Multi-Omics Integration For Disease Prediction Via Multi-Level Graph Attention Network And Adaptive Fusion

Haoran Luo, BS[1,2], Hong Liang, PhD[1], Miao Jia, BS[2], Yanhong Li, BS[2], Xiaohui Yao, PhD[1,2,#], Shan Cong, PhD[1,2,#]
[1]College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, Heilongjiang 150001, China; [2]Qingdao Innovation and Development Center, Harbin Engineering University, Qingdao, Shandong 266000, China

Abstract

By capturing complementary information from multiple omics data, multi-omics integration has demonstrated promising performance in disease prediction. However, with the increasing number of omics data, accurately quantifying the characterization ability of each omics and avoiding mutual interference becomes challenging due to the intricate relationships among omics data. Here, we propose a novel multi-omics integration framework that improves diagnostic prediction. Specifically, our framework involves constructing co-expression and co-methylation networks for each subject, followed by applying multi-level graph attention to incorporate biomolecule interaction information. The true-class-probability strategy is employed to evaluate omics-level confidence for classification, and the loss is designed using an adaptive mechanism to leverage both within- and across-omics information. Extensive experiments demonstrate that the proposed framework outperforms state-of-the-art methods on classification tasks, and indicate that the integration of three omics yields superior performance compared to employing only one or two data types. Code and data are available at https://github.com/Yaolab-fantastic/GRAMI-NET.

Introduction

In light of recent advancements in the acquisition of high-throughput omics data, multi-omics integration has emerged as a rapidly expanding research field, aiming to provide a more comprehensive understanding of the underlying biological processes and molecular mechanisms involved in complex diseases[1,2]. Compared to single-omics studies, the integration of multiple types of omics data enables the capture of complementary information across various molecular layers, leading to a more holistic view of biological systems[3]. Traditional approaches often involve statistical tools, which may have limited capacity to fully capture the complex, non-linear relationships present in multi-omics datasets. In recent years, the application of artificial neural networks (ANN) to multi-omics studies has emerged as a promising avenue for addressing these limitations[4–6]. ANN approaches can learn intricate patterns within the data, enabling more accurate predictions and classifications, as well as the identification of previously unexplored relationships between molecular entities. These methods have shown great potential in various biomedical applications, such as disease prediction, patient stratification, and the discovery of novel biomarkers[7].

While these advanced ANN techniques have shown promise in multi-omics studies, effectively harnessing the complex and implicit information within and across omics layers continues to pose a formidable challenge[7–9]. Specifically, one major challenge is to effectively learn the inherent structure and complex relationships of omics data. Numerous current omics investigations tend to disregard the functional interactions among molecules, treating each element as an independent entity[10,11]. Such studies may overlook crucial aspects of the underlying biological processes and result in a limited understanding of the complex molecular networks that drive various biological systems[12]. Incorporating the consideration of functional interactions into multi-omics analyses can enable a more comprehensive and accurate representation of biological mechanisms[13]. Thus, several studies have considered the importance of integrating functional interaction information, introducing a protein-based network[14,15], HINT protein-protein interaction (PPI) network[16,17], and STRING PPI database[18,19] in their proposed models. These approaches have demonstrated improved performance and more accurate insights into the complex biological processes under investigation. However, as functional interactions are usually relevant to the disease context, the aforementioned networks still face the challenge of lacking disease-specific information. As a result, these generic interaction networks may not fully capture the nuances and complexities associated with a particular disease or condition. Therefore, recent efforts have increasingly focused

Correspondence should be addressed to shan.cong@hrbeu.edu.cn or xiaohui.yao@hrbeu.edu.cn

Table 1: Summary of data used in this study

Dataset	Feature Selection	Categories	Number of features mRNA, meth, miRNA	Number of features for training mRNA, meth, miRNA
ROSMAP	200	NC: 169, AD: 182	55,889, 23,788, 309	200, 200, 200
BRCA	1000	Normal-like: 115, Basal-like: 131, HER2-enriched: 46, Luminal A: 436, Luminal B: 147	20,531, 20,106, 503	1000, 1000, 503

on constructing disease-specific networks by transforming omics data into graph structures, reflecting a growing recognition of the importance of contextualized molecular interactions in understanding disease mechanisms. For example, Ramirez et al.[20] adopted graph convolutional network (GCN) for cancer classification by using the gene co-expression network, which is built from Spearman correlation. Xing et al.[21] employed weighted correlation network analysis to create a co-expression network, then applied multi-level attention graph neural network to conduct disease predictions. The co-expression networks discussed above are sample-wise, meaning that each sample corresponds to a single graph in which nodes represent genes (or other molecules) and edges represent expression correlations. Alternatively, Wang et al.[22] constructed one network for each omics with nodes denoting samples and edges denoting expression similarity between samples, then applied GCNs on each graph for initial omics-specific prediction. Drawing inspiration from the success of network representations in aforementioned multi-omics studies, we are motivated to adopt a network-based representation. This approach is intuitive, easy to interpret, and has the potential to better preserve the inherent structure and capture functional interactions from omics data, resulting in improved disease prediction performance.

