Dual-Graph Learning Convolutional Networks for Interpretable Alzheimer's Disease Diagnosis

Tingsong Xiao[1], Lu Zeng[1], Xiaoshuang Shi[1(✉)], Xiaofeng Zhu[1,2(✉)], and Guorong Wu[3,4]

[1] Department of Computer Science and Engineering, University of Electronic Science and Technology of China,Chengdu, China
{xsshi2013,seanzhuxf}@gmail.com
[2] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, China
[3] Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[4] Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract. In this paper, we propose a dual-graph learning convolutional network (dGLCN) to achieve interpretable Alzheimer's disease (AD) diagnosis, by jointly investigating subject graph learning and feature graph learning in the graph convolution network (GCN) framework. Specifically, we first construct two initial graphs to consider both the subject diversity and the feature diversity. We further fuse these two initial graphs into the GCN framework so that they can be iteratively updated (*i.e.,* dual-graph learning) while conducting representation learning. As a result, the dGLCN achieves interpretability in both subjects and brain regions through the subject importance and the feature importance, and the generalizability by overcoming the issues, such as limited subjects and noisy subjects. Experimental results on the Alzheimer's disease neuroimaging initiative (ADNI) datasets show that our dGLCN outperforms all comparison methods for binary classification. The codes of dGLCN are available on https://github.com/xiaotingsong/dGLCN.

1 Introduction

Neuroimaging techniques (*e.g.,* magnetic resonance imaging (MRI)) is an effective way to monitor the progression of AD, so they are widely used for early AD diagnosis. Machine learning methods for early AD diagnosis usually include

T. Xiao and L. Zeng—Equal contribution.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16452-1_39.

L. Wang et al. (Eds.): MICCAI 2022, LNCS 13438, pp. 406–415, 2022.
https://doi.org/10.1007/978-3-031-16452-1_39

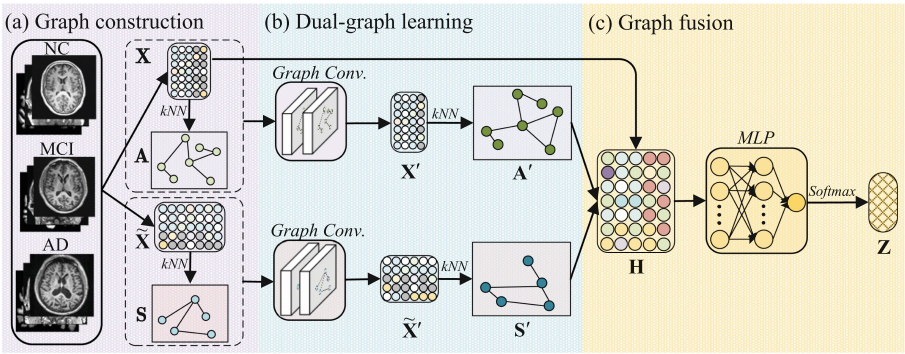


Fig. 1. The flowchart of the proposed dGLCN framework.

traditional methods and deep learning methods. Deep learning methods obtain higher diagnosis performance than traditional methods because of their powerful feature extraction capability with the help of large-scale subjects. For example, GCN generates discriminative representation for early AD diagnosis by aggregating the neighbor information of subjects [18,25,27]. However, previous methods for medical image analysis (including traditional methods and deep learning methods) still suffer from issues as follows, *i.e.,* limited and noisy subjects, and the interpretability.

Medical image analysis often suffers from the influence of limited and noisy subjects [1,20] because either subject labelling or feature extraction requires prior knowledge while experienced clinicians are always lacked. As a result, deep learning methods built on limited and noisy subjects easily result in the overfitting issue, and thus they influence the generalizability of the diagnosis model and the clinician's judgement [15]. However, few studies of medical image analysis have exploited them in the same framework.

Poor interpretability is a well-known drawback of deep learning methods. Existing methods mostly focus on post-hoc interpretation, where the interpretation model is built after the diagnosis model. In this way, the interpretation model will be re-built if the input changes. Hence, the post-hoc interpretation easily results in inflexible interpretation. Inspired by the interpretability of traditional methods, some deep learning methods employ either self-paced learning [26] or feature selection [16] to explore the subject diversity or the feature diversity. However, they usually require to have clean samples (*i.e.,* without noise) in the training process, so this might restrict their applications. In addition, these methods fail to jointly consider the subject diversity and the feature diversity in a unified framework, so that they cannot comprehensively capture the inherent correlations of the data and interpret the model.

