Biomedical Signal Processing and Control 86 (2023) 105212

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Research on early diagnosis of Alzheimer's disease based on dual fusion cluster graph convolutional network

Lu Meng *, Qianqian Zhang

College of Information Science and Engineering, Northeastern University, China

ARTICLE INFO

Keywords:
Graph convolution neural network
Multi-modal
Feature extraction
Adjacency matrix
ADNI dataset

ABSTRACT

Objective: Mild Cognitive Impairment (MCI) is an early stage of Alzheimer's Disease (AD), often mistaken for natural aging. Early detection and treatment of MCI are crucial for effective treatment, but the condition can be difficult to diagnose. In recent years, multi-modal data and deep learning methods have shown promise in this field. The objective of this study is to develop a computer-aided MCI diagnosis system that effectively processes multi-modal data using deep learning methods.

Method: We proposed a Dual Fusion Cluster Graph Convolution Network (DFCGCN) model, which combines two channels of feature extraction, one adjacency matrix, and the Cluster GCN in series. Brain imaging is down-sampled using graph pooling and flattened into sparse vectors, from which advanced features are extracted. Similarity between connectivity matrices is calculated using the Gaussian kernel function and combined with non-imaging details to construct a population graph that better represents inter-subject variability. Finally, features are assigned to subjects in the population graph, and node embeddings are learned using Cluster GCN to output diagnostic results.

Result: We tested the proposed algorithm on the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, achieving an accuracy, sensitivity, and specificity of 90.7%, 91.1%, and 94.0%, respectively.

Conclusion: The DFCGCN model presented in this study enhances the diagnosis of MCI and outperforms other state-of-the-art algorithms. This approach has potential to be a valuable tool for early detection and treatment of MCI.

1. Introduction

Alzheimer's Disease (AD) is a devastating and progressive neurodegenerative condition that has a profound impact on individuals and society at large. According to the International Alzheimer's Association, the number of individuals affected by AD worldwide is expected to reach 131.5 million by 2050. Given the significant personal and societal burden associated with this condition, early and accurate diagnosis, as well as effective treatment, are of critical importance. Therefore, research efforts aimed at developing innovative approaches for early detection and intervention are of great significance in the fight against AD.

Computer-aided diagnosis has been extensively studied to achieve timely and accurate detection and treatment of AD. However, some studies have focused on single-modal diagnosis. In this regard, Kanghan et al. [1] used a total of 694 structural MRI scans and proposed a volumetric convolutional neural network model for the pairwise classification of AD, MCI, and NC in four stages. The gradient visualization method was employed to study biomarker information related to progressive mild cognitive impairment (PMCI) and stable mild cognitive impairment (SMCI). The experiment was conducted on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, achieving classification accuracies of 73.95% and 86.60% for PMCI and AD, respectively, which outperformed other models while also identifying distinct regions in the temporal and parietal lobes. In another study, Yigit [2] used structural magnetic resonance imaging (SMRI) as input, and achieved 3D to 2D simplification by splitting slice data from axial, sagittal, and coronal dimensions. Then, they applied convolutional neural network (CNN) to diagnose AD and MCI on the open-access imaging studies in the aging series (OASIS) dataset. Their model was also tested on data from the Alzheimer's Disease Minimum Interval Resonance Imaging in AD (MIRIAD) dataset, achieving an average accuracy of 80.0% for AD and MCI classification. Yu et al. [3] introduced a connectivity-weighted penalty to construct functional connectivity, replacing the traditional

* Corresponding author.
E-mail address: menglu@mail.neu.edu.cn (L. Meng).

https://doi.org/10.1016/j.bspc.2023.105212
Received 7 April 2023; Received in revised form 4 June 2023; Accepted 23 June 2023
Available online 29 June 2023

l_1 norm regularization. They supposed that the higher correlation of the BOLD signals, the stronger link between ROIs. The adaptive graph pooling part of our paper is inspired by this.

In a separate study, Li et al. [4] used the group-constrained Kalman filter (gKF) algorithm to construct dynamic effective connectivity (dEC) and employed the virtual adversarial training convolutional neural network (VAT-CNN) to extract the local features of dEC. Finally, they proposed the high-order connectivity weight-guided graph attention networks (cwGAT) to aggregate features of dEC. The results achieved the classification accuracy of 90.9%, 89.8%, and 82.7% for NC vs. early MCI (EMCI), EMCI vs. late MCI (LMCI), and NC vs. EMCI vs. LMCI respectively. Although the results outperform the state-of-the-art methods significantly, the structural information from DTI is not utilized in the work. Several studies demonstrated that the fusion of functional and structural information can enhance model performance for MCI identification [5].

In addition to this, Yee et al. [6] and Pan et al. [7] analyzed single-modal data: 18F-FDG PET. Single-modal research has achieved remarkable results, but in contrast, multi-modal can provide complementary information.

Research on computer-aided diagnosis of Alzheimer's disease has increasingly focused on multi-modal data analysis [8]. AD patients have been shown to display a marked decrease in the number of brain network connections and weakened interactions between brain regions within the network. Additionally, alterations in the small-world characteristics of the brain network suggest a disruption in network integrity. Functional magnetic resonance imaging (fMRI) measures changes in brain blood flow caused by neural activity, while diffusion tensor imaging (DTI) enables observation of the trajectory of nerve fiber bundles. Both modalities are closely linked to AD brain lesions and have been extensively studied in the context of AD diagnosis.

Li et al. [9] proposed a multi-modal hyperconnected functional network for MCI identification by integrating the hyperconnected networks of BOLD fMRI and ASL fMRI, which led to an 11.5% increase in classification accuracy compared to using BOLD fMRI alone. But it is a major limitation that the proposed method ignores the effect of DTI on functional network formation. Neuroscience studies have reported that stronger structural connections between ROIs indicate higher functional interactions [10].

Lei et al. [11] proposed a feature selection method from multi-modal brain networks and used support vector machines for prediction, resulting in a 3.76% improvement in classification accuracy compared to using fMRI alone. Li et al. [12] employed the hyper weighted LASSO algorithm to construct fMRI brain networks and incorporated DTI as a penalty parameter for regularization, which increased the classification accuracy by 5.5% when using fMRI and DTI data together. However, a limitation of the above two algorithms is that the impact of non-imaging information and the correlation between subjects are not considered. Wolz et al. [13] and Ktena et al. [14] highlighted that non-imaging such as gender, age, and the type of equipment used for data acquisition, can introduce heterogeneity in the extracted features, resulting in a non-uniform distribution and limiting the generalization ability of the trained models.

Research in the field of neuroimaging has demonstrated the potential of deep learning-based diagnosis to improve the accuracy and efficiency of AD diagnosis [15,16,17,18]. One such approach was proposed by Yao et al. [18], who proposed a mutual multi-scale triplet graph convolutional network (MMTGCN) to analyze fMRI and DTI for brain disorder diagnosis. First, three sets of functional and structural connectivity networks were generated using three different brain segmentation templates, respectively. And a triplet GCN (TGCN) module is proposed to learn the relationship between each GCN. The accuracy of NC vs. MCI was 86.6% and the AUC was 90.3%. But one shortcoming of the paper is global connectivity is ignored by using k-nearest neighbors (k-NN) to extract nodes.

It can be found that there are some limitations in the existing multi-modal methods: (1) There is still no effective multi-modal fusion method [11,12,18]. (2) Some feature extraction methods ignore the relationship of long-distance brain regions [18]. (3) A popularly input for diagnostic models is the brain connectivity networks [9,11,19]. To reduce redundancy, common methods focus on eliminating a certain number of connections, which ignores the physiological functional connections between brain regions. (4) The impact of non-imaging information is ignored [11,12].

