2023. 11. 09 BIBS Seminar

Integration of multi-omics data using adaptive graph learning and attention mechanism for patient classification and biomarker identification

Kyulhee Han

Contents Table

Introduction

Multi-omics data

- High-throughput biological data generation technologies have evolved rapidly
  - Genomics, epigenomics, transcriptomics, proteomics, and metabolomics
  - Various omics data can help understanding complex diseases at the molecular level for better disease treatment and more accurate clinical decision-making

- Curse of dimensionality
  - When the number of molecular features in a dataset is very large compared to the number of samples, making it difficult to analyze and interpret the data effectively

Multi-omics integration

- Simple approaches
  - Unsupervised
    - perform clustering and classification tasks by integrating multi-omics data into a single low-dimensional embedding space
    - lack of additional information on sample labels, so cannot achieve end-to-end training, which often leads to sub-optimal results
  - Supervised
    - feature concatenation (combining data from each omics), or analyze each data type independently and combine the prediction results (combining prediction result from each omics)
    - cannot effectively consider the correlation between different omics data types and prediction results may be biased towards certain omics

Multi-omics integration

- Considering correlation of each omics type
  - Focus on the correlation between different omics
  - **GRridge (van de Wiel et al.)**: adaptive group-regularized (logistic) ridge regression method to classify cervical cancer using methylation microarray data
  - **DIABLO (Singh et al.):** extended the sparse generalized canonical correlation analysis into a supervised classification framework
  - **SMSPL (Yang et al.):** interactively recommends high-confidence samples from different data types in a soft-weighted manner to predict cancer subtypes
  - These methods assumed simple linearity among omics features, which is not applicable for complex biological study

Multi-omics integration

- Deep learning approaches
  - Focus on the capturing nonlinear relationship, especially using Graph Convolutional Network (GCN)
  - **MoGCN (Li et al.):** graph convolutional network which adopts autoencoder (AE) and similarity network fusion (SNF) methods to obtain multi-omics integrated embedding information and similarity network, respectively
  - **MOGONET (Wang et al.):** extended the sparse generalized canonical correlation analysis into a supervised classification framework

Multi-omics integration

- Limitation of deep learning approaches
  - Using fixed sample similarity networks to learn sample embedding information → incorrect graph makes incorrect results
  - Simply treating the embedding information of different omics equally in the process of omics integration → fails to obtain more reasonable and rich integrated information
  - Hard to select the biomarkers during the model training

Article

Computers in Biology and Medicine 164 (2023) 107303

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Integration of multi-omics data using adaptive graph learning and attention mechanism for patient classification and biomarker identification

Dong Ouyang [a,b], Yong Liang [a,*], Le Li [b], Ning Ai [b], Shanghui Lu [b], Mingkun Yu [b], Xiaoying Liu [c], Shengli Xie [d]

[a] Peng Cheng Laboratory, Shenzhen, 518055, China
[b] School of Computer Science and Engineering, Faculty of Innovation Engineering, Macau University of Science and Technology, 999078, Macao Special Administrative Region of China
[c] Computer Engineering Technical College, Guangdong Polytechnic of Science and Technology, Zhuhai, 519090, China
[d] Guangdong-HongKong-Macao Joint Laboratory for Smart Discrete Manufacturing, Guangzhou, 510000, China

