UNIVERSITY OF LEEDS

This is a repository copy of *Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/178624/

Version: Accepted Version

Article:
Song, X, Zhou, F, Frangi, AF orcid.org/0000-0002-2675-528X et al. (5 more authors) (2021) Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction. Medical Image Analysis, 69. 101947. ISSN 1361-8415

https://doi.org/10.1016/j.media.2020.101947

© 2020, Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

1

Graph Convolution Network with Similarity Awareness and Adaptive Calibration for Disease-induced Deterioration Prediction

Xuegang Song[a], Feng Zhou[b], Alejandro F Frangi[a, c, d, e], *Fellow, IEEE*, Jiuwen Cao[f], Xiaohua Xiao[g], Yi Lei[g], Tianfu Wang[a], Baiying Lei[a]*

[a] *National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Centre, Shenzhen University, Shenzhen, China, 518060.*

[b] *Department of Industrial and Manufacturing, Systems Engineering, The University of Michigan, Dearborn, MI, USA,42185*

[c] *CISTIB Centre for Computational Imaging & Simulation Technologies in Biomedicine, School of Computing, University of Leeds, Leeds LS2 9LU, United Kingdom*

[d] *LICAMM Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, Leeds LS2 9LU, United Kingdom*

[e] *Medical Imaging Research Center (MIRC) – University Hospital Gasthuisberg, KU Leuven, Herestraat 49, 3000 Leuven. Belgium*

[f] *Artificial Intelligence Institute, Hangzhou Dianzi University, Zhejiang, China, 310010*

[g] First Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen University, Shenzhen, China,518050

Tel: 86-755-26534314. Fax: 86-755-26534940, Correspondence Baiying Lei (e-mail:leiby@szu.edu.cn)

Abstract

Graph convolution networks (GCN) have been successfully applied in disease prediction tasks as they capture interactions (i.e., edges and edge weights on the graph) between individual elements. The interactions in existing works are constructed by fusing similarity between imaging information and distance between non-imaging information, whereas disregarding the disease status of those individuals in the training set. Besides, the similarity is being evaluated by computing the correlation distance between feature vectors, which limits prediction performance, especially for predicting significant memory concern (SMC) and mild

cognitive impairment (MCI). In this paper, we propose three mechanisms to improve GCN, namely similarity-aware adaptive calibrated GCN (SAC-GCN), for predicting SMC and MCI. First, we design a similarity-aware graph using different receptive fields to consider disease status. The labelled subjects on the graph are only connected with those labelled subjects with the same status. Second, we propose an adaptive mechanism to evaluate similarity. Specifically, we construct initial GCN with evaluating similarity by using traditional correlation distance, then pre-train the initial GCN by using training samples and use it to score all subjects. Then, the difference between these scores replaces correlation distance to update similarity. Last, we devise a calibration mechanism to fuse functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) information into edges. The proposed method is tested on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Experimental results demonstrate that our proposed method is useful to predict disease-induced deterioration and superior to other related algorithms, with a mean classification accuracy of 86.83% in our prediction tasks.

Key words: Disease prediction, Graph convolution network, Similarity awareness, Adaptive mechanism, Calibration mechanism, Dual-modal information

1. Introduction

Alzheimer's disease (AD) is a severe brain disorder, which is yet incurable, and no effective medicine exists for now (Association, 2018; Wang et al., 2013). The early stage of AD, i.e., mild cognitive impairment (MCI), has an annual 10%-15% conversion rate and an over 50% conversion rate within 5 years to AD (Hampel and Lista, 2016). In MCI stages, with specific cognitive training and pharmacological treatment, the deterioration process can be delayed or stopped (Gauthier et al., 2006). Therefore, it is essential to detect MCI and its earlier stage, significant memory concerns (SMC). However, the accurate disease prediction of SMC and MCI is still a challenging task due to their subtle differences in neuroimaging features (Li et al., 2019b; Wee et al., 2014; Zhang et al., 2018).

To overcome the limitation of subtle differences in neuroimaging features, it is increasingly popular to use multi-modal data to describe or strengthen features from multiple sources (Lei et al., 2020; Li et al., 2019a, 2020b; Tong et al., 2017; Zhu et al., 2019). For example, Zhu et al. (2019) proposed a multi-modal rank minimisation method to combine magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF). They then predicted AD with a linear regression classifier. Experi-

mental results showed that the classification accuracy based on the above three modalities increased by 6% compared to that based on CSF. Li et al. (2019a) proposed a sparse regression algorithm for inference of the integrated hyper-connectivity networks from BOLD functional MRI (fMRI) and arterial spin labelling (ASL). Finally, they used a support vector machine (SVM) to predict MCI. Experimental results showed that the classification accuracy based on the above two modalities increased by 11.5% compared to that based on BOLD fMRI. Integrating fMRI and diffusion tensor imaging (DTI) is shown to achieve good performance by integrating their complementary cues (Lei et al., 2020; Li et al., 2020b). Lei et al. (2020) developed a multi-task learning method to select features from fMRI functional and DTI structural brain networks, and then the selected features were sent into an SVM for final prediction. Experimental results showed that the classification accuracy based on fMRI and DTI data increased by 3.76% compared to that based on fMRI data. Li et al. (2020b) used the DTI tractography as penalty parameters in an ultra-weighted-lasso algorithm to construct more accurate fMRI functional brain networks and finally used SVM for prediction. Experimental results showed that the classification accuracy based on fMRI and DTI data increased by 5.5% compared to that based on fMRI data. These works show that the performance of using multi-modal neuroimaging is better than using single modal neuroimaging for disease prediction. However, these studies were limited to use traditional machine learning methods for feature learning or as a classifier, which limited their performance to some extent.

As a deep learning method, graph convolution network (GCN) has witnessed great success in disease prediction recently (Kazi et al., 2019; Ktena et al., 2018; Parisot et al., 2018; Zhang et al., 2019), which is based on the graph theory (Bapat et al., 2010). On a graph, a node represents a subject's data, and the edges establish connections between each pair of nodes. Parisot et al. (2018) integrated similarity between imaging information and distance between phenotypic information (e.g., gender, equipment type, and ages) into edges for the prediction of Autism Spectrum Disorder (ASD) and conversion to AD. Kazi et al. (2019) designed different kernel sizes in spectral convolution to learn cluster-specific features for predicting MCI and AD. Experimental results showed that their method performed better when the classes had large and different variances. All these studies validate the effectiveness of GCN and show its convolution operation is the key to prediction performance.

However, there are still limitations in the effectiveness of multi−modal fusion and GCNs. First, existing

GCN studies (Kazi et al., 2019; Kipf and Welling, 2017; Ktena et al., 2018; Parisot et al., 2018; Zhang et al., 2019) for disease prediction use whole population (including labelled subjects in the training set and unlabeled subjects in the test set) to construct a graph, but fail to consider the difference between disease status in those labelled subjects. Ignoring disease status on graph affects convolution performance and eventually deteriorates system training. Second, the existing works estimate edge weights by fusing similarity between imaging information and distance between non-imaging information. However, the similarity between imaging information are roughly computed based on the correlation distance between feature vectors, which affects convolution performance, especially when SMC and MCI have subtle differences among feature vectors. Third, the existing multi-modal GCN (Zhang et al., 2019), composed of multiple GCN frameworks for feature learning and then concatenating multi-modal features for disease prediction, ignores the complementary relationship between fMRI and DTI data in graph construction.

To overcome the above limitations, we design a similarity-aware adaptive calibrated GCN, which uses two GCN models corresponding to fMRI and DTI data and balances their outputs via a combined weight mechanism. Three mechanisms are proposed in this paper. First, similarity-aware receptive fields are designed on graphs to consider the difference of disease status. Specifically, every labelled node representing a training sample is only connected with those labelled nodes with the same disease status. Every unlabeled node representing a test sample may connect with every node on a graph. Second, we propose an adaptive mechanism, which uses the difference between pre-scores to replace correlation distance to estimate more accurate similarity. Specifically, we use the initial similarity calculated based on correlation distance to construct an initial graph and pre-train GCN using training samples. Then we use the pre-trained GCN to score all subjects. The difference between these pre-scores is used to form the updated similarity. This is motivated by pre-trained GCNs leading to similarity metrics better than correlation distance. Third, based on the relevant and complementary relationship between fMRI functional network and DTI structural network, we propose a calibration mechanism to fuse functional and structural information into edges. We validate our method by using the ADNI (https://ida.loni.usc.edu) public database. Experimental results show that our method achieves promising performance for predicting SMC and MCI.

2. Methodology

Figure 1 shows an overview of our proposed prediction framework. Our objective is to predict the status of an individual described as a node binary classification problem, where each node is assigned as a label $l \in \{0, 1\}$. For n subjects, each subject is represented by fMRI, DTI and phenotypic information (e.g., gender and equipment type). Based on fMRI and DTI data, we construct a functional connection (FC) brain network and a structural connection (SC) brain network for every subject. To fuse fMRI and DTI information, we develop two graphs corresponding to two GCN models, and each GCN model is trained and utilised independently. A graph is described as $\mathcal{G} = \{\boldsymbol{\mathcal{V}}, \mathcal{E}, \boldsymbol{A}\}$. $\mathcal{V}$ represents vertices, and each vertex represents a subject, $\mathcal{E}$ represents edges and each edge models the similarity between the corresponding subjects, and all edges compose adjacency matrix $\boldsymbol{A}$. In this paper, we use feature matrix $\boldsymbol{X}$ to represent features of all subjects on the graph. Each row of $\boldsymbol{X}$ represents the selected features of its corresponding subject, and the number of matrix rows matches with the number of total subjects on a graph.

Generally, we divide our framework into four parts. First, we construct FC and SC brain networks for every subject. Second, we construct functional and structural graphs. Our similarity-aware receptive fields are proposed in this part. Third, we design an adaptive calibrated GCN to output scores of subjects. We propose an adaptive mechanism and a calibration mechanism to improve the adjacency matrix in this part. Last, we employ a combined weight mechanism to balance functional scores and structural scores to accomplish our classification task.

Table 1:The notation.

Notation	Size	Description
n		Number of subjects
N		Number of brain ROIs
m		Number of selected features by using recursive feature elimination (RFE)
K		Polynomial order
r_G		Distance of gender
r_E		Distance of equipment type
w_1		Combined weight coefficient for functional score
w_2		Combined weight coefficient for structural score
$\rho(\cdot)$		Calculation of correlation distance
$Sim(\cdot)$		Calculation of similarity
$Score_v^f$		Functional score of subject v
$Score_v^s$		Structural score of subject v
$\boldsymbol{F}_v^f$	$1 \times m$	Functional feature vector of subject v
$\boldsymbol{F}_v^s$	$1 \times m$	Structural feature vector of subject v
$\boldsymbol{F}_u^f$	$1 \times m$	Functional feature vector of subject u

$\boldsymbol{F_u^s}$	$1 \times m$	Structural feature vector of subject u
$\boldsymbol{X^f}$	$n \times m$	Functional feature matrix
$\boldsymbol{X^s}$	$n \times m$	Structural feature matrix
$\boldsymbol{A_s^f}$	$n \times n$	Similarity-aware functional adjacency matrix
$\boldsymbol{A_s^s}$	$n \times n$	Similarity-aware structural adjacency matrix
$\boldsymbol{A_{sa}^f}$	$n \times n$	Similarity-aware adaptive functional adjacency matrix
$\boldsymbol{A_{sa}^s}$	$n \times n$	Similarity-aware adaptive structural adjacency matrix
$\boldsymbol{A_{sac}}$	$n \times n$	Similarity-aware adaptive calibrated adjacency matrix
$\boldsymbol{Scores^f}$	$n \times 1$	Functional score vector
$\boldsymbol{Scores^s}$	$n \times 1$	Structural score vector

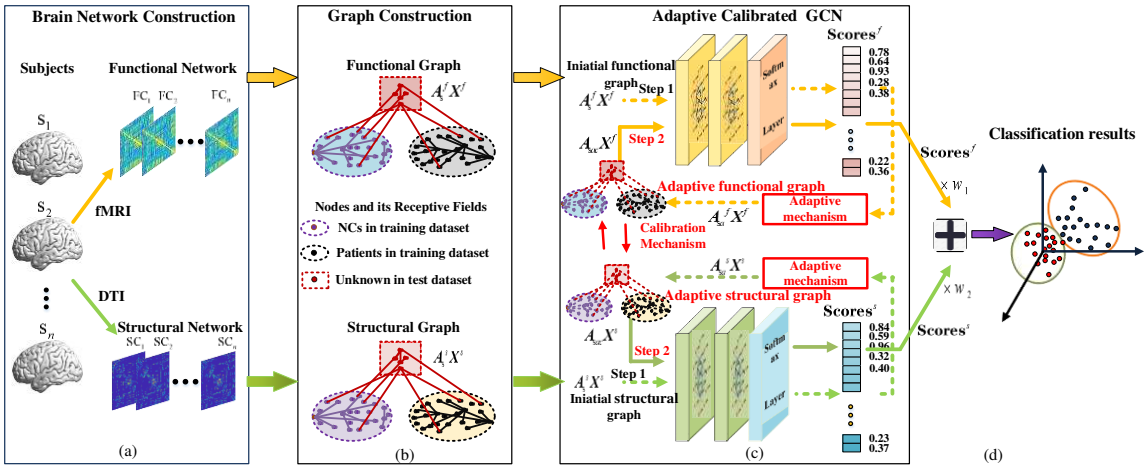


Figure 1: General framework of our proposed disease deterioration prediction algorithm. (a) Supposing there are total n subjects in our classification task. We get n functional networks, and n structural networks, with every subject, has a functional network and a structural network. (b) There are n nodes on a graph with every node representing a subject, and we construct the functional graph with every node represented by functional features and construct the structural graph with every node represented by structural features. (c) After adaptive calibrated GCN, we get a $n \times 1$ functional score vector $\boldsymbol{Scores^f}$ and a $n \times 1$ structural score vector $\boldsymbol{Scores^s}$. Every functional score represents the predicted result of its corresponding subject based on its functional features, and a structural score represents the predicted result based on a subject's structural features. (d) We use a combined weight mechanism to finally form a $n \times 1$ score vector as the final predicted results.

2.1 Dataset description and brain network construction

2.1.1 Dataset

A total of 170 subjects from the ADNI database are used for training and testing, including SMC, early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and normal control (NC). The gender, age and equipment type are used as phenotypic information in this paper, and the detailed information is shown in Table 2.

Table 2: Detailed information about the used dataset.

Group	SMC(44)	EMCI(44)	LMCI(38)	NC(44)
Male/Female	17M/27F	22M/22F	19M/19F	22M/22
Age (mean±SD)	76.3±5.4	76.5±6.1	76.0±7.7	76.5±4
GE/SIEMENS/PHILIPS	21/21/2	9/30/5	26/9/3	14/25/5

Our prediction task is a node binary classification problem. Therefore, we carry out our method on the six tasks, including NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, and EMCI vs. LMCI.