Our proposed solutions aim to address the limitations identified in the existing literature on multi-modal methods for the computer-aided diagnosis of MCI:

(1) We propose a highly effective multi-modal fusion mechanism that takes into account the complementary nature of fMRI and DTI data. We achieve feature fusion by concatenating the two modalities on a per-subject basis, followed by constructing connections between adjacency matrices to expand the receptive field of graph convolutional networks (GCN).
(2) We propose the use of graph pooling algorithms to extract global features that have a larger receptive field compared to convolutional methods. We leverage structural information between nodes through connection reconstruction and preserve nodes with greater differences within clusters by sampling.
(3) We propose a novel feature selection algorithm based on the multilayer perceptrons (MLP) learning model. We remove redundant features while selecting higher-quality features through labeled learning with MLP. This approach ensures that the selected features are highly relevant to the MCI diagnosis.
(4) We design a similarity fusion method that integrates non-imaging information into topological similarity calculation. This approach addresses the non-uniform distribution of features caused by non-imaging differences among individuals, such as gender, age, and sampling equipment. Our proposed solution helps to improve the generalization ability of the models.

Overall, our proposed solutions aim to enhance the accuracy of MCI diagnosis by improving the effectiveness of multi-modal methods and overcoming the limitations identified in existing literature. These proposed solutions could also be extended to other related applications in the field of neuroimaging.

2. Method

2.1. Algorithm overview

The presented algorithm in this paper follows a multi-stage approach for optimizing feature extraction from multi-modal data and expanding the receptive field of graph convolutional neural networks to improve disease diagnosis results. Fig. 1 shows the overall flowchart of the algorithm. The algorithm consists of the following steps:

(a) Image preprocessing is performed to minimize the differences in image acquisition and enhance the effectiveness of subsequent statistical analysis.
(b) Dual-channel feature extraction is conducted using the same processing for multi-modal data. Graph pooling and MLP are used to extract higher-level information from subgraphs, and finally, features from the two channels are fused.
(c) A population graph is constructed that uses the correlations between brain imaging data and non-imaging information to extend the receptive field of GCN.
(d) A diagnostic model is built using graph convolutional neural networks, which integrates multi-modal information and stabilizes model performance while demonstrating its efficiency in an intuitive manner.

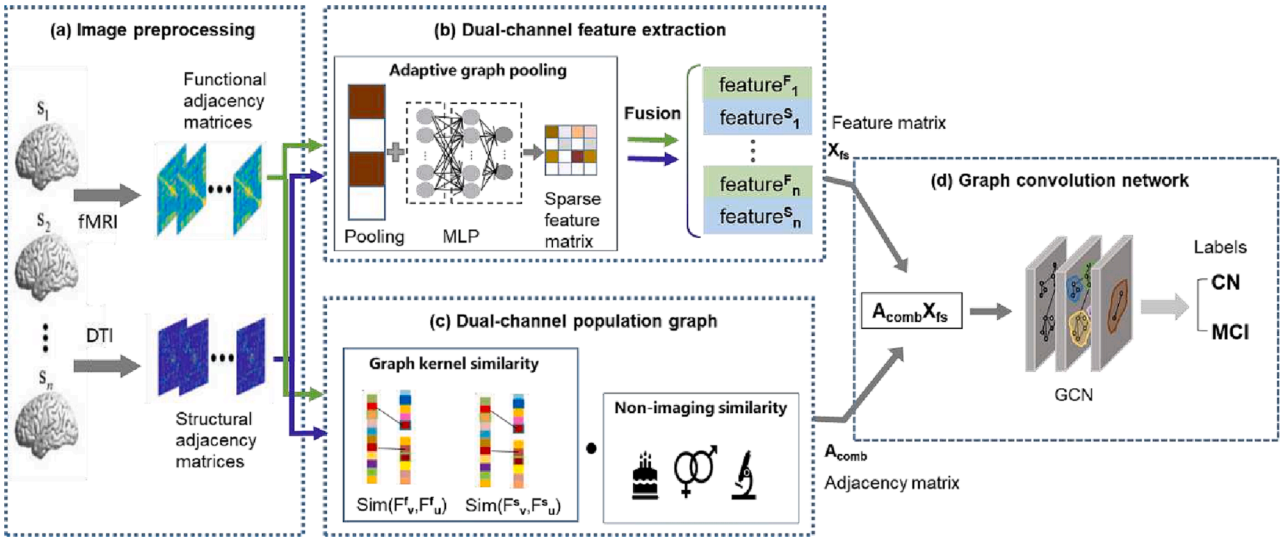

Fig. 1. Overall algorithm flow chart.

2.2. Brain image preprocessing

DTI and fMRI are considered the most effective neuroimaging modalities for diagnosing AD and MCI. DTI can provide information on structural changes in the brain, while fMRI reflects spontaneous functional activity, both of which are closely related to brain lesions in AD patients. Combining these two modalities can help to understand the underlying pathological mechanisms of AD, leading to better diagnosis and treatment for AD and MCI patients.

In fMRI imaging, standard preprocessing steps are performed using specialized software such as Statistical Parametric Mapping (SPM) and Data Processing Assistant for Resting-State fMRI (DPARSF). Slice timing correction is performed first to eliminate any temporal discrepancies in the images. Head motion correction is then conducted by realigning the images, followed by normalization and smoothing to reduce noise and align the images to a common template. Local average time series regression is applied to remove noise and artifacts from the data. Finally, the brain is segmented into 90 regions of interest (ROIs) using the Automated Anatomical Labeling atlas (AAL90), and an average time series matrix is generated for each individual.

In DTI imaging, specialized software such as Pipeline for Analyzing Brain Diffusion Images (PANDA) and FMRIB Software Library (FSL) are commonly used. Motion artifacts and eddy current distortion are corrected by aligning diffusion-weighted images with a reference b0 image. The b-matrix is then reoriented, and a transformation is applied to calculate the diffusion tensor, which provides information on the direction and magnitude of water diffusion. The eigenvalues ($\lambda_1,\lambda_2,\lambda_3$) and corresponding eigenvectors are calculated, which are used to estimate fiber tracts in the brain. Similar to fMRI preprocessing, the brain is segmented into 90 ROIs, and global deterministic fiber imaging is generated using the default settings of the Fiber Assignment by Continuous Tracking (FACT) algorithm. Finally, a fiber count matrix of 90 ROIs is obtained for each individual.

Overall, the combined use of DTI and fMRI allows for a comprehensive understanding of the pathological mechanisms of AD and MCI, which can lead to improved diagnosis and treatment for these conditions. The use of specialized software and preprocessing techniques enables the extraction of meaningful information from the raw imaging data.

2.3. Dual-channel feature extraction

Given the intricate nature of neuroimaging images and the complementary relationship between functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI), the present study employed an adaptive graph pooling method to extract deep features from the connectivity matrices of the two modalities, as obtained in Section 2.2. These deep features were subsequently fused, as illustrated in Fig. 1(b), leading to an effective feature extraction method that improved diagnostic accuracy to some extent.

2.3.1. Adaptive graph pooling

The proposed approach involves selecting key subgraphs of the brain network for each subject individually. Specifically, an unsupervised graph pooling method is first utilized to downsample the brain image into a sparse matrix. Subsequently, a MLP is trained using the sparse feature matrix and labels to extract high-order features for further processing. Fig. 2 illustrates the adaptive graph pooling method, which consists of a hierarchical graph pooling module with structure learning (HGP-SL) and an MLP module capable of learning complex data representations.

HGP-SL can be broken down into two main steps, as shown in Fig. 2 (b) and Fig. 2(c). The first step involves selecting nodes to minimize the loss of graph information. This is achieved by computing an information score, represented by the L1 norm of the Manhattan distance between the features of a node and its neighbors. The nodes are then sorted according to their information scores and the user-defined pooling rate.

The second step involves using an unsupervised edge prediction algorithm to connect isolated subgraphs and partially correct initial brain region connections caused by node selection. This step is crucial in ensuring that the information lost during node selection is partially recovered through the addition of edges between subgraphs. By incorporating both node and edge information, the adaptive graph pooling method enables the extraction of deep features from the brain network for subsequent analysis.