September 2023

Main contribution

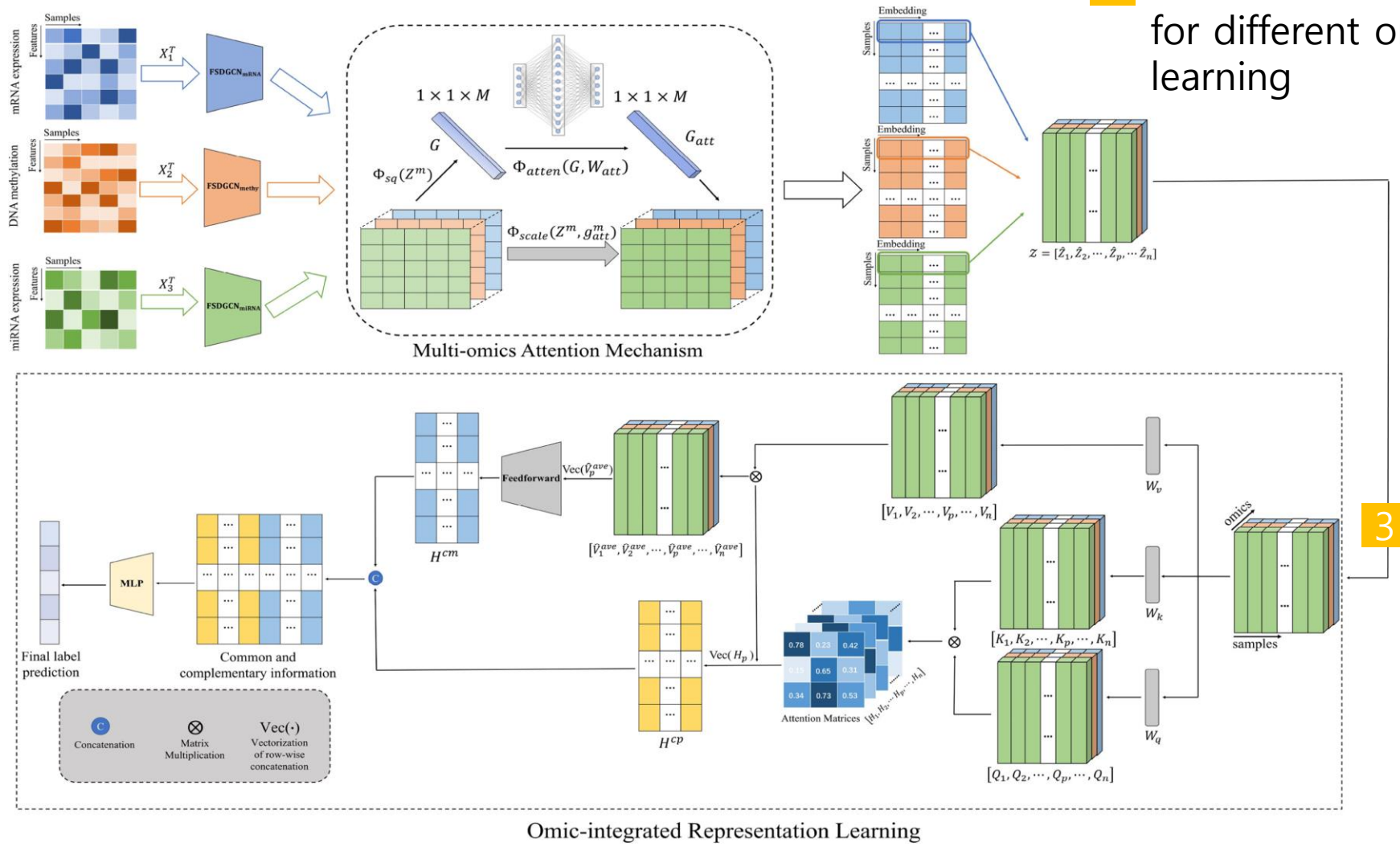
- Learning an omic-specific optimal sample similarity network in adaptive graph learning manner
- Contribution of each omics can be treated differently, and effectively capture common and complementary information
- Selecting important biomarkers during end-to-end training without using independent feature selection methods

Method

Workflow

1 FSDGCN for omic-specific learning

2 multi-omics attention mechanism for different omics importance learning



3 omic-integrated representation learning for multi-omics integration

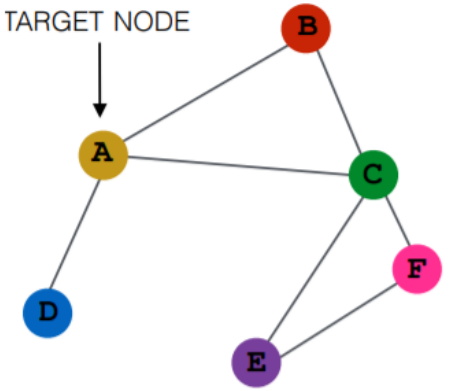
Overview

- Basic information
  - Graph Convolutional Network (GCN)
  - Attention mechanism
- FSDGCN (dynamic graph convolutional network with feature selection)
  - Architecture
  - Feature selection
- MOAM (multi-omics attention mechanism)
- OIRL (omic-integrated representation learning)
- Optimization

Basic information

Graph Convolutional Network (GCN)

- Architecture
  - Use neighborhood information to predict the states of target node
  - Inputs are 1) adjacency matrix (determine graph structure) and 2) node feature matrix (determine the value(s) of each node)



INPUT GRAPH

- 3 steps to train the model
  - Aggregate & Combining the neighbors' information
  - Generate final output for specific purpose (e.g. averaging node features to get node level prediction)

(2) Combining

Update rule: $$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{h}_i^{(l)}\mathbf{W}_0^{(l)} + \sum_{j\in\mathcal{N}_i}\frac{1}{c_{ij}}\mathbf{h}_j^{(l)}\mathbf{W}_1^{(l)}\right)$$

(1) Aggregates

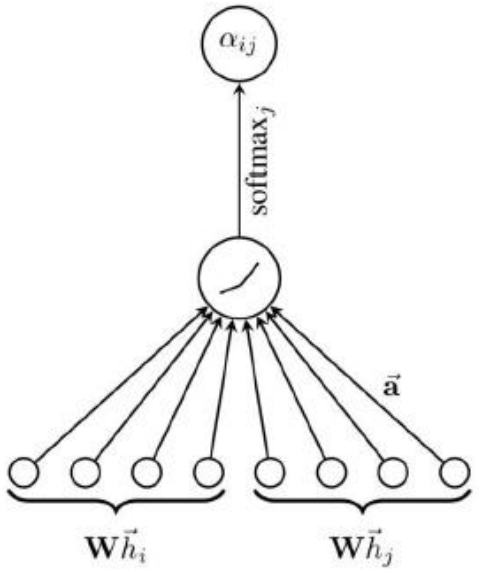
(3) Readout
make node level prediction, graph level prediction, etc.

