting in minimal inter-class confusion. These results confirm the effectiveness of the proposed framework in accurately distinguishing LGG molecular subtypes.

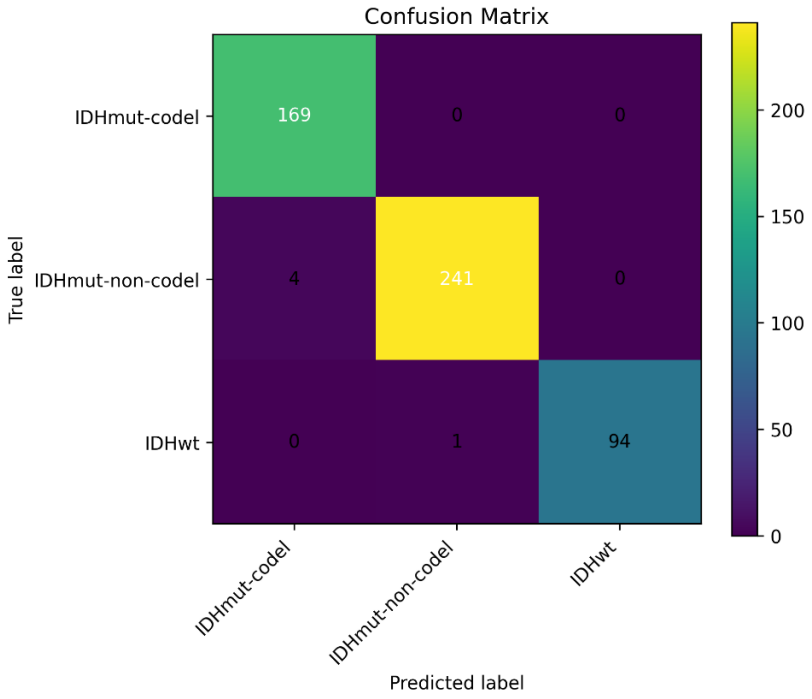


Figure2. Confusion matrix showing the aggregated cross-validated classification results of the proposed KAN-based framework across the three LGG molecular subtypes (IDHmut-codel, IDHmut-non-codel, and IDHwt).

The robustness of the proposed KAN-based framework was further evaluated through class-wise performance analysis. As reported in Table3, the model achieved consistently high precision, recall, and F1-scores across all LGG molecular subtypes. Specifically, the IDHmut-codel subtype attained a precision of 0.98, recall of 1.00, and F1-score of 0.99, while the IDHmut-non-codel subtype achieved a precision of 1.00, recall of 0.98, and F1-score of 0.99. Similarly, the IDHwt subtype demonstrated a precision of 1.00, a recall of 0.99, and an F1-score of 0.99.

Overall, the model achieved macro-averaged and weighted precision, recall, and F1-scores of 0.99, indicating balanced and stable predictive performance across all classes. These results confirm the ability of the proposed KAN-based model to reliably capture subtype-specific molecular patterns from integrated multi-omics data.

Table 3. Cross-validated classification performance of the proposed KAN-based model

Class	**Precision**	**Recall**	**F1-score**
IDHmut-codel	0.98	1.00	0.99
IDHmut-non-codel	1.00	0.98	0.99
IDHwt	1.00	0.99	0.99
Macro Average	0.99	0.99	0.99
Weighted Average	0.99	0.99	0.99

4.2 ROC Curve and AUC Analysis

Receiver operating characteristic (ROC) curves for the three LGG molecular subtypes were created using a one-vs-rest (OvR) technique to assess the discriminative power of the suggested KAN-based framework. With ROC curves that are very close to the upper-left corner of the ROC space, Figure 3 illustrates the model's exceptional discrimination across all classes.

Strong subtype separability was demonstrated by the continuously high class-specific AUC values, which were 0.9989 for IDHmut-codel, 0.9965 for IDHmut-non-codel, and 0.9926 for IDHwt. Furthermore, the model's overall discriminative performance across all samples was reflected in the micro-averaged ROC curve, which produced an AUC of 0.9974.