2.3.1.1. Select subgraph nodes. In the adaptive graph pooling process, the information score S for each node is determined by combining its features with those of its neighbors and computing the L1 norm of the distance between these two components. These scores are then sorted, and a subset of nodes with a user-defined pooling rate and corresponding adjacency matrices for subgraphs are selected based on the highest information scores.

$$S = \gamma(g) = \left\| \left(\mathrm{I} - \left(D^{(l)} \right)^{-1} A^{(l)} \right) H^{(l)} \right\|_1 \tag{1}$$

In Formula (1), I is the identity matrix, $A^{(l)}$ and $H^{(l)}$ respectively the

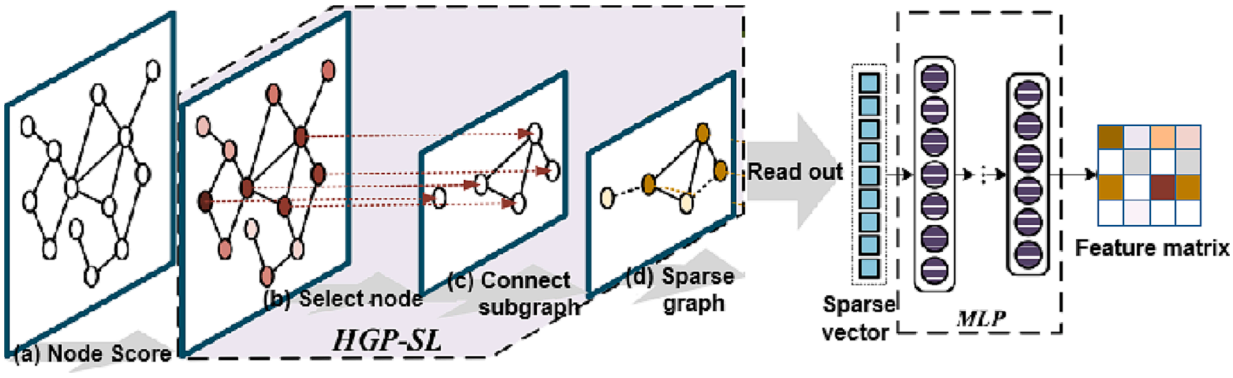

Fig. 2. Adaptive graph pooling feature extraction.

adjacency matrix and node feature matrix for the l-th layer. $D^{(l)}$ is a diagonal degree matrix. Vector S contains the calculated scores for each node. Based on these information scores, nodes within the top-K range are selected in order.

To determine a node's information score, this method calculates the difference between the node's features and the average features of its neighboring nodes. The larger this difference, the higher the information score assigned to the node, which indicates a lower likelihood of being removed during node selection. This approach ensures that nodes with unique features and important connections are more likely to be retained, which contributes to the preservation of essential graph information.

2.3.1.2. Structure learning. To address the potential issue of isolated subgraphs and the limitations of the initial brain graph structure in the node selection method, this paper employed an adaptive approach to preserve the coherence of subgraphs and predict possible connections between the selected nodes:

Firstly, an unsupervised edge prediction algorithm is utilized to connect the isolated subgraphs and enhance the initial brain region connections. Then, the obtained subgraphs are combined into a hierarchical graph structure, and a structure learning method is applied to adaptively learn the connections between subgraphs. This process allows the model to explore and learn new connections beyond the initial brain graph structure and improve the overall representation of the brain network.

$$E^{(l)}(p,q) = \frac{H^{(l)}(p,:)\cdot H^{(l)}(q,:)}{\|H^{(l)}(p,:)\|\|H^{(l)}(q,:)\|} + A^{(l)}(p,q) \tag{2}$$

Here, $E^{(l)}(p,q)$ is the required similarity between the two nodes, p and q, $H^{(l)}$ is the node feature matrix in the previous section. The adjacency matrix $A^{(l)}(p,q)$ is also included in computing cosine similarity to enhance the strength of connections between nodes. Then, a sparse attention mechanism is used to normalize the similarity scores.

$$Sim^{(l)}(p,:) := \underset{n\in\Delta^{K-1}}{argmin}\|n - E^{(l)}(p,:)\|^2 \tag{3}$$

$$where\Delta^{K-1} = \left\{n \in R^K \middle| 1^T n = 1, n \geq 0\right\} \tag{4}$$

Here, the simplex Δ^{K-1} is a K-1 dimensions probability simplex that corresponds to K, the number of nodes in the brain graph. For each node p, the softmax function is used to normalize the similarity scores between it and other nodes, resulting in a probability distribution Δ^{K-1}. However, after the process of normalization, small non-zero values may be preserved, increasing the complexity of down-sampling subgraphs. So, this paper projected target vectors onto the Δ^{K-1} simplex and achieved sparsity upon reaching boundaries. Specifically, it updated adjacency matrices to their optimal values for this quadratic-constrained optimization problem:

$$A^{(l+1)}(p,q) = \left[E^{(l)}(p,q) - \tau\left(E^{(l)}(p:)\right)\right]_+ \tag{5}$$

$$\tau(n) = \frac{\left(\sum j \in Q(n) n_j\right) - 1}{|Q(n)|} \tag{6}$$

$$\text{where}Q_{(n)} = \left\{j \in [K] \middle| n_j > 0\right\} \tag{7}$$

Here, $[K] = \{1, \cdots, K\}$, $[t]_+ := max\{0, t\}$ To maintain the sparsity of $A^{(l+1)}(p,q)$, any values in the matrix that are lower than the threshold $\tau(\cdot)$ are set to 0.

In this study, the graph pooling operations are conducted on an individual basis, and the outcomes are saved as sparse vectors. This approach enables the downsampling of brain imaging into a compact set of essential features while simultaneously preserving high-performance standards in terms of storage and computational requirements. Moreover, this technique retains a greater amount of information in the feature matrix than previous methodologies.

2.3.2. Feature fusion

After pooling features from both fMRI and DTI modalities, the resulting sparse feature matrices 'feature^F' and 'feature^S' have the same dimensions. In this paper, a fusion algorithm similar to feature concatenation is proposed to effectively enhance brain imaging feature expression for each subject. Fig. 3 illustrates the schematic diagram of the proposed multi-modal feature fusion. The functional sparse feature matrix 'feature^F' and structural sparse feature matrix 'feature^S' are concatenated to represent each subject's features.

The "concat" operation used in this article is not simply a concatenation of the feature matrices, but a way to expand the information of the brain images. This method is based on subject-wise concatenation, meaning that the features of each subject are concatenated together. For instance, in Fig. 3, the features of Subject 1 are represented by the joint sparse feature matrices "feature $1^{F\prime\prime}$ and "feature $1^{S\prime\prime}$. The resulting fused feature matrix X_{fs} is then used as input for the graph convolutional network.

As shown in Table 1, the network structure of dual-channel feature extraction is depicted. In the HGP-SL block, HGP-SL are identical and consist of two parts: Select subgraph nodes and Structure learning. After

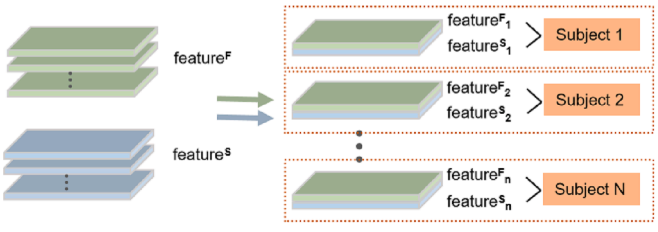

Fig. 3. Multi-modal feature fusion.

Table 1
The network structure of dual-channel feature extraction.

	Type	Input size	Output size
HGP-SL block	Relu	590,90	590,90
	Global_mean_pool	590,90	590,90
	Global_max_pool	590,90	590,90
	Concat pool	590,90	590,180
HGP-SL block	Relu	118,90	118,90
	Global_mean_pool	118,90	118,90
	Global_max_pool	118,90	118,90
	Concat pool	118,90	118,180
	Sum Concat	118,180	118,180
	Block output	/	118,180
MLP block	Liner	118,180	118,128
	Relu	118,128	118,128
	Dropout	118,128	118,128
	Liner	118,128	118,64
	Relu	118,64	118,64
	Dropout	118,64	118,64
	Concat	118,64	118,128
	Block output	/	118,128

HGP-SL, new graph is represented using "Global_mean_pool and Global_max_pool" operations. The MLP block is composed of two fully connected layers and is responsible for fusing fMRI and DTI data by "Concat".