To address the above issues, we propose to conduct dual-graph learning in the GCN framework (shown in Fig. 1), including graph construction, dual-graph learning and graph fusion. Specifically, graph construction constructs two initial

graphs, *i.e.,* the subject graph and the feature graph, from the original feature space, to consider the subject diversity and the feature diversity. Dual-graph learning separately outputs two kinds of representation of the original features by graph convolution layers and updates two initial graphs. Graph fusion fuses two updated graphs to generate the final representation of the original data for early AD diagnosis. As a result, the optimal dual-graph is captured. The updated graphs containing the subject importance and the feature importance are able to interpret the subjects and the features. Moreover, the subject graph and the feature graph, respectively, contain the correlations among the subjects and the correlations among the features. Furthermore, such correlations are learnt from the new feature space (rather than the original feature space), thereby the correct correlations among the data are captured and can improve the generalizability of the model on the dataset with limited and noisy subjects. The contributions of this paper are summarized as follows:

- Our method is the first work to consider dual-graph learning by comprehensively capturing the correlations among the data to improve the generalizability and contain interpretability. In the literature, [7] focused on graph learning on a subject graph. [29] focused on learning a fixed feature graph and [4,5] focused on constructing a fixed dual-graph. In particular, these mentioned methods seldom consider the interpretability.
- Our method is able to identify the subjects and the brain regions related to the AD. Moreover, it can be easily applied to interpret other disease diagnosis on neuroimaging data.

2 Method

2.1 Graph Construction

We select the GCN framework as our backbone. The quality of its initial graph is very important as the graph stores the correlations among the data. In particular, if the correct correlations are captured, the quality of the graph is guaranteed. Recently, it is popular to construct the subject graph, which contains the correlations among subjects to indicate the subject diversity. That is, different subjects have distinct characteristics and contributions to the diagnosis model. Moreover, the higher the correlation between two subjects is, the larger the edge weight is. In this work, we extract node features based on region-of-interests (ROIs), which have structural or functional correlations to AD [19], so it is obvious that the features (*i.e.,* ROIs) are relevant. Moreover, distinct brain regions have different influence to AD. However, to the best of our knowledge, few methods focused on taking into account the correlations among ROIs (*i.e.,* the features) and the ROI diversity, but the feature graph can achieve both of them in this paper.

The graph is seldom provided in medical image analysis, so the graph should be constructed based on the information among the data. In this work, we construct the graphs where node features are the subject/features characteristics and the edge weight measures the correlation between two subjects/features. In

particular, the higher the correlation between two subjects/features is, the larger the edge weight is. For simplicity, we employ the kNN method to construct both the subject graph $\mathbf{A} \in \mathbb{R}^{n\times n}$ and the feature graph $\mathbf{S} \in \mathbb{R}^{d\times d}$, where n and d denote the number of subjects and features.

2.2 Dual-Graph Learning

The initial graph obtained from the original feature space usually contains noise or redundancy, so its quality cannot be guaranteed. In this case, graph learning is an effective solution to improve its quality by iteratively updating it and the representation from the new feature space rather than the original feature space. As a result, graph learning is able to correctly capture the correlations among the data. For example, the graph learning convolutional network (GLCN) in [7] combines graph learning on the subject graph with the graph convolution in a unified network. In this paper, we conduct dual-graph learning to simultaneously update two initial graphs and the representation in this paper. To do this, we first separately update each graph by graph convolution layers and then fuse them for representation learning.

Subject Graph Update. Given the feature matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ and the initial subject graph $\mathbf{A}$, we employ graph convolutions to obtain

$$\mathbf{X}^{l+1} = \sigma(\mathbf{D}^{-1/2}\tilde{\mathbf{A}}\mathbf{D}^{-1/2}\mathbf{X}^{l}\mathbf{\Theta}^{l}) \tag{1}$$

where $\mathbf{X}^{l} \in \mathbb{R}^{n\times d_l}$ is a d_l-dimensional representation in the l-th layer and σ denotes the activation function. $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$, where $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix, $\mathbf{D}$ and $\mathbf{\Theta}^{l}$, respectively, represents the degree matrix of $\tilde{\mathbf{A}}$ and the trainable parameters of the l-th layer. After obtaining the new representation $\mathbf{X}'$, we employ the kNN method to generate a subject graph $\mathbf{A}' \in \mathbb{R}^{n\times n}$, *i.e.,* the update of the initial subject graph $\mathbf{A}$.