These findings show that the suggested KAN-based strategy preserves a strong sensitivity–specificity trade-off throughout a broad range of choice thresholds. The model's outstanding generalisation ability and efficacy for multi-class LGG subtype classification using integrated multi-omics data are further supported by the consistently high AUC values across all subtypes.

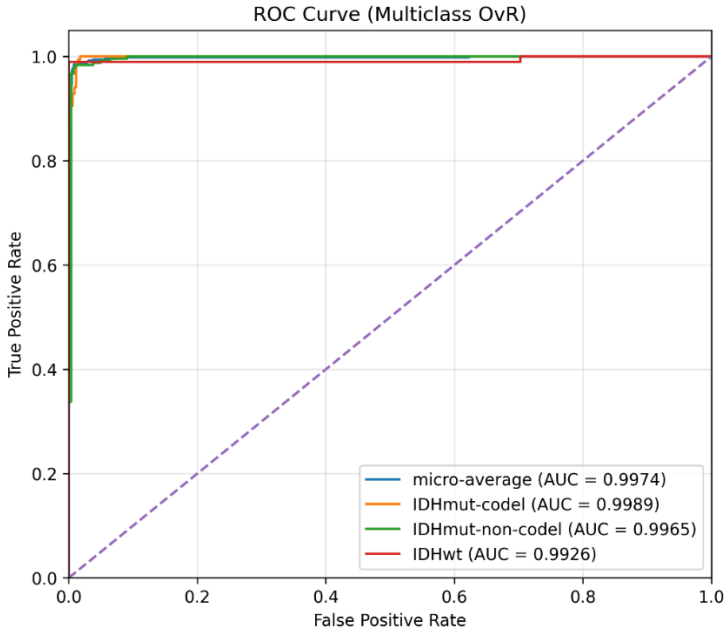


Figure3. Multiclass one-vs-rest (OvR) ROC curves for LGG molecular subtype classification using the proposed KAN-based model. The figure shows ROC curves for IDHmut-codel, IDHmut-non-codel, and IDHwt subtypes, along with the micro-averaged ROC curve, demonstrating strong discriminative performance across all classes.

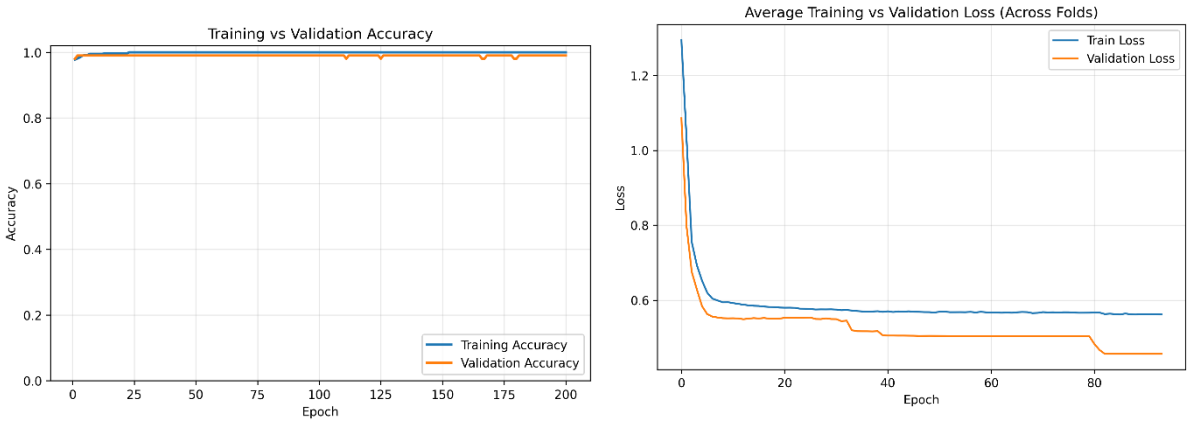
4.3 Cross-Validation Stability and Convergence

The stability and convergence behaviour of the proposed KAN-based framework were examined using stratified 10-fold cross-validation. As illustrated in Figure 4(a), the training and validation accuracy curves rapidly converge to high values and remain closely aligned throughout the training process. This alignment indicates stable learning behaviour and minimal performance variation across epochs, with no evidence of overfitting.