2.4. Personalized population graph

In the context of disease diagnosis, graph-based methods typically focus only on pairwise similarity between subjects, relying on their specific imaging feature vectors, rather than modeling relationships between subjects. To address this limitation, this paper constructs a population graph based on individual-level data [20], as shown in Fig. 4. In this graph, the nodes represent individual subjects, while the edge weights represent the correlations between the subjects. This representation allows for the integration of rich imaging and non-imaging information in diagnostic tasks.

To construct the population graph, the preprocessed data from Section 2.2 is used to calculate imaging similarity using a Gaussian kernel function (Formula (9)) and non-imaging similarity using cosine similarity. The Hadamard product of the two similarity matrices is then

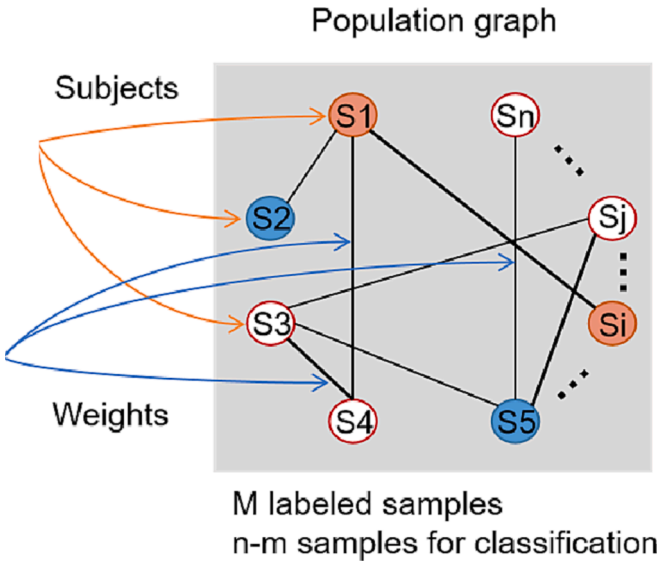

Fig. 4. Diagram of the adjacency matrix.

computed to obtain the adjacency matrix. Additionally, the fMRI and DTI adjacency matrices are fused to expand the receptive field of the graph convolutional network (GCN), which is beneficial for better feature extraction.

2.4.1. Graph kernel similarity

The definition of graph edges is a critical step in capturing the underlying structures in data and interpreting the similarities between feature vectors. This paper employs a graph kernel to directly measure the topological similarity between time series, which preserves the structural information while computing similarity [21]. The use of a Gaussian kernel function is proposed to calculate similarity, which is given as:

$$\kappa\left(r_a^i, r_b^i\right) = exp\left(-\frac{\|r_a^i - r_b^i\|}{2\sigma}\right) \tag{8}$$

This article suggests that indicators calculated by subjects in the same category are more similar, while those from different classes are dissimilar. A threshold T (hyperparameter) is set as a boundary to remove redundancy and better observe substantial similarity. When the similarity is less than T, set $\kappa\left(r_a^i, r_b^i\right)$ to 1; otherwise, it is set to 0.

The final formula for calculating the Gaussian kernel similarity of a population network is as follows:

$$S_k\left(N_i, N_j\right) = \frac{\sum_{a=1}^{M}\sum_{b=1}^{M} w_a^i w_b^j \kappa\left(x_a^i, x_b^j\right)}{\sum_{M}^{n_i} w_a^i \sum_{b=1}^{M} w_b^j} \tag{9}$$

In formula (9), the Gaussian kernel function $\kappa\left(x_a^i, x_b^j\right)$ represents the Gaussian distance between different brain regions of diverse individuals, where $x_a^i = \sum_u^M \widehat{A}_i(a,u)$ represents the local topology of subject a, and $w_a^i = \frac{1}{\sum_{u=1}^{M} \kappa\left(x_a^i, x_u^i\right)}$ is the Gaussian distance between all brain regions of person i.

2.4.2. Phenotypic information similarity

Using cosine similarity to calculate phenotypic similarity:

$$\mathrm{Sim}(i,j) = \left|\frac{M_i \cdot M_j}{\|M_i\|\|M_j\|}\right| \tag{10}$$

$$S_T(i,j) = \begin{cases} 1, \text{if} \mathrm{Sim}(i,j) > 0.5 \\ 0, otherwise \end{cases} \tag{11}$$

For the obtained similarity matrix, a filtering process can be performed by setting a threshold of 0.5 to retain connections between subjects with similarities. Combined with formula (9) and formula (11), the adjacency matrices A^F and A^S describing the correlation in fMRI and DTI can be obtained, "∘" represents the Hadamard product of two adjacency matrices:

$$A^F = S_k^F \circ S_T^F$$

$$A^S = S_k^S \circ S_T^S \tag{12}$$

2.4.3. Receptive field fusion

To construct the population graph A_{comb}, in addition to calculating the similarity, it is also necessary to realize the fusion of the adjacency matrices (A^F and A^S) of the two modalities.

In this paper, three receptive fields are designed. The receptive field 1 is A^F calculated from the fMRI data in the training set according to the formula (12), and the edge weight is the element in A^F. The receptive field 2 is the A^S calculated by the DTI data in the training set according to the formula (12), and the edge weight is also the element in A^S. Receptive field 3 is the connection between the test and the train, and the edge weight is 1, which ensures that new nodes can be replaced during testing. As shown in Fig. 5, the number of subjects in the training set is m, the number in the test set is n-m, and the matrix dimensions (n

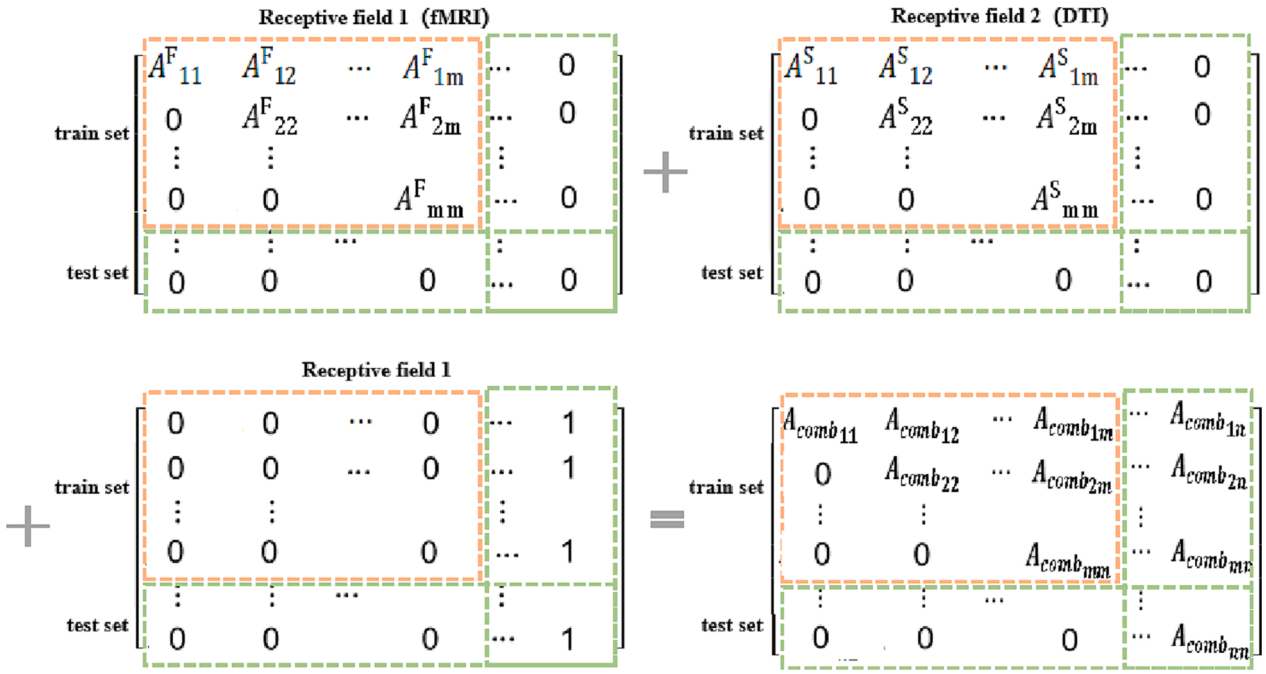

Fig. 5. Similarity receptive field.