The other major challenge is the efficient integration of various omics data. Different data distributions and the heterogeneity among multi-omics can hinder effective integration and produce an imbalance in the learning process[7,23]. Integration strategies can be categorized into three classes, including early, intermediate, and late approaches[7,24,25]. Both early and intermediate strategies involve using a single model to handle multiple omics data, either by concatenating the omics data into a single input dataset or by learning a joint representation of the multiple omics[26–29]. A common issue of one-model methods is the learning imbalance introduced by different data sizes, distributions, mutual interference among multi-omics data, etc.[7]. In contrast, late integration models treat each omics data separately and combines their respective predictions. By taking advantage of omics-specific models and well-designed cross-omics integration, the late strategy has shown superior performance on different biological problems. For example, Wang et al.[22] proposed MOGONET, which trains three GCNs separately on each omics dataset and generates three omics-specific predictions. They then leverage a view correlation discovery network to integrate these initial classifications and produce a final prediction. Han et al.[30] employs encoder networks to extract feature-level omics-specific informativeness and adopts true-class probability to evaluate the classification confidence of different omics. They show that confidence-based integration has the potential to improve classification performance.

Based on the above observations, we propose a multi-omics integration framework for disease prediction that leverages graph attention networks and incorporates a confidence-adaptive mechanism. Specifically, graph structures have been shown to be significantly effective in representing large-scale, unstructured and irregular data like omics data, especially when disease specificity is taken into consideration. Furthermore, the confidence mechanism is capable of assessing the reliability of omics data and adapting sample variations across different omics datasets. We demonstrate that the combination of graph-based feature representation and confidence-based adaptive fusion outperforms state-of-the-art methods on complex disease classification tasks.

Materials and Methods

The proposed method is illustrated in Figure 1. First, we construct a co-expression matrix for one type of omics data of each subject. Second, the initial feature for each type of omics data is generated by the multi-level GAT process, which utilizes three graph attention network layers to extract topological and hierarchical features from the GAT network 0, network 1, and network 2. Third, the aggregated features of each type of omics data are parameterized training based on the confidence-adaptive mechanism respectively. The final decision features are inferred by TCP module. Finally, after the fusion of all the decision features, a final label prediction is obtained with neural network.