Similarly, the average training and validation loss curves aggregated across all folds (Figure 4(b)) exhibit a smooth and monotonic decrease during the early training stages, followed by gradual stabilisation. Notably, the validation loss remains low and closely tracks the training loss, confirming the effectiveness of the employed regularisation strategy and early stopping mechanism.

Quantitatively, the model achieved a mean classification accuracy of 99.02% with a standard deviation of 1.32% across folds, demonstrating strong robustness to data partitioning. Collectively, the tightly aligned accuracy trajectories, stable loss convergence, and low cross-fold variance provide clear evidence of the proposed framework's strong generalisation capability and reliable optimisation behaviour in high-dimensional multi-omics classification tasks.

Figure 4. *(a)* Training and validation accuracy across epochs, showing closely aligned curves and



stable generalization. *(b)* Average training and validation loss aggregated across cross-validation folds, illustrating smooth convergence and effective regularization.

4.4 Biomarker Analysis Based on Attention Weights

In addition to strong classification performance, the proposed KAN-based framework provides intrinsic interpretability through its feature-level attention mechanism. The learned attention weights were quantitatively analysed and used as indicators of feature importance to identify subtype-associated molecular biomarkers.

As summarised in Table 4, the model identified the top 20 biomarkers based on their average attention weights across all validation folds. These biomarkers span multiple omics layers, including mRNA expression, miRNA regulation, and DNA methylation, highlighting the framework's ability to integrate heterogeneous molecular information. Notably, several highly ranked mRNA biomarkers, such as CDH4, CDKN2C, PTGFRN, EGFR, and GRB10, have been previously reported to play important roles in glioma development, progression, and prognosis. In addition, the identified miRNA MIR196A2 and the DNA methylation probe cg03513762 have been implicated in glioma-related molecular regulation.

The consistency between the biomarkers highlighted by the attention mechanism and previously reported glioma-associated molecular markers demonstrates that the proposed model effectively captures subtype-specific biological signals. These findings confirm the capability of the KAN-based framework to support transparent and biologically meaningful biomarker discovery directly from multi-omics data, without reliance on predefined biological interaction networks.

Table 4. Top 20 KAN-Based Biomarkers Identified by Attention Weights

Rank	Biomarker ID	Gene Name	Omics Type	REF
1	ENSG00000179242.16	CDH4	mRNA	PMID: 37237299[26]
2	ENSG00000123080.12	CDKN2C	mRNA	PMID: 37237299 [27]
3	ENSG00000134247.10	PTGFRN	mRNA	PMID: 33484965 [28]
4	ENSG00000146648.19	EGFR	mRNA	PMID: 35427766 [29]
5	ENSG00000187514.17	PTMA	mRNA	PMID: 35488273[30]
6	ENSG00000008130.15	NADK	mRNA	PMID: 39159134[31]
7	ENSG00000133789.15	SWAP70	mRNA	PMID: 21360565[32]
8	ENSG00000134121.10	CHL1	mRNA	PMID: 34718336[33]
9	ENSG00000106070.20	GRB10	mRNA	PMID: 37179120 [34]
10	hsa-mir-196a-2	MIR196A2	miRNA	PMID: 29802998[35]
11	ENSG00000132341.12	RAN	mRNA	PMID: 25367065 [36]
12	ENSG00000138095.19	LRPPRC	mRNA	PMID: 38372449 [37]
13	cg03513762	HumanMethylation450 -	DNA methylation	PMID: 23867710[38]
14	ENSG00000050426.16	LETMD1	mRNA	PMID: 35818268[39]
15	ENSG00000165795.23	NDRG2	mRNA	PMID: 29285747[40]
16	ENSG00000162873.15	KLHDC8A	mRNA	PMID: 39327249 [41]
17	ENSG00000028137.19	TNFRSF1B	mRNA	PMID: 37712139 [42]
18	ENSG00000117450.15	PRDX1	mRNA	PMID: 39430237[43]
19	ENSG00000011454.18	RABGAP1	mRNA	PMID: 36685876[44]
20	ENSG00000070047.13	PHRF1	mRNA	PMID: 38030167[45]