Feature Graph Update. Given the feature matrix $\mathbf{X}$, we obtain its transpose as $\tilde{\mathbf{X}} = \mathbf{X}^{T} \in \mathbb{R}^{d\times n}$. In this paper, we design the feature graph $\mathbf{S} \in \mathbb{R}^{d\times d}$ to explore the correlations among features as AD is influenced by the complex correlations among ROIs (*i.e.,* features) [21]. Specifically, we regard the n-dimensional representation as the characteristics of the node (*i.e.,* the ROI information) and the edge weight as the correlation between two n-dimensional representation. After the two-layer graph convolution, the new representation of $\tilde{\mathbf{X}}$ is

$$\tilde{\mathbf{X}}^{l+1} = \sigma(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{S}}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{X}}^{l}\tilde{\mathbf{\Theta}}^{l}) \tag{2}$$

where $\tilde{\mathbf{X}}^{l} \in \mathbb{R}^{d\times n_l}$ is a n_l-dimensional representation in the l-th layer. $\tilde{\mathbf{S}} = \mathbf{S}+\mathbf{I}_d$, where $\mathbf{I}_d \in \mathbb{R}^{d\times d}$ is the identity matrix, $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{\Theta}}^{l}$, respectively, represent the degree matrix of $\tilde{\mathbf{S}}$ and the trainable parameters of the l-th layer. After obtaining the new representation $\tilde{\mathbf{X}}'$, we utilize kNN method to generate a new feature graph $\mathbf{S}' \in \mathbb{R}^{d\times d}$, *i.e.,* the update of the initial feature graph $\mathbf{S}$.

2.3 Graph Fusion and Objective Function

Equation (1) and Eq. (2) learn the new representations of the original graph, which are further utilized to construct the subject graph and the feature graph separately. Meanwhile, the two graphs output the subject importance and the feature importance, respectively. In particular, the subject importance contains the weight of every subject while the feature importance involves the weight of every feature. Such importance can be used to interpret subjects and ROIs as well as can be fused into the original feature matrix for representation learning. To do this, we combine them (*i.e.,* $\mathbf{A}'$ and $\mathbf{S}'$) with $\mathbf{X}$ to have $\mathbf{H} \in \mathbb{R}^{n \times d}$ as

$$\mathbf{H} = \mathbf{A}'\mathbf{X}\mathbf{S}' \tag{3}$$

Compared with the original feature matrix $\mathbf{X}$, the new feature matrix $\mathbf{H}$ takes into account both the subject importance and the feature importance. Inspired by [12] and [11], we define $\mathbf{H}$ as follows by adding the original graphs (*i.e.,* $\mathbf{A}$ and $\mathbf{S}$) so that the updated graphs (*i.e.,* $\mathbf{A}'$ and $\mathbf{S}'$) have tiny variation in every iteration to achieve robust classification model.

$$\mathbf{H} = (\mathbf{A}' + \lambda_1 \mathbf{A})\mathbf{X}(\mathbf{S}' + \lambda_2 \mathbf{S}) \tag{4}$$

where λ_1 and λ_2 are parameters. After that, the new representation can be obtained by the MLP

$$\mathbf{H}^{l+1} = \sigma(\mathbf{H}^l \mathbf{W}^l) \tag{5}$$

where the $\mathbf{H}^l \in \mathbb{R}^{n \times d_l}$ denotes the representation of the l-th layer while $\mathbf{H^0} = \mathbf{H}$. $\mathbf{W}^l \in \mathbb{R}^{d_l \times d_{l+1}}$ is the MLP parameters in the l-th layer.

After conducting the MLP, the output matrix $\mathbf{H}^L$ is derived. We further use the softmax function to obtain the label prediction $\mathbf{Z} = [\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_{n-1}] \in \mathbb{R}^{n \times c}$, where $\mathbf{z}_i \in \mathbb{R}^c$ denotes the label prediction for the i-th subject, *i.e.,*

$$z_{ij} = \text{softmax}\left(h_{ij}^L\right) = \frac{\exp\left(h_{ij}^L\right)}{\sum_{m \in c} \exp\left(h_{im}^L\right)}. \tag{6}$$

where h_{ij} and z_{ij} represent the elements in the i-th row and the j-th column of $\mathbf{H}$ and $\mathbf{Z}$, respectively. The cross-entropy function is used to calculate the loss.

$$\mathcal{L}_{ce} = -\sum_{i \in N} \sum_{j=1}^{c} y_{ij} \ln z_{ij} \tag{7}$$

where y_{ij} is the element of the i-th row and the j-th column of the real label $\mathbf{Y}$.