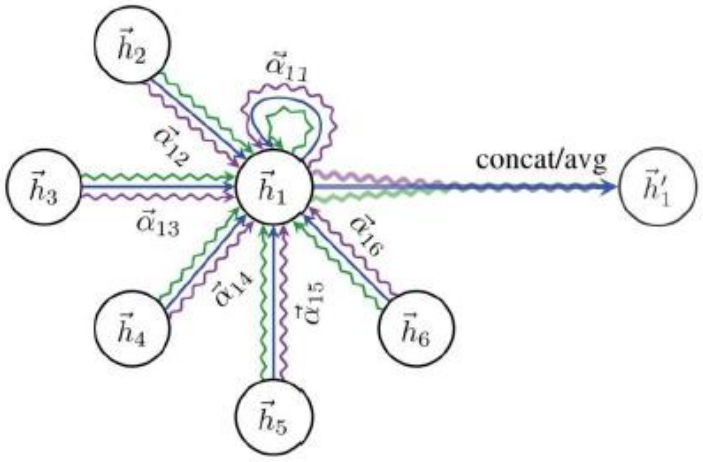
https://medium.com/neuralspace/graphs-neural-networks-in-nlp-dc475eb089de

Basic information

Attention mechanism

- Definition
  - Determining which neighbor's information should more considered to update the hidden states of target node
  - Similar node should be considered as important information
  - To calculate 'attention score', we need to define the attention mechanism to assess the similarity between target node and neighbor node (e.g. dot product, using trainable parameters, etc.)



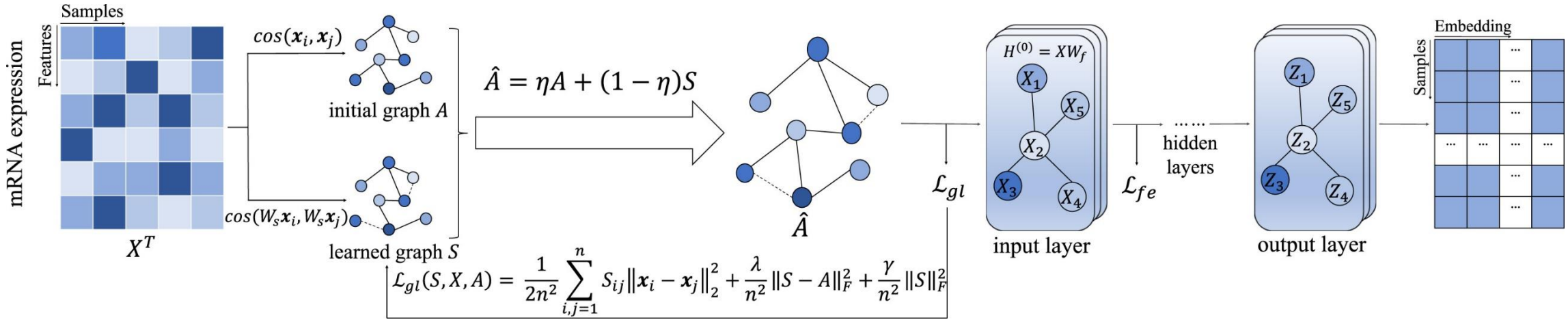


[Figure from Veličković et al. (ICLR 2018)]

FSDGCN (dynamic graph convolutional network with feature selection)

Architecture

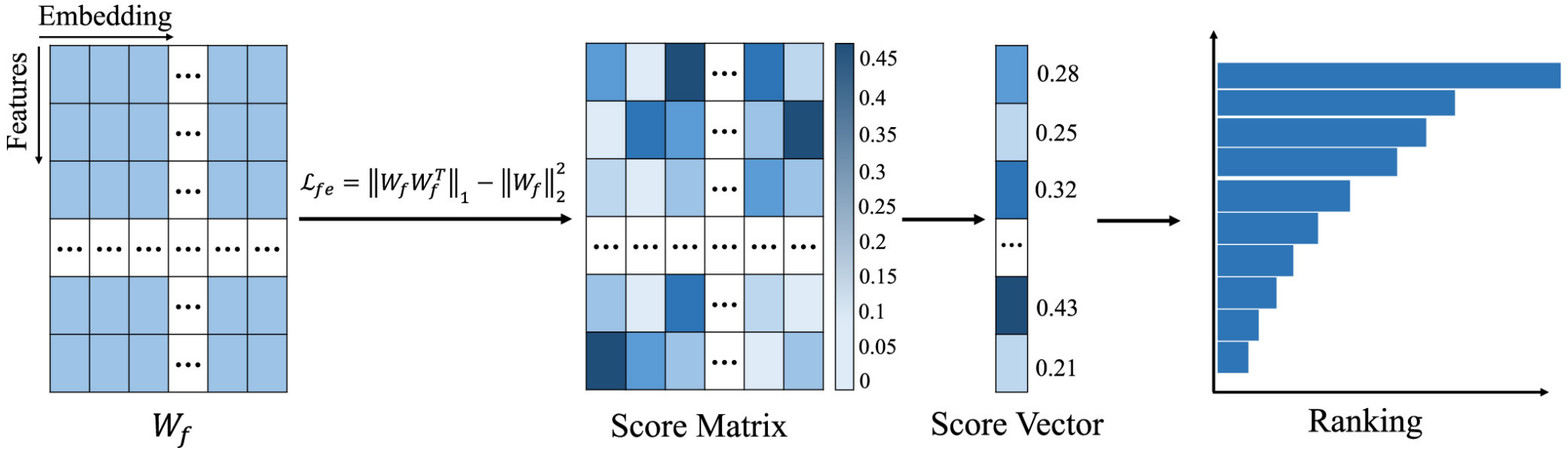
- Graph structure learning
  - Initial graph (A) is cosine similarity matrix of original input data
  - Learned graph (S) is cosine similarity matrix of weighted cosine similarity
  - Updated graph ($\hat{A}$) is weighted sum of those two graph
  - Constraints for smoothness (to make edge between similar node) and sparsity (to avoid trivial solution (S=0), limit total edges and make learned graph closer to initial graph) were considered