4.5 Comparison with Previous Work

The performance of the proposed KAN-based framework was compared with representative LGG classification methods reported in the literature, as summarised in Table 5. The proposed model achieved a classification accuracy of 99.02%, outperforming existing multi-omics LGG frameworks, including BioGAT-LGG (98.03%) and DeepAutoGlioma (98.03%), while m aintaining full feature-level interpretability.

Compared to graph-based approaches such as BioGAT-LGG and MOGONET, which rely on explicit graph construction and node-level message passing, the proposed framework learns nonlinear interactions directly at the feature level without requiring predefined or data-derived graph structures. This graph-free learning strategy reduces architectural complexity while preserving effective joint integration of multi-omics data.

In contrast to autoencoder-driven and unsupervised frameworks (e.g., i-Modern and DeepMoIC), which either report lower subtype discrimination performance or focus on grade-level classification, the proposed KAN-based model demonstrates superior subtype-level accuracy for LGG. Furthermore, unlike image-based approaches such as DeepGlioma, which rely on single-modality histopathological data, the proposed method exploits complementary molecular information across multiple omics layers.

Overall, the comparative results in Table 5 indicate that the proposed attention-enhanced KAN framework achieves state-of-the-art classification performance for LGG, while offering improved interpretability and flexibility relative to existing graph-based and deep learning models.

Table5. Summary of representative studies for LGG classification across different data modalities and learning paradigms.

Study	Data Type(s)	Method	Accuracy	Year	Cancer Type	Graph-Based / Omics Integration
KAN-based LGG (Proposed)	mRNA + miRNA + DNA meth.	KAN (Feature-level learning)	99.02	2025	LGG	No / Joint
BioGAT-LGG	mRNA + miRNA + DNA meth.	Correlation-based GATv2	98.03%	2025	LGG	Yes / Joint
DeepAutoGlioma [16]	mRNA + DNA meth.	Autoencoder + CNN	98.03%	2023	LGG	No / Joint
MOGONET [14]	mRNA + miRNA + DNA meth.	Multi-view GCN	—	2021	Multiple Cancers	Yes / Joint
i-Modern [46]	mRNA + miRNA + DNA meth. + CNV + mutation	Autoencoder + FC layers (Unsupervised)	—	2022	Glioma	No / Joint
DeepMoIC [47]	mRNA + miRNA + DNA meth.	Autoencoder + Deep GCN	73.2% (Grade II vs III)	2024	LGG	Yes / Joint
DeepGlioma [48]	SRH histology images	CNN + Transformer	91.5%	2023	LGG	No / Single-Modality

5. Discussion

This study demonstrates the effectiveness of an attention-enhanced Kolmogorov–Arnold Network (KAN) for accurate molecular subtype classification of lower-grade glioma (LGG) using integrated multi-omics data. The KAN-based model achieved a high cross-validated classification accuracy of 99.02%, confirming its strong generalisation capability and robustness in modelling heterogeneous molecular profiles.

Unlike graph-based deep learning approaches that rely on pre-defined or data-derived interaction networks, the proposed method operates directly on tabular multi-omics features, enabling the

capture of complex nonlinear relationships across molecular layers without explicit graph construction. This feature-centric learning strategy reduces architectural complexity while maintaining competitive, and in this case superior, superior predictive performance.

A key strength of the proposed framework lies in its intrinsic interpretability. The feature-level attention mechanism assigns importance weights during training, allowing direct identification of subtype-associated biomarkers across mRNA, miRNA, and DNA methylation layers. The presence of highly ranked biomarkers that are consistent with previously reported glioma-related molecular markers supports the biological relevance of the learned attention patterns and highlights the utility of KAN-based models for transparent biomarker discovery.