We obtain optimal $\mathbf{A}'$ and $\mathbf{S}'$, where every element in $\mathbf{A}'$ represents the correlations between two subjects and every element in $\mathbf{S}'$ denotes the correlations between two ROIs. Moreover, they are symmetric matrices. Following [30], we first calculate the ℓ_2-norm value of every row in $\mathbf{S}'$ and then rank their ℓ_2-norm values. In this way, the features whose corresponded rows in $\mathbf{S}'$ with the top ℓ_2-norm values are regarded as important features. We can also obtain the important subjects by calculating the ℓ_2-norm value of every row in $\mathbf{A}'$. We did not follow [17] to set a sparse constraint on $\mathbf{S}'$ as the two ways output similar results in terms of feature importance while [17] is time-consuming.

3 Experiments

3.1 Experimental Settings

Our dataset from the ADNI (www.loni.usc.edu/ADNI) includes 186 ADs, 393 mild cognitive impairment patients (MCIs) and 226 normal controls (NCs). Moreover, 393 MCIs include 226 MCI non-converts (MCIn) and 167 MCI converts (MCIc). We use them to form four binary classification tasks, *i.e.,* AD-NC (186 vs. 226), AD-MCI (186 vs. 393), NC-MCI (226 vs. 393) and MCIn-MCIc (226 vs. 167) on the MRI data. The MRI data is first dealt with by the steps, *i.e.,* spatial distortion correction, skull-stripping, and cerebellum removal, and then segmented into gray matter, white matter, and cerebrospinal fluid. Finally, we use the AAL template [23] to obtain 90 ROIs for every subject.

The comparison methods include three traditional methods (*i.e.,* ℓ_1-norm support vector machines (ℓ_1-SVM) [28], self-paced learning (SPL) [10] and boosted random forest (BRF) [13]) and five deep methods (*i.e.,* graph convolutional networks (GCN) [9], graph attention networks (GAT) [24], dual graph convolutional networks (DGCN) [31], interpretable dynamic graph convolutional networks (IDGCN) [30] and sample reweight (SR) [22]).

We obtain the author-verified codes for all comparison methods and let them to achieve their best performance. Since the datasets used in this experiment do not provide a predefined subject graph, we construct it with the kNN method by setting $k = 30$. To avoid the over-fitting issue on the datasets with limited subjects, in all experiments, we repeat the 5-fold cross-validation scheme 100 times with different random seeds on all datasets for all methods. We finally report the average results and the corresponding standard deviation (std). We adopt four commonly used metrics to evaluate all methods, including classification accuracy, sensitivity, specificity and AUC.

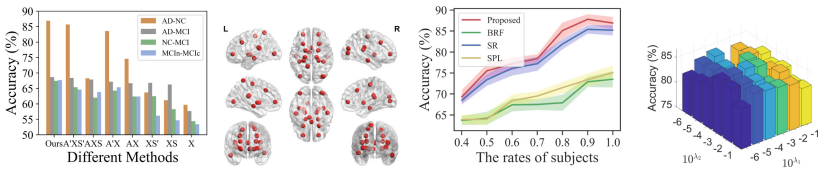
3.2 Result Analysis

Table 1 shows the classification results of all methods on all datasets. First, our method achieves the best results, followed by SR, GCN, IDGCN, GAT, DGCN, BRF, SPL and ℓ_1-SVM, on all datasets for four evaluation metrics. Moreover, our improvement has statistically significant difference (with $p \leq 0.05$ by t-test) from every comparison method. For example, our method on average improves by 3.2% and 6.9%, respectively, compared to the best comparison method, *i.e.,* SR, and the worst comparison method, *i.e.,* ℓ_1-SVM. It indicates that it is feasible for our method to take into account both the subject diversity and the feature diversity as they benefit our method to comprehensively capture the complex correlations among the data, and thus improving the generalizability. Second, all deep learning methods (*i.e.,* SR, GCN, IDGCN, GAT, DGCN and our method) outperform traditional methods, *i.e.,* BRF, SPL and ℓ_1-SVM. For example, our method on average improves by 6%, compared to the best traditional method (*i.e.,* BRF), on all datasets in terms of four evaluation metrics. The reason is that (1) traditional methods are worse for representation learning than deep learning

methods and (2) deep learning methods are still effective on the datasets with limited subjects and our method achieves the best generalizability.

