Journal of Biomedical Informatics 140 (2023) 104326

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Diagnosis of Alzheimer's disease using hypergraph p-Laplacian regularized multi-task feature learning☆

Yanjiao Ban [a], Huan Lao [b,c,*], Bin Li [a], Wenjun Su [a], Xuejun Zhang [a,d,**]

[a] School of Computer, Electronics and Information, Guangxi University, Nanning 530004, Guangxi, China
[b] Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin 541004, Guangxi, China
[c] School of Artificial Intelligence, Guangxi Minzu University, Nanning 530006, Guangxi, China
[d] Guangxi Key Laboratory of Multimedia Communications and Network Technology, Guangxi University, Nanning 530004, Guangxi, China

ARTICLE INFO

Dataset link: www.adni.loni.usc.edu

Keywords:
Multi-task learning
Hypergraph learning
Multimodal classification
Alzheimer's disease

ABSTRACT

Multimodal data-based classification methods have been widely used in the diagnosis of Alzheimer's disease (AD) and have achieved better performance than single-modal-based methods. However, most classification methods based on multimodal data tend to consider only the correlation between different modal data and ignore the inherent non-linear higher-order relationships between similar data, which can improve the robustness of the model. Therefore, this study proposes a hypergraph p-Laplacian regularized multi-task feature selection (HpMTFS) method for AD classification. Specifically, feature selection for each modal data is considered as a distinct task and the common features of multimodal data are extracted jointly by group-sparsity regularizer. In particular, two regularization terms are introduced in this study, namely (1) a hypergraph p-Laplacian regularization term to retain higher-order structural information for similar data, and (2) a Frobenius norm regularization term to improve the noise immunity of the model. Finally, using a multi-kernel support vector machine to fuse multimodal features and perform the final classification. We used baseline sMRI, FDG-PET, and AV-45 PET imaging data from 528 subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) to evaluate our approach. Experimental results show that our HpMTFS method outperforms existing multimodal-based classification methods.

1. Introduction

Alzheimer's disease (AD) [1] is an insidious neurodegenerative disease that develops with loss of memory, language and mobility. Many surveys [2–4] have shown that the number of people with AD and the costs of treatment and care associated with AD is increasing rapidly. AD has become a major threat to patient safety and socio-economic well-being. Early diagnosis of AD and its prodromal stage, mild cognitive impairment (MCI), can contribute significantly to improving the clinical outcome of people with AD. Neuroimaging techniques, such as structural magnetic resonance imaging (sMRI), fluorodeoxyglucose positron emission tomography (FDG-PET) and AV-45 PET, have proven to be effective tools for the early diagnosis of AD/MCI.

In recent years, with the development of machine learning, combining neuroimaging data with machine learning for AD/MCI classification has become a hot research topic. In AD/MCI classification studies based on image data and machine learning, most studies use only single modal image data. For example, Klöppel et al. [5], Chu et al. [6], Ahmed et al. [7], and Liu et al. [8] used only MRI images and support vector machines (SVM) to distinguish AD/MCI from normal controls (NC). Similarly, Illán et al. [9], Padilla et al. [10], and Jiang et al. [11] used only PET images and SVM for AD/MCI classification. However, more and more studies have confirmed that data from different modalities can provide complementary information [12–15], with the combined modality of MRI and PET being the most commonly used multimodal AD/MCI diagnostic option. Since MRI and PET images can provide structural and metabolic information about the patient's brain in respect, the simultaneous use of MRI and PET images can complement each other and improve the accuracy of AD diagnosis. For example, Liu et al. [12] and Suk et al. [13] combined MRI images and PET images to classify AD and obtained an accuracy of 91.40% and

☆ This work was supported by the Guangxi Key Projects of Science and Technology (Grant No. 2020AA21077007), the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (MIMS22-09) and the Project of Improving the Basic Scientific Research Ability of Young and Middle-Aged Teachers in Guangxi Universities (2023KY0170).

* Corresponding author at: School of Artificial Intelligence, Guangxi Minzu University, Nanning 530006, Guangxi, China.
** Corresponding author at: School of Computer, Electronics and Information, Guangxi University, Nanning 530004, Guangxi, China.
E-mail addresses: laohuanlh@gxmzu.edu.cn (H. Lao), xjzhang@gxu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.jbi.2023.104326
Received 8 December 2022; Received in revised form 1 February 2023; Accepted 1 March 2023
Available online 3 March 2023

95.35%, respectively; Zhu et al. [14] and Gray et al. [15] introduced MRI- and PET-based cerebrospinal fluid (CSF) data and genetic data, respectively, to further improve the classification performance of AD/MCI.