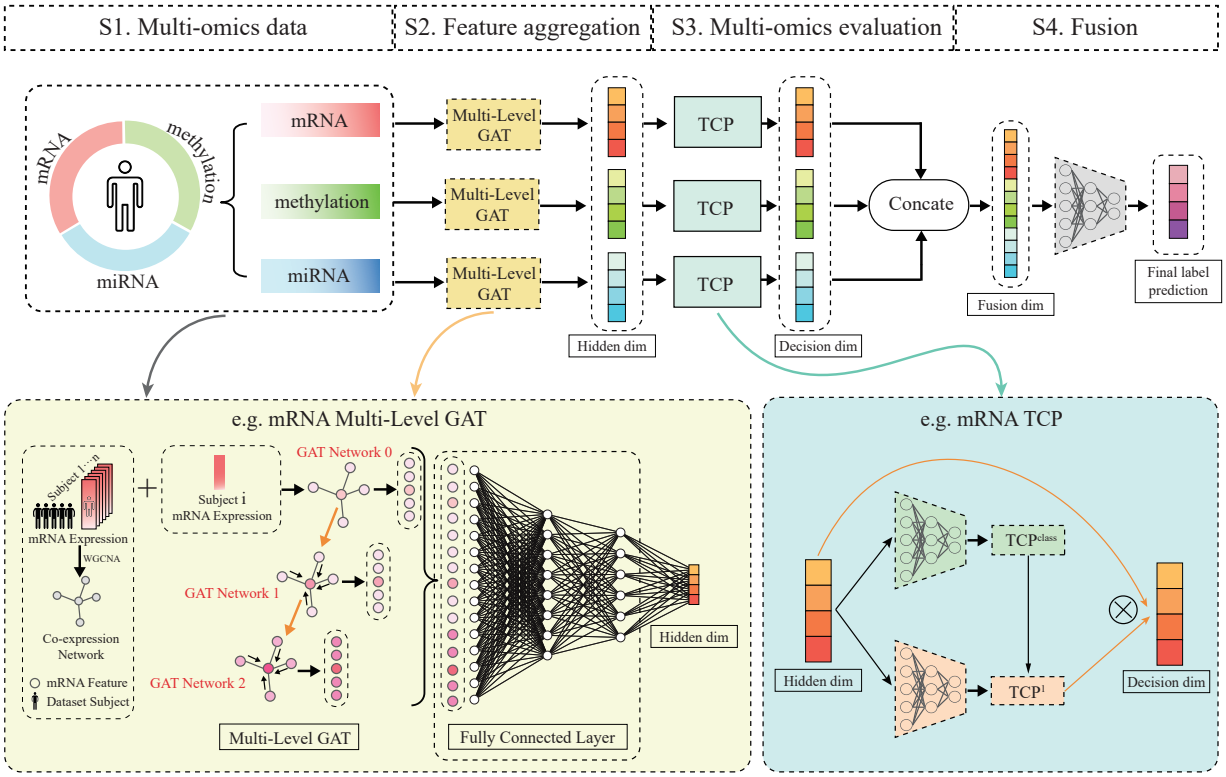

Figure 1: Architecture of the proposed method. **(S1)** For each subject, multi-omics data is input into the model after preprocessing. **(S2)** The initial feature is generated by the multi-level GAT for each type of omics data respectively. **(S3)** The dicision feature of each type of omics data is generated by the TCP module. **(S4)** In the fusion process, the decision features of each omics are concatenated into one fusion feature. Then this fusion feature is performed to conduct downstream classification task.

S1. Multi-omics Datasets

We conduct extensive experiments on two publicly available datasets: ROSMAP for binary classification (ie., Alzheimer's disease vs. healthy controls (HC)), and BRCA for multiclass classification (ie., breast invasive carcinoma PAM50 subtype classification). Our study uses three types of omics data, including mRNA expression, miRNA expression, and DNA methylation. Table 1 presents the details of the two datasets. For more information regarding data acquisition and processing, please refer to Wang et al.[22].

Co-expression and Co-methylation Network Construction

Our objective is to capture the inherent topological characteristics of omics data to construct co-expression and co-methylation networks, thereby facilitating a more comprehensive understanding of the molecular interactions within biological systems. To achieve this, we employ the WGCNA[31], which is adept at processing various types of omics data, and resulting in the identification of related entities that offer valuable insights into the complex interplay among molecular components. As shown in Figure 1 (S1), for each patient, an initial co-expression network can be represented by $G_0 = G(V^{K\times 1}, E^{K\times K})$, where K represents the number of features, $V^{K\times 1}$ represents the vector of the nodes in G, and $E^{K\times K}$ represents the edge matrix. The edge matrix $E^{K\times K}$ is calculated based on the co-expression computation by WGCNA analysis. The WGCNA is performed by R package "WGCNA". Specifically, for each subject, a $1 \times K$ dimension vector is generated. For N subjects belonging to one class of omics data, a $N \times K$ matrix is

generated to calculate the co-expression matrix $A^{K\times K}$. The correlation A_{ij} between node v_i and v_j is calculated as

$$A_{ij} = (\frac{1}{2}(1 + \frac{\sum_{n=1}^{N}(v_{i,n} - \bar{v}_i)(v_{j,n} - \bar{v}_j)}{\sqrt{\sum_{n=1}^{N}(v_{i,n} - \bar{v}_i)^2}\sqrt{\sum_{n=1}^{N}(v_{j,n} - \bar{v}_j)^2}}))^{\beta}, \tag{1}$$

where $\bar{v}_i$ and $\bar{v}_j$ are the average features of the node v_i and v_j, β is a soft threshold. Then, the edge matrix $E^{K\times K}$ is generated from the matrix $A^{K\times K}$ by binarizing the values through

$$E_{ij} = \begin{cases} 1, A_{ij} > adj_threshold \\ 0, otherwise \end{cases}. \tag{2}$$

S2. Feature aggregation module

To optimally exploit the information contained within nodes and their relationships in the co-expression and co-methylation matrices, we introduce the Graph Attention Network (GAT) to further refine and extract the characteristic information of omics data. GAT employs masked self-attentional layers, which allows for flexible weighting of neighbor contributions. By integrating the attention mechanism with a weighted sum of neighboring node attributes, GAT is able to adaptively emphasize more relevant neighboring nodes while attenuating the influence of less relevant ones. Consequently, GAT demonstrates an enhanced capacity for capturing complex and non-linear relationships and patterns within the graph structure[32]. Additionally, the attention weights provide an interpretable measure, offering insights into the relative importance of neighboring nodes and the underlying structure of the graph. For each type of omics data, an initial network generated in Sect.2.2 $G_0 = G(V^{K\times 1}, E^{K\times K})$ is passed through a GAT layer. The computation process for GAT can be divided into two primary steps. First step is to calculate attention coefficient parameter. For a given node i, the similarity coefficients between the node itself and its neighboring nodes are calculated as follows:

$$e_{ij} = a([Wh_i \parallel Wh_j]), j \in N_i. \tag{3}$$

The feature enhancement process involves augmenting the node features by applying a linear transformation using a shared parameter matrix W. This transformation adds dimensions to the node features, enabling more expressive and nuanced representations of the graph structure. Then the transformed features of the node i and j are stitched by the $[Wh_i \parallel Wh_j]$. The concatenated high-dimensional features are mapping to a real number and the attention coefficients are normalized as follows:

$$\alpha_{ij} = \frac{\exp(\text{LeakyRelu}(e_{ij}))}{\sum_{k\in N}\exp(\text{LeakyRelu}(e_{ik}))}. \tag{4}$$

For the Second step, the features are weighted and aggregated according to the calculated attention coefficients as follows:

$$h'_i = \sigma(\sum\nolimits_{j\in N}\alpha_{ij}Wh_j), \tag{5}$$

where h'_i denotes the generated feature output by the GAT layer for each node i.