× n) of the three receptive fields are the same. Because it is a symmetric matrix, only the upper triangular data of the matrix should be considered. The final A_{comb} can be obtained by summing the three matrices.

2.5. Graph convolution network

In this article, a two-layer graph convolutional network (GCN) was implemented, which consists of two different layers. The first layer is the GCN layer, while the second layer uses a variant of GCN called Cluster-GCN [23] to accelerate ordinary GCN blocks. The structure of the GCN in this article is illustrated in Fig. 6.

Cluster-GCN proposed by Chiang et al. [23] reduced computational costs by clustering nodes. During the backpropagation phase, the model only needs to calculate gradients for the minimum subset.

$$\frac{1}{|B|}\sum_{i\in B}\nabla \mathrm{loss}\left(y_i, z_i^L\right) \tag{13}$$

Among them, B represents the subset of nodes, y_i is the correct label, and z_i^L is the predicted label. The loss function is defined as follows:

$$\mathrm{loss}\left(y_i, z_i^L\right) = -\left[y_i \log\left(z_i^L\right) + (1-y_i)\log\left(1-z_i^L\right)\right] \tag{14}$$

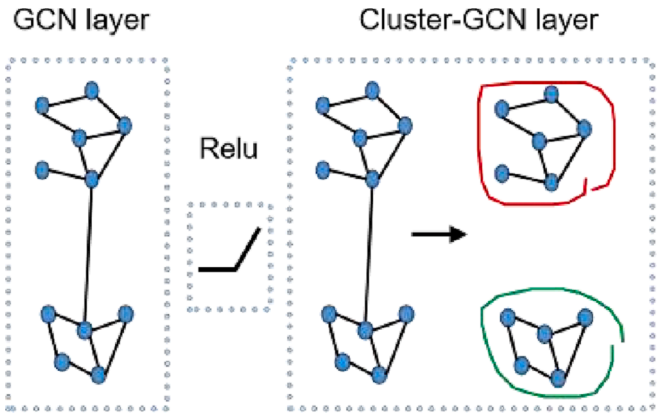

Fig. 6. Graph convolution network in this paper.

To add on, the Cluster-GCN algorithm is able to improve scalability and reduce computation cost for large graphs, making it a suitable choice for processing brain imaging data with a large number of subjects. By partitioning the graph into smaller subgraphs, the algorithm reduces the number of nodes and edges that need to be processed at once, allowing for more efficient computation. The use of Cluster-GCN in this paper enables the processing of large-scale brain imaging data, which is essential for accurate disease diagnosis and prediction.

Considering the imbalance of positive and negative samples in this paper, we introduced a balance factor α_i in the loss function [24]. For label 1, introduce weight α, and for label 0, introduce weight $1-\alpha$.

$$\alpha_i = \begin{cases} \alpha, if\, y_i = 1 \\ 1-\alpha, otherwise \end{cases} \tag{15}$$

Table 2 depicts the module parameters of GCN. When training the GCN model, the test data is NAN, and the probability of MCI for all training data will be output. According to the characteristics of the output data, when the output is greater than 0, the label is 1, which means MCI state, otherwise is the NC state. Similarly, when testing the model, the labels of the test in the population graph will be output.

3. Experiments and results

3.1. Experimental environmental parameters

The present study was conducted within a Linux system environment that utilized MATLAB2017b software and an Intel(R) Core(TM) i7-10700F CPU @ 2.90 GHz, in conjunction with 128 GB memory. The algorithm program was implemented through Python 3.7 and the PyTorch deep learning framework, which offered a high-performance

Table 2
The network structure of GCN.

Block	Type	Input size	Output size
GCN block	GCNConv	118,128	118,64
	Cluster -GCNConv	118,64	118,1
	Block output	/	118,1

platform for constructing convolutional neural networks and leveraging GPU parallel computing to expedite the training process.

The primary neural network utilized in this study is a graph convolutional neural network, which is well-suited to handling graph-structured data. The training parameters for this network were established based on previous experience and a series of iterative experiments, as detailed in Table 3. By carefully selecting these parameters, the network was optimized for optimal performance in the context of the present study.

3.2. Data sets

In this paper, the primary dataset utilized is the open-source Alzheimer's Disease Neuroimaging Initiative (ADNI) database, which was previously employed in a study by Song [19]. The ADNI database was created as a collaborative effort between academic and private researchers, and includes data from over 1,700 adult participants recruited from more than 50 locations across the United States and Canada for three separate studies (ADNI-1, ADNI-GO, and ADNI-2). The present study employed 118 sets of functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) data from the ADNI database, which encompassed samples of MCI and NC, for both training and testing. Additionally, non-imaging information such as gender, age, and equipment type were incorporated into the analysis. Further details regarding the characteristics of the dataset are provided in Table 4. By utilizing this comprehensive and well-established dataset, the present study is able to draw upon a wealth of high-quality data to investigate the research questions at hand.

In this study, all voxels of brain structures were utilized that were obtained using a segmentation criterion of C = 138. The inclusion of all brain structure voxels is particularly effective in predicting AD, as changes in brain structure are a hallmark of the disease and result in differences in brain structure voxels between healthy individuals, those with MCI, and those with AD. By incorporating these voxels into the analysis, the present study is able to leverage the full richness of the neuroimaging data to better characterize the differences between these populations and ultimately improve predictive accuracy. The use of this comprehensive approach to voxel selection represents a novel contribution to the field and has the potential to yield important insights into the underlying mechanisms of AD.

3.3. Training and verification of the model

The training process of the proposed model is evaluated using a loss curve, as shown in Fig. 7. The curve demonstrates a gradual reduction in the loss value during training, indicating a gradual improvement in the model's learning ability. To ensure the reliability of the results, the ADNI dataset was divided into ten parts using a 10-fold cross-validation method, with each part serving as a test set while the remaining nine parts were used for training. Since the ADNI dataset consists of 118 samples, each experiment had a training set size of 106 and a test set size of 12. The average Accuracy and AUC values for the ten experiments were then calculated to obtain final evaluation metrics. This approach allows for a more comprehensive evaluation of the model's performance, providing a more reliable indication of its classification ability.

Table 3
Training parameter setting.

Hyperparameter	Value
epoch	100,000
patience	5000
batch_size	128
Dropout	0.01
learning rate	0.001
pooling_ratio	0.05
balance factor	0.3

Table 4
Detailed information about the used dataset.

Category	NC(37)	MCI(81)
Female/Male	15/22	45/36
Age (average)	75.4	75.2
GE/SIEMENS/PHILIPS	17/18/2	25/51/5

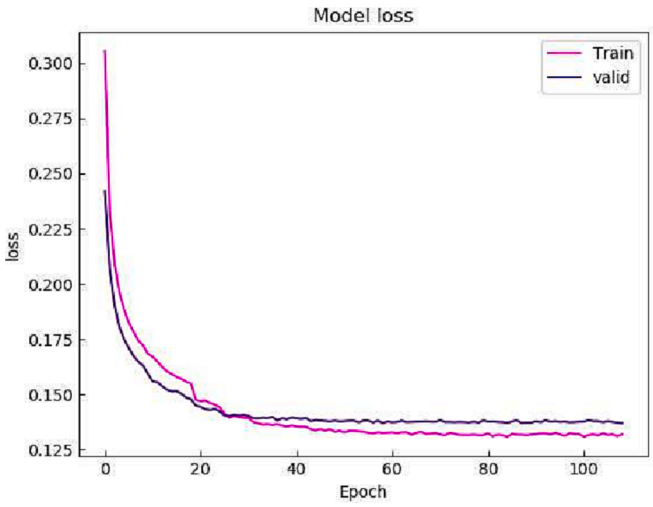

Fig. 7. Loss curve of model training.