Although the multimodal AD/MCI approaches described above have produced good results, they still face two main challenges. First, different modalities have different distributions of data, and integrating these multimodal data to understand AD pathology effectively is still arduous. Many early studies directly combine feature matrices of distinct modalities of data, which is not enough to mine the hidden information related to pathology. To address such a problem, Zhang et al. [16] proposed a combination of multi-kernel combination and SVM (called multi-kernel SVM, MKSVM) that integrates multimodal data effectively and embeds it naturally into a traditional SVM classifier. Second, there is an inherent correlation between multimodal data, and removing redundant or irrelevant features and extracting common features from multimodal data is still challenging. To solve this problem, Zhang et al. [17] proposed a multi-task feature selection (MTFS) method, which treats feature selection of different modal data as different tasks, extracts common features of multimodal data by group-sparsity regularizer, and finally classifies them using MKSVM. Based on the MTFS method, Liu et al. [18] and Jie et al. [19] retained structural information of similar data by introducing different constraints to extract more discernible common features, respectively. Among them, Liu et al. [18] selected complementarity information from multimodal data through the multimodal relational constrained multi-tasking feature selection (IMTFS for short) method. Jie et al. [19] capture the intrinsic correlation of multiple modal data through a manifold regularized multi-task feature learning (M2TFS for short) method. In general, the IMTFS and M2TFS methods preserve the underlying structural information of the data by modeling the relationships between like data. Compared to the underlying structural information of the data, the higher-order structural information of the data is more discriminative, which can further improve the classification performance of the model. Hypergraphs are a usual method of capturing information about the higher-order structure of data and can be used to model complex relationships between individuals through connections between hyperedges and vertices [20]. Hypergraph methods are now widely used in computer vision [21,22], bioinformatics [23,24], and medical image analysis [25,26]. Inspired by the M2TFS method and the hypergraph method, this study proposes a multi-task feature selection method, called hypergraph p-Laplacian regularized multi-task feature selection (abbreviated as HpMTFS), to better capture the correlation between sMRI images, FDG-PET images, and AV-45 PET images. Specifically, we treat feature selection of different modal data as distinct tasks and jointly extract common features of multimodal data by group-sparsity regularizer. Furthermore, we introduced two regularization terms simultaneously: (1) a hypergraph p-Laplacian regularization term. In this study, the complex relationships between brain images are described by hypergraphs, and the hypergraph Laplacian is usually used to capture the features of hypergraphs. We choose the hypergraph p-Laplacian, which has better performance than the ordinary hypergraph Laplacian, to capture the features of the hypergraph and better preserve the higher-order structural information of similar data; (2) a Frobenius norm regularization term. The model is constrained from the perspective of spatial constraints to enhance the sparsity of elements, simplify the model, and improve the noise resistance of the model. Finally, the extracted multimodal features are fused and classified using MKSVM. Experiments on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset have shown that the proposed approach can capture correlations between different modality data well and help improve the accuracy of AD/MCI classification.

The rest of the paper is organized as follows. In Section 2, we describe in detail the dataset and data pre-processing methods used for the study. In Section 3, we describe in detail the overall framework of this study, including hypergraph construction, multimodal feature selection, and multi-kernel learning. In Section 4, the evaluation of our proposed approach is presented. In Section 5, we discuss the effects of different regularization conditions and regularization parameters on classification performance and compare our approach with existing methods. In Section 6, we analyze the limitations of this study. Finally, we conclude Section 7.

Table 1
Demographic information of the subjects in this study.

Diagnosed	Gender(M/F)	Age (Mean ±SD)	MMSE (Mean ±SD)
AD	83/61	74.56 ±8.16	23.06 ±2.10
sMCI	80/52	71.62 ±7.19	28.26 ±1.62
pMCI	37/34	73.85 ±6.36	26.16 ±2.14
NC	87/94	71.79 ±6.96	27.87 ±1.70

Note. Abbreviations: MMSE=Mini-Mental State Examination; SD=Standard Deviation; AD=Alzheimer's disease; sMCI=stable mild cognitive impairment; pMCI=progressive mild cognitive impairment; NC=normal control.