FSDGCN (dynamic graph convolutional network with feature selection)

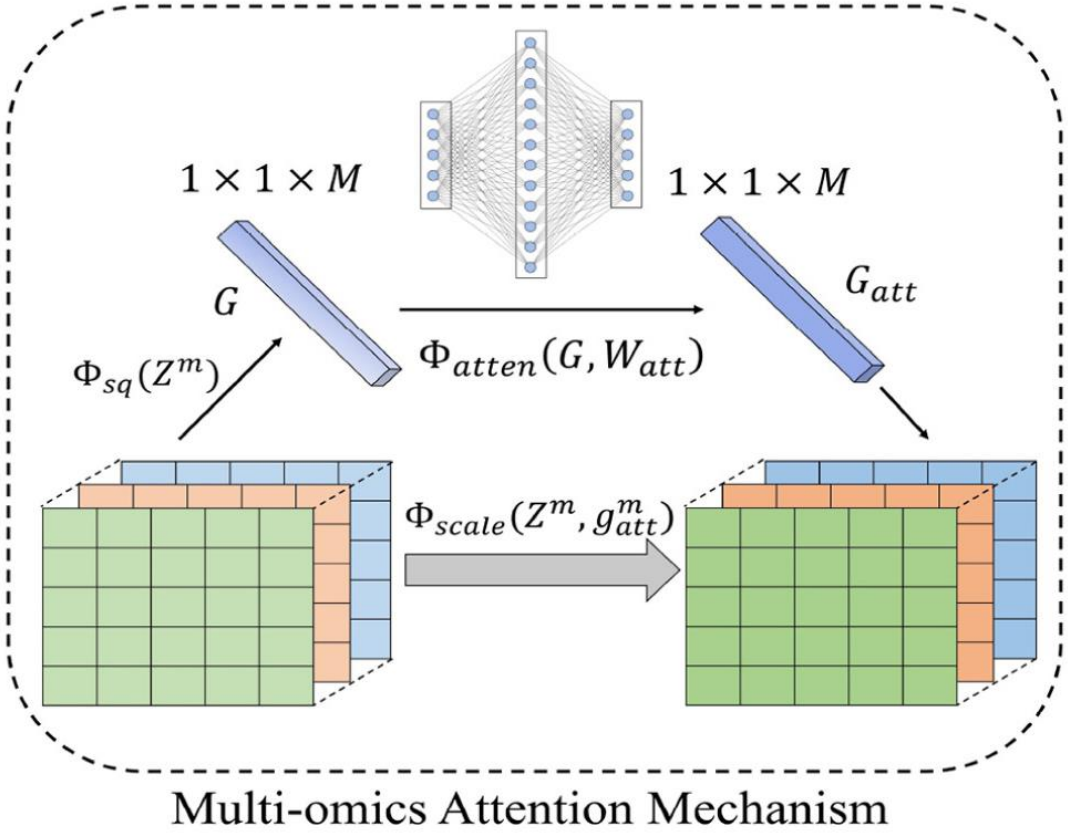
Feature selection

- Feature selection with correlation
  - Inner product regularization on the feature indicator matrix to select features with high similarity
  - This regularization makes feature indicator matrix sparser (removing uninformative features)



MOAM (Multi-omics attention mechanism)