The receiver operating characteristic (ROC) curve is a widely used tool for visualizing the performance of binary classification models at various decision thresholds. The area under the ROC curve (AUC) is a commonly used metric to evaluate the predictive performance of a model, with higher AUC values indicating better performance. In this study, Fig. 8(a) presents the ROC curve for our model, which achieved an AUC value of 0.97, indicating its high accuracy and good predictive ability for MCI detection.

The present study evaluated the performance of the proposed model by analyzing its confusion matrix, which provides a comprehensive view of the classification results. The confusion matrix, shown in Fig. 8(b), reveals the model's ability to accurately classify MCI patients and normal individuals. The false positive rate (FPR) indicates the likelihood of misclassifying a normal individual as an MCI patient, while the true positive rate (TPR) reflects the proportion of actual MCI patients correctly identified as such by the model. In this experiment, the FPR was found to be 0.056, indicating a relatively low probability of misclassification, while the TPR was 0.91, indicating that the model accurately identified a large proportion of MCI patients. The false negative rate (FNR) represents the proportion of actual MCI patients who were incorrectly classified, and the true negative rate (TNR) represents the proportion of actual normal individuals who were correctly classified as such by the model. In this experiment, the FNR was 0.095, suggesting that a small proportion of MCI patients were misclassified, while the TNR was 0.94, indicating a high accuracy in correctly identifying normal individuals. Overall, these results suggest that the proposed model has good classification performance for MCI detection.

3.4. Validation of the training parameters

To solve the problem of imbalanced positive and negative samples in the data set, we introduced a balance factor to bias the model towards the minority class. Additionally, we introduced Sensitivity (%) and Specificity (%) as evaluation indicators to understand the performance of the model. We selected the optimal parameters through ablation experiments, and the experimental results are presented in Table 5.

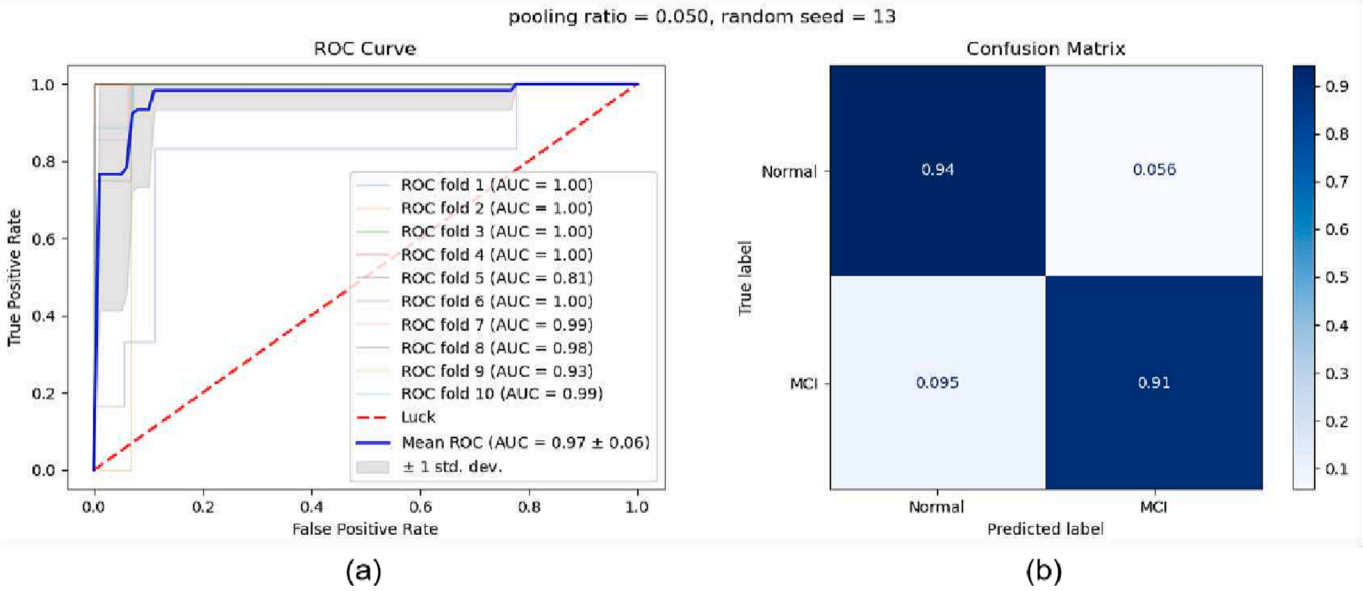

Fig. 8. ROC curve and Confusion matrix.

From Table 5, we can observe that when the balance factor is set to 0.1 and 0.2, the true negative rate (Specificity) is significantly higher than the true positive rate (Sensitivity). This indicates that the model is overly biased towards the minority class. In contrast, when the balance factor is set to 0.4, the model does not exhibit a preference for the minority class. Overall, the experimental results when the balance factor is set to 0.3 are more balanced.

The pooling_ratio refers to retaining the top pooling_ratio% nodes in each sub-graph. As shown in Fig. 9, when we set pooling_ratio as 0.05, in the overall 10-fold experiment, the performance is relatively stable and the average accuracy is the highest, indicating that the sparse features extracted by this pooling_ratio have typical characteristics and can effectively represent information.

3.5. Validation of the model and multi-modal data

Table 6 highlights the significant improvement in diagnostic accuracy achieved by utilizing GCN compared to traditional machine learning algorithms, including MLP, RF, and SVM. The results demonstrate that the average Accuracy for the GCN model increased by 3.39%, 3.54%, and 3.55% for fMRI, DTI, and fMRI + DTI, respectively. Additionally, the average Sensitivity increased by 14.18%, 7.68%, and 4.09%, while the average specificity decreased by −5.43%, −0.38%, and increased by 4.53%. These findings indicate that incorporating information from different modalities can provide complementary insights into brain function and structure, resulting in more accurate diagnoses.

Table 6 presents the comparison between the proposed DFCGCN model and other machine learning algorithms, including GCN, MLP, RF, and SVM. The results show that the DFCGCN model outperforms GCN in terms of diagnostic accuracy, sensitivity, and specificity. For example, on fMRI data, the DFCGCN model achieved an average accuracy improvement of 7.57%, compared to GCN. The model also showed improved performance when incorporating multiple modalities, with an average accuracy improvement of 16.23% on fMRI + DTI data. Furthermore, the DFCGCN model also outperformed traditional machine learning algorithms such as MLP, RF, and SVM. The comparison results suggest that multi-modal data can provide complementary information to improve diagnostic accuracy, and the proposed DFCGCN model, with its dual-fusion method and perception receptive fields, is effective in integrating information from different modalities.

Table 7 compares the performance of the proposed DFCGCN method with other state-of-the-art algorithms on fMRI + DTI multi-modal data. The results demonstrate that the DFCGCN method achieves comparable or better performance than other state-of-the-art methods in terms of Accuracy, Sensitivity, Specificity, AUC, and other evaluation metrics. The proposed method achieves an Accuracy of 90.47%, Sensitivity of 89.17%, Specificity of 91.76%, and AUC of 0.97, which are highly competitive compared to other methods. By comparison, we conclude that the proposed DFCGCN model can effectively integrate complementary information between fMRI and DTI, enhance inter-subject correlation, and achieve better feature extraction performance. These results further confirm the effectiveness and robustness of the proposed DFCGCN method in the diagnosis of MCI using multi-modal neuroimaging data.

Table 5
Effect of different balance factors.