Table 1. Classification results (average ± std) of all methods on four datasets.

Dataset	Metrics	ℓ_1-SVM	SPL	BRF	GCN	GAT	DGCN	IDGCN	SR	Proposed
AD-NC	ACC	$74.5_{\pm 2.9}$	$75.1_{\pm 1.3}$	$73.5_{\pm 2.7}$	$82.5_{\pm 1.8}$	$83.9_{\pm 2.0}$	$79.6_{\pm 3.0}$	$83.3_{\pm 2.5}$	$82.8_{\pm 2.7}$	$\mathbf{86.9}_{\pm 1.6}$
	SEN	$76.8_{\pm 1.1}$	$81.5_{\pm 1.5}$	$79.7_{\pm 2.3}$	$82.9_{\pm 1.7}$	$84.1_{\pm 1.1}$	$74.4_{\pm 3.3}$	$82.6_{\pm 2.4}$	$85.2_{\pm 3.3}$	$\mathbf{85.9}_{\pm 1.6}$
	SPE	$73.7_{\pm 1.3}$	$70.7_{\pm 2.0}$	$66.7_{\pm 2.0}$	$83.7_{\pm 2.7}$	$79.1_{\pm 1.5}$	$85.8_{\pm 2.4}$	$83.2_{\pm 3.0}$	$78.4_{\pm 1.3}$	$\mathbf{87.2}_{\pm 1.7}$
	AUC	$74.7_{\pm 1.5}$	$72.2_{\pm 1.2}$	$73.2_{\pm 2.6}$	$84.3_{\pm 2.5}$	$81.6_{\pm 1.9}$	$75.2_{\pm 2.9}$	$83.5_{\pm 2.8}$	$80.1_{\pm 2.8}$	$\mathbf{90.6}_{\pm 2.5}$
AD-MCI	ACC	$66.4_{\pm 3.3}$	$67.1_{\pm 2.3}$	$65.3_{\pm 2.2}$	$67.3_{\pm 2.0}$	$67.2_{\pm 2.2}$	$68.0_{\pm 2.7}$	$66.8_{\pm 2.4}$	$67.6_{\pm 2.8}$	$\mathbf{68.7}_{\pm 1.5}$
	SEN	$66.7_{\pm 4.2}$	$63.6_{\pm 2.2}$	$67.6_{\pm 2.4}$	$67.4_{\pm 3.1}$	$65.2_{\pm 1.9}$	$60.6_{\pm 2.5}$	$64.2_{\pm 2.5}$	$67.2_{\pm 2.6}$	$\mathbf{68.5}_{\pm 1.6}$
	SPE	$65.5_{\pm 2.5}$	$72.1_{\pm 2.5}$	$63.6_{\pm 2.6}$	$65.8_{\pm 2.2}$	$75.2_{\pm 3.7}$	$\mathbf{77.3}_{\pm 3.1}$	$74.4_{\pm 1.7}$	$76.0_{\pm 1.4}$	$74.8_{\pm 2.4}$
	AUC	$59.7_{\pm 3.0}$	$60.3_{\pm 1.3}$	$62.1_{\pm 2.3}$	$62.4_{\pm 1.5}$	$60.6_{\pm 2.4}$	$62.0_{\pm 2.1}$	$61.6_{\pm 2.6}$	$60.6_{\pm 1.7}$	$\mathbf{65.5}_{\pm 2.8}$
NC-MCI	ACC	$61.5_{\pm 1.7}$	$63.9_{\pm 1.4}$	$62.4_{\pm 2.4}$	$62.4_{\pm 1.7}$	$64.7_{\pm 1.7}$	$63.5_{\pm 3.1}$	$64.2_{\pm 2.6}$	$64.4_{\pm 2.5}$	$\mathbf{67.5}_{\pm 2.4}$
	SEN	$57.5_{\pm 2.2}$	$59.1_{\pm 1.7}$	$61.4_{\pm 1.7}$	$58.7_{\pm 2.4}$	$55.4_{\pm 1.6}$	$61.4_{\pm 1.8}$	$58.0_{\pm 2.5}$	$61.2_{\pm 1.9}$	$\mathbf{61.7}_{\pm 1.5}$
	SPE	$62.3_{\pm 2.5}$	$64.5_{\pm 2.0}$	$65.9_{\pm 2.4}$	$64.3_{\pm 1.5}$	$65.7_{\pm 2.5}$	$58.2_{\pm 1.4}$	$64.6_{\pm 2.3}$	$65.6_{\pm 2.2}$	$\mathbf{67.7}_{\pm 1.3}$
	AUC	$59.5_{\pm 2.3}$	$60.2_{\pm 2.8}$	$65.6_{\pm 2.4}$	$65.6_{\pm 2.2}$	$62.4_{\pm 2.8}$	$60.8_{\pm 2.1}$	$58.5_{\pm 1.9}$	$66.4_{\pm 3.3}$	$\mathbf{69.2}_{\pm 2.8}$
MCIn-MCIc	ACC	$62.1_{\pm 2.4}$	$63.8_{\pm 2.6}$	$63.7_{\pm 3.1}$	$64.4_{\pm 1.6}$	$63.9_{\pm 2.4}$	$62.6_{\pm 1.8}$	$63.9_{\pm 2.6}$	$63.7_{\pm 1.8}$	$\mathbf{67.7}_{\pm 2.4}$
	SEN	$65.9_{\pm 2.5}$	$61.1_{\pm 2.1}$	$67.8_{\pm 2.3}$	$\mathbf{67.9}_{\pm 1.9}$	$66.6_{\pm 1.4}$	$62.2_{\pm 2.2}$	$62.9_{\pm 1.6}$	$61.6_{\pm 1.7}$	$64.0_{\pm 2.5}$
	SPE	$63.0_{\pm 2.7}$	$64.9_{\pm 3.3}$	$62.2_{\pm 3.5}$	$61.9_{\pm 2.6}$	$55.9_{\pm 2.2}$	$67.9_{\pm 2.8}$	$66.5_{\pm 1.6}$	$66.0_{\pm 1.4}$	$\mathbf{69.9}_{\pm 3.1}$
	AUC	$59.8_{\pm 3.7}$	$62.2_{\pm 3.2}$	$62.6_{\pm 3.3}$	$62.8_{\pm 2.0}$	$61.2_{\pm 2.7}$	$61.7_{\pm 1.6}$	$62.3_{\pm 2.4}$	$62.5_{\pm 2.5}$	$\mathbf{63.7}_{\pm 3.0}$