2.1.2 Functional brain network construction

For fMRI data preprocessing, we apply the standard procedures including using the GRETNA toolbox (Wang et al., 2015) to preprocess our fMRI time-series signal. We discard the first ten acquired fMRI volumes and correct the remaining 170 volumes by applying mean-subtraction. We apply head movement correction, perform spatial normalisation with DARTEL, and perform smooth filtering by employing the Gaussian kernel. Finally, we regress the local mean time-series, and use the automated anatomical labelling (AAL) (Tzourio-mazoyer et al., 2002) to segment brain space into 90 regions of interests (ROIs). After the above process, we obtain the time-series of 90 ROIs for each individual.

For constructing a functional brain network, Pearson's correlation (PC) is used, which captures the relationship between pair ROIs, and sparse representation (SR) method, which establishes multi-ROI relationship. Based on SR method, many popular methods have been proposed and applied, such as weighted sparse representation (WSR) (Yu et al., 2017), strength-weighted sparse group representation (WSGR), Group sparse representation (GSR)(Zhang et al., 2017), strength and similarity guided GSR (SSGSR)(Zhang et al., 2018), and sparse low-rank (SLR) graph learning (Qiao et al., 2016). The reviewed literature (Qiao et al., 2018) summarises the above methods. In this paper, we do not focus on the methods of brain network construction and use the reliable and straightforward PC method to construct our FC network. After brain network construction, we finally get a 90×90 brain functional network for every subject.

2.1.3 Structural brain network construction

For DTI structural brain network, we use PANDA Toolbox (Goto et al., 2013) to get the global brain

deterministic fibre bundle. We obtain the fractional anisotropy (FA) as feature vectors and use the AAL template on DTI image to divide the brain space into 90 ROIs. For SC network construction from DTI data, the average FA of links between network nodes is defined as the connection weight in the DTI network, and then we get a 90×90 SC network for every subject.

2.1.4 Feature selection method

After brain network construction, we finally have a 90×90 FC network and a 90×90 SC network for every subject. To reduce the dimension of FC and SC brain networks, we extract upper triangular matrix elements to form a 1×4005 feature vector for every brain network. Then we use recursive feature elimination (RFE) (Guyon et al., 2002) to select features. Finally, a low-dimensional feature vector is used to represent an FC or SC brain network. For example, for subject v, we have a low-dimensional functional feature vector $\boldsymbol{F}_v^f$ and a low-dimensional structural feature vector $\boldsymbol{F}_v^s$.

2.2 Graph construction

The above low-dimensional feature vectors and acquired phenotypic information (e.g., gender, age, and equipment type) are used to construct graphs. We develop two GCN models with a functional graph and a structural graph, respectively. Graphs include nodes and edges, where nodes represent subjects and edges establish their connections. Specifically, every node on the functional graph is represented by its corresponding subject's functional feature vector. Every node on the structural graph is represented by its corresponding subject's structural feature vector. Edge connections and edge weights are the keys in graph theory as they decide which nodes are used to perform convolutions and corresponding convolutions coefficients, therefore they attract much attention (Liu et al., 2019; Xu et al., 2018). The two-layer network with a graph (Kipf and Welling, 2017) can be described as the equation $\boldsymbol{Z} = softmax(\boldsymbol{A}ReLU(\boldsymbol{AXW}^{(0)})\boldsymbol{W}^{(1)})$ and the filtering principle of graph theory is illustrated in Figure 2, where $\boldsymbol{A}$ is the adjacency matrix with normalization. We can see that a big convolution coefficient means big filtering effect in its corresponding feature..

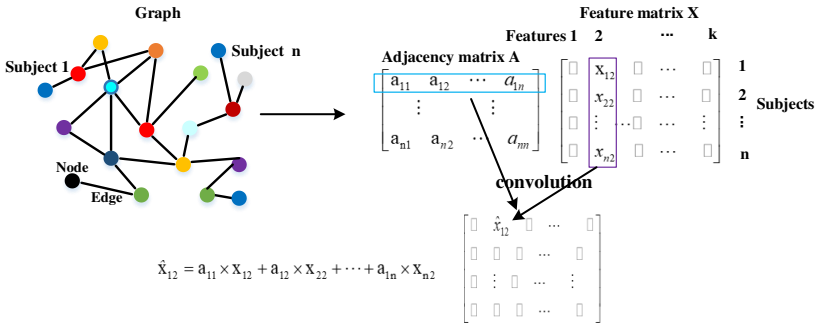


Figure 2. Filtering principle of the graph theory.

In existing methods, edge connections consider gender and equipment type with ignoring the disease status of those subjects in the training set, and edge weights are evaluated by a computed correlation coefficient of feature vectors. In this subsection, we design similarity-aware receptive fields to consider disease status of those subjects in training set in terms of edge connections. In the next subsection, we design an adaptive mechanism and calibration mechanism to improve edge weights. For edge weights, we first use an existing method to initialise them.

2.2.1 Edge connections based on similarity-aware receptive fields

Previous work considers gender and equipment type to establish edge connections by assigning bigger edge weights between those subjects with the same gender and same equipment type. Still, it fails to consider disease status of those subjects in the training set. As disease status results in differences on subjects' features and status of most subjects on the graph (a graph includes those subjects in both training set and test set) are known, it is necessary to consider disease status in edge connections. Hence, we design three receptive fields that incorporate knowledge on disease status. Two receptive fields are for labelled subjects in the training set, and one receptive field is for unlabeled subjects in the test set. For a labelled patient, we establish its connections with all labelled patients. For a labelled NC, we establish its connections with all labelled NCs. For every unlabeled subject in the test set, we ignore to consider its disease status and establish its connections with all other subjects. The detailed description of three receptive fields is shown in Figure 3.

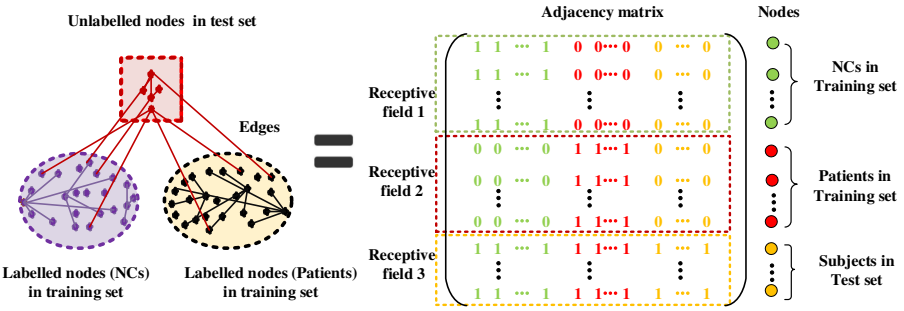


Figure 3: Detailed description of similarity-aware receptive fields. We describe our similarity-aware fields by classifying NC and Patient. In the adjacency matrix, '1' represents connection is established, and '0' represents connection is not established.

2.2.2 Edge weights initialisation

Initial edge weights are estimated based on previous works (Kazi et al., 2019; Kipf and Welling, 2017; Ktena et al., 2018; Parisot et al., 2018; Zhang et al., 2019), which fuse similarity between imaging information and distance between non-imaging information. We use $Sim(\cdot)$ to denote similarity between paired subjects, r_G represents the distance of gender, and r_E represents the distance of equipment type. Based on the edge connections in similarity-aware receptive fields in Figure 3, the initial similarity-aware functional adjacency matrix $\boldsymbol{A}_s^f$ and the initial similarity-aware structural adjacency matrix $\boldsymbol{A}_s^s$ are calculated as:

$$\boldsymbol{A}_s^f(v,u) = Sim\left(\boldsymbol{F}_v^f, \boldsymbol{F}_u^f\right) \times \left(r_G(G_v, G_u) + r_E(E_v, E_u)\right), \tag{1}$$

$$\boldsymbol{A}_s^s(v,u) = Sim\left(\boldsymbol{F}_v^s, \boldsymbol{F}_u^s\right) \times \left(r_G(G_v, G_u) + r_E(E_v, E_u)\right), \tag{2}$$

where $\boldsymbol{F}_v^f$ and $\boldsymbol{F}_u^f$ are functional feature vectors of subject v and subject u, $\boldsymbol{F}_v^s$ and $\boldsymbol{F}_u^s$ are their structural feature vectors, G_v and $G_u$represent their gender information, E_v and E_u represent their equipment type information, r_G and r_E are defined as:

$$r_G(G_v, G_u) = \begin{cases} 1, & G_v = G_u, \\ 0, & G_v \neq G_u. \end{cases}, r_E(E_v, E_u) = \begin{cases} 1, & E_v = E_u, \\ 0, & E_v \neq E_u. \end{cases} \tag{3}$$

The initial similarity is estimated by calculating the correlation distance between feature vectors as (Parisot et al., 2018):

$$Sim\left(\boldsymbol{F}_v^f, \boldsymbol{F}_u^f\right) = \exp\left(-\frac{\left[\rho\left(\boldsymbol{F}_v^f, \boldsymbol{F}_u^f\right)\right]^2}{2\sigma^2}\right), Sim\left(\boldsymbol{F}_v^s, \boldsymbol{F}_u^s\right) = \exp\left(-\frac{[\rho(\boldsymbol{F}_v^s, \boldsymbol{F}_u^s)]^2}{2\sigma^2}\right), \tag{4}$$

where ρ(·) is the correlation distance function, and σ is the width of the kernel.

The above initial similarity $Sim(\cdot)$ is used to construct the edge weight, which plays the role as a convolution coefficient in graph theory as shown in Figure 2. In the work (Parisot et al., 2018), the final classi-

fication performance gets significant improvement by combing $Sim(\cdot)$ with phenotypic information. The edge weight is doubled when its corresponding two subjects have the same gender and equipment type, and the edge weight is set to zero when the corresponding two subjects have different gender and equipment type. The method of integrating phenotypic information increases the difference between edge weights and the final classification results validate this effectiveness

After establishing edge connections based on our similarity-aware receptive fields and above initial edge weights, we get the initial similarity-aware functional adjacency matrix $\boldsymbol{A}_s^f$ and the initial similarity-aware structural adjacency matrix $\boldsymbol{A}_s^s$.

2.3 Adaptive calibrated GCN

In this subsection, we develop two GCN models. One model is used to predict disease status based on functional data, and the other is used based on structural data. Each model is trained and utilised independently. Specifically, we use functional data in the training set and their corresponding labels to train a GCN model, and then use the trained model to predict the status of all subjects. After the process, we get a functional score vector $\boldsymbol{Scores}^f \epsilon \mathbb{R}^{n\times 1}$ to represent the predicted scores. Besides, we use the structural data in the training set and their corresponding labels to train the other GCN model, and also use the model to predict the status of all subjects. After the process, we get a structural score vector $\boldsymbol{Scores}^s \epsilon \mathbb{R}^{n\times 1}$ to represent the predicted scores. The above two GCN models can accomplish prediction tasks independently. As integrating fMRI functional data and DTI structural data shows better performance (Lei et al., 2020; Li et al., 2020b), we use a combined weight mechanism method to combine their predicted results to perform the final prediction. The corresponding two combined weight coefficients are set as 0. 5 in this paper according to the experimental results.

Using the correlation distance to compute similarity in Eq. (4) is inaccurate enough since SMC and MCI have subtle differences among feature vectors. We propose an adaptive mechanism to improve the similarity measure in view that GCN has better capability to extract in-depth features than the correlation distance. We develop a calibration mechanism to fuse functional and structural data into edges. By using our adaptive calibrated mechanism, we update our initial GCN models by pre-training and finally use the updated GCN models to predict disease status. Our model is not trained end-to-end, and there are two steps in our adaptive calibrated GCN. First, based on initial graphs, we train GCN models and then use them to score every subject.

Based on these scores, we use our adaptive mechanism and calibration mechanism to construct a new adjacency matrix and then form new graphs. Second, based on new graphs, we train GCN models again and finally use them to predict disease status.

2.3.1 Adaptive mechanism

Random forest-derived similarity evaluation methods (Shi et al., 2005; Shi and Horvath, 2006) use machine learning to evaluate similarity in unsupervised clustering tasks, which inspire us to propose an adaptive mechanism in GCN for our disease prediction. Compared with the initial adjacency matrices, the adaptive adjacency matrices use score difference to replace correlation distance for constructing more accurate edge weights. First, we construct dual-modal GCN models with initial graphs and then pre-train GCN models using training samples. Second, we input all subjects to the pre-trained GCN to get their scores. We use $\boldsymbol{Scores}^f$ to represent functional score vector and use $\boldsymbol{Scores}^s$ to represent structural score vector. Last, we re-compute edge weights with updated similarity based on scores. The adaptive similarity based on scores are calculated:

$$Sim\left(\boldsymbol{F}_v^f, \boldsymbol{F}_u^f\right) = \exp\left(-\frac{\left[Score_v^f - Score_u^f\right]^2}{2\sigma^2}\right), Sim(\boldsymbol{F}_v^s, \boldsymbol{F}_u^s) = \exp\left(-\frac{[Score_v^s - Score_u^s]^2}{2\sigma^2}\right), \quad (5)$$

where $Score_v^f$ and $Score_u^f$ denote the scores of subject v and subject u on functional data, whereas $Score_v^s$ and $Score_u^s$ denote their scores on structural data. Every score is a scalar and ranges from 0 to 1, which is used to represent the predicted disease status of a subject based on functional or structural features. In labels, we use 0 or 1 to represent the status of the subject. σ is also the width of the kernel. By Eqs. (1), (2), (3) and (5), we finally get a more accurate similarity-aware adaptive functional adjacency matrix $\boldsymbol{A}_{sa}^f$ and a more accurate similarity-aware adaptive structural adjacency matrix $\boldsymbol{A}_{sa}^s$.