2. Materials

2.1. Subject selection

The multimodal image data used in this study were obtained from the ADNI database (adni.loni.usc.edu). ADNI is a global study to explore the pathogenesis of AD and develop treatments to delay or prevent the progression of AD. As of 2022, the ADNI dataset has been studied in four phases, namely ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. In this study, we selected 528 subjects from the ADNI-2 stage with concurrent baseline sMRI images, FDG PET images, and AV-45 PET images, including 144 AD subjects, 132 sMCI subjects, 71 pMCI subjects, and 181 NC subjects. sMCI stands for stable MCI, where patients do not develop AD within 0–36 months, and pMCI stands for progressive MCI, where patients develop AD within 6–36 months. In particular, all sMRI images (labeled "MT1; GradWarp; N3m"), FDG PET images (labeled "Coreg, Avg, Std Img and Vox Siz, Uniform Resolution"), and AV-45 PET images (labeled "AV45 Coreg, Avg, Std Img and Vox Siz, Uniform Resolution") were collected from the ADNI dataset server in the most complete pre-processed format. Demographic information on all subjects is shown in Table 1.

2.2. Data pre-processing

In this study, for all sMRI images, we first performed skull stripping, tissue segmentation, alignment, and modulation using the "segmented data" module of the CAT12 toolkit [27] (neuro.uni-jena.de/cat/) to obtain grey matter images (called GM-MRI) with dimensions of $121 \times 145 \times 121$. Then divided the GM-MRI images of each subject into 90 regions of interest (ROIs) according to the automatic anatomical labeling (AAL) atlas, and finally calculated the grey matter volume of each ROI as a feature [28]. For all FDG PET images and AV-45 PET images, we first used the FLIRT module in the FSL software [29] (fsl.fmrib.ox.ac.uk/fsl/fslwiki) to register each subject's FDG PET images and AV-45 PET images to their respective corresponding pre-processed sMRI images, and then calculated the average intensity of each ROI as a feature by AAL atlas. Thus, 90 features were obtained for each modal image of each subject. In addition, as in the literature [30], we performed z-score normalization on the obtained feature data.

3. Method

The overall framework of this study is shown in Fig. 1, which mainly includes four steps: data pre-processing, hypergraph construction, multimodal feature selection, and classification. Specifically, the sMRI images, FDG PET images, and AV-45 PET images were first pre-processed to extract the grey matter volume or mean intensity of the 90 ROIs as pre-selected features (see Material 2 for detailed steps). Next,