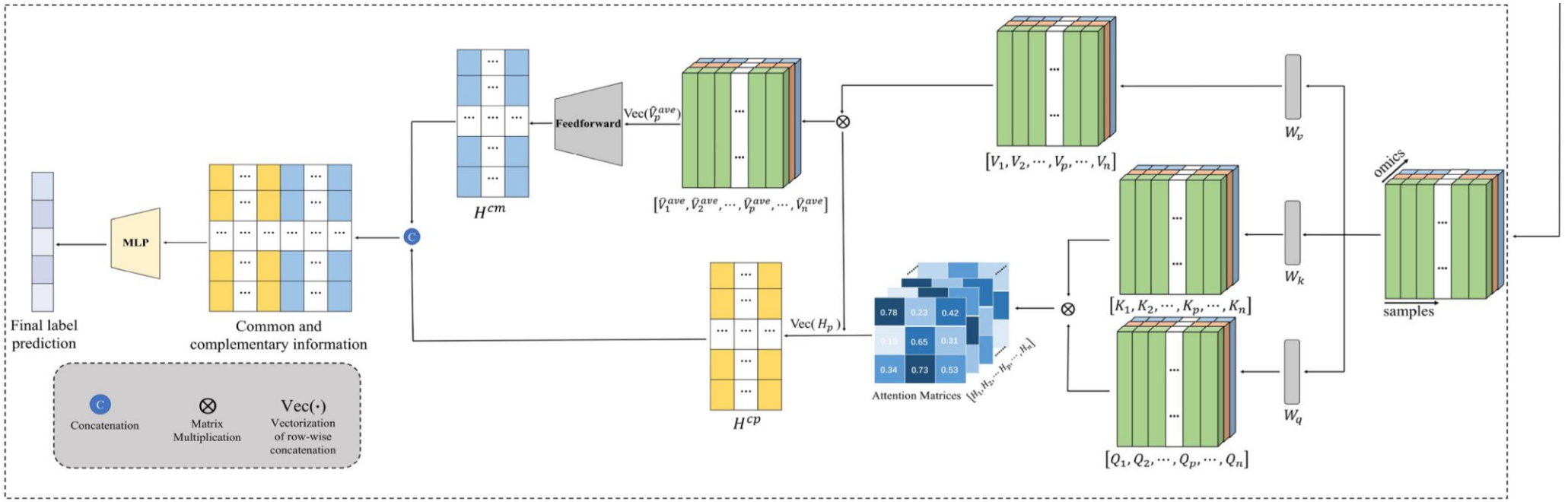
- Purpose
  - different types of omics data will generate the embedding representations of different quality and have different contributions for downstream classification tasks
- Procedure
  - Squeeze all information for each omics data (summarize all embedded values) (G)
  - Using attention mechanism to compute attention weights of embedding information (G^{att})
  - Attention weights are combined with embedding information ($\tilde{Z}^m = g_{att}^m \cdot Z^m$)



Multi-omics Attention Mechanism

OIRL (Omic-integrated representation learning)

- Purpose
  - Considering 1) commonalities between omics and 2) complementarity of different omics
- Preparation
  - Key, query, value matrix are made from embedding matrix with learnable parameters
    - E.g. Query matrix for specific patient $\mathrm{Q_p} = W_q \hat{Z}_p$



Omic-integrated Representation Learning

OIRL (Omic-integrated representation learning)

- Procedure
  - Calculate inter-omic attention matrix (H_p), which is how much concern the ith omic has for the jth omic of patient p using query value of ith omics and key values of jth omics data
  - Calculating the attention score matrix $\hat{V}_p$ by multiplying inter-omic attention matrix H_p and value matrix V_p^T for the value vector of each omics
    - Multi head attention (different type of view) can capture different perspective information
  - Common representation
    - Vectorize (row-wise concatenate) all attention score matrix, and uses them as input of 2-layer neural network
    - Output of feedforward network is omics common representation (h_p^{cm})
  - Complementary representation
    - Attention matrix itself can be regarded as more complex complementary information between omics for patient
    - Vectorize (row-wise concatenate) all attention score matrix to get complementary representation (h_p^{cp})

Optimization

- Objective function
  - Minimizing loss for both of 1) FSDGCN, and 2) prediction error of MLP model
  - There are hyper parameter μ to balance both errors

$$\mathscr{L} = \sum_{m=1}^{M} \mathscr{L}_{FSDGCN}^{(m)} + \mu \mathscr{L}_{MLP}\left(Y, \widehat{Y}_H\right)$$

  - Loss for FSDGCN also have two hyper parameters α, β to balance each sub-loss

$$\mathscr{L}_{FSDGCN}^{(m)} = \mathscr{L}_{ce}^{(m)}\left(Y, \widehat{Y}_{FSDGCN}\right) + \alpha \mathscr{L}_{gl}^{(m)}\left(S, X, A\right) + \beta \mathscr{L}_{fe}^{(m)}\left(W_f\right)$$

- Training strategy – alternate update
  1) Pretrain the omic-specific FSDGCN
  2) Fix both of MOAM and OIRL and updating minimize the loss of FSDGCN
  3) Fix FSDGCN part and update other two modules
  4) Repeat 2~3 until convergence

Result

Evaluation & Experimental details

Evaluation metrics

- accuracy (ACC)
  - # correctly predicted / # total samples

- Precision / Recall

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

- F1 score (for the binary classification)

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Predicted Value (predicted by the test)		Actual Value (as confirmed by experiment): positives	Actual Value (as confirmed by experiment): negatives
	positives	**TP** True Positive	**FP** False Positive
	negatives	**FN** False Negative	**TN** True Negative

https://devopedia.org/confusion-matrix, https://www.v7labs.com/blog/f1-score-guide