2.3.2 Calibration mechanism

As functional and structural information is complementary, we propose a calibration mechanism to integrate fMRI functional and DTI structural information. Let the symbol ∘ represent the Hadamard product, based on the above similarity-aware adaptive functional adjacency matrix $\boldsymbol{A}_{sa}^f$ and similarity-aware adaptive structural adjacency matrix $\boldsymbol{A}_{sa}^s$, the similarity-aware adaptive calibrated adjacency matrix $\boldsymbol{A}_{sac}$ is defined as:

$$\boldsymbol{A}_{sac} = \boldsymbol{A}_{sa}^f \circ \boldsymbol{A}_{sa}^s. \quad (6)$$

After using the calibration mechanism, we form a similarity-aware adaptive calibrated adjacency matrix $\boldsymbol{A}_{sac}$. It is worth mentioning that the adjacency matrix is further normalized using Eq. (7). After this, in the normalized adjacency matrix, the sum of every row of elements is set to 1

$$\boldsymbol{A}_{sac}(i,j) = \boldsymbol{A}_{sac}(i,j) / \sum_{k=1}^{n} \boldsymbol{A}_{sac}(i,k)\,, \tag{7}$$

2.3.3 Graph convolutional network architecture

In GCN, spectral theory improves adjacency matrix $\boldsymbol{A}_{sac}$ by applying the convolution of Fourier transform and Taylor's expansion formula to achieve an excellent filtering effect and computational efficiency. The spectral convolution (Defferrard et al., 2016; Shuman et al., 2013) on graphs can be described as the multiplication of a signal $\boldsymbol{x} \in \mathbb{R}^n$ (a scalar for every node) with a filter $g_{\boldsymbol{\theta}} = diag(\boldsymbol{\theta})$ by:

$$g_{\boldsymbol{\theta}} * \boldsymbol{x} = \boldsymbol{U} g_{\boldsymbol{\theta}}(\boldsymbol{\Lambda}) \boldsymbol{U}^T \boldsymbol{x} = \sum_{k=0}^{K} \boldsymbol{\theta}_k T_k\left(\tilde{\boldsymbol{L}}\right)\boldsymbol{x}, \tag{8}$$

where $\boldsymbol{U}$ is the matrix of eigenvectors and is computed from formula $\boldsymbol{L} = \boldsymbol{I}_N - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}_{ac}\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$. $\boldsymbol{I}_N$ and $\boldsymbol{D}$ are, respectively, the identity matrix and the diagonal degree matrix. $g_\theta(\boldsymbol{\Lambda})$ is well approximated by a truncated expansion in terms of Chebyshev polynomials to the K^{th}-order. $\boldsymbol{\theta}_k$ is a vector of Chebyshev coefficients, T_k is Chebyshev polynomials function, $\tilde{\boldsymbol{L}} = 2/\lambda_{max}\boldsymbol{\Lambda} - \boldsymbol{I}_N$.

After spectral convolution, similarity-aware adaptive calibrated adjacency matrix $\boldsymbol{A}_{sac}$ is approximated by $\sum_{k=0}^{K} \boldsymbol{\theta}_k T_k\left(\tilde{\boldsymbol{L}}\right)$. By adjusting polynomial order K, it can get a different filter effect. For example, the performance reaches the best with $K = 3$ or 4 in prediction tasks (Kipf and Welling, 2017; Parisot et al., 2018).

Our dual-modal GCN structure is illustrated in Figure 1. Every GCN model consists of two graph convolution layers activated by rectified linear unit (ReLU) function and one softmax output layer. The functional and structural GCN models are trained using the whole population graph as input. After dual-modal adaptive calibrated GCN, we get an updated functional score and structural score for every subject. Namely, we use a combined weight mechanism to combine the two scores to perform the final prediction. Specifically, the final predicted score for a subject v is denoted by $w_1 \times Score_v^f + w_2 \times Score_v^s$. According to our experimental results in the experimental section, we set $w_1 = 0.5$ and $w_2 = 0.5$ for our all prediction tasks. For example, for NC vs. SMC, the label of an SMC subject is set 1, and the label of an NC is set to zero. The predicted result of a subject after GCN models is represented by a score which ranges from 0 to 1. A subject

with a predicted score ranging from 0 to 0.5 is regarded as an NC, and a subject with a predicted score ranging from 0.5 to 1 is regarded as an SMC.

3. Experiments and results

We evaluate the proposed method on the ADNI database using a 10-fold cross-validation strategy. As our main contribution is to improve traditional GCN for predicting SMC and MCI, the GCN parameters of all strategies in this paper are fixed and chosen according to previous work (Parisot et al., 2018). Parameters details are as below: dropout rate is 0.1, regularisation is 5×10^{-4}, the learning rate is 0.005, the number of epochs is 200, and the default polynomial order is 3. Different from (Parisot et al., 2018), to reduce the number of parameters in GCN and avoid overfitting, the number of neurons per layer is set as 8 and the number of the selected features is set as 50. For dual-modal GCN, $w_1 = 0.5$ and $w_2 = 0.5$. In this section, we refer to the graph constructed from the phenotypic data, including gender and equipment type information. Given the small size of our dataset and that age reduces the performance (Parisot et al., 2018), we ignore age information in GCN. Prediction accuracy (ACC), sensitivity (SEN), specificity (SPE) and area under the curve (AUC) are used as evaluation metrics. Six binary classification experiments including NC vs. SMC, NC vs. EMCI, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI and EMCI vs. LMCI validate our prediction performance.

We divide this section into three parts. First, we test the performance of our three mechanisms and compare them with other popular traditional algorithms. Second, we describe the effect of our similarity-aware receptive fields and adaptive mechanism on the adjacency matrix. Third, we describe the effect of our adjacency matrix on feature values. The critical parameters of the proposed method are described in the discussion section.

3.1 Classification performance of our method

The proposed prediction framework is compared to other four related popular frameworks, including GCN (Parisot et al., 2018), multiple layer perception (MLP), random forest (RF) (Breiman, 2001) and SVM(Cortes and Vapnik, 1995). The parameters are set according to work by (Parisot et al., 2018), the parameters of MLP are the same with GCN implementation, RF and SVM use the scikit-learn library implementation (Pedregosa et al., 2011). The parameters of RF are: The number of trees is 500, and the max-

imum depth is three. The parameters of SVM are: The kernel is 'sigmoid', the kernel coefficient is 0.1, the regularisation parameter is 0.1, and the maximum number of iterations is 200.

To describe our three mechanisms in detail, similarity-aware receptive fields, adaptive mechanism and calibration mechanism are named as 'S', 'A' and 'C', respectively. For example, the GCN with similarity-aware receptive fields is represented by S-GCN, SA-GCN represents the GCN with similarity-aware receptive fields and adaptive mechanism, and SAC-GCN represents similarity-aware adaptive calibrated GCN. The results of the experiment are shown in Table 3. ROC curves comparison is shown in Figure 4.

Table 3: Disease prediction performance of different methods in our six tasks.

Modal	Method	NC vs. SMC				NC vs. EMCI				NC vs. LMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
fMRI	MLP	59.09	61.36	56.81	63.58	62.50	61.36	62.06	68	65.85	65.78	65.90	72.97
	RF	60.22	65.91	54.54	68.34	65.90	52.27	79.54	70.51	68.29	60.52	75	70.87
	SVM	63.63	68.18	59.09	69.21	64.77	63.63	63.63	68.75	69.51	63.15	75	79.67
	GCN	70.45	84.09	56.81	76.39	68.18	79.54	56.81	73.61	71.95	71.05	72.72	76.67
	S-GCN	72.72	77.27	68.18	81.66	69.31	52.27	86.36	74.12	73.17	71.05	75	78.77
	SA-GCN	76.13	79.54	72.72	84.81	71.59	79.54	65.90	79.44	80.48	76.31	84.09	91.27
	SAC-GCN	77.27	81.81	72.72	80.37	75	84.09	65.91	80.94	84.14	78.94	88.63	92.64
DTI	MLP	67.63	68.18	59.09	74.07	70.45	63.63	77.27	84.95	73.17	71.05	75	84.99
	RF	65.63	70.45	56.81	69.32	69.31	70.45	68.18	72.52	73.17	73.68	72.72	71.79
	SVM	71.59	86.36	56.81	84.35	69.31	72.72	65.90	71.82	71.95	71.05	72.72	80.32
	GCN	72.72	75	70.45	83.88	72.72	77.27	68.18	80.94	76.82	78.94	75	87.86
	S-GCN	75	88.63	61.36	84.81	73.86	77.27	70.45	82.90	76.82	78.94	75	90.43
	SA-GCN	79.54	86.36	72.72	90.03	77.27	86.36	68.18	85.80	84.14	84.21	84.09	91.09
	SAC-GCN	81.81	88.63	75	89.36	81.81	86.36	77.27	88.89	87.80	86.84	86.63	91.33
Dual	MLP	68.18	81.81	54.54	75.83	71.59	70.45	72.72	77.69	75.60	73.68	77.27	86.42
	RF	67.04	72.72	61.36	71.95	72.72	75	70.45	73.33	76.82	76.31	77.27	84.15
	SVM	73.86	86.36	61.36	76.76	71.59	75	68.18	73.14	73.17	73.68	72.72	80.08
	GCN	76.13	86.36	65.90	88.22	75	77.27	75.55	80.73	79.26	78.94	79.54	89.71
	S-GCN	78.40	88.63	68.18	86.00	76.13	79.54	72.72	83.32	82.92	81.57	84.09	89.83
	SA-GCN	81.81	86.36	77.27	90.29	79.54	88.63	70.45	86.67	85.36	81.57	88.64	89.53
	SAC-GCN	84.09	88.63	79.54	89.67	85.22	90.90	79.54	89.82	89.02	89.47	88.63	92.88

Modal	Method	SMC vs. EMCI				SMC vs. LMCI				EMCI vs. LMCI			
		ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
fMRI	MLP	60.22	65.90	54.54	63.43	58.83	44.73	70.45	64.35	65.85	71.05	61.36	70.22
	RF	63.63	65.90	61.36	66.99	61.97	57.07	77.27	60.19	62.19	65.78	59.09	66.33
	SVM	64.77	56.81	72.72	67.98	64.63	63.15	65.90	71.11	67.07	55.26	77.27	71.65
	GCN	72.72	77.27	68.18	83.37	71.95	55.26	86.36	82.06	73.17	97.73	54.54	79.67
	S-GCN	75	79.54	70.45	84.64	73.17	55.26	88.63	82.83	76.82	92.10	63.63	89.11
	SA-GCN	77.27	84.09	70.45	86.57	76.82	63.15	88.63	85.89	78.04	94.73	63.63	82.48
	SAC-GCN	80.68	79.54	81.81	89.31	76.82	63.15	88.63	85.89	79.26	84.21	75	90.67
DTI	MLP	68.18	68.18	68.18	75	70.73	68.42	72.72	81.16	67.07	60.52	72.72	69.08
	RF	70.45	81.81	59.09	79.60	73.17	65.78	79.54	79.13	68.29	68.42	68.18	70.10
	SVM	70.45	65.90	75.00	75.26	74.39	68.42	79.54	79.01	73.17	68.42	77.27	75.54
	GCN	79.54	79.54	79.54	93.39	81.70	78.94	84.09	84.39	74.39	89.47	61.36	78.95
	S-GCN	80.68	84.09	77.27	89.88	82.92	78.94	86.36	93.90	78.04	94.73	63.63	82.48

	SA-GCN	84.09	84.09	84.09	91.58	84.14	81.36	82.66	89.71	80.48	89.47	72.72	88.10
	SAC-GCN	85.22	88.63	81.81	92.05	86.58	84.21	88.63	95.69	82.92	94.73	72.72	94.14
Dual	MLP	69.31	70.45	68.18	73.86	71.95	76.31	68.18	83.07	69.51	65.78	72.72	70.57
	RF	71.59	70.45	72.72	79.34	75.60	71.05	79.54	80.74	71.95	73.68	70.45	72.13
	SVM	72.72	77.27	68.18	76.39	75.60	68.42	81.81	80.14	75.60	65.78	84.09	77.57
	GCN	80.09	77.27	81.31	88.79	82.70	84.21	79.54	86.90	79.26	94.73	65.90	89.35
	S-GCN	82.95	86.36	79.54	94.32	84.14	81.57	86.36	88.82	81.70	92.10	72.72	83.55
	SA-GCN	85.22	90.90	79.54	94.73	86.58	84.21	88.63	95.69	82.92	94.73	72.72	94.14
	SAC-GCN	88.63	95.45	81.81	95.56	87.80	84.21	90.90	90.25	86.58	92.10	81.81	94.26

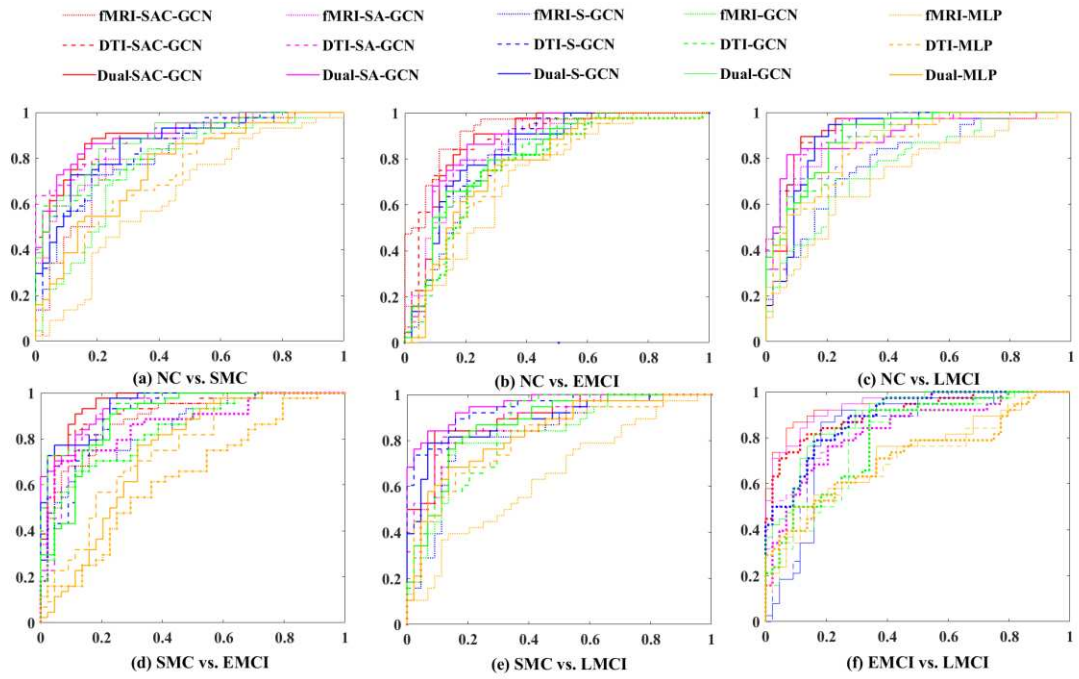


Figure 4: ROC curves comparison of different scenarios.

We use the most common approach to construct a brain network in this paper. As shown in Table 3, the performance of traditional classifiers (MLP, RF, SVM) based on our brain networks is poor, and there is only a few variation with less than 2.73% difference in mean ACC of six tasks between the best and the worst performance. SVM shows the best performance with mean ACC of six tasks based on dual-modal data reaching to 73.75%. Compared with the above traditional methods, the performance of GCN is much improved. Specifically, compared with the best performance in traditional classifiers based on fMRI, DTI and dual modalities, the mean ACC of six tasks increase by 5.67%, 4.50% and 4.95%, and the mean AUC of six tasks increase by 7.23%, 7.18% and 9.93%. The performance comparison follows the previous work (Parisot et al., 2018), and it validates the effectiveness of graph theory on classification. For the above six classification tasks based on dual-modal GCN, the performance of NC vs. SMC is the worst, and the performance of

NC vs. LMCI is the best.