(a) (b) (c) (d)

Fig. 2. Results of (a) Ablation study of 8 methods, (b) Top 20 ROIs selected by our method on AD-NC, (c) Subject interpretability of our method on AD-NC, and (d) our method with different parameter settings (*i.e.,* λ_1 and λ_2) on AD-NC.

3.3 Ablation Analysis

We use Eq. (4) (*i.e.,* $\mathbf{H} = (\mathbf{A}' + \lambda_1 \mathbf{A})\mathbf{X}(\mathbf{S}' + \lambda_2 \mathbf{S})$) to make Eq. (3) have a tiny variation in every iteration, so we generate 7 comparison methods to investigate the effectiveness of both the subject importance and the feature importance, *i.e.,* $\mathbf{H} = \mathbf{A}'\mathbf{X}\mathbf{S}'$, $\mathbf{H} = \mathbf{A}\mathbf{X}\mathbf{S}$, $\mathbf{H} = \mathbf{A}'\mathbf{X}$, $\mathbf{H} = \mathbf{A}\mathbf{X}$, $\mathbf{H} = \mathbf{X}\mathbf{S}'$, $\mathbf{H} = \mathbf{X}\mathbf{S}$, and $\mathbf{H} = \mathbf{X}$. We list the classification results of all eight methods in Fig. 2.(a) and Appendix A. First, dual-graph learning is very important in our method as the subject graph $\mathbf{A}'$ plays a greater role than the initial subject graph $\mathbf{A}$. Moreover, the feature graph has the same scenario. For example, the method of $\mathbf{A}'\mathbf{X}$ improves by 9% compared to that of $\mathbf{A}\mathbf{X}$, while the method of $\mathbf{X}\mathbf{S}'$ outperforms about 3%, compared to that of $\mathbf{X}\mathbf{S}$, in terms of classification accuracy, on dataset

AD-NC. Second, dual-graph learning is more effective than graph learning on one graph as more graphs could find more complex correlations among the data. Third, it is of great importance to add a portion of the initial graphs into the process of graph learning, mentioned in [3].