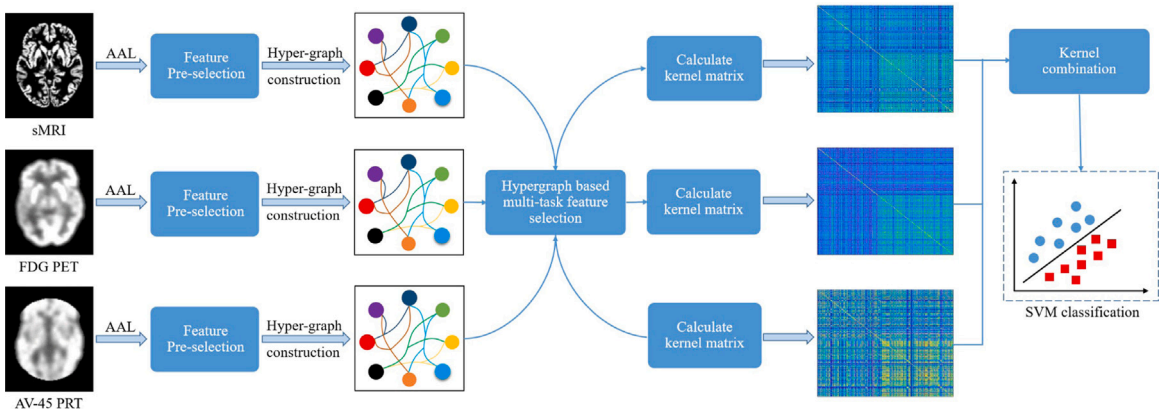
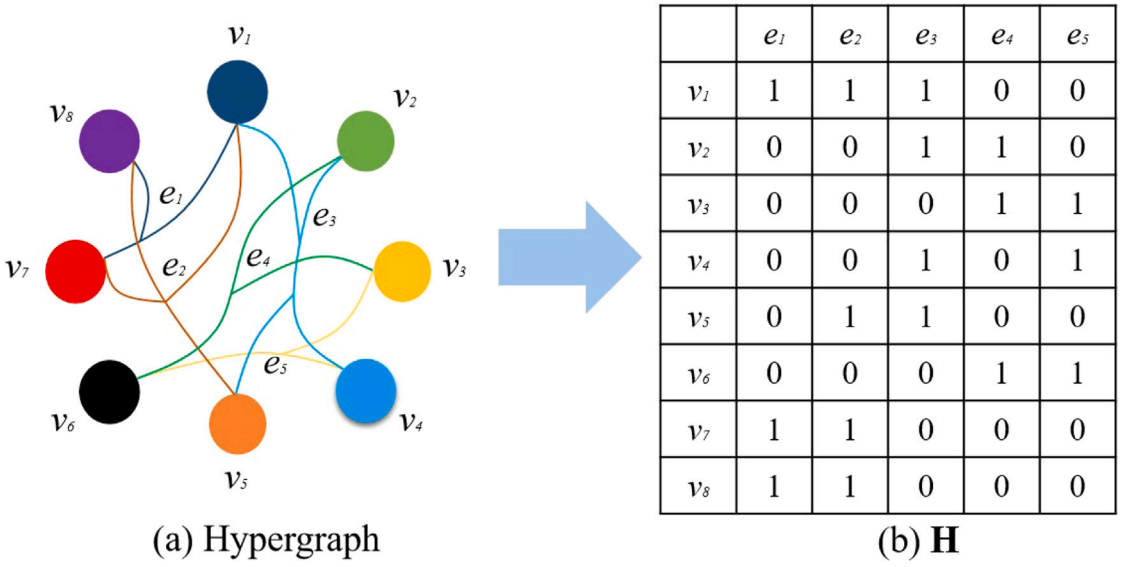


Fig. 1. The framework of multi-modality feature selection with hypergraph method.



	e_1	e_2	e_3	e_4	e_5
v_1	1	1	1	0	0
v_2	0	0	1	1	0
v_3	0	0	0	1	1
v_4	0	0	1	0	1
v_5	0	1	1	0	0
v_6	0	0	0	1	1
v_7	1	1	0	0	0
v_8	1	1	0	0	0

Fig. 2. Hypergraph model and corresponding incidence matrix.

based on the extracted features, a hypergraph was constructed for each modality of data to capture higher-order structural information of the data. Then, the most discriminative common features were selected from the pre-selected features by hypergraph p-Laplacian regularized multi-task feature selection method. Finally, MKSVM was used for multimodal feature fusion and classification. In the following, we describe the proposed method.

3.1. Hypergraph construction

Hypergraphs were first proposed by Berge [31] to describe complex relationships between data. Hypergraphs can connect $n(n \geq 2)$ vertices simultaneously via hyperedge and can therefore describe higher-order relationships between individuals, while traditional graph structures where each edge can connect only two vertices. Mathematically, a hypergraph G can be represented as $G(V, E, w)$, where V denotes vertexs, E denotes hyperedges, and w denotes the weight of the hyperedges. In the AD classification tasks, the specific structure of the hypergraph can be shown in Fig. 2(a), each vertex v_i represents a subject, and its value is the subject's feature vector. Each hyperedge e_i represents a connection between multiple vertices, and the weight $w(e)$ represents the importance of the connection. The structure of the hypergraph, i.e. the relationship between the hyperedges and the vertices, is usually represented by an incidence matrix $\mathbf{H} \in \{0,1\}^{|V|\times|E|}$ (Fig. 2(b)):

$$\mathbf{H}(v,e) = \begin{cases} 1, & if\ v \in e \\ 0, & if\ v \notin e \end{cases} \tag{1}$$

In the AD classification tasks, we can use the k-nearest neighbor strategy to generate a hypergraph. For subject v_i, the k nearest neighbors can be found by calculating the Euclidean distance between it and the other subjects, and then connecting these k+1 subjects via the hyperedge e_i. At this point, the v_i rows and e_i columns of the $\mathbf{H}$ matrix have a value of 1, indicating that the vertex v_i is connected to the hyperedge e_i (as shown on Fig. 2(b)). Thus, a complete hypergraph is a set of N (number of subjects) vertices and N hyperedges, all with a hyperedge weight of 1.