Because of the effectiveness of GCN and shortcomings of existing researches, we propose three mechanisms to improve GCN in this paper. First, we propose similarity-aware receptive fields to consider disease status in edge connections. As Table 3 shows, the performance of S-GCN improves performance compared with GCN. Specifically, based on fMRI, DTI and dual modalities, the mean ACC of S-GCN of our six tasks increase by 1.96%, 1.57% and 2.30%, the mean SEN increase by -6.24%, 3.90% and 1.83%, the mean SPE increase by 9.47%, -0.75% and 2.64%, and the mean AUC increase by 3.22%, 2.49% and 0.35%. The above comparison results validate that considering disease status is essential in graph construction. By using similarity-aware receptive fields on dual modalities, the final performance of NC vs. LMCI gets the highest improvement with ACC increased by 3.66%. In contrast, the ACC of the remaining tasks increased by 2.27%, 1.13%, 2.86%, 1.44%, and 2.44%.

Second, we propose an adaptive mechanism to improve edge weights. As shown in Table 3, based on similarity-aware receptive fields, adaptive mechanism yields improved results. Specifically, based on fMRI, DTI and dual modalities, the mean ACC of SA-GCN compared with S-GCN increase by 3.35%, 3.72% and 2.53%, the mean SEN increase by 8.37%, 1.54% and 2.77%, the mean SPE increase by 1.13%, 5.06% and 2.27%, and the mean AUC increase by 3.22%, 1.98% and 4.20%. The above comparison results show that combined our adaptive mechanism with similarity-aware receptive fields further improves performance. By using the adaptive mechanism on dual modalities, the final performance of NC vs. SMC and NC vs. EMCI gets the most significant improvement with ACC increased by 3.42% and 3.41%. The ACC of the other tasks increases by 2.44%, 2.27%, 2.43% and 1.22%. After using similarity-aware receptive fields and adaptive mechanism, we can get the mean ACC of 83.57% for our six tasks.

Third, we propose a calibration mechanism to fuse functional and structural information into the adjacency matrix. As shown in Table 3, SAC-GCN yields improved results compared with SA-GCN. Specifically, based on fMRI, DTI and dual modalities, the mean ACC of SAC-GCN compared with SA-GCN increase by 2.14%, 2.74% and 3.31%, the mean SEN increase by -0.93%, 2.92% and 2.39%, the mean SPE increase by 4.54%, 2.93% and 4.16%, and the mean AUC increase by 1.56%, 2.52% and 0.23%. The above comparison results show that our calibration mechanism can improve performance when functional adjacency matrix and structural adjacency matrix have high precision. Eventually, the mean ACC, SEN, SPE and

AUC of SAC-GCN of our six tasks is 86.89%, 90.12%, 83.70% and 92.07%, respectively.

Compared with the results based on fMRI data, it shows better prediction performance based on DTI data. Specifically, for the three traditional methods (MLP, RF and SVM), the mean ACC of our six tasks increase by 7.48%, 6.30% and 6.08%, and the mean AUC of our six tasks increase by 11.11%, 6.53% and 6.32%. For GCN series methods (GCN, S-GCN, SA-GCN and SAC-GCN), the mean ACC of our six tasks increases by 4.91%, 4.52%, 4.88% and 5.49%, and the mean AUC of our six task increases by 6.27%, 5.54%, 4.30% and 5.27%. We employ a combined weight mechanism to fuse the results of dual-modal data for the final disease prediction. Compared with the prediction results based on single modal DTI data, the prediction results based on dual-modal data show improvement. Specifically, for GCN methods (GCN, S-GCN, SA-GCN and SAC-GCN), the mean ACC of our six tasks increase by 2.42%, 3.15%, 1.96% and 2.53%, respectively.

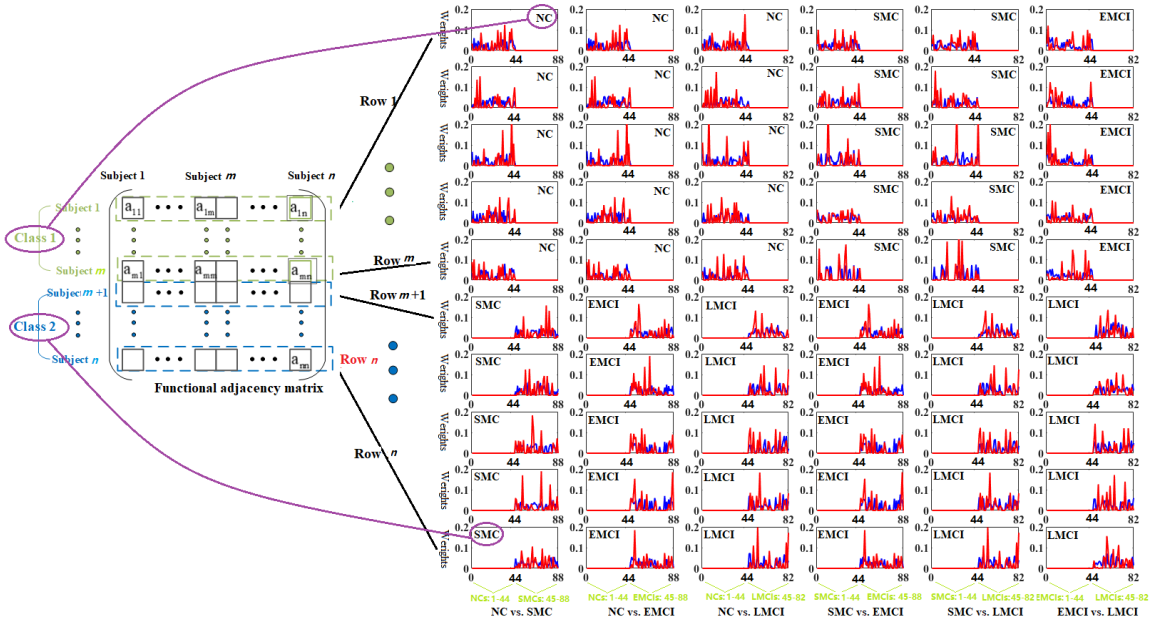
For our three mechanisms, similarity-aware receptive fields consider disease status in graph construction and adaptive mechanism uses scores difference to replace correlation distance for constructing a more accurate adjacency matrix. The two appealing mechanisms are not limited to our tasks, and they may extend to other prediction tasks (e.g., AD, ASD and PD).

3.2 Effect of similarity-aware receptive fields and adaptive mechanism on adjacency matrix

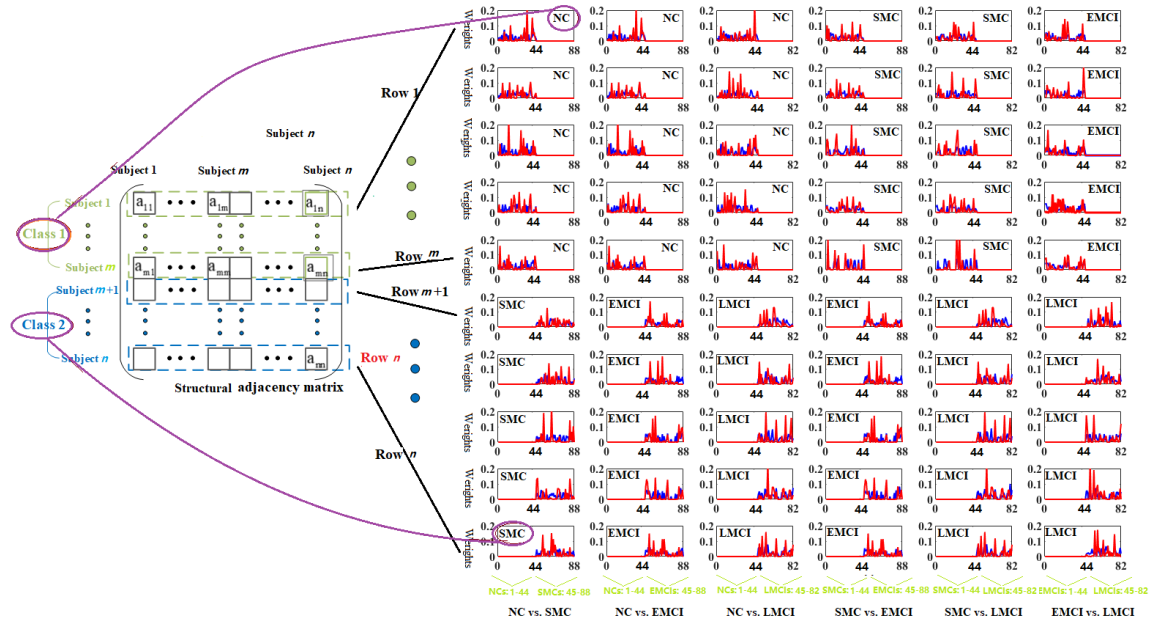
The adjacency matrix is the key of graph theory, which is a mathematical description of edges and edge weights, and plays the role as a filter (Kipf and Welling, 2017; Parisot et al., 2018). Specifically, after applying spectral convolution as Eq. (8), similarity-aware adaptive calibrated adjacency matrix $\boldsymbol{A}_{sac}$ is further approximated by $\sum_{k=0}^{K} \boldsymbol{\theta}_k T_k(\tilde{\boldsymbol{L}})$. A row of elements of the approximated matrix $\sum_{k=0}^{K} \boldsymbol{\theta}_k T_k(\tilde{\boldsymbol{L}})$ can be regarded as the convolution coefficients of its related subjects. Our three mechanisms play the role to improve the adjacency matrix and therefor improve the convolution coefficients, and experimental results in the above subsection validate their effectiveness. In this subsection, we describe how similarity-aware receptive fields and adaptive mechanism affect the adjacency matrix.

The proposed similarity-aware receptive fields consider the disease status and constrain the receptive field of labelled nodes to those nodes with the same status, which means we are establishing connections only between those subjects with the same status. Different from similarity-aware receptive fields focuing on edge connections, the adaptive mechanism is proposed to improve edge weights. Edge weights represent convolution coefficients, where a considerable weight means its corresponding two subjects have better similarity

and a significant impact on each other. To describe the effect of similarity-aware receptive fields and adaptive mechanism, we pick up five subjects from the training set randomly for every disease status in every prediction task. Our prediction task is a node binary classification problem, so there are ten subjects to be picked up for every prediction task. Figure 5 visualises their corresponding edge weights in an adaptive functional adjacency matrix and adaptive structural adjacency matrix. The two adaptive adjacency matrices have been processed by normalisation.



(a) Effect on edge weights in fMRI functional adjacency matrix.



(b) Effect on edge weights in DTI structural adjacency matrix.

Figure 5: Effect of similarity-aware receptive fields and adaptive mechanism on edge weights in our six prediction tasks. In our six tasks, we pick up ten subjects randomly from the training set (five subjects for each disease status) and show their edge weights with all subjects on the graph. In every subfigure, the abscissa represents subjects' indices on the graph, and the ordinate represents a subject's edge weights. Blue lines represent the edge weights constructed by using the traditional method, and red lines represent the edge weights constructed by using our similarity-aware receptive fields and adaptive mechanism.

Figure 5 shows that parts of edge weights are zeros, which is the effect of similarity-aware receptive fields that establish edge connections only between those subjects with the same status. For example, for NC vs. SMC, in the first subfigure, we describe an NC subject' edge weights with all 88 subjects on the graph. As abscissa represents subject's indices where indices 1-44 represent 44 NCs and indices 45-88 represent 44 SMCs, the NC's edge weights with subjects 1-44 are mostly non-zeros whereas its edge weights with subjects 45-88 are all zeros. Part of subjects are test samples, and edge weights with these test samples are all set to zero.

Compared with a little difference between edge weights computed by traditional methods (Kazi et al., 2019; Ktena et al., 2018; Parisot et al., 2018; Zhang et al., 2019), our adaptive mechanism increases the difference seen in every subfigure in Figure 5. Specifically, the red lines, which represent edge weights based on our adaptive mechanism, show large fluctuations, whereas the blue lines show small fluctuations. The standard deviations of these fluctuations are described in Table 4. The standard deviations based on our adaptive mechanism are larger than those based on the traditional method. In the work (Parisot et al., 2018), by including phenotypic information as Eqs. (1) and (2), the edge weight is doubled when its corresponding two subjects have the same gender and equipment type, and the edge weight is set to zero when the corresponding two subjects have different gender and equipment type. This increases the difference between edge weights, which is validated to be useful to improve the final classification performance. Similar to the work (Parisot et al., 2018), our adaptive mechanism also increases the difference and the final performance also gets improvement as shown in Table 3. This suggests that our adaptive mechanism has a better ability to explore the similarity relationship between subjects. Comparing edge weights in the DTI structural adjacency matrix with those edge weights in the fMRI functional adjacency matrix for the same subject, they show obvious differences. In Table 4, we use "Difference" to represent the differences between edge weights in fMRI functional adjacency matrix and DTI structural adjacency matrix. Standard deviations show there are many differences between edge weights in fMRI functional adjacency matrix and DTI structural adjacency

matrix. Our adaptive mechanism usually increases the differences. The differences support the viewpoint that fMRI functional information and DTI structural information have good complementarity (Lei et al., 2020; Li et al., 2020b), and it also agrees with the excellent performance of our calibration mechanism and dual-modal GCN.

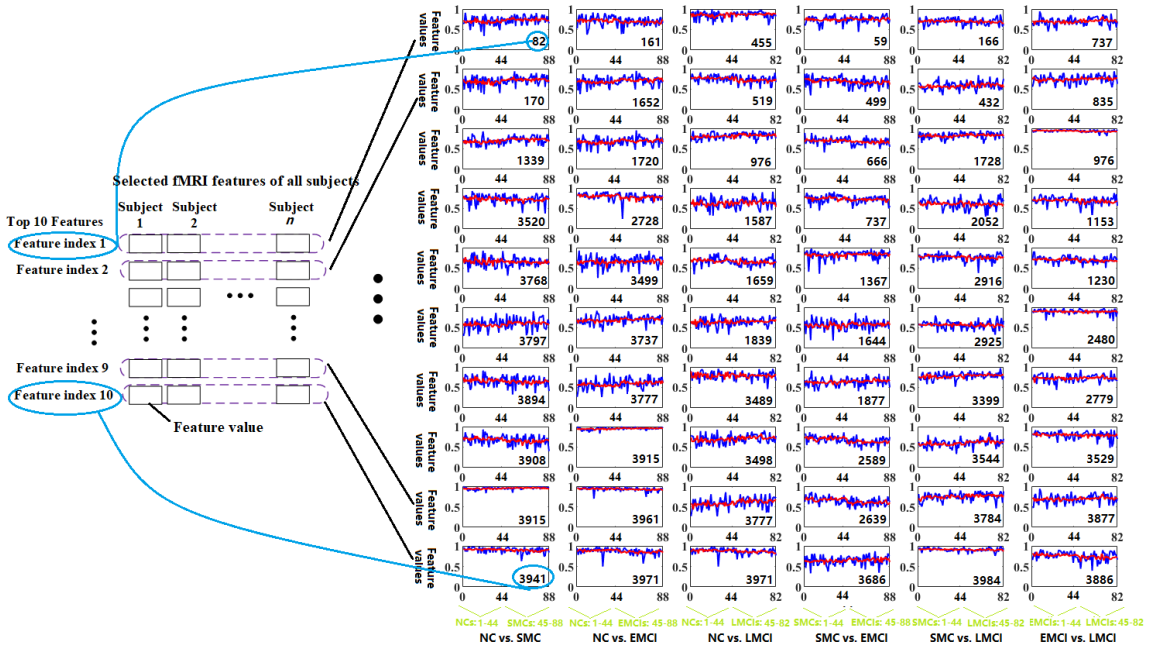
Table 4: The standard deviations of the edge weights with and without our adaptive mechanism across our six tasks. ($\times 10^{-2}$). Cases 1-10 represent ten subjects in the corresponding task, and the ten subjects are the selected subjects in Figure 5. "Difference (i.e., $\boldsymbol{A}^f - \boldsymbol{A}^s$)" represents the difference of edge weights between fMRI functional adjacency matrix and DTI structural adjacency matrix.