To stabilize the learning process of self-attention, we employed multi-head attention proposed by Vaswani et al.[32]. Namely, the operations of the layer are independently replicated K times that each replica with different parameters, and outputs are feature-wise aggregated by concatenating as follows:

$$h'_i = \|_{k=1}^{K}\, \sigma(\sum\nolimits_{j\in N_i}\alpha_{ij}^{k}W_1^{k}h_j), \tag{6}$$

where $\parallel$ denotes feature concatenation, α_{ij}^{k} denotes the attention coefficients derived by the k-th replica, and W^k denotes the weight matrix specifying the linear transformation of the k-th replica.

To further exploit the relationships between internal features, we adopt the multi-level graph feature fusion method proposed by Xing et al.[21], as illustrated in Figure 1(S2). The initial network, G_0, encapsulates the information pertaining to the original features. Then, G_1 is updated from G_0 through the application of a multi-head GAT layer. In a

similar manner, G_2 is derived from G_1. Consequently, these three distinct levels of graph networks establish a multi-level fusion framework, aggregating feature information step by step through the GAT layers. Following this process, the multiple features of nodes on each graph are reduced by linear transformation layers, which serve to retain the information of individual nodes. Ultimately, the vectors obtained after the linear transformation layers are concatenated, resulting in enriched, multi-level feature representations that effectively capture the underlying complexities within the omics data.

In summary, for each generated network G, the output features comprise weighted combinations of the co-expression or co-methylation modules. Employing the multi-level method in conjunction with WGCNA analysis, the network is further optimized through the multi-head GAT. This approach enables an adaptive refinement of feature interactions within the network, which in turn enhances the performance of the model. This design systematically ensures that the feature information of each node is effectively utilized, contributing to a more robust and accurate representation of the underlying biological processes.

S3. Multi-omics evaluation module

The mutual interference between omics data has always been one of the major challenges in the integration of multi-omics data[33]. To mitigate this issue, we implement a late integration method, which requires that the internal characterization information of each type of omics data is thoroughly extracted, ensuring that each data type yields sufficiently accurate prediction results. However, the heterogeneity-induced imbalance may adversely impact the performance of multi-omics integration. In practice, the decision-making process for each omics data should be transparent and provide feedback for its own predictions. Consequently, we introduce a confidence-adaptive technique capable of delivering more accurate predictions[30,34], addressing the aforementioned challenges and facilitating a more effective integration of multi-omics data.

Typically, many studies adopt the classifier which consists of a softmax output layer and a cross-entropy loss function. This configuration can lead to overly confident predictions, which, in turn, may result in erroneous classification outcomes[35,36]. For example, given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i^*)\}_{i=1}^N$ where $\mathbf{x}_i$ presents feature vector and $y_i^* \in \mathcal{Y} = \{1, ..., k\}$ presents a ground-truth label. Then, a final prediction label can be defined as

$$\hat{y} = \underset{k \in \mathcal{Y}}{\operatorname{argmax}}\, P(Y = k|\mathbf{w}, \mathbf{x}), \tag{7}$$

where P presents the label probability distribution, w presents the parameters of the neural network, and k presents class. Usually, minimizing the cross-entropy loss is used to train the network. The cross-entropy loss is defined as

$$\mathcal{L}_{\mathrm{CE}}(\mathbf{w}; \mathcal{D}) = -\frac{1}{N} \sum_{i=1}^{N} y_i^* \log P(Y = y_i^*|\mathbf{w}, \mathbf{x}_i), \tag{8}$$

where N denotes the number of samples. Thus, maximum class probability (MCP) can be computed as follows:

$$\mathrm{MCP}(\mathbf{x}) = \max_{k \in \mathcal{Y}} P(Y = k|\mathbf{w}, \mathbf{x}) = P(Y = \hat{y}|\mathbf{w}, \mathbf{x}). \tag{9}$$

From Equations 8 and 9, an observation is the maximum softmax function exaggerates the differences between logits, particularly when one logit is significantly higher than the others. This results in an increased probability for the most likely class and diminished probabilities for the remaining classes, leading to overconfidence in predictions. In addition, the cross-entropy loss function can cause the model to focus on increasing the probability of the correct class at the expense of other classes, which can also lead to overconfident predictions.