Evaluation & Experimental details

Evaluation metrics

- average F1 score weighted by support (F1_weighted)
  - Taking weighted average of f1-scores for specific label
  - Weight is proportional to number of label
- macro-averaged F1 score (F1_macro)
  - Taking simple average of f1-score for specific label

	TP	FP	FN	Precision	Recall	f1-score
A	1	2	1	0.33	0.50	0.40
				= 1 / (1+2)	= 1 / (1+1)	= 2*0.33*0.5 / (0.33+0.5)
B	1	2	2	0.33	0.33	0.33
C	3	1	2	0.75	0.60	0.67
계	5	5	5			

Evaluation & Experimental details

Dataset & Experimental detail

- KIPAN dataset for kidney cancer type classification
- SCC dataset for pan-cancer classification related to squamous cells carcinomas
- BRCA dataset for breast invasive carcinoma PAM50 subtype classification
- 5-fold CV approach was used

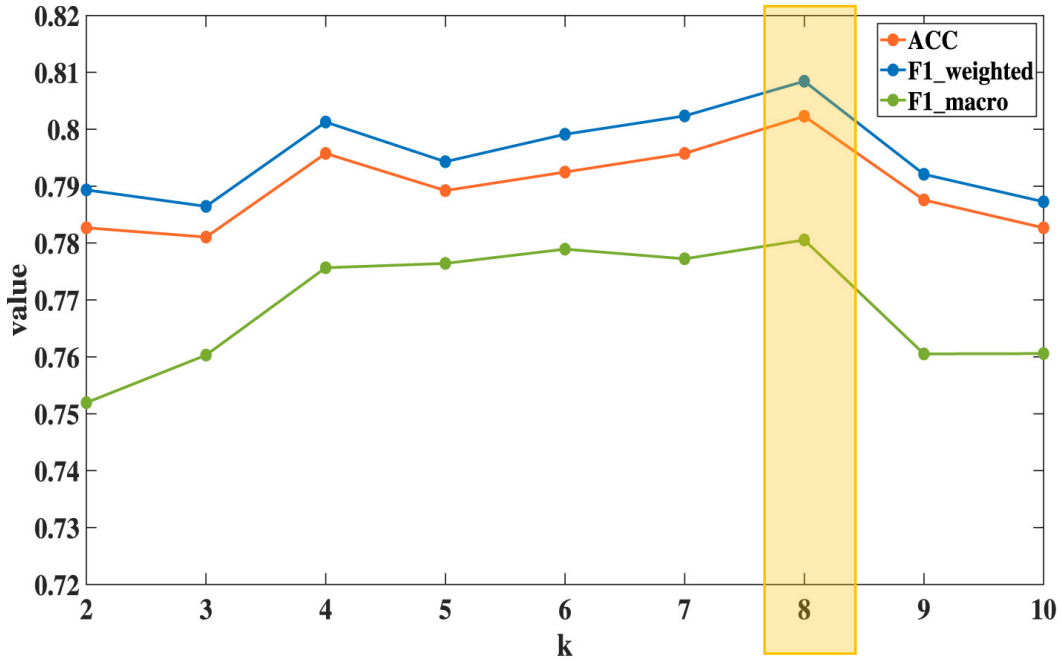
Dataset	Categories	Number of features for training mRNA, meth, miRNA
KIPAN	KICH:65, KIRC:321, KIRP:274	3000, 3000, 1535
SCC	CESC:306, HNSC:497, LUSC:365	2000, 2000, 1652
BRCA	Normal-like:115, Basal-like:131, HER2-enriched:46, Luminal A:436, Luminal B:147	1000, 1000, 503

Note: mRNA refers to mRNA expression data, meth refers to DNA methylation data, miRNA refers to miRNA expression data.