Case	Modality	NC vs. SMC	NC vs. EMCI	NC vs. LMCI	SMC vs. EMCI	SMC vs. LMCI	EMCI vs. LMCI
		None/Adapt	None/Adapt	None/Adapt	None/Adapt	None/Adapt	None/Adapt
1	fMRI	1.83/1.86	1.79/1.81	1.81/1.87	1.66/1.62	1.62/1.70	1.56/1.59
	DTI	3.07/4.13	3.07/4.13	3.16/4.00	2.83/3.15	2.98/3.50	2.90/3.33
	Difference	2.51/3.88	2.61/3.95	2.68/3.69	2.58/2.88	3.03/3.14	2.61/2.96
2	fMRI	1.67/1.69	1.69/1.68	1.71/1.72	1.66/1.72	1.66/1.69	1.64/1.67
	DTI	2.97/3.39	3.09/3.37	3.39/3.96	2.95/2.97	3.13/3.71	3.45/3.79
	Difference	3.08/2.32	3.20/2.36	2.94/3.52	2.52/2.59	2.71/3.34	3.06/3.76
3	fMRI	1.87/1.98	1.85/1.89	1.85/1.98	1.84/1.88	1.67/1.71	1.82/1.87
	DTI	4.59/4.65	4.59/4.65	3.43/4.93	4.12/4.69	3.04/3.91	3.21/5.07
	Difference	4.01/4.31	4.04/4.32	4.66/4.05	4.53/3.93	2.51/3.65	3.91/4.49
4	fMRI	1.92/1.95	1.89/1.90	1.93/1.94	1.59/1.62	1.59/1.67	1.62/1.67
	DTI	3.01/3.44	3.01/3.44	3.25/4.01	2.15/2.88	2.78/3.13	2.90/3.52
	Difference	2.52/2.87	2.46/2.87	2.50/3.47	1.71/2.27	2.32/3.21	2.35/3.06
5	fMRI	1.84/1.94	1.85/1.86	1.91/1.92	2.42/2.54	1.60/1.64	2.49/2.53
	DTI	2.71/3.22	2.71/3.22	3.62/3.92	3.76/4.16	2.93/3.76	4.89/5.27
	Difference	2.38/2.86	2.02/2.68	3.22/3.48	2.80/3.43	2.16/3.61	4.45/4.52
6	fMRI	1.57/1.62	1.94/1.99	2.06/1.96	2.04/2.07	0.24/0.31	1.96/2.05
	DTI	4.08/4.71	3.27/3.15	3.14/3.73	3.79/3.95	3.14/3.73	3.14/3.73
	Difference	3.65/3.83	2.50/2.54	2.86/3.58	3.45/3.33	3.24/3.73	2.96/3.55
7	fMRI	1.71/1.75	2.03/2.09	2.15/2.13	2.07/2.09	0.27/0.55	2.09/2.15
	DTI	3.18/4.51	3.25/3.66	3.54/4.78	3.03/3.84	3.54/4.78	3.54/4.78
	Difference	2.73/4.29	2.65/3.13	3.27/4.48	2.38/3.41	3.71/4.72	3.13/4.63
8	fMRI	1.89/1.93	1.63/1.69	1.93/2.00	1.62/1.67	1.93/1.96	1.93/2.06
	DTI	2.92/3.11	2.93/3.49	3.68/3.61	3.15/3.56	3.61/3.68	3.68/4.12
	Difference	2.39/2.68	2.59/3.27	3.14/3.16	2.78/3.32	2.90/3.09	2.91/3.16
9	fMRI	1.66/1.68	1.62/1.65	1.17/1.74	1.56/1.63	1.72/1.85	1.78/1.83
	DTI	2.76/3.06	2.80/3.73	3.01/3.67	3.27/3.48	3.01/3.67	3.01/3.67
	Difference	3.11/2.43	2.65/3.39	2.41/3.07	2.91/3.07	2.61/3.16	2.71/3.28
10	fMRI	1.94/1.97	1.95/2.11	1.97/2.00	1.89/1.99	1.88/1.96	1.91/1.95
	DTI	3.40/3.44	3.43/3.63	3.72/2.78	3.12/3.66	3.72/3.78	3.72/3.78
	Difference	2.55/2.59	3.06/2.66	3.60/3.42	2.35/2.86	3.20/3.45	3.29/3.34

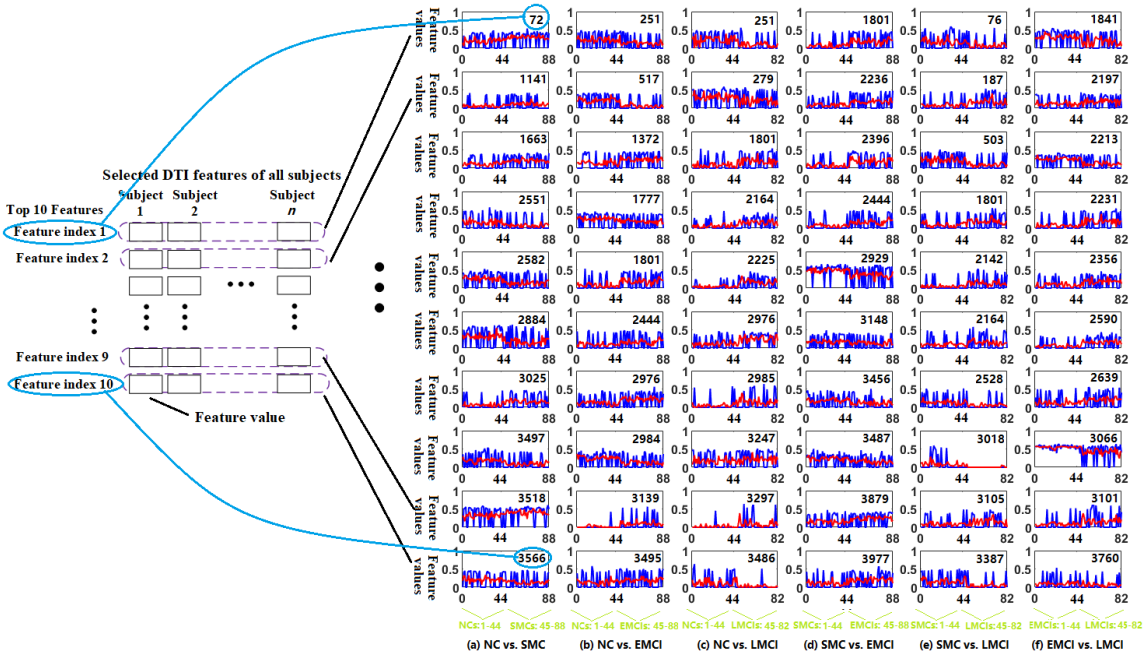
3.3 Effect of our adjacency matrix on feature values

Figure 6 visualises the top 10 most discriminative functional features and the top 10 most discriminative structural features and visualises feature values after pre-multiplying adjacency matrix. Figure 7 shows

t-SNE visualisation results of feature maps, and the detailed effect on the mean and standard deviation of feature values is shown in Tables 5-6. As FC and SC brain networks are usually represented by the selected most discriminative features from 1×4005 feature vectors, we use the indices of selected features in 1×4005 vector to represent them in this subsection. A features' index represents the relationship between pair ROIs whereas corresponding feature value represents the relationship weight.



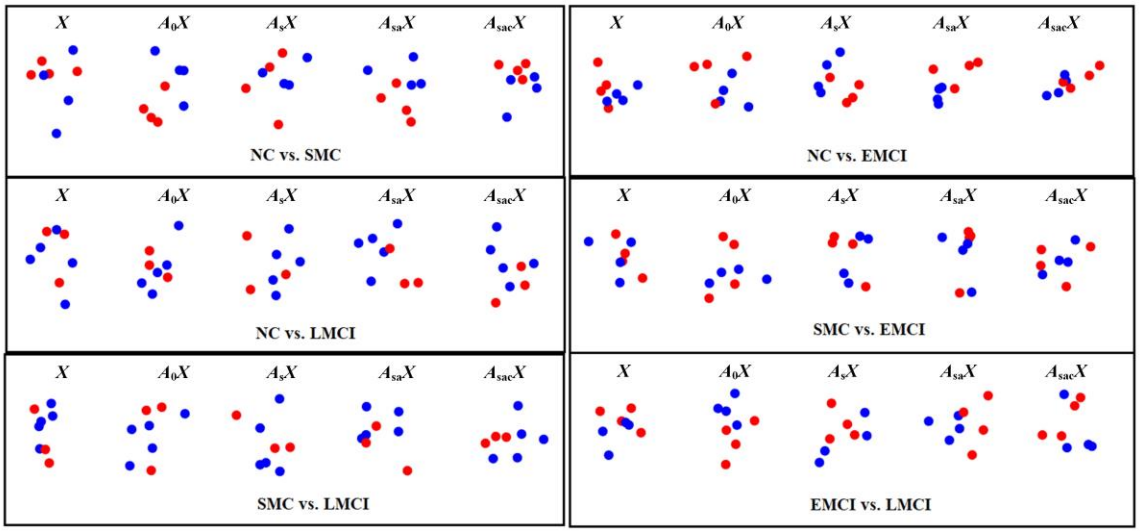
(a) fMRI feature values with and without pre-multiplying adjacency matrix.



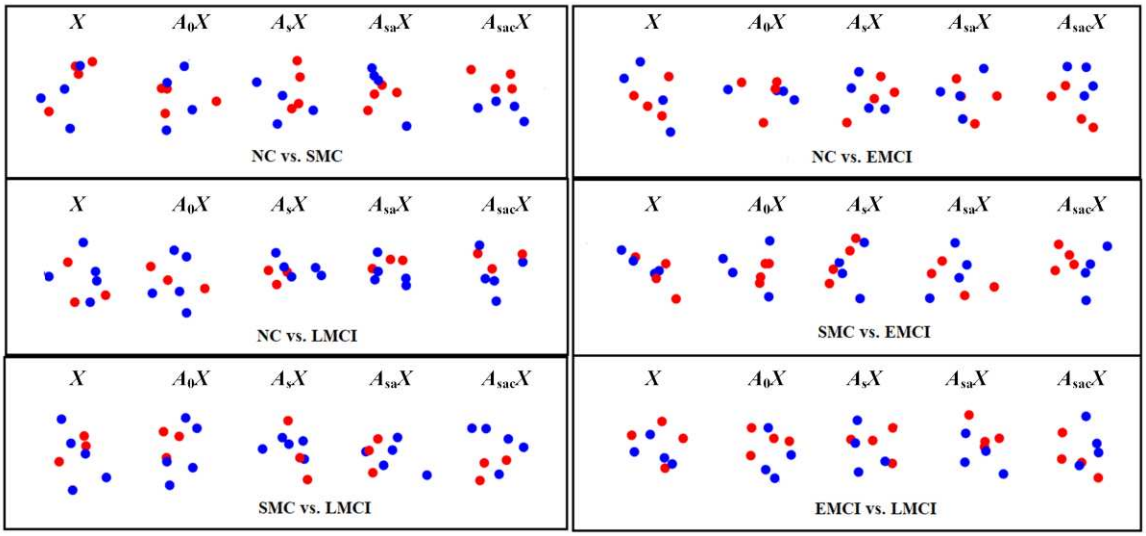
(b) DTI feature values with and without pre-multiplying adjacency matrix.

Figure 6: The top 10 most discriminative fMRI and DTI features in our six prediction tasks. The abscissae represent subjects' indices for prediction, and ordinates represent feature values. The blue line represents original feature values, and the red line represents feature values after pre-multiplying adjacency matrix $\boldsymbol{A}_{sac}$.

As shown in Figure 6, there are different noise levels among different features. For example, the noise in the number 3915 fMRI feature for NC vs. SMC is small, whereas the noise in the number 3797 fMRI feature is big. The noise in the number 3886 fMRI feature for SMC vs. LMCI is small, whereas the noise in the number 1153 fMRI feature is big. The noise level of the same feature between different disease statuses is consistent. For example, the noise level in the number 3519 fMRI feature for NC vs. SMC follows its noise level for NC vs. EMCI. The noise level in the number 251 DTI feature for NC vs. EMCI follows its noise level for NC vs. LMCI. By pre-multiplying our adjacency matrix $\boldsymbol{A}_{sac}$, the noises in all fMRI and DTI features are suppressed, as shown in Figure 6 that red line has a small fluctuation.



(a) t-SNE visualisation results based on fMRI data.



(b) t-SNE visualisation results based on DTI data.

Figure 7: The t-SNE visualisation results of fMRI and DTI feature maps in different tasks. The effect is shown by pre-multiplying the adjacency matrices $\boldsymbol{A}_0, \boldsymbol{A}_s, \boldsymbol{A}_{sa},$ and $\boldsymbol{A}_{sac}$ on $\boldsymbol{X}$. $\boldsymbol{X}$ is a feature matrix, which includes feature values of test subjects. As there are 82 or 88 subjects in our tasks and we use the 10-fold cross-validation strategy, there are usually eight subjects in the test set for every fold, Hence, the t-SNE visualisation results are based on the eight test samples. $\boldsymbol{A}_0$ represents the adjacency matrix constructed based on the traditional method, $\boldsymbol{A}_s$ represents the adjacency matrix constructed based on the traditional method and our similarity-aware receptive fields, $\boldsymbol{A}_{sa}$ represents the adjacency matrix constructed based on our similarity-aware receptive fields and our adaptive adjacency matrix, and $\boldsymbol{A}_{sac}$ represents the adjacency matrix constructed based on our similarity-aware receptive fields, adaptive mechanism and calibration mechanism.