To mitigate the false prediction caused by high confidence, as shown in Figure 1(S3), true class probability (TCP) is introduced to obtain more trustworthy prediction results[34]. TCP differs from MCP in the way confidence is determined. While MCP selects the maximum softmax output as the confidence, TCP considers the softmax output probability corresponding to the true label as the confidence. In this study, TCP is defined as TCP : $(\mathbf{x}, y^*) \rightarrow P(Y = y^*|\mathbf{w}, \mathbf{x})$, and the confidence criterion is formulated as the ratio between TCP and MCP:

$$\mathrm{TCP}^r(\mathbf{x}, y^*) = \frac{P(Y = y^*|\mathbf{w}, \mathbf{x})}{P(Y = \hat{y}|\mathbf{w}, \mathbf{x})}, \tag{10}$$

where the correct prediction will be 1 and the wrong one will be between [0,1]. During training process, another confidence network is introduced to produce this confidence score $\hat{c}(\mathbf{x}, \theta) \in [0, 1]$. The $\hat{c}(\mathbf{x}, \theta) \in [0, 1]$ is trained to be close to the TCP confidence $c^*(\mathbf{x}, y^*) = P(Y = y^*|\mathbf{w}, \mathbf{x})$. To obtain the decision-level feature $\mathbf{x}_{\mathbf{tcp}}^{i \in \mathbf{m}}$ with TCP confidence, a confidence neural network TCP^1 is introduced to obtain a confidence score s^m to approximate the TCP confidence c^*, where m denotes the number of class of omics data. The TCP^1 further reduces the dimension of the output layer to one logit and activate with sigmoid function to obtain s^m. Specifically, each type of omics data will generate aggreagated features $\mathbf{x}_{\mathbf{GAT}}^{i \in \mathbf{m}}$ after the multi-GAT module. The $\mathbf{x}_{\mathbf{tcp}}^{i \in \mathbf{m}}$ is calculated as $\mathbf{x}_{\mathbf{tcp}}^{i \in \mathbf{m}} = \mathbf{x}_{\mathbf{GAT-1}}^{i \in \mathbf{m}} * s^m$. The $\mathbf{x}_{\mathbf{GAT-1}}^{i \in \mathbf{m}}$ refers to the $GAT_{i \in \mathbf{m}}(\cdot)$ with the last fully connected layer removed.

S4. Fusion module

As shown in Figure 1(S4), the initial characterization information of each type of omics data is transformed into corresponding feature vectors, which are defined as decision-level features. Aiming to improve the fusion's ability to focus on important features, while reducing the impact of distractors or irrelevant features, a selective attention mechanism is introduced into the fusion process. For each decision feature F^m, a final $F^m_{attention}$ is calculated as $F^m_{attention} = \sigma(F^m)$, where σ denotes an attention activate fuction. After acquiring the feature $F^m_{attention}$ of decision level, the fusion operation is performed by concatenating vectors of each type of omics data. Then the final classification is completed through neural network.

The loss function consists of two parts. First, for each type of omics data, the loss function of $GAT_i(\cdot)$ can be written as:

$$L^{(i)}_{GAT} = \sum_{j=1}^{n} L_{CE}(\hat{\mathbf{y}}^{(i)}_j, \mathbf{y}_j) = \sum_{j=1}^{n} -\log\left(\frac{e^{\hat{\mathbf{y}}^{(i)}_j \cdot \mathbf{y}_j}}{\sum_k e^{\hat{\mathbf{y}}^{(i)}_{j,k}}}\right), \tag{11}$$

where $L_{CE}(\cdot)$ denotes the cross-entropy loss, $\hat{\mathbf{y}}^{(i)}_j$ denotes the prediction probability distribution of the jth training sample from the ith type of specific omics data. Second, according to Equation 10, ℓ_2 is performed to train the TCP module as follows:

$$\mathcal{L}_{\mathrm{conf}}(\theta; \mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{c}(\mathbf{x}_i, \theta) - c^*(\mathbf{x}_i, y^*_i))^2 \tag{12}$$

Based on this, we adjust the loss function as:

$$\mathcal{L}_{\mathrm{overall}} = \sum_{i=1}^{m} \mathcal{L}_{\mathrm{conf}} + \mathcal{L}_{ce}, \tag{13}$$

where c denotes the number of class of omics data.

To summarize, the total loss function can be defined as follows in our experiments using three different types of omics data:

$$L_{global} = \sum_{i=1}^{m} L^{(i)}_{GAT} + \gamma L_{overall}, \tag{14}$$

where γ denotes an adjustable parameter between the early-stage classification loss and the final classification loss. We set $\gamma = 1$ in all our experiments.

Results

We assess the performance of our proposed model in comparison to state-of-the-art methods across various classification tasks. Additionally, we investigate the potential benefits of integrating three omics data than one or two data types in terms of classification accuracy. To compare the performance of different methods, we employ three metrics including accuracy (ACC), F1 score (F1), and area under the receiver operating characteristic curve (AUC) for the binary classification task, and use accuracy (ACC), average F1 score weighted (F1_weighted), and macro-averaged F1 score (F1_macro) for multiclass classification. We implement the same experimental settings as those used in Mogonet[22] and five random experiments to report the mean and standard deviation of evaluation metrics.