While the current evaluation focuses on the TCGA-LGG cohort using stratified cross-validation, the strong and consistent results suggest that the proposed approach provides a solid and scalable foundation for multi-omics integration in precision oncology. Future research directions include extending the framework to addadditional independent cohorts, exploring broader multi-omics modalities, and expanding the use of clinical and longitudinal data to further enhance translational applicability.

6. Conclusion and Recommendations

In this study, we proposed an interpretable deep learning framework based on the KAN-based model for accurate molecular subtype classification of LGG using integrated multi-omics data. Experimental evaluation using stratified cross-validation demonstrated strong and consistent performance, indicating robust generalisation and high classification accuracy across LGG subtypes. Furthermore, the attention mechanism enabled intrinsic identification of subtype-associated molecular biomarkers across multiple omics layers, enhancing model transparency without reliance on predefined biological networks.

Based on these findings, the following recommendations can be drawn:

- KAN architectures are recommended for LGG molecular subtyping, as the proposed model demonstrated high classification accuracy and robust generalisation under stratified cross-validation.

- Feature-level attention mechanisms are recommended for biomarker discovery, given their effectiveness in identifying subtype-associated molecular markers across mRNA, miRNA, and DNA methylation layers with intrinsic interpretability.
- KAN-based multi-omics integration is recommended as an interpretable alternative to graph-based models that rely on predefined interaction networks, enabling direct learning from gene-level features while avoiding potential bias and limited adaptability introduced by stat ic biological graphs.
- Application of the proposed framework to independent Chinese glioma cohorts (e.g, CGGA) is recommended to further assess robustness and generalisability across different populations and sequencing platforms.
- Extension of the proposed framework to additional multi-omics modalities (such as proteomics, metabolomics, copy number variation, or single-cell omics) is recommended to evaluate scalability and the ability to capture complementary molecular mechanisms.
- Further expansion of the clinical and longitudinal data dimensions within the proposed framework is recommended, supporting more comprehensive outcome prediction and long-term disease modelling in precision oncology.

7. Declarations

7.1 Ethical approval

This study did not involve experiments on human participants or animals. All data were obtained from publicly available repositories (TCGA), and therefore ethical approval was not required.

7.2 Consent to participate

Not applicable. No human subjects were directly involved, and all datasets were fully anonymised and publicly accessible.

7.3 Consent to publish

Not applicable. No individual or identifiable human data are included, and therefore no consent for publication was required.

7.4 Institutional Review Board Statement

Not applicable.

7.5 Clinical Trial Registration

Not applicable.

7.6 Informed Consent Statement

Not applicable.

7.7 Acknowledgments

The authors would like to acknowledge all contributors who provided invaluable support in data collection, analysis, and manuscript preparation.

7.8 Conflicts of Interest

The authors declare no conflicts of interest. The funders had no role in the design of the study, data collection, analysis, or decision to publish the results.

References

1. Louis, D.N., et al., *The 2021 WHO classification of tumors of the central nervous system: a summary.* Neuro-oncology, 2021. **23**(8): p. 1231-1251.
2. Menyhárt, O. and B. Győrffy, *Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis.* Computational and structural biotechnology journal, 2021. **19**: p. 949-960.
3. Sartori, F., et al., *A Comprehensive Review of Deep Learning Applications with Multi-Omics Data in Cancer Research.* Genes, 2025. **16**(6): p. 648.
4. Muzio, G., L. O'Bray, and K. Borgwardt, *Biological network analysis with deep learning.* Briefings in bioinformatics, 2021. **22**(2): p. 1515-1530.
5. Zohari, P. and M.H. Chehreghani, *Graph Neural Networks in Multi-Omics Cancer Research: A Structured Survey.* arXiv preprint arXiv:2506.17234, 2025.
6. Kourou, K., et al., *Machine learning applications in cancer prognosis and prediction.* Computational and structural biotechnology journal, 2015. **13**: p. 8-17.
7. Eraslan, G., et al., *Deep learning: new computational modelling techniques for genomics.* Nature reviews genetics, 2019. **20**(7): p. 389-403.
8. Alipanahi, B., et al., *Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning.* Nature biotechnology, 2015. **33**(8): p. 831-838.
9. Danaee, P., R. Ghaeini, and D.A. Hendrix. *A deep learning approach for cancer detection and relevant gene identification*. in *Pacific symposium on biocomputing. pacific symposium on biocomputing*. 2016.