Balance factor	Sensitivity(%)	Specificity(%)
0.1	69.3	97.5
0.2	78.0	93.3
0.3	**91.1**	**94.0**
0.4	96.8	85.0

3.6. Validation of adaptive graph pooling

The proposed method uses adaptive graph pooling method to extract the deep features from the brain network for subsequent analysis. In the experiment, our method is compared with another state-of-the-art pooling algorithm, which combines connectivity matrices with the RFE (Recursive Feature Elimination) dimensionality reduction strategy, with a more detailed feature extraction algorithm. The results, depicted in Fig. 10, show that the proposed feature extraction algorithm outperforms the connectivity matrics + RFE algorithm in terms of Accuracy, regardless of whether single-modal or multi-modal data are used. The connectivity matrices + RFE algorithm may neglect regional brain activity details, and it maps two vectors to floating-point values between −1 and 1, resulting in a loss of more structural information within the image. In comparison, the proposed algorithm directly extracts features from the entire graph structure and observes the activity between brain regions in subjects, preserving more spatial information and enabling a more comprehensive understanding of brain function and structure, which results in improved accuracy in diagnosis or classification tasks.

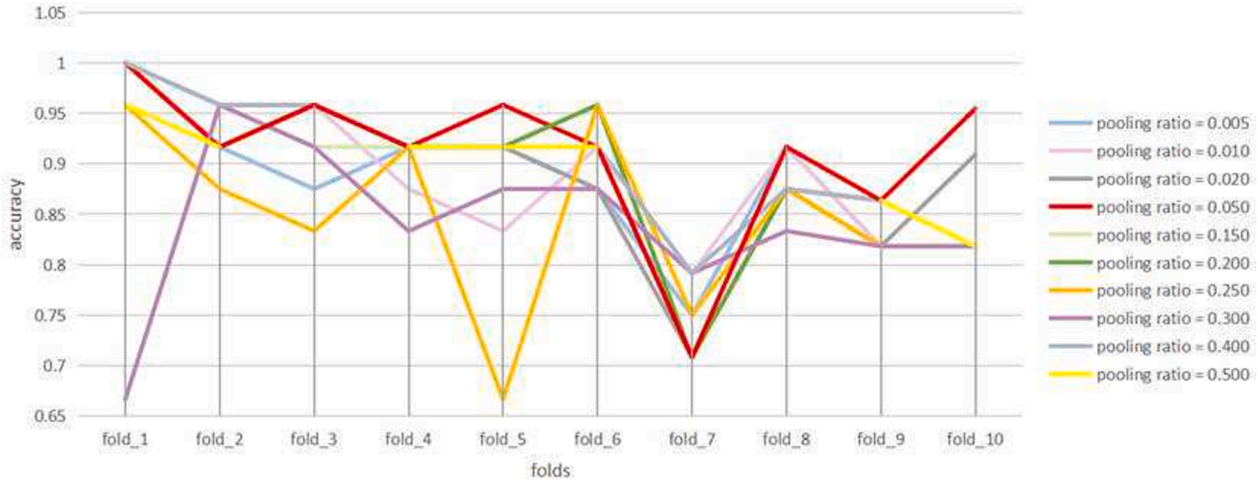

Fig. 9. Effect of different pooling_ratio.

Table 6
Diagnosis performance of different methods for NC vs. MCI.

Data Modality	Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
fMRI	MLP	64.18	63.57	63.98
	RF[25]	67.10	56.40	77.27
	SVM[26]	67.14	63.39	69.32
	GCN[20]	70.07	**75.30**	64.76
	CGCN(Ours)	**74.69**	62.13	**94.88**
DTI	MLP	71.81	67.34	76.14
	RF[25]	71.24	72.07	70.45
	SVM[26]	70.63	71.89	69.31
	GCN[20]	74.77	**78.11**	71.59
	CGCN(Ours)	**79.62**	78.05	**88.13**
fMRI + DTI	MLP	73.60	72.07	75.00
	RF[25]	74.77	75.66	73.86
	SVM[26]	72.38	74.34	70.19
	GCN[20]	77.13	78.11	77.55
	DFCGCN (Ours)	**90.70**	**91.10**	**94.07**

Table 7
Algorithm comparison with the related works.

Author	Accuracy(%)	Sensitivity(%)	Specificity(%)
Lei et al.(2020)[11]	86.5	85.3	87.5
Li et al. (2020)[12]	87.7	88.9	86.5
Yao et al (2021)[18]	86.6	–	–
Song et al. (2021)[19]	87.1	90.2	84.1
Zhou et al.(2022)[27]	89.4	**93.8**	84.9
Lei et al.(2023)[5]	89.44	90.3	88.46
Ours	**90.7**	91.10	**94.07**

3.7. Validation of graph kernel and non-imaging information

The Gaussian kernel function plays a fundamental role in measuring the "similarity" between samples and projecting them into a space that captures this similarity. Specifically, the Gaussian kernel function is capable of clustering similar samples effectively, and it can transform linearly inseparable data into linearly separable data.

Fig. 11 presents the population graph that is constructed using the Gaussian kernel function and non-imaging information. Although the fMRI illustration is somewhat unclear, the images of DTI and fMRI + DTI reveal conspicuous clustering characteristics. Notably, participants with similar features in the graph exhibit strong clustering effects before conducting graph convolution operations, which demonstrates the efficacy of constructing graphs using Gaussian kernels and non-imaging information.

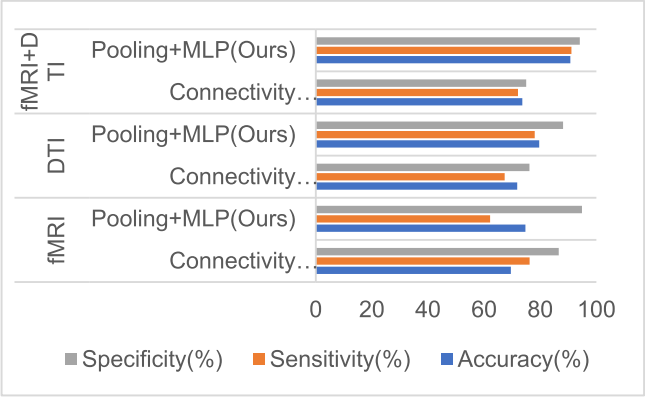

Fig. 10. Comparison of adaptive graph pooling and connectivity matrices plus RFE.

The findings presented in Fig. 12 indicate that across the three modalities of fMRI, DTI, and fMRI + DTI, the application of non-imaging + graph kernel methods consistently results in higher Accuracy and Sensitivity than using only graph kernels or only non-imaging information algorithms. Notably, multi-modal data consistently achieved higher Accuracy than single-modality data. Furthermore, the non-imaging + graph kernel method showed the highest Specificity in the single-modality fMRI setting.

The study results revealed that the combination of graph kernels and non-imaging information significantly enhanced the performance. This finding highlights the complementary roles of these two methods in accurately characterizing population graph.

4. Conclusion

This paper presents an efficient feature extraction algorithm using graph convolutional networks, achieving an accuracy of 90.7% and outperforming similar algorithms. The proposed algorithm is validated through comparative experiments, demonstrating its superiority in preserving spatial features in multi-modal data that consider the brain's functional and structural connectivity between regions. By utilizing graph pooling operations and clustering sparse features, the algorithm

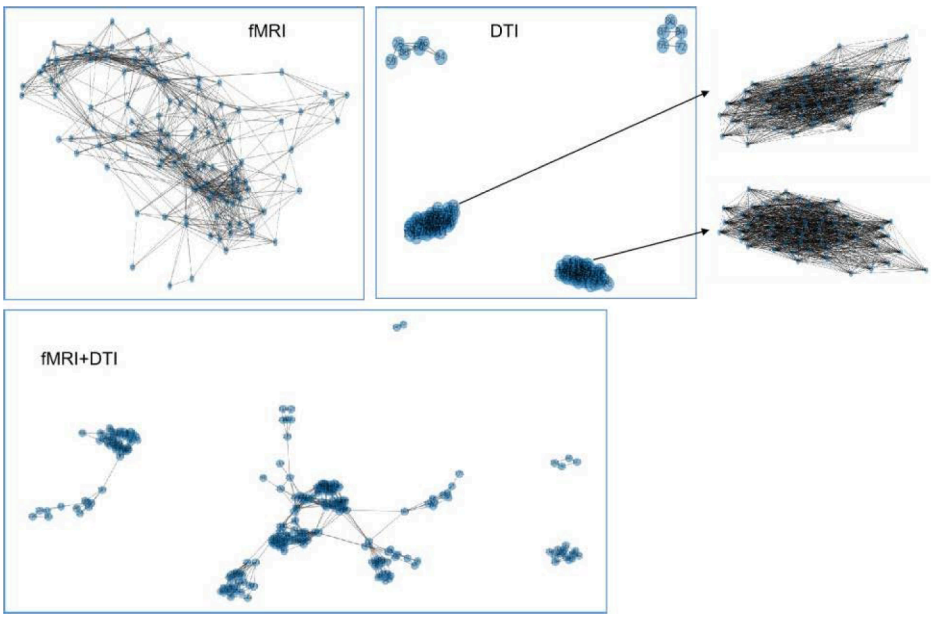

Fig. 11. Graphical representation of clustering effects for the three modal data.