Table 2: Comparison with state-of-the-art methods on the ROSMAP (AD vs. HC) and BRCA (five subtypes according to PAM50) datasets, where the best results are in bold.

	ROSMAP (two categories)			BRCA (five categories)		
Method	ACC	F1	AUC	ACC	F1_weighted	F1_macro
KNN	0.657±0.036	0.671±0.044	0.709±0.045	0.742±0.024	0.730±0.023	0.682±0.025
SVM	0.770±0.024	0.778±0.016	0.770±0.026	0.729±0.018	0.702±0.015	0.640±0.017
Lasso	0.694±0.037	0.730±0.033	0.770±0.035	0.732±0.012	0.698±0.015	0.642±0.026
RF	0.726±0.029	0.734±0.021	0.811±0.019	0.754±0.009	0.733±0.010	0.649±0.013
XGBoost	0.760±0.046	0.772±0.045	0.837±0.030	0.781±0.008	0.764±0.010	0.701±0.017
NN	0.755±0.021	0.764±0.021	0.827±0.025	0.754±0.028	0.740±0.034	0.668±0.047
GRridge	0.760±0.034	0.769±0.029	0.841±0.023	0.745±0.016	0.726±0.019	0.656±0.025
BPLSDA	0.742±0.024	0.755±0.023	0.830±0.025	0.642±0.009	0.534±0.014	0.369±0.017
BSPLSDA	0.753±0.033	0.764±0.035	0.838±0.021	0.639±0.008	0.522±0.016	0.351±0.022
CF	0.784±0.011	0.788±0.005	0.880±0.005	0.815±0.008	0.815±0.009	0.771±0.009
GMU	0.776±0.025	0.784±0.016	0.869±0.016	0.800±0.039	0.798±0.058	0.746±0.058
Mogonet	0.815±0.023	0.821±0.022	0.874±0.012	0.829±0.018	0.825±0.016	0.774±0.017
Dynamic	0.842±0.013	0.846±0.007	**0.912±0.007**	0.877±0.003	0.880±0.005	0.845±0.005
Ours	**0.869±0.012**	**0.867±0.021**	0.898±0.023	**0.880±0.007**	**0.883±0.007**	**0.850±0.021**

Classification performance comparison

A total of fourteen computational methods are compared, encompassing six single-omics classifiers that employ early integration (K-nearest neighbors (KNN)[37], support vector machine (SVM)[38], Lasso[39], random forest (RF)[40], eXtreme Gradient Boosting (XGboost)[41], and fully connected neural networks (NN)[42]), and seven multi-omics classifiers including group-regularized ridge regression[43], BPLSDA that projects data to latent structures with discriminant analysis[44], block PLSDA with additional sparse constraints (BSPLSDA)[44], concatenation of final representations (CF) that concatenates late-stage multi-omics representations[45,46], gated multimodal fusion (GMU) that implements fusion with intermediate representation[47], and two state-of-the-art methods (Mogonet[22] and Dynamic[30]).

Table 2 shows the comparison results on ROSMAP and BRCA. We can observe that our proposed model exhibits superior performance in both binary and multiclass classification tasks. Specifically, compared with the second-best method (Dynamic), our model achieves improvements in most evaluation metrics, except for the AUC of ROSMAP classification. This highlights the informativeness of disease-specific networks as well as the efficacy of multi-level graph attention in information consolidation. Additionally, our model significantly outperforms Mogonet, demonstrating the benefits of utilizing confidence-based adaptive fusion to identify and filter informative modalities that more accurately characterize each subject.

Performance comparison under different omics types

In order to illustrate the added value and complementary nature of integrating multiple types of omics data, we conducted additional comparison experiments by applying the proposed methods to various subsets of omics data. Specifically, we compare the classification performance of utilizing three types of omics data, namely mRNA, methylation, and miRNA, with the performance of using two types of omics data (mRNA+methylation, mRNA+miRNA, and methylation+miRNA), as well as with the performance of using only one type of omics data.

As illustrated in Figure 2, the utilization of all three types of omics data (mRNA + methylation + miRNA) outperforms other combinations of omics data (ie., two omics and single omics data). This observation underscores the distinct roles played by various omics data in the underlying biological processes and complex diseases, as well as the complementary effects of combining more types of omics information in sample characterization.