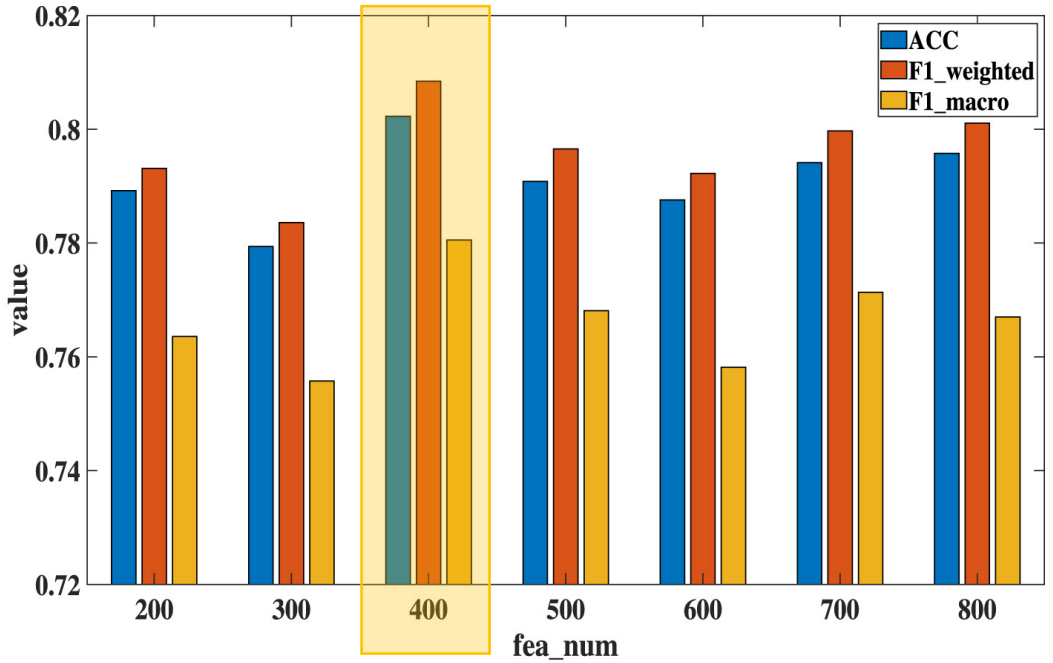
Results 1

Parameters analysis

- Choosing best hyper parameters
  - k: determines the average number of aggregated neighbor nodes for the FSDGCN module
  - fea_num: determines the embedding dimension of feature indicator matrix



(a)



(b)

Results 2

Comparison experiments

- vs. GCN methods
  - MOGLAM outperformed always outperformed MOGONET and MoGCN
- vs. machine learning methods
  - XGBoost and SMSPL worked better for simple task (e.g. pan-cancer classification)
  - However, these method seems like sensitive with data type (Low performance on BRCA data)

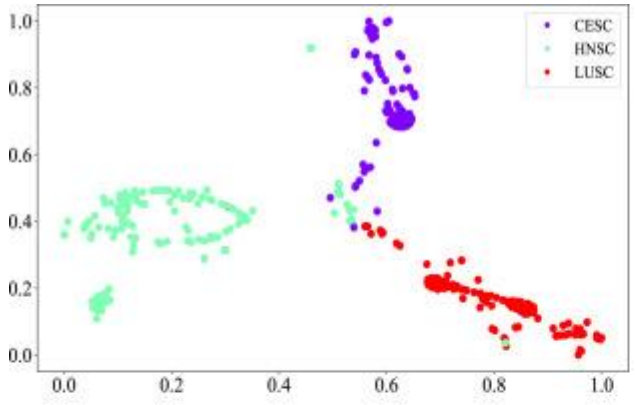
Table 1. Classification performance of all methods on BRCA dataset.

Method	ACC	F1_weighted	F1_macro
XGBoost	0.7566 ± 0.029	0.7488 ± 0.030	0.6876 ± 0.038
NN	0.7604 ± 0.018	0.7574 ± 0.019	0.7030 ± 0.027
BPLSDA	0.6234 ± 0.006	0.4906 ± 0.006	0.3074 ± 0.010
BSPLSDA	0.6266 ± 0.005	0.4938 ± 0.005	0.3146 ± 0.007
SMSPL	0.7310 ± 0.031	0.7468 ± 0.028	0.7104 ± 0.031
MOGONET	0.7886 ± 0.021	0.7740 ± 0.029	0.7254 ± 0.037
MoGCN	0.8190 ± 0.025	0.8196 ± 0.027	0.7930 ± 0.026
MMGL	0.8030 ± 0.050	0.7912 ± 0.069	0.7398 ± 0.082
MOGLAM	**0.8380 ± 0.023**	**0.8456 ± 0.022**	**0.8124 ± 0.028**

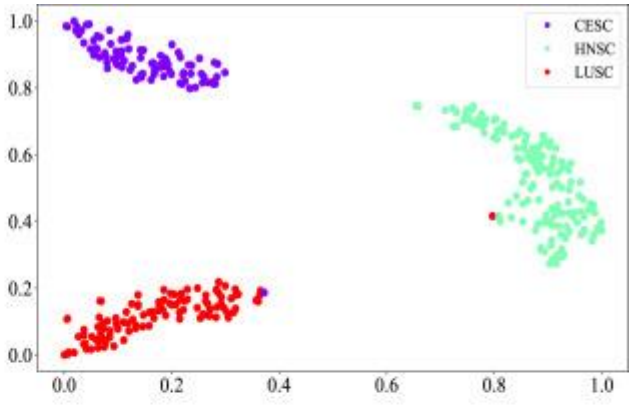
Results 3

Visualization of the embedding representation

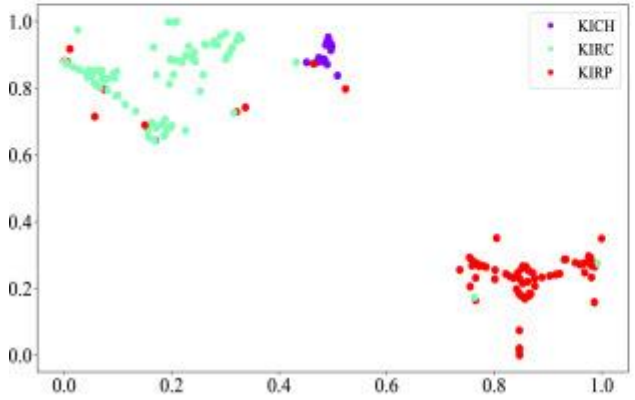
- tSNE analysis
  - MOGLAM can preserve original class
  - Also, it generates clusters with small intra-class scatter and large inter-class scatter