Figure 7 describes the feature visualisation results of graph theory on the test set, and we have compared the effect of four kinds of adjacency matrices on feature values. As there are 82 or 88 subjects for every task and we use a 10-fold cross-validation strategy, there are typically eight subjects in the test set. As shown in Figure 7, compared with $\boldsymbol{X}$, $\boldsymbol{A}_0\boldsymbol{X}$ has a better visualisation result for some tasks. Specifically, for NC vs. SMC, SMC vs. EMCI, EMCI vs. LMCI based on fMRI data and for NC vs. SMC, NC vs. LMCI, SMC vs. EMCI, SMC vs. LMCI, EMCI vs. LMCI based on DTI data, it has a better visualisation result. For NC vs. EMCI, NC vs. LMCI, SMC vs. LMCI based on fMRI data and for NC vs. EMCI based on DTI data, the improvement is not obvious. Compared with $\boldsymbol{X}$, $\boldsymbol{A}_{sac}\boldsymbol{X}$ has a better visualisation result for our six tasks.

Tables 5-6 show the details of the experimental results. In the feature index column, we list the top 10 features' indices, which are selected by using RFE method. The feature's index represents the feature's position in the 1×4005 feature vector, which are formed by extracting upper triangular matrix elements from

474 the 90×90 brain network. We can see there are many differences in the top 10 features' indices between
475 different prediction tasks. Most of fMRI features' indices are different from DTI features' indices in the same
476 prediction task. For example, the top 10 fMRI features' indices for NC vs. SMC is [82, 170, 1339, 3520, 3768,
477 3797, 3894, 3908, 3915, 3941], whereas the top 10 DTI features' indices for NC vs. SMC is [72, 1141, 1663,
478 2551, 2582, 2884, 3025, 3497, 3518, 3566].

479 Tables 5-6 also describe the mean values and standard deviations of the top 10 feature values. Standard
480 deviations show the different noise levels of the top 10 features. For example, the number 3915 fMRI feature
481 in NC vs. SMC has a small standard deviation, which follows its appearance for NC vs. EMCI. This result
482 also follows in Figure 6. The number 2976 DTI feature for NC vs. EMCI has a big standard deviation, which
483 also follows its appearance for NC vs. LMCI. This result is also consistent with Figure 6. The consistency of
484 mean value and standard deviation for the same feature in different prediction tasks shows the stability of our
485 fMRI and DTI data, but also shows there is a little fluctuation between the same features in different subjects
486 although they have same disease status.

487 Tables 5-6 also describe the effect of disease status on feature values. Tables 5-6, show different disease
488 states have different mean values in all prediction tasks. For example, in Table 5, the mean value of the
489 number 82 fMRI feature of all NC subjects is 0.69, whereas its mean value of SMC subjects is 0.73. This
490 difference between different disease statuses provides the foundation to predict disease. Compared with the
491 effect of disease status on fMRI feature values in Table 5, the effect on DTI feature values in Table 6 appears
492 much more apparent. For example, for NC vs. SMC, the mean difference of mean values of the top 10 fMRI
493 features is 0.04, whereas the mean difference of the top 10 DTI features is 0.1. The more obvious discrimi-
494 native DTI features make the prediction tasks easier, and this follows the results in Tables 3, whereas the
495 performance of our method and traditional methods based on DTI data is much better than the performance
496 based on fMRI data.

497 The effectiveness of t-test method (Arbabshirani et al., 2017; Dietterich, 1998) for feature selection and
498 the work (Huang et al., 2020) suggest that big mean difference and small standard deviation are beneficial for
499 classification. As shown in Figure 6, Table 5 and Table 6, by pre-multiplying adjacency matrix $\boldsymbol{A}_{sac}$, the
500 standard deviations become smaller, and the results in Figure 7 validate that pre-multiplying adjacency
501 matrix can improve final classification performance.

Table 5: Effect of our adjacency matrix $\boldsymbol{A}_{sac}$ on the top 10 most discriminative fMRI feature values in our six classification tasks. We compare fMRI features' mean values and standard deviations with or without pre-multiplying adjacency matrix $\boldsymbol{A}_{sac}$, and compare fMRI features' mean values between different disease status. The mean column is measured on $\boldsymbol{A}_{sac}\boldsymbol{X}$, $\boldsymbol{A}_{sac}$ represents our adaptive calibrated adjacency matrix, and $\boldsymbol{X}$ represents the top 10 fMRI feature values of all subjects on the graph.

NC vs. SMC				NC vs. EMCI				NC vs. LMCI			
Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (NC/SMC)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (NC/EMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (NC/LMCI)
82	0.71±0.11	0.71±0.04	0.69/0.73	161	0.70±0.11	0.70±0.04	0.72/0.67	455	0.85±0.09	0.86±0.03	0.85/0.87
170	0.70±0.11	0.70±0.04	0.67/0.73	1652	0.69±0.11	0.70±0.04	0.67/0.72	519	0.74±0.09	0.74±0.04	0.77/0.71
1339	0.70±0.09	0.70±0.04	0.67/0.73	1720	0.67±0.11	0.68±0.04	0.65/0.70	976	0.81±0.08	0.81±0.03	0.79/0.83
3520	0.72±0.11	0.71±0.03	0.73/0.70	2728	0.78±0.09	0.78±0.04	0.81/0.76	1587	0.63±0.13	0.62±0.05	0.62/0.63
3768	0.64±0.12	0.64±0.04	0.65/0.64	3499	0.66±0.14	0.65±0.04	0.64/0.66	1659	0.66±0.10	0.67±0.04	0.68/0.64
3797	0.59±0.14	0.59±0.04	0.56/0.61	3737	0.69±0.10	0.68±0.03	0.67/0.70	1839	0.63±0.11	0.63±0.04	0.62/0.65
3894	0.65±0.12	0.66±0.03	0.67/0.64	3777	0.59±0.12	0.58±0.04	0.55/0.61	3489	0.79±0.09	0.78±0.05	0.79/0.78
3908	0.67±0.11	0.67±0.04	0.70/0.64	3915	0.94±0.03	0.94±0.01	0.94/0.95	3498	0.70±0.11	0.70±0.05	0.67/0.73
3915	0.94±0.03	0.94±0.01	0.94/0.95	3961	0.93±0.04	0.93±0.01	0.94/0.92	3777	0.59±0.13	0.60±0.05	0.56/0.64
3941	0.89±0.06	0.89±0.02	0.91/0.88	3971	0.86±0.08	0.87±0.03	0.89/0.85	3971	0.87±0.07	0.87±0.03	0.89/0.85
SMC vs. EMCI				SMC vs. LMCI				EMCI vs. LMCI			
Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (SMC/EMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (SMC/LMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (EMCI/LMCI)
59	0.74±0.09	0.74±0.03	0.73/0.75	166	0.70±0.09	0.70±0.03	0.68/0.71	737	0.69±0.08	0.69±0.02	0.70/0.68
499	0.68±0.12	0.68±0.04	0.71/0.66	432	0.74±0.10	0.74±0.03	0.72/0.75	835	0.59±0.10	0.58±0.03	0.56/0.60
666	0.68±0.08	0.68±0.03	0.70/0.66	1728	0.93±0.03	0.93±0.01	0.94/0.92	976	0.82±0.08	0.82±0.02	0.81/0.83
737	0.72±0.09	0.72±0.03	0.74/0.71	2052	0.69±0.10	0.69±0.02	0.70/0.67	1153	0.61±0.12	0.61±0.03	0.62/0.61
1367	0.82±0.11	0.82±0.03	0.81/0.82	2916	0.70±0.08	0.70±0.02	0.71/0.68	1230	0.76±0.09	0.77±0.03	0.78/0.75
1644	0.57±0.12	0.57±0.04	0.55/0.59	2925	0.89±0.06	0.89±0.01	0.89/0.89	2480	0.57±0.09	0.56±0.02	0.57/0.56
1877	0.63±0.10	0.63±0.03	0.62/0.65	3399	0.72±0.09	0.72±0.03	0.72/0.73	2779	0.78±0.08	0.78+0.03	0.76/0.80
2589	0.66±0.10	0.65±0.05	0.71/0.62	3544	0.79±0.09	0.79±0.02	0.80/0.78	3529	0.59±0.10	0.60+0.04	0.57/0.63
2639	0.63±0.10	0.63±0.05	0.67/0.60	3784	0.69±0.11	0.69±0.03	0.68/0.70	3877	0.74±0.10	0.75+0.03	0.73/0.77
3686	0.64±0.11	0.64±0.03	0.62/0.66	3984	0.75±0.10	0.75±0.04	0.77/0.72	3886	0.90±0.04	0.91+0.01	0.91/0.90

Table 6: Effect of our adjacency matrix $\boldsymbol{A}_{sac}$ on the top 10 most discriminative DTI feature values in our six classification tasks. We compare DTI features' mean values and standard deviations with or without pre-multiplying adjacency matrix $\boldsymbol{A}_{sac}$, and compare DTI features' mean values between different disease status. The mean column is measured on $\boldsymbol{A}_{sac}\boldsymbol{X}$, $\boldsymbol{A}_{sac}$ represents our adaptive calibrated adjacency matrix, and $\boldsymbol{X}$ represents the top 10 DTI feature values of all subjects on the graph.

NC vs. SMC				NC vs. EMCI				NC vs. LMCI			
Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{ac}\boldsymbol{X}$ (Mean±std)	Means (NC/SMC)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{ac}\boldsymbol{X}$ (Mean±std)	Means (NC/EMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{ac}\boldsymbol{X}$ (Mean±std)	Means (NC/LMCI)
72	0.24±0.20	0.24±0.07	0.20/0.29	251	0.18±0.21	0.17±0.07	0.22/0.13	251	0.17±0.21	0.16±0.10	0.22/0.08
1141	0.08±0.15	0.08±0.06	0.05/0.11	517	0.13±0.18	0.13±0.08	0.21/0.05	279	0.25±0.25	0.25±0.10	0.30/0.19
1663	0.15±0.17	0.15±0.06	0.11/0.19	1372	0.17±0.20	0.18±0.07	0.14/0.23	1801	0.13±0.20	0.12±0.09	0.07/0.18
2551	0.11±0.19	0.11±0.07	0.15/0.07	1777	0.21±0.18	0.21±0.06	0.25/0.18	2164	0.10±0.15	0.10±0.07	0.05/0.15
2582	0.19±0.21	0.19±0.08	0.25/0.13	1801	0.13±0.20	0.13±0.08	0.07/0.18	2225	0.09±0.14	0.09±0.07	0.03/0.16
2884	0.24±0.26	0.23±0.10	0.30/0.16	2444	0.13±0.20	0.13±0.06	0.09/0.16	2976	0.17±0.19	0.17±0.09	0.11/0.24
3025	0.10±0.18	0.10±0.06	0.06/0.15	2976	0.17±0.19	0.16±0.08	0.11/0.22	2985	0.08±0.20	0.09±0.08	0.04/0.15
3497	0.14±0.20	0.13±0.07	0.18/0.08	2984	0.18±0.18	0.19±0.08	0.25/0.12	3247	0.20±0.22	0.19±0.07	0.17/0.21
3518	0.37±0.22	0.36±0.07	0.32/0.40	3139	0.04±0.14	0.05±0.05	0.01/0.09	3297	0.04±0.15	0.05±0.07	0.01/0.08

3566	0.16±0.20	0.16±0.06	0.20/0.13	3495	0.16±0.22	0.16±0.06	0.13/0.19	3486	0.07±0.18	0.07±0.08	0.11/0.02
SMC vs. EMCI				SMC vs. LMCI				EMCI vs. LMCI			
Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (SMC/EMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (SMC/LMCI)	Feature index	$\boldsymbol{X}$ (Mean±std)	$\boldsymbol{A}_{sac}\boldsymbol{X}$ (Mean±std)	Means (EMCI/LMCI)
1801	0.13±0.20	0.12±0.07	0.06/0.18	76	0.25±0.23	0.24±0.10	0.30/0.16	1841	0.12±0.21	0.12±0.08	0.17/0.06
2236	0.13±0.18	0.12±0.06	0.08/0.17	187	0.14±0.17	0.14±0.06	0.11/0.18	2197	0.11±0.19	0.11±0.07	0.07/0.16
2396	0.12±0.19	0.13±0.09	0.06/0.19	503	0.18±0.17	0.18±0.08	0.25/0.11	2213	0.14±0.20	0.15±0.09	0.22/0.09
2444	0.11±0.19	0.11±0.07	0.06/0.16	1801	0.12±0.20	0.12±0.08	0.06/0.18	2231	0.12±0.19	0.13±0.07	0.08/0.17
2929	0.42±0.24	0.41±0.09	0.48/0.34	2142	0.13±0.17	0.12±0.07	0.08/0.17	2356	0.08±0.17	0.08±0.06	0.04/0.13
3148	0.15±0.18	0.15±0.05	0.17/0.13	2164	0.10±0.15	0.10±0.06	0.05/0.16	2590	0.11±0.19	0.11±0.05	0.09/0.14
3456	0.13±0.18	0.14±0.06	0.18/0.10	2528	0.15±0.21	0.15±0.07	0.11/0.19	2639	0.09±0.17	0.09±0.06	0.13/0.05
3487	0.19±0.17	0.19±0.06	0.23/0.15	3018	0.48±0.18	0.48±0.08	0.55/0.41	3066	0.04±0.15	0.04±0.06	0.08/0.00
3879	0.19±0.19	0.19±0.07	0.14/0.24	3105	0.11±0.21	0.12±0.07	0.07/0.17	3101	0.10±0.17	0.11±0.07	0.08/0.14
3977	0.15±0.20	0.15±0.07	0.10/0.20	3387	0.07±0.16	0.07±0.04	0.09/0.06	3760	0.11±0.19	0.10±0.08	0.16/0.04

4. Discussion

4.1 Effect of phenotypic information

Non-imaging phenotypic information (e.g., equipment type and gender) is a factor to affect imaging. For example, different equipment types probably use different imaging parameters, and this finally results in some differences in the extracted image features. An advantage of GCN algorithms is integrating non-imaging phenotypic information into edge weights on graphs, as shown in Eqs. (1) and (2). For a subject on a graph, there is a convolution filter as shown in Figure 2. The convolution filter uses the features from other subjects to update the features of the subject being analysed, and edge weights are corresponding to the convolution coefficients. In view the differences resulted by equipment type and gender on image features, we assign a bigger edge weight between the pair subjects with the same equipment type and gender, as shown in Eqs. (1) and (2). The non-imaging phenotypic information is not used as a biomarker to supplement extracted features. In contrast, it is used to establish a more adequate and practical graph. As shown by Parisot et al. (Parisot et al., 2018), the gender and equipment type is vital information for graph construction in AD and ASD prediction, which result in 3% improvement on the final accuracy. Considering the characteristics of our tasks, we also investigate the effect of phenotypic information on final prediction accuracy, and the results in our six prediction tasks are shown in Figure 8. The combination of phenotypic information and a similarity function is shown in Eqs. (1) and (2).

In this experiment, we observe apparent variations on accuracy. Specifically, the performance based on the only one similarity is the worst, whereas the performance based on similarity of both phenotypic information (gender and equipment type) is the best. The difference between the best and the worst performing

graphs in our six prediction tasks are 12.1%, 8.1%, 4.8%, 8.2%, 8.5% and 4.6%, respectively. Gender appears to have a more considerable influence on accuracy than the imaging equipment used. This shows that features with different gender in our tasks have many differences. These findings are consistent with the previous study by (Parisot et al., 2018).