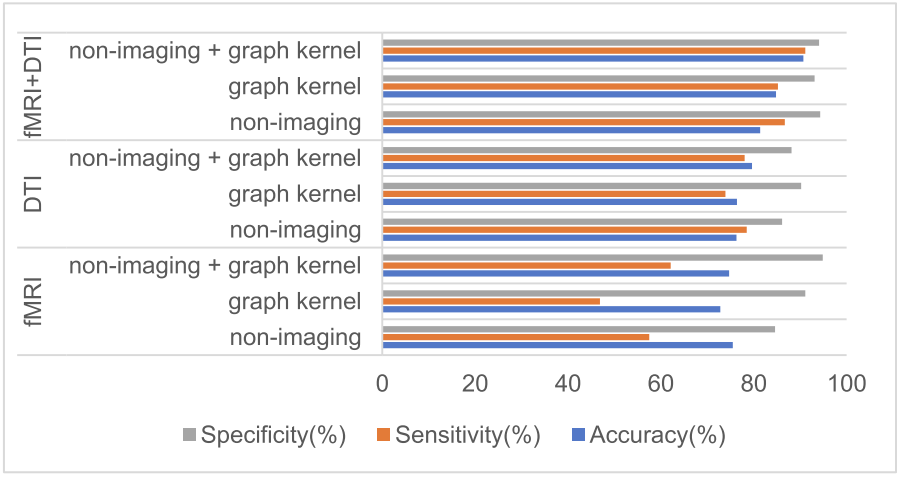

Fig. 12. Comparison of Gaussian kernel, non-imaging information, and Gaussian kernel plus non-imaging information.

maintains performance superiority compared to conventional methods. The population graph calculation using graph kernels can effectively represent individual correlations. Thus, the proposed multi-modal graph convolutional algorithm thoroughly explores the brain's connectivity characteristics and avoids the complex construction of a brain connection network, making it highly applicable in other diagnostic tasks.

Funding

This study was supported by National Natural Science Foundation of China (62073061), the Fundamental Research Funds for the Central Universities (N2204009), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2022-JKCS-21), Chongqing Science and Health Joint Medical Research Project (2023MSXM137), and Liaoning Provincial Natural Science Foundation Joint Fund for Medical-Industrial Crossover (2022-YGJC-31).

CRediT authorship contribution statement

Lu Meng: Supervision, Conceptualization, Methodology, Funding acquisition, Investigation, Writing - review & editing. **Qianqian Zhang:** Data curation, Investigation, Visualization, Software, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] K. Oh, Y.-C. Chung, K.W. Kim, W.-S. Kim, I.-S. Oh, Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, Sci. Rep. 9 (1) (2019).
[2] A. Yiğit, Z. Işik, Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease, Turk. J. Electr. Eng. Comput. Sci. 28 (1) (2020) 196–210.

[3] R. Yu, H. Zhang, L. An, X. Chen, Z. Wei, D. Shen, Connectivity strength-weighted sparse group representation-based brain network construction for MCI classification, Hum. Brain Mapp. 38 (5) (2017) 2370–2383.
[4] Y. Li, J. Liu, Y. Jiang, Y. Liu, B. Lei, Virtual adversarial training-based deep feature aggregation network from dynamic effective connectivity for MCI identification, IEEE Trans. Med. Imaging 41 (1) (2021) 237–251.
[5] B. Lei, Y. Zhu, S. Yu, H. Hu, Y. Xu, G. Yue, T. Wang, C. Zhao, S. Chen, P. Yang, X. Song, X. Xiao, S. Wang, Multi-scale enhanced graph convolutional network for mild cognitive impairment detection, Pattern Recogn. 134 (2023), 109106.
[6] E. Yee, K. Popuri, M.F. Beg, A.D.N. Initiative, Quantifying brain metabolism from FDG-PET images into a probability of Alzheimer's dementia score, Hum. Brain Mapp. 41 (1) (2020) 5–16.
[7] X. Pan, T.L. Phan, M. Adel, C. Fossati, T. Gaidon, J. Wojak, E. Guedj, Multi-view separable pyramid network for AD prediction at MCI stage by 18 F-FDG brain PET imaging, IEEE Trans. Med. Imaging 40 (1) (2020) 81–92.
[8] S. Rathore, M. Habes, M.A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, Neuroimage 155 (2017) 530–548.
[9] Y. Li, J. Liu, X. Gao, B. Jie, M. Kim, P. Yap, C.Y. Wee, D. Shen, Multi-modal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification, Med. Image Anal. 52 (2019) 80–96.
[10] D. Zhu, K. Li, C.C. Faraco, F. Deng, D. Zhang, L. Guo, T. Liu, Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles, Neuroimage 59 (2) (2012) 1382–1393.
[11] B. Lei, N. Cheng, A.F. Frangi, E.L. Tan, J. Cao, P. Yang, T. Wang, Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease, Med. Image Anal. 61 (2020), 101652.
[12] Y. Li, J. Liu, Z. Tang, B. Lei, Deep spatial-temporal feature fusion from adaptive dynamic functional connectivity for MCI identification, IEEE Trans. Med. Imaging 39 (9) (2020) 2818–2830.
[13] R. Wolz, P. Aljabar, J.V. Hajnal, J. Lötjönen, D. Rueckert, Nonlinear dimensionality reduction combining MR imaging with non-imaging information, Med. Image Anal. 16 (4) (2012) 819–830.
[14] S.I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, D. Rueckert, Metric learning with spectral graph convolutions on brain connectivity networks, Neuroimage 169 (2018) 431–442.
[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, 1–9.
[16] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[17] T. Zhou, K.H. Thung, X. Zhu, D. Shen, Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis, Hum. Brain Mapp. 40 (3) (2019) 1001–1016.
[18] D. Yao, J. Sui, M. Wang, E. Yang, Y. Jiaerken, N. Luo, D. Shen, A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity, IEEE Trans. Med. Imaging 40 (4) (2021) 1279–1289.
[19] X. Song, F. Zhou, A.F. Frangi, et al., Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction, Med. Image Anal. 69 (2021), 101947.
[20] S. Parisot, S.I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, D. Rueckert, Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease, Med. Image Anal. 48 (2018) 117–130.
[21] H. Jiang, et al., Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction, Comput. Biol. Med. 127 (2020), 104096.
[23] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, C.-J. Hsieh, Cluster-GCN, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.
[24] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP (99), 2999-3007.
[25] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[26] C. Cortes, V. Vapnik, Support-vector network, Mach. Learn. 20 (3) (1995) 273–297.
[27] Y. Zhou, X. Si, Y.-P. Chao, Y. Chen, C.-P. Lin, S. Li, X. Zhang, Y. Sun, D. Ming, Q. Li, Automated classification of mild cognitive impairment by machine learning with hippocampus-related white matter network, Front. Aging Neurosci. 14 (2022).

Further reading

[22] J. He, C.L.X. Zhang, D. Zhang, A review of multi-modal fusion technology for deep learning, Comput. Eng. (05) (2020) 1–11.