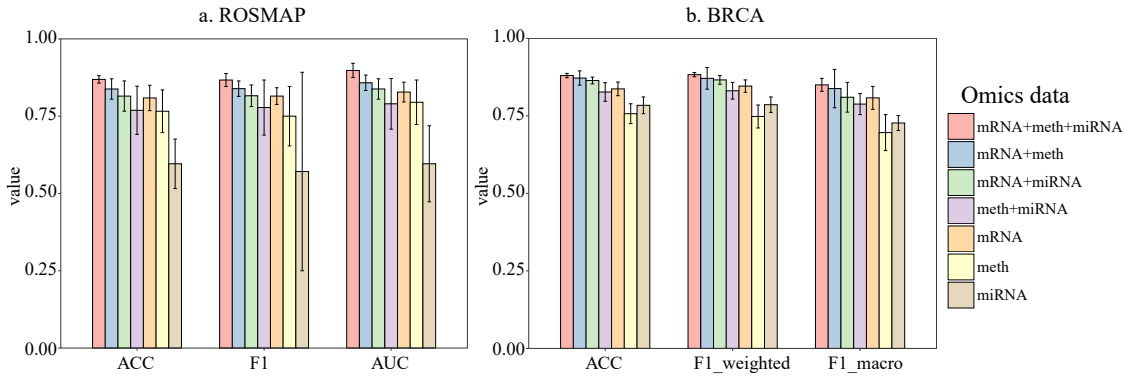

Figure 2: Performance comparison of multi-omics data classification via the proposed method

Discussion and Conclusions

The advancement of omics technologies and personalized medicine brings an abundance of supervised datasets that are profitable to achieve prediction tasks, such as disease diagnosis, grading of tumors, and cancer subtypes. Effective integration of multi-omics data has been shown to outperform single omics data in disease prediction. To achieve optimal multi-omics integration, three expectations would be considered. Firstly, the valid information of each type of omics data should be fully utilized to capture the maximum characterization. Secondly, imbalances due to omics heterogeneity should be overcome to improve integration performance. Lastly, a relatively greater number of omics types would result in better prediction performance.

There are several insights we can gain from the experimental results. Our approach involves constructing individual graphs for each subject within each omics data, rather than creating a subject-similarity network per omics. This data formation strategy transforms the node classification task to graph classification, which allows us to leverage the inner topological information among molecules obtained from disease-specific data to improve model performance. Moreover, our use of confidence-adaptive integration at the decision level helps to address the issue of multi-omics imbalance and interference. Experiments on subsets data highlight the potential of integrating more omics data types for improving prediction performance. However, it is worth noting that even if we can obtain more types of omics data and clinical data in the future, blindly increasing the number of data types may not necessarily improve performance and may instead result in more interference, leading to a decrease in prediction.

In conclusion, we propose an innovative multi-omics integration model that can adaptively integrate diverse omics data while capturing the intricate relationships among various molecular entities, ultimately leading to more accurate and clinically relevant insights into disease mechanisms and patient stratification. Specifically, it can fully exploit the inner information of omics data through the co-expression and co-methylation networks and perform confidence-adaptive fusion overcoming the heterogeneity of multi-omics data. By hierarchically exploiting topological information from co-expressed networks, the multi-level GAT module provides an initial feature for each subject of each type of omics data. The TCP module makes each initial feature more adaptive to produce more trusted decisions. Implemented different classification tasks on two different disease data, the proposed method achieved state-of-the-art performance on the main classification metrics in all experiments. Meanwhile, it also performed well in the integration experiments of different numbers of multi-omics data.

We would like to acknowledge that we have not yet obtained another two datasets (ie., LGG and KIPAN) mentioned in Mogonet and Dynamics. In the future, we will expand our analysis to include more datasets to demonstrate the robustness and generalizability of our model. Furthermore, we plan to expand this work in the direction of model interpretation, such as biomarker identification in terms of molecular subnetworks by taking advantage of the graph structure of the proposed model.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (62103116, 62102115), Shandong Provincial Natural Science Foundation (2022HWYQ-093), the Natural Science Foundation of Heilongjiang Province (LH2022F016 and LH2022H067), and the Fundamental Research Funds for the Central Universities (3072022TS2614).