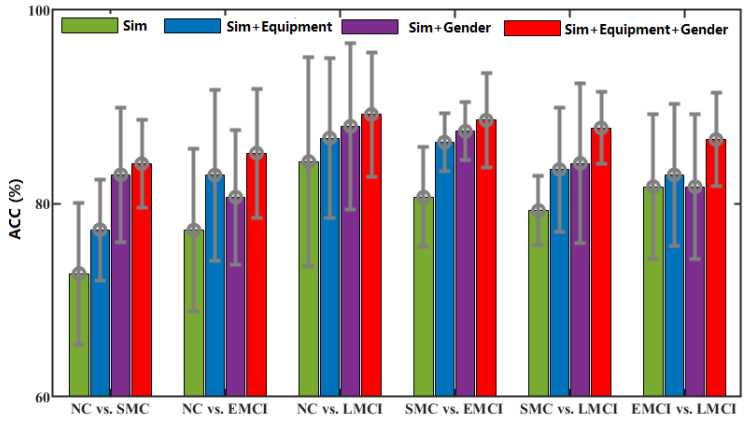


Figure 8: Influence of phenotypic information on the prediction accuracy in our six prediction tasks.

4.2 Effect of the number of the selected features

RFE is adopted to select features in the paper due to its promising performance. As it recursively removes attributes and builds the model using the remaining attributes, the number of features needs to be set to a reasonable value. We test the influence of the selected features' number through experiment, and its influence in all classification tasks on ACC is shown in Figure 9. In Figure 9, the number of the selected features varies from 0 to 300 with a step 10. The ACC values in all classification tasks increase as the number increases starting from zero, then the performance maintains a little fluctuation with the number further increasing. Eventually, after exceeding a specific value, the further increase in the number results in performance deterioration. In our six prediction tasks, the ACC values reach the best with the number varying about from 40 to 80. For NC vs. SMC, the performance deteriorates rapidly with the number increasing over about 80. For EMCI vs. LMCI, the performance deteriorates rapidly with the number over 160. These results validate that the number of the selected features need to be set as a reasonable value. A large number can increase system burden and cause performance deterioration, while a small number cannot represent the subject's information. Therefore, we set the number of the selected features in all tasks as 50 in this paper.

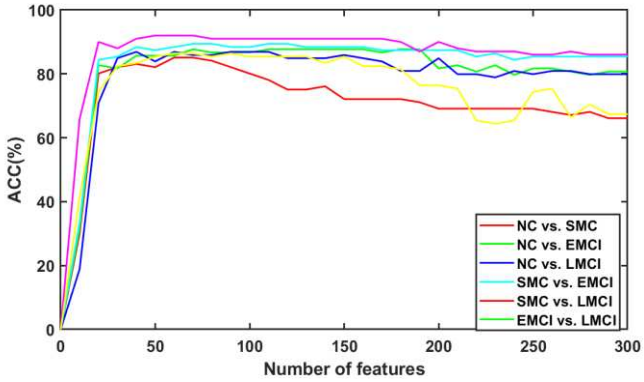


Figure 9: Effect of the number of the selected features on prediction accuracy in our six prediction tasks.

4.3 Parameters of weight mechanism

We have developed two GCN models according to functional data and structural data. After our dual-modal GCN, we get a functional score and structural score for every subject. Namely, we use a combined weight mechanism to combine the two scores to perform the final prediction. For example, the final predicted score for a subject v is denoted as $w_1 \times Score_v^f + w_2 \times Score_v^s$. The parameters w_1 and w_2 are selected according to our experimental results. In this subsection, we show the effect of different weight parameters on performance in Table 7.

Table 7: Effect of different weight parameters on accuracy in our six classification tasks.

Parameters	NC vs. SMC	NC vs. EMCI	NC vs. LMCI	SMC vs. EMCI	SMC vs. LMCI	EMCI vs. LMCI
w_1=0.1, w_2=0.9	81.82±5.97	82.95±6.52	87.80±3.88	86.39±4.38	86.58±6.11	82.92±3.51
w_1=0.2, w_2=0.8	82.95±6.44	84.09±6.44	87.80±5.39	87.51±4.09	86.58±6.11	84.14±3.51
w_1=0.3, w_2=0.7	82.95±5.97	82.95±6.52	89.02±6.00	88.63±5.23	87.80±5.23	84.14±4.57
w_1=0.4, w_2=0.6	84.09±4.57	84.09±6.44	89.02±7.02	89.75±4.38	89.02±4.38	85.36±4.57
w_1=0.5, w_2=0.5	84.09±4.57	85.22±6.65	89.02±6.44	88.63±4.86	87.80±3.74	86.58±4.86
w_1=0.6, w_2=0.4	82.97±4.72	85.22±6.65	86.58±7.05	88.63±3.74	86.58±3.92	84.14±5.36
w_1=0.7, w_2=0.3	79.54±6.92	84.09±7.43	85.36±6.52	87.51±6.41	84.14±4.38	81.70±6.95
w_1=0.8, w_2=0.2	79.54±6.95	79.54±6.65	84.14±5.52	84.09±7.76	81.70±5.52	80.48±8.46
w_1=0.9, w_2=0.1	77.27±7.52	76.12±6.30	84.14±6.11	82.97±7.05	79.26±5.8	79.26±7.76

As Table 7 shows, different combined weight coefficients have an obvious influence on the final prediction accuracy. According to the above results, we set w_1=0.5 and w_2=0.5 in our six tasks.

4.4 Visualisation of the adjacency matrix

The proposed similarity-aware receptive fields, adaptive mechanism and calibration mechanism play the role to improve adjacency matrix and eventually result in better performance. To describe the effect of the above methods on the adjacency matrix, we use imagesc() function in MATLAB to show four kinds of adjacency matrices. In Figure 10, there are four functional adjacency matrices and four structural adjacency matrices, where $\boldsymbol{A}_0$ represents the adjacency matrix constructed based on the traditional method, $\boldsymbol{A}_s$ represents the adjacency matrix constructed based on the traditional method and our similarity-aware receptive fields, $\boldsymbol{A}_{sa}$ represents the adjacency matrix constructed based on our similarity-aware receptive fields and our adaptive adjacency matrix, and $\boldsymbol{A}_{sac}$ represents the adjacency matrix constructed based on our similarity-aware receptive fields, adaptive mechanism and calibration mechanism.

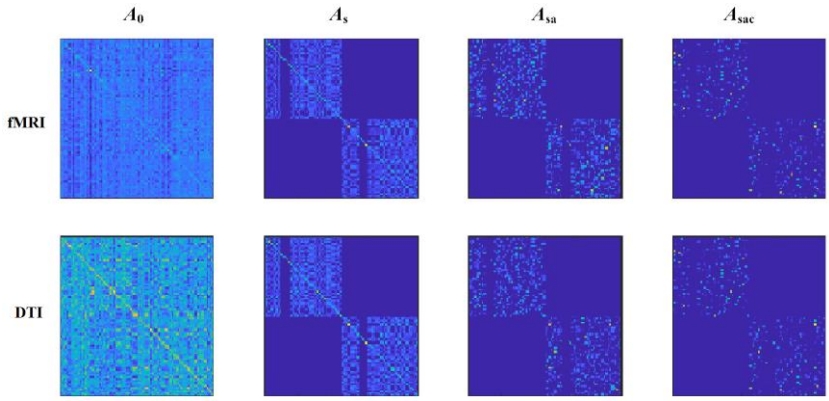


Figure 10: Visualisation results of kinds of adjacency matrices.

As shown in Figure 10, the adjacency matrix $\boldsymbol{A}_0$ constructed by using the traditional method is a dense matrix. After using our similarity-aware receptive fields, it becomes much sparse as the similarity-aware receptive fields ignore a part of connections. For the adjacency matrix $\boldsymbol{A}_0$ constructed by using the traditional method, there are many differences between functional and structural adjacency matrices. After using our three mechanisms, we finally get a stable and united adjacency matrix $\boldsymbol{A}_{sac}$.

4.5 Most discriminative connectivity features

Tables 8-9 list the top 10 most discriminative connectivity features and related ROI brain regions in six classification tasks. For fMRI data, we can see that many of these selected brain regions follow the observations reported in the previous studies. For example, the right olfactory cortex (OLF.R) (Li et al., 2020a; Sun et al., 2012; Tekin and Cummings, 2002; Vasavada et al., 2015; Yu et al., 2019; Zhang et al., 2018), left

hippocampus (HIP.L) (Salvatore et al., 2015; Zhang et al., 2018), left calcarine cortex(CAL.L) (Li et al., 2020a; Xu et al., 2016) are usually reported as highly associated with AD/MCI pathology. However, there are many differences in the top 10 most discriminative connectivity features between our six prediction tasks and two modalities. As shown in Figure 9, the performance of our six prediction tasks is saturated when the number of the selected features is set as 30. Therefore, we show the top 30 discriminative connectivity features for the FC network and SC network in Figure 11. As shown in Figure 11, there are many differences in the top 30 most discriminative connectivity features between different prediction tasks and different modalities. In the literature (Li et al., 2019b, 2020a; Wee et al., 2014; Yu et al., 2019; Zhang et al., 2018), there are also many differences in the top 10 most discriminative connectivity features and the top 10 most discriminative ROIs for SMC vs. NC. Based on above differences in our paper and literature, the different noise levels of the top 10 feature values in Tables 5-6, and the influence of selected features' number in Figure 9, we conclude there are several hundred connectivity features are associated with prediction tasks. This conclusion follows the results in the literature (Parisot et al., 2018), where GCN obtains the best performance when using RFE to select 2000 features, or using MLP to select 250 features, or using Autoencoder (AE) to select 500 features. The above results also show that different construction methods of brain network and feature selection methods can cause obvious difference in most discriminative connectivity features.

Table 8: The top 10 most discriminative fMRI features and their corresponding ROIs in our six classification tasks.

NC vs. SMC			NC vs. EMCI			NC vs. LMCI		
Feature	ROI index	ROI name	Feature	ROI index	ROI name	Feature	ROI index	ROI name
82	1,83	PreCG.L, TPOsup.L	161	2,74	PreCG.R, PUT.R	455	6,26	ORBsup.R, ORBsupmed.R
170	2,83	PreCG.R, TPOsup.L	1652	21,83	OLF.L, TPOsup.L	519	6,90	ORBsup.R, ITG.R
1339	17,52	ROL.L, MOG.R	1720	22,83	OLF.R, TPOsup.L	976	12,64	IFGoperc.R, SMG.R
3520	59,70	SPG.L, PCL.R	2728	39,88	PHG.L, TPOmid.R	1587	20,87	SMA.R, TPOmid.L
3768	68,84	PCUN.R, TPOsup.R	3737	67,75	PCUN.L, PAL.L	1659	21,90	OLF.L, ITG.R
3797	70,72	PCL.R, CAU.R	3777	69,72	PCL.L, CAU.R	1839	24,69	SFGmed.R, PCL.L
3894	75,84	PAL.L, TPOsup.R	3915	77,78	THA.L, THA.R	3489	58,70	PoCG.R, PCL.R
3908	76,84	PAL.R, TPOsup.R	3499	58,80	PoCG.R, HES.R	3498	58,79	PoCG.R, HES.L
3915	77,78	THA.L, THA.R	3961	81,82	STG.L, STG.R	3777	69,72	PCL.L, CAU.R
3941	79,81	HES.L, STG.L	3971	82,84	STG.R, TPOsup.R	3971	82,84	STG.R, TPOsup.R
SMC vs. EMCI			SMC vs. LMCI			EMCI vs. LMCI		
Feature	ROI index	ROI name	Feature	ROI index	ROI name	Feature	ROI index	ROI name
59	1, 60	PreCG.L, SPG.R	737	9, 62	ORBmid.L, IPL.R	166	2, 79	PreCG.R, HES.L
499	6, 70	ORBsup.R, PCL.R	835	10, 88	ORBmid.R,TPOmid.R	432	5, 87	ORBsup.L, TPOmid.L
666	8, 72	MFG.R, CAU.R	976	12, 64	IFGoperc.R, SMG.R	1728	23, 24	SFGmed.L, SFGmed.R
737	9, 62	ORBmid.L, IPL.R	1153	14, 88	IFGtriang.R,TPOmid.R	2052	27, 90	REC.L, ITG.R
1367	17, 80	ROL.L, HES.R	1230	15, 90	ORBinf.L, ITG.R	29	43, 82	CAL.L, STG.R
1644	21, 75	OLF.L, PAL.L	2480	35, 50	PCG.L, SOG.R	2925	44, 45	CAL.R, CUN.L
1877	25, 42	ORBsupmed.L,AMYG.R	2779	40, 89	PHG.R, ITG.L	3399	55, 79	FFG.L, HES.L

2589	37, 52	HIP.L, MOG.R	3529	59, 79	SPG.L, HES.L	3544	60, 64	SPG.R, SMG.R
2639	38, 50	HIP.R, SOG.R	3877	74, 82	PUT.R, STG.R	3784	69, 79	PCL.L, HES.L
3686	65, 71	ANG.L, CAU.L	3886	75, 76	PAL.L, PAL.R	3984	83, 90	TPOsup.L, ITG.R

Table 9: The top 10 most discriminative DTI features and their corresponding ROIs in our six classification tasks.