10. Yang, W. and G. Jin, *Origin-independent analysis links SARS-CoV-2 local genomes with COVID-19 incidence and mortality.* Briefings in Bioinformatics, 2021. **22**(2): p. 905-913.
11. Song, J., W. Peng, and F. Wang, *Identifying cancer patient subgroups by finding co-modules from the driver mutation profiles and downstream gene expression profiles.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021. **19**(5): p. 2863-2872.
12. Singh, A., et al., *DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays.* Bioinformatics, 2019. **35**(17): p. 3055-3062.
13. Sun, Q., et al., *SADLN: Self-attention based deep learning network of integrating multi-omics data for cancer subtype recognition.* Frontiers in Genetics, 2023. **13**: p. 1032768.
14. Wang, T., et al., *MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification.* Nature communications, 2021. **12**(1): p. 3445.
15. Wu, J., et al., *DeepMoIC: multi-omics data integration via deep graph convolutional networks for cancer subtype classification.* BMC genomics, 2024. **25**(1): p. 1209.
16. Munquad, S. and A.B. Das, *DeepAutoGlioma: a deep learning autoencoder-based multi-omics data integration and classification tools for glioma subtyping.* BioData Mining, 2023. **16**(1): p. 32.
17. Ryan, B., R.E. Marioni, and T.I. Simpson, *Multi-Omic Graph Diagnosis (MOGDx): a data integration tool to perform classification tasks for heterogeneous diseases.* Bioinformatics, 2024. **40**(9): p. btae523.
18. Elbashir, M.K., A. Alanazi, and M.A. Mahmood, *Decoding Multi-Omics Signatures in Lower-Grade Glioma Using Protein–Protein Interaction-Informed Graph Attention Networks and Ensemble Learning.* Diagnostics, 2025. **15**(22): p. 2894.
19. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique.* Journal of artificial intelligence research, 2002. **16**: p. 321-357.
20. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data.* Nucleic acids research, 2016. **44**(8): p. e71-e71.
21. Jiang, G., et al., *A comprehensive workflow for optimizing RNA-seq data analysis.* BMC genomics, 2024. **25**(1): p. 631.
22. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47-e47.
23. Law, C.W., et al., *voom: Precision weights unlock linear model analysis tools for RNA-seq read counts.* Genome biology, 2014. **15**: p. 1-17.
24. Tibshirani, R., *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996. **58**(1): p. 267-288.
25. Qi, L., et al., *Multi-omics data fusion for cancer molecular subtyping using sparse canonical correlation analysis.* Frontiers in genetics, 2021. **12**: p. 607817.
26. Xie, J., et al., *CDH4 inhibits ferroptosis in oral squamous cell carcinoma cells.* BMC Oral Health, 2023. **23**(1): p. 329.
27. Pereira, M.J., et al., *CDKN2C expression in adipose tissue is reduced in type II diabetes and central obesity: impact on adipocyte differentiation and lipid storage?* Translational Research, 2022. **242**: p. 105-121.