References

1. Tarazona S, Arzalluz-Luque A, Conesa A. Undisclosed, unmet and neglected challenges in multi-omics studies. Nature Computational Science. 2021;1(6):395-402.
2. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. Nature Reviews Cardiology. 2021;18(5):313-30.
3. Hampel H, Nisticò R, Seyfried NT, Levey AI, Modeste E, Lemercier P, et al. Omics sciences for systems biology in Alzheimer's disease: State-of-the-art of the evidence. Ageing Research Reviews. 2021;69:101346.
4. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances. 2021;49:107739.
5. Kang M, Ko E, Mersha TB. A roadmap for multi-omics data integration using deep learning. Briefings in Bioinformatics. 2022;23(1):bbab454.
6. Krassowski M, Das V, Sahu SK, Misra BB. State of the field in multi-omics research: from computational needs to data mining and sharing. Frontiers in Genetics. 2020;11:610798.
7. Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. Computational and Structural Biotechnology Journal. 2021;19:3735-46.
8. Zhou G, Li S, Xia J. Network-based approaches for multi-omics integration. Computational methods and data analysis for metabolomics. 2020:469-87.
9. Wörheide MA, Krumsiek J, Kastenmüller G, Arnold M. Multi-omics integration in biomedical research–A metabolomics-centric review. Analytica chimica acta. 2021;1141:144-62.
10. Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. Nature Reviews Genetics. 2017;18(9):551-62.
11. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome biology. 2017;18(1):1-15.
12. Muzio G, O'Bray L, Borgwardt K. Biological network analysis with deep learning. Briefings in bioinformatics. 2021;22(2):1515-30.
13. Wang H, Sham P, Tong T, Pang H. Pathway-based single-cell RNA-seq classification, clustering, and construction of gene-gene interactions networks using random forests. IEEE Journal of Biomedical and Health Informatics. 2019;24(6):1814-22.
14. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Molecular systems biology. 2007;3(1):140.
15. Wei P, Pan W. Incorporating gene networks into statistical tests for genomic data via a spatially correlated mixture model. Bioinformatics. 2008;24(3):404-11.
16. Das J, Yu H. HINT: High-quality protein interactomes and their applications in understanding human disease. BMC systems biology. 2012;6(1):1-12.
17. Kong Y, Yu T. A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data. Bioinformatics. 2018;34(21):3727-37.
18. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research. 2015;43(D1):D447-52.
19. Rhee S, Seo S, Kim S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence; 2018. p. 3527-34.

20. Ramirez R, Chiu YC, Hererra A, Mostavi M, Ramirez J, Chen Y, et al. Classification of cancer types using graph convolutional neural networks. Frontiers in physics. 2020;8:203.
21. Xing X, Yang F, Li H, Zhang J, Zhao Y, Gao M, et al. Multi-level attention graph neural network based on co-expression gene modules for disease diagnosis and prognosis. Bioinformatics. 2022;38(8):2178-86.
22. Wang T, Shao W, Huang Z, Tang H, Zhang J, Ding Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nature Communications. 2021;12(1):1-13.
23. Lee B, Zhang S, Poleksic A, Xie L. Heterogeneous multi-layered network model for omics data integration and analysis. Frontiers in genetics. 2020;10:1381.
24. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. Information Fusion. 2019;50:71-91.
25. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics. 2022;23(2):bbab569.
26. El-Manzalawy Y, Hsieh TY, Shivakumar M, Kim D, Honavar V. Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data. BMC medical genomics. 2018;11(3):19-31.
27. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401(6755):788-91.
28. Zhang S, Liu CC, Li W, Shen H, Laird PW, Zhou XJ. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. Nucleic acids research. 2012;40(19):9379-91.
29. Yang Z, Michailidis G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. Bioinformatics. 2016;32(1):1-8.
30. Han Z, Yang F, Huang J, Zhang C, Yao J. Multimodal Dynamics: Dynamical Fusion for Trustworthy Multi-modal Classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022. p. 20707-17.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9(1):1-13.
32. Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph Attention Networks. In: International Conference on Learning Representations; 2018. .
33. Hawe JS, Theis FJ, Heinig M. Inferring interaction networks from multi-omics data. Frontiers in genetics. 2019;10:535.
34. Corbière C, Thome N, Bar-Hen A, Cord M, Pérez P. Addressing failure prediction by learning model confidence. Advances in Neural Information Processing Systems. 2019;32.
35. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. In: Proceedings of the 34th International Conference on Machine Learning; 2017. p. 1321-30.
36. Neumann L, Zisserman A, Vedaldi A. Relaxed softmax: efficient confidence auto-calibration for safe pedestrian detection. OpenReview; 2018. .
37. Evelyn F, JL H. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. International Statistical Review/Revue Internationale de Statistique. 1989;57(3):238.
38. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20:273-97.
39. Ranstam J, Cook J. LASSO regression. Journal of British Surgery. 2018;105(10):1348-8.
40. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. vol. 1. IEEE; 1995. p. 278-82.
41. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016. p. 785-94.

42. Schwenker F, Trentin E. Pattern classification and clustering: A review of partially supervised learning approaches. Pattern Recognition Letters. 2014;37:4-14.

43. Van De Wiel MA, Lien TG, Verlaat W, van Wieringen WN, Wilting SM. Better prediction by use of co-data: adaptive group-regularized ridge regression. Statistics in medicine. 2016;35(3):368-81.

44. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019;35(17):3055-62.

45. Hong D, Gao L, Yokoya N, Yao J, Chanussot J, Du Q, et al. More diverse means better: Multimodal deep learning meets remote-sensing imagery classification. IEEE Transactions on Geoscience and Remote Sensing. 2020;59(5):4340-54.

46. Huang Y, Du C, Xue Z, Chen X, Zhao H, Huang L. What makes multi-modal learning better than single (provably). Advances in Neural Information Processing Systems. 2021;34:10944-56.

47. Arevalo J, Solorio T, Montes-y Gómez M, González FA. Gated multimodal units for information fusion. arXiv preprint arXiv:170201992. 2017.