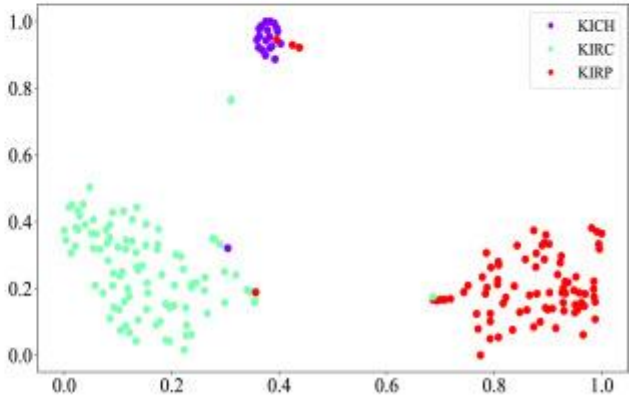
(a) MOGONET



(b) MOGLAM



(c) MOGONET



(d) MOGLAM

Results 4

Ablation studies

- Using 4 variants by removing specific module of MOGLAM
  - **(1) FSNN_MOAM_OIRL:** using fully connected network instead of GCN part in FSDGCN module
  - **(2) FSGCN_MOAM_OIRL:** removing feature selection part
  - **(3) FSDGCN_MOAM_concat:** concatenate the omics embeddings directly instead of considering both of complementary and common information
  - **(4) FSDGCN_OIRL: r**emoving multi-omics attention mecahnism

Results 4

Ablation studies

- Comparison results
  - The performance of MOGLAM was superior to other variants
  - FSDGCN_MOAM_concat works poorly, which shows simple concatenation of embedded information cannot provide meaningful insight

Table 3. The results of ablation study on BRCA dataset.

	ACC	F1_weighted	F1_macro
FSNN_MOAM_OIRL	0.8198 ± 0.004	0.8280 ± 0.004	0.7978 ± 0.005
FSGCN_MOAM_OIRL	0.8282 ± 0.028	0.8372 ± 0.027	0.8116 ± 0.033
FSDGCN_MOAM_concat	0.8122 ± 0.017	0.8194 ± 0.015	0.7914 ± 0.020
FSDGCN_OIRL	0.8342 ± 0.020	0.8420 ± 0.019	0.8060 ± 0.018
MOGLAM	**0.8380 ± 0.023**	**0.8456 ± 0.022**	**0.8124 ± 0.028**

Table 4. The results of ablation study on KIPAN dataset.

	ACC	F1_weighted	F1_macro
FSNN_MOAM_OIRL	0.9608 ± 0.008	0.9610 ± 0.008	0.9500 ± 0.012
FSGCN_MOAM_OIRL	0.9616 ± 0.010	0.9620 ± 0.010	0.9534 ± 0.014
FSDGCN_MOAM_concat	0.9606 ± 0.012	0.9610 ± 0.012	0.9518 ± 0.014
FSDGCN_OIRL	0.9618 ± 0.010	0.9620 ± 0.009	0.9528 ± 0.013
MOGLAM	**0.9650 ± 0.007**	**0.9650 ± 0.007**	**0.9566 ± 0.011**

Results 5

Important biomarkers identified by MOGLAM

- Biomarker discovery
  - Using feature indicator matrix, scores for each features were calculated and ordered
  - Top 10 markers were considered as biomarkers of each omic

Table 5. The top 10 biomarkers of each omic were selected by MOGLAM on BRCA dataset.

Omics data type	Biomarkers
mRNA expression	ESR1, SOX11, DEK, FABP7, ABCC11, C1orf106, DNMBP, ANKRD30A, AGR3, SPDEF
DNA methylation	ACSM2A, NLRP8, TKTL2, SNORA42, PIK3C2A, ATP8B1, KSR1, C1orf110, MIR128-1, ZNF516
miRNA expression	hsa-mir-375, hsa-mir-187, hsa-mir-190b, hsa-mir-29a, hsa-mir-135b, hsa-mir-25, hsa-mir-9-1, hsa-mir-577, hsa-mir-149, hsa-mir-183

Also applying further analysis (GO enrichment test, survival analysis, etc.)

Conclusion

Conclusion

- MOGLAM includes three modules: FSDGCN for omic-specific learning, MOAM to consider different importance of each omics data, OIRL to integrate the omics specific embedding with considering both of complementary and common information
- MOGLAM performs better than other GCN based method or simple machine learning based methods
- MOGLAM can also efficiently identify meaningful potential biomarkers for each omics data without additional feature selection methods

Questions

Thank you for listening