NC vs. SMC			NC vs. EMCI			NC vs. LMCI		
Feature	ROI index	ROI name	Feature	ROI index	ROI name	Feature	ROI index	ROI name
72	1,73	PreCG.L, PUT.L	251	3,77	SFGdor.L, THA.L	251	3,77	SFGdor.L, THA.L
1141	14,76	IFGtriang.R, PAL.R	517	6,88	ORBsup.R, TPOmid.R	279	4,19	SFGdor.R, SMA.L
1663	22,26	OLF.R, ORBsupmed.R	1372	17,85	ROL.L, MTG.L	1801	24,31	SFGmed.R, ACG.L
2551	36,67	PCG.R, PCUN.L	1777	23,73	SFGmed.L, PUT.L	2164	29,79	INS.L, HES.L
2582	37,45	HIP.L, CUN.L	1801	24,31	SFGmed.R, ACG.L	2225	30,80	INS.R, HES.R
2884	43,50	CAL.L, SOG.R	2444	34,69	DCG.R, PCL.L	2976	45,51	CUN.L+R, MOG.L
3025	46,56	CUN.R, FFG.R	2976	45,51	CUN.L+R, MOG.L	2985	45,60	CUN.L+R, SPG.R
3497	58,78	PoCG.R, THA.R	2984	45,59	CUN.L+R, SPG.L	3247	51,73	MOG.L, PUT.L
3518	59,68	SPG.L, PCUN.R	3139	48,85	LING.R, MTG.L	3297	52,85	MOG.R, MTG.L
3566	60,86	SPG.R, MTG.R	3495	58,76	PoCG.R, PAL.R	3486	58,67	PoCG.R, PCUN.L
SMC vs. EMCI			SMC vs. LMCI			EMCI vs.LMCI		
Feature	ROI index	ROI name	Feature	ROI index	ROI name	Feature	ROI index	ROI name
1801	24,31	SFGmed.R, ACG.L	76	1,77	PreCG.L, THA.L	1841	24,71	SFGmed.R, CAU.L
2236	31,32	ACG.L, ACG.R	187	3,13	SFGdor.L, IFGtriang.L	2197	30,52	INS.R, MOG.R
2396	33,77	DCG.L, THA.L	503	6,74	ORBsup.R, PUT.R	2213	30,68	INS.R, PCUN.R
2444	34,69	DCG.R, PCL.L	1801	24,31	SFGmed.R, ACG.L	2231	30,86	INS.R, MTG.R
2929	44,49	CAL.R, SOG.L	2142	29,57	INS.L, PoCG.L	2356	33,37	DCG.L, HIP.L
3148	49,53	SOG.L, IOG.L	2164	29,79	INS.L, HES.L	2590	37,53	HIP.L, IOG.L
3456	57,69	PoCG.L, PCL.L	2528	36,44	PCG.R, CAL.R	2639	38,50	HIP.R, SOG.R
3487	58,68	PoCG.R, PCUN.R	3018	46,49	CUN.R, SOG.L	3066	47,54	LING.L, IOG.R
3879	74,84	PUT.R, TPOsup.R	3105	48,51	LING.R, MOG.L	3101	47,89	LING.L, ITG.L
3977	82,90	STG.R, ITG.R	3387	55,67	FFG.L, PCUN.L	3760	68,76	PCUN.R, PAL.R

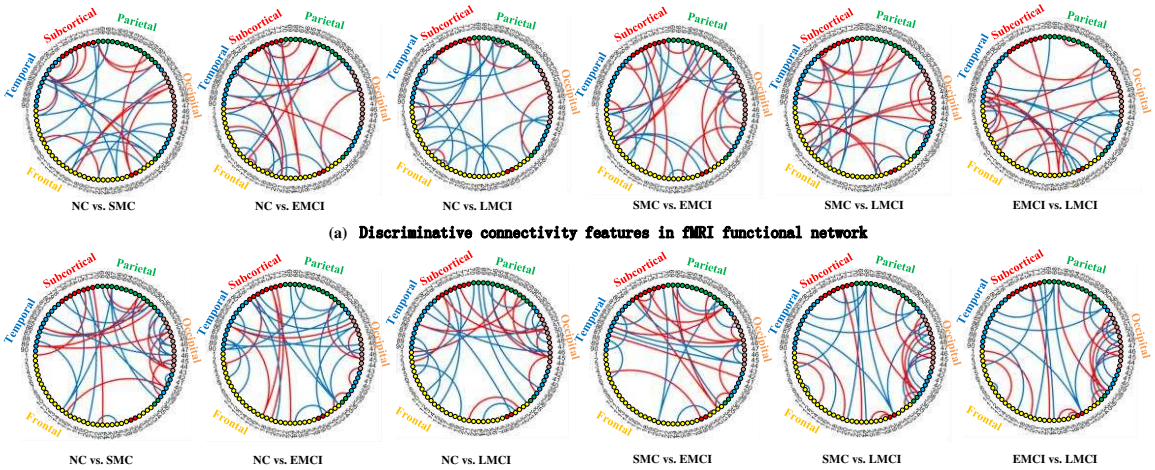


Figure 11: Top 30 discriminative connectivity features in fMRI and DTI brain connection networks in our six prediction tasks.

Table 10: Algorithm comparison with the related works.

References	Modality	Subject	Method	Task	ACC	SEN	SPE
(Wee et al., 2016)	fMRI	29 EMCI, 30 NC	Fused multiple graphical lasso	EMCI vs. NC	79.6	75.8	70.0

(Yu et al., 2017)	fMRI	50 MCI, 49 NC	Weighted Sparse Group Representation	MCI vs. NC	84.8	91.2	78.5
(Guo et al., 2017)	fMRI	33 EMCI, 32 LMCI, 28 NC	Multiple Features of Hyper-Network	EMCI vs. NC	72.8	78.2	67.1
				LMCI vs. NC	78.6	82.5	72.1
(Li et al., 2020b)	fMRI+DTI	36MCI, 37NC	Adaptive dynamic functional connectivity	MCI vs. NC	87.7	88.9	86.5
(Zhu et al., 2019)	MRI+PET+ CSF	99MCI, 53NC	SPMRM model	MCI vs. NC	83.5	95.0	62.8
(Lei et al., 2020)	fMRI+DTI	40 LMCI, 77 EMCI, 67 NC	Low-Rank Self-calibrated Brain Network, Joint Non-Convex Multi-Task Learning	NC vs. SMC	82.9	88.6	77.2
				NC vs. EMCI	85.2	86.3	84.1
				NC vs. LMCI	87.8	84.2	90.9
				SMC vs. EMCI	84.0	81.8	86.3
				SMC vs. LMCI	90.2	89.4	90.9
				EMCI vs. LMCI	81.7	78.9	84.0
Ours	fMRI+DTI	40 LMCI, 77 EMCI, 67 NC	Similarity-aware adaptive calibrated GCN	NC vs. SMC	84.9	88.6	79.5
				NC vs. EMCI	85.2	90.9	79.5
				NC vs. LMCI	89.0	89.4	88.6
				SMC vs. EMCI	88.6	95.4	81.8
				SMC vs. LMCI	87.8	84.2	90.9
				EMCI vs. LMCI	85.5	92.1	81.8

4.6 Comparison to the related prior works

Besides investigating our three mechanisms and parameters of GCN impact prediction performance, we further compare our SAC-GCN method with other different competing methods in the corresponding papers. Table 10 shows the comparison results. We can observe that our proposed method has achieved promising performance. Apart from good prediction performance, our proposed method does not need to construct complex brain connection networks. Hence, it has a good application prospect in other prediction tasks.

In our earlier work (Lei et al., 2020), we proposed to use self-calibrated low-rank regularisation to construct fMRI functional network, concatenated fMRI and DTI features. We used a multi-task learning framework to select the most discriminative features for final prediction. Although the work archived good performance, it ignores to integrate phenotypic information and the interactions between subjects. Compared to it, our SAC-GCN has good performance without constructing complicated brain connection networks. The proposed method is not limited to the tasks in this paper, and can flexibly be adapted to other multi-modal tasks.

5. Conclusion

In this paper, we propose three mechanisms to improve GCNs for SMC and MCI prediction. These mechanisms improve prediction performance significantly by establishing a more accurate adjacency matrix.

In the adjacency matrix, the similarity-aware receptive fields consider the disease status of those subjects in the training set and constrain the receptive field of labelled subjects to those subjects with the same status. The adaptive mechanism uses pre-trained GCNs to score all subjects and then uses score difference to replace correlation distance to update similarity. Besides, the calibration mechanism fuses dual-modal information into the adjacency matrix. Our experimental results on SAC-GCNs show significant improvement over traditional GCNs. To reveal the reason for good performance, we describe how our mechanisms improve the adjacency matrix and then describe its filtering effect by analysing feature values. Despite the superior performance, our SAC-GCN has a more straightforward structure and practical application prospect in other prediction tasks. In our future work, we will improve our calibration mechanism and extend this work to multi-task classification.

Acknowledgements

This work was supported partly by China Postdoctoral Science Foundation (Nos. 2019M653014), National Natural Science Foundation of China (Nos. U1902209, U1902209, 61871274, and 61801305), Guangdong Pearl River Talents Plan (2016ZT06S220), Shenzhen Peacock Plan (Nos. KQTD2016053112051497 and KQTD2015033016104926), and Shenzhen Key Basic Research Project (Nos. JCYJ20180507184647636, JCYJ20170818142347251, JCYJ20170818094109846 and JCYJ20170413152804728), Royal Academy of Engineering Chair in Emerging Technologies Scheme (CiET1819/19), Pengcheng Visiting Scholars Programme from the Shenzhen Government. *(Corresponding authors: Baiying Lei* leiby@szu.edu.cn). The asterisk indicates corresponding authors.

References

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 145, 137-165.

Association, A., 2018. 2018 Alzheimer's disease facts and figures. Alzheimers Dement. 14(3), 367-429.

Bapat, R.B., 2010. Graphs and matrices. Springer, New York.

Breiman, L., 2001. Random forests. Mach. Learn. 45(1), 5-32.

Cortes, C., Vapnik, V., 1995. Support-vector network. Mach. Learn. 20(3), 273-297.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems, pp. 3844-3852.

Dietterich, T.G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput. 10(7), 1895-1923.

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K, Belleville, S., Brodaty, H., Bennett, D., Chertkow, H, et al., 2006. Mild cognitive impairment. Lancet 367, 1262-1270.

Guo, H., Zhang, F., Chen, J., Xu, Y., Xiang, J., 2017. Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer's disease. Front. Neurosci. 11, 615-636.

Goto, M., Abe, O., Aoki, S., Hayashi, N., Miyati, T., Takao, H., Iwatsubo, T., Yamashita, F., Matsuda, H., Mori, H., et al., 2013. Diffeomorphic anatomical registration through exponentiated lie algebra provides reduced effect of scanner for cortex volumetry with atlas-based method in healthy subjects. Neuroradiology 55, 869-875.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389-422.

Hampel, H., Lista, S., 2016. Dementia: The rising global tide of cognitive impairment. Nat. Rev. Neurol. 12, 131-132.

Huang, Z., Zhu, Z., Yau, C., Tan, K., 2020. Identifying autism spectrum disorder from resting-state fMRI using deep belief network. IEEE T. Neur. Net. Lear. 14(8), 1-15.

Kazi, A., Shekarforoush, S., Krishna, S.A., Burwinkel, H., Vivar, G., Kortüm, K., Ahmadi, S.A., Albarqouni, S., Navab, N., 2019. InceptionGCN: Receptive field aware graph convolutional network for disease prediction. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 73-85.

Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representat*ions*, arXiv:1609.02907.

Ktena, S.I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., Rueckert, D., 2018. Metric learning with spectral graph convolutions on brain connectivity networks. Neuroimage 169, 431-442.

Lei, B., Cheng, N., Frangi, A.F., Tan, E., Cao, J., Yang, P., Elazab, A., Du, J., Xu, Y., Wang, T., 2020. Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. Med. Image Anal. 61, 101652.

Li, Y., Liu, J., Gao, X., Jie, B., Kim, M., Yap, P., Wee, C.Y., Shen, D., 2019a. Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification. Med. Image Anal. 52, 80-96.

Li, Y., Liu, J., Peng, Z., Sheng, C., Kim, M., Yap, P., Wee, C.Y., Shen, D., 2020a. Fusion of ULS group constrained high- and low-order sparse functional connectivity networks for MCI classification. Neuroinformatics 18, 1-24.

Li, Y., Liu, J., Tang, Z., Lei, B., 2020b. Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification. IEEE Trans. Med. Imaging 39(9), 2818-2830.

Li, Y., Yang, H., Lei, B., Liu, J., Wee, C.Y., 2019a. Novel effective connectivity inference using ultra-group constrained orthogonal forward regression and elastic multilayer perceptron classifier for MCI identification. IEEE Trans. Med. Imaging 38(5), 1227-1239.

Liu, Z., Chen, C., Li, L., Zhou, J., Li, X., Song, L., Qi, Y., 2019. Geniepath: Graph neural networks with adaptive receptive paths. In: Proceedings of the AAAI Conference on Artificial Intelligence, 33, pp. 4424-4431.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., Rueckert, D., 2018. Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. 48, 117-130.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825-2830.

Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L., Shen, D., 2016. Estimating functional brain networks by incorporating modularity prior. NeuroImage 141, 399-407.

Qiao, L., Zhang, L., Chen, S., Shen, D., 2018. Data-driven graph construction and graph learning: A review. Neurocomputing 312, 336-351.

Salvatore, C., Cerasa, A., Battista, P., Gilardi, M.C., Quattrone, A., Castiglioni, I., the Alzheimer's Disease Neuroimaging Initiative 2015. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach. Front. Neurosci. 9, 307.

Shi, T., Horvath, S., 2006. Unsupervised learning with random forest predictors. J. Comput. Graph. Stat. 15(1), 118-138.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Modern Pathol. 18, 547-557.

Shuman, D.I., Narang, S.K., Frossard, P., Ortega, A., Vandergheynst, P., 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process. Mag. 30(3), 83-98.

Sun, G.H., Raji, C.A., Maceachern, M., Burke, J.F., 2012. Olfactory identification testing as a predictor of the development of Alzheimer's dementia: A systematic review. Laryngoscope 122(7), 1455-1462.

Tekin, S., Cummings, J.L., 2002. Frontal-subcortical neuronal circuits and clinical neuropsychiatry: An update. J. Psychosom. Res. 53(2), 647-654.

Tong, T., Gray, K.R., Gao, Q., Chen, L., Rueckert, D., the Alzheimer's Disease Neuroimaging Initiative, 2017. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. Pattern Recognit. 63, 171-181.

Tzourio-mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273-289.

Vasavada, M., Wang, J., Eslinger, P.J., Gill, D.J., Sun, X., Karunanayaka, P., Yang, Q., 2015. Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. J. Alzheimer's Dis. 45(3), 947-958.

Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., He, Y., 2015. Gretna: A graph theoretical network analysis toolbox for imaging connectomics. Front. Hum. Neurosci. 9, 386.

Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., He, Y., 2013. Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. Biol. Psychiat. 73(5), 472-481.

Wee, C.Y., Yang, S., Yap, P., Shen, D., the Alzheimer's Disease Neuroimaging Initiative, 2016. Sparse temporally dynamic resting-state functional connectivity networks for early MCI identification. Brain Imaging Behav. 10, 342-356.

Wee, C.Y., Yap, P., Zhang, D., Wang, L., Shen, D., 2014. Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification. Brain Struct. Funct. 219, 641-656.

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., Jegelka, S., 2018. Representation learning on graphs with jumping knowledge networks. In: International Conference on Machine Learning, pp. 5449-5458.

Xu, L., Wu, X., Li, R., Chen, K., Long, Z., Zhang, J., Guo, X., Yao, L., the Alzheimer's Disease Neuroimaging Initiative, 2016. Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. J. Alzheimer's Dis. 51(4), 1045-1056.

Yu, R., Qiao, L., Chen, M., Lee, S., Fei, X., Shen, D., 2019. Weighted graph sregularised sparse brain network construction for MCI identification. Pattern Recognit. 90, 220-231.

Yu, R., Zhang, H., An, L., Chen, X., Wei, Z., Shen, D., 2017. Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification. Hum. Brain Mapp. 38(5), 2370-2383.

Zhang, Y., Zhan, L., Cai, W., Thompson, P.M., Huang, H., 2019. Integrating heterogeneous brain networks for predicting brain disease conditions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 214-222.