28. Dooley, K., et al., *A versatile platform for generating engineered extracellular vesicles with defined therapeutic properties.* Molecular Therapy, 2021. **29**(5): p. 1729-1743.
29. Levantini, E., et al. *EGFR signaling pathway as therapeutic target in human cancers*. in *Seminars in cancer biology*. 2022. Elsevier.
30. Zhang, Z., et al., *Integrated analysis of single-cell and bulk RNA sequencing data reveals a pan-cancer stemness signature predicting immunotherapy response.* Genome Med, 2022. **14**(1): p. 45.
31. Lin, W. and N. Wang, *NUAK1-Mediated Phosphorylation of NADK Mitigates ROS Accumulation to Promote Osimertinib Resistance in Non-Small Cell Lung Carcinoma.* 2024. **84**(23): p. 4081-4098.
32. Chiyomaru, T., et al., *SWAP70, actin-binding protein, function as an oncogene targeting tumor-suppressive miR-145 in prostate cancer.* Prostate, 2011. **71**(14): p. 1559-67.
33. Tian, W., et al., *Exosomal miR-338-3p suppresses non-small-cell lung cancer cells metastasis by inhibiting CHL1 through the MAPK signaling pathway.* 2021. **12**(11): p. 1030.
34. Ren, L.L., et al., *GRB10 is a novel factor associated with gastric cancer proliferation and prognosis.* Aging (Albany NY), 2023. **15**(9): p. 3394-3409.
35. Haerian, M.S., et al., *MIR196A2 rs11614913 contributes to susceptibility to colorectal cancer in Iranian population: A multi-center case-control study and meta-analysis.* Gene, 2018. **669**: p. 82-90.
36. Sun, Y., et al., *Evaluation of superparamagnetic iron oxide-polymer composite microcapsules for magnetic resonance-guided high-intensity focused ultrasound cancer surgery.* BMC Cancer, 2014. **14**: p. 800.
37. Yu, Y. and H. Deng, *LRPPRC promotes glycolysis by stabilising LDHA mRNA and its knockdown plus glutamine inhibitor induces synthetic lethality via m(6) A modification in triple-negative breast cancer.* 2024. **14**(2): p. e1583.
38. Naumov, V.A., et al., *Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips.* Epigenetics, 2013. **8**(9): p. 921-34.
39. Wen, N., J. Zhang, and Q. Zhang, *MiR-494 Inhibits the Proliferation, Migration and Invasion of Cervical Cancer Cells by Regulating LETMD1.* Cell Mol Biol (Noisy-le-grand), 2022. **67**(5): p. 81-87.
40. Chen, X.L., et al., *Potential role of NDRG2 in reprogramming cancer metabolism and epithelial-to-mesenchymal transition.* Histol Histopathol, 2018. **33**(7): p. 655-663.
41. Peng, Y., et al., *KLHDC8A Regulates M1/M2 Macrophage Polarization Through the PD-1/STAT3 Pathway to Promote Papillary Thyroid Cancer Development.* Discov Med, 2024. **36**(188): p. 1858-1868.
42. Gao, Y., et al., *Single-cell transcriptomics identify TNFRSF1B as a novel T-cell exhaustion marker for ovarian cancer.* Clin Transl Med, 2023. **13**(9): p. e1416.
43. Song, Y., et al., *PRDX1 inhibits ferroptosis by binding to Cullin-3 as a molecular chaperone in colorectal cancer.* Int J Biol Sci, 2024. **20**(13): p. 5070-5086.
44. Bingen, J.M., et al., *Differential DNA methylation associated with multiple sclerosis and disease modifying treatments in an underrepresented minority population.* Front Genet, 2022. **13**: p. 1058817.
45. Lin, H.W., et al., *PHRF1 Promotes Cell Invasion by Modulating SOX4 Expression in Colorectal Cancer HCT116-p53(-/-) Cells.* Anticancer Res, 2023. **43**(12): p. 5437-5446.

46. Pan, X., et al., *i-Modern: Integrated multi-omics network model identifies potential therapeutic targets in glioma by deep learning with interpretability.* Computational and Structural Biotechnology Journal, 2022. **20**: p. 3511-3521.
47. Wu, J., et al., *DeepMoIC: multi-omics data integration via deep graph convolutional networks for cancer subtype classification.* BMC genomics, 2024. **25**(1): p. 1-13.
48. Hollon, T., et al., *Artificial-intelligence-based molecular classification of diffuse gliomas using rapid, label-free optical imaging.* Nature medicine, 2023. **29**(4): p. 828-832.