3.4 Interpretability

Feature Interpretability. We compare our method with three methods (*i.e.,* ℓ_1-SVM, BRF and IDGCN) to investigate their feature interpretability. To do this, since we repeat the 5-fold cross validation scheme 100 times to obtain 500 matrices of $\mathbf{S}'$, we select the top 20 features (*i.e.,* ROIs) from every experiment to obtain the frequency of every feature within 500 times. We list the ROIs selected by every method on all four datasets (*i.e.,* the third column) as well as the ROIs selected by all methods on all datasets (*i.e.,* the second column) in Appendix B and visualize top 20 ROIs selected by our method in Fig. 2.(b) and Appendix B. Obviously, ROIs selected by every method on all datasets are related to AD. For example, the hippocampal formation right is selected by all methods on all datasets and has been shown to be highly related to AD in [2]. In addition, compared to all comparison methods, our method selects much more ROIs, *e.g.,* the frontal lobe white matter and temporal lobe white matter, which are highly related to AD in [6,8,14]. Above observations suggest that our method is superior to all comparison methods, in terms of feature interpretability.

Subject Interpretability. We investigate the subject interpretability by evaluating the classification performance with different training rates. To do this, we first employ four methods with subject interpretability (*i.e.,* SPL, BRF, SR and Proposed) to output the subject importance within 500 experiments and then rank their importance in a descending order. With such an order, we increase the proportion of the training subjects from 40% to 100% and report the classification accuracy in Fig. 2.(c) and Appendix C. Obviously, all four methods achieve the highest accuracy when the training rates are between 70% and 90%. This indicates that there are noisy subjects in the datasets to affect the training process, and these methods can overcome this issue by learning the subject importance. In addition, our method always achieves the best results on all datasets with different rates as it is the most effective method to identify the noisy subjects, compared to all comparison methods.

3.5 Parameter Sensitivity Analysis

We investigate the parameter sensitivity by setting $k \in \{5, 10, ..., 35\}$, and $\lambda_1, \lambda_2 \in \{10^{-6}, 10^{-5}, ..., 10^{-1}\}$ and report the results in Fig. 2.(d) and Appendix D. First, our method is sensitive to the settings of k, but it achieves good performance while setting $k = 30$ on all datasets. Second, our method achieves the best performance with $\lambda_1 = 10^{-5}$ and $\lambda_2 = 10^{-3}$ on AD-NC, while is insensitive to them on other datasets. These cases are consistent with [3] of adding a portion of the initial graph into the process of graph learning.

4 Conclusion

In this paper, we propose a dual-graph learning method in the GCN framework to achieve the generalizability and the interpretability for medical image analysis. To do this, we consider the subject diversity and the feature diversity to conduct subject graph learning and feature graph learning in the same framework. Experimental results on the ADNI verify the effectiveness of our proposed method, compared to the state-of-the-art methods. In medical image analysis, imbalanced datasets are very common, *e.g.,* the number of normal controls (*i.e.,* majority class) is larger than the number of patients (*i.e.,* minority class), in the future, we will extend our method for interpretable medical image analysis on imbalanced datasets with limited subjects.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (Grant No. 61876046), Medico-Engineering Cooperation Funds from University of Electronic Science and Technology of China (No. ZYGX2022YGRH009 and ZYGX2022YGRH014) and the Guangxi "Bagui" Teams for Innovation and Research, China.

References

1. Adeli, E., Li, X., Kwon, D., Zhang, Y., Pohl, K.M.: Logistic regression confined by cardinality-constrained sample and feature selection. IEEE Trans. Pattern Anal. Mach. Intell. **42**(7), 1713–1728 (2019)
2. Beal, M.F., Mazurek, M.F., Tran, V.T., Chattha, G., Bird, E.D., Martin, J.B.: Reduced numbers of somatostatin receptors in the cerebral cortex in Alzheimer's disease. Science **229**(4710), 289–291 (1985)
3. Chen, Y., Wu, L., Zaki, M.: Iterative deep graph learning for graph neural networks: better and robust node embeddings. In: NeurIPS, pp. 19314–19326 (2020)
4. Feng, J., et al.: Dual-graph convolutional network based on band attention and sparse constraint for hyperspectral band selection. Knowl.-Based Syst. **231**, 107428 (2021)
5. Fu, X., Qi, Q., Zha, Z.J., Zhu, Y., Ding, X.: Rain streak removal via dual graph convolutional network. In: AAAI, pp. 1–9 (2021)
6. Ihara, M., et al.: Quantification of myelin loss in frontal lobe white matter in vascular dementia, Alzheimer's disease, and dementia with Lewy bodies. Acta Neuropathol. **119**(5), 579–589 (2010)
7. Jiang, B., Zhang, Z., Lin, D., Tang, J., Luo, B.: Semi-supervised learning with graph learning-convolutional networks. In: CVPR, pp. 11313–11320 (2019)
8. Karas, G., et al.: Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. Neuroradiology **49**(12), 967–976 (2007)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)
10. Kumar, M.P., Packer, B., Koller, D.: Self-paced learning for latent variable models. In: NeurIPS, pp. 1189–1197 (2010)
11. Liu, W., He, J., Chang, S.F.: Large graph construction for scalable semi-supervised learning. In: ICML (2010)

12. Liu, X., Lei, F., Xia, G.: MulStepNET: stronger multi-step graph convolutional networks via multi-power adjacency matrix combination. J. Ambient Intell. Human Comput., 1–10 (2021). https://doi.org/10.1007/s12652-021-03355-x
13. Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., Fujiyoshi, H.: Boosted random forest. IEICE Trans. Inf. Syst. **98**(9), 1630–1636 (2015)
14. Mizuno, Y., Ikeda, K., Tsuchiya, K., Ishihara, R., Shibayama, H.: Two distinct subgroups of senile dementia of Alzheimer type: quantitative study of neurofibrillary tangles. Neuropathology **23**(4), 282–289 (2003)
15. Morgado, P.M., Silveira, M., Alzheimer's Disease Neuroimaging Initiative, et al.: Minimal neighborhood redundancy maximal relevance: application to the diagnosis of Alzheimer s disease. Neurocomputing **155**, 295–308 (2015)
16. Muñoz-Romero, S., Gorostiaga, A., Soguero-Ruiz, C., Mora-Jiménez, I., Rojo-Álvarez, J.L.: Informative variable identifier: expanding interpretability in feature selection. Pattern Recogn. **98**, 107077 (2020)
17. Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and robust feature selection via joint 2, 1-norms minimization. In: NeurIPS, pp. 1813–1821 (2010)
18. Parisot, S., et al.: Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. **48**, 117–130 (2018)
19. Petersen, R.C., et al.: Memory and MRI-based hippocampal volumes in aging and AD. Neurology **54**(3), 581 (2000)
20. Qiu, S., et al.: Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain **143**(6), 1920–1933 (2020)
21. Reijmer, Y.D., et al.: Disruption of cerebral networks and cognitive impairment in Alzheimer disease. Neurology **80**(15), 1370–1377 (2013)
22. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. In: ICML, pp. 4334–4343 (2018)
23. Tzourio-Mazoyer, N., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage **15**(1), 273–289 (2002)
24. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: ICLR (2018)
25. Wang, C., Samari, B., Siddiqi, K.: Local spectral graph convolution for point set feature learning. In: ECCV, pp. 52–66 (2018)
26. Yun, Y., Dai, H., Cao, R., Zhang, Y., Shang, X.: Self-paced graph memory network for student GPA prediction and abnormal student detection. In: Roll, I., McNamara, D., Sosnovsky, S., Luckin, R., Dimitrova, V. (eds.) AIED 2021. LNCS (LNAI), vol. 12749, pp. 417–421. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78270-2_74
27. Zeng, L., Li, H., Xiao, T., Shen, F., Zhong, Z.: Graph convolutional network with sample and feature weights for Alzheimer's disease diagnosis. Inf. Process. Manag. **59**(4), 102952 (2022)
28. Zhu, J., Rosset, S., Tibshirani, R., Hastie, T.J.: 1-norm support vector machines. In: NeurIPS, pp. 49–56 (2003)
29. Zhu, X., et al.: A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. Med. Image Anal. **38**, 205–214 (2017)
30. Zhu, Y., Ma, J., Yuan, C., Zhu, X.: Interpretable learning based dynamic graph convolutional networks for Alzheimer's disease analysis. Inf. Fusion **77**, 53–61 (2022)
31. Zhuang, C., Ma, Q.: Dual graph convolutional networks for graph-based semi-supervised classification. In: WWW, pp. 499–508 (2018)