The hypergraph p-Laplacian operator is typically used to preserve higher-order structural features of the data [32–34], so we construct the hypergraph p-Laplacian matrix according to the method of Saito et al. [34]. First, calculate the degree of each vertex v and each hyperedge e according to the $\mathbf{H}$ matrix:

$$d(v) = \sum_{e\in E} w(e)\mathbf{H}(v,e) \tag{2}$$

$$\delta(e) = \sum_{v\in V} \mathbf{H}(v,e) \tag{3}$$

Secondly, let $\mathbf{D}v \in R^{|V|\times|V|}$, $\mathbf{D}e \in R^{|E|\times|E|}$, and $\mathbf{W} \in R^{|E|\times|E|}$ denote the degree matrices of vertexs, hyperedges and the weight of

hyperedges respectively (where $\mathbf{D}v_{i,i} = d(v_i)$, $\mathbf{D}e_{i,i} = \delta(e_i)$, $\mathbf{W}_{i,i} = w(e_i)$), then the adjacency matrix $\mathbf{A}$ can be defined as:

$$\mathbf{A} = \mathbf{A}_p - diag(diag(\mathbf{A}_p)) \tag{4}$$

where $diag(.)$ denotes the diagonal matrix, and

$$\mathbf{A}_p = \mathbf{H} \times \mathbf{W} \times (\mathbf{D}e - \mathbf{W})^{-1} \times \mathbf{H}^T \tag{5}$$

Finally, the hypergraph p-Laplacian matrix (p=2) can be calculated as:

$$\mathbf{L}^{ph} = \mathbf{D}v^{-1/2} \times (\mathbf{D}v - \mathbf{A}) \times \mathbf{D}v^{-1/2} \tag{6}$$

3.2. Multi-task feature selection

Multi-task learning [16,19] is a method for improving the performance of a particular model by simultaneously learning relevant information between M related but not identical tasks. In the AD classification tasks, considering the selection of features for each modal data as one task, the number of learning tasks M is the type of modal data. Let the feature matrices of the multimodal data and their corresponding class label matrices be $\mathbf{X}_m = [x_m(1), x_m(2), \ldots, x_m(N)]^T \in R^{N\times d}$ and $\mathbf{Y} = [y_1, y_2, \ldots, y_N]^T \in R^{N\times 1}$, where $x_m(i)$ is a column vector of size $d\times 1$ representing the d features of subject i; y_i is the class of subject i with a value of 1 or -1. Then the objective function of the linear multi-task feature selection model [35] can be defined as:

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{Y} - \mathbf{X}_m \mathbf{w}_m\|_2^2 + \mu \|\mathbf{W}\|_{2,1} \tag{7}$$

Where the first term is the empirical error and $\mathbf{w}_m \in R^d$ is the discriminant function parameter for task m. The second term is the group-sparsity regularization term, $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M] \in R^{d\times M}$ is a weight matrix of M modal vectors, and μ is the coefficient of the group-sparsity regularization term. The "group-sparsity" can be achieved by increasing the number of non-zero rows in the weight matrix by taking the $l_{2,1}$ norm of the matrix $\mathbf{W}$:

$$\|\mathbf{W}\|_{2,1} = \sum_{i=1}^{d} \|\mathbf{w}_i\|_2 \tag{8}$$

Since the feature selection task retains only those features whose coefficients are not zero, the introduction of the $l_{2,1}$ norm not only allows for the joint extraction of common features across multiple tasks but also further reduces the number of common features. It is worth noting that these common features were extracted from the same brain regions in different modality data so that we can interpret the effects of AD on brain regions from different perspectives (MRI: structural, PET: functional) based on these common features.

3.2.1. Hypergraph p-Laplacian regularized multi-task feature learning

In the M2TFS method, Jie et al. preserve information about the underlying structure of similar data by introducing manifold regularization term of the form $(\mathbf{Xw})^T\mathbf{L}(\mathbf{Xw})$. Inspired by this, we introduce the hypergraph p-Laplacian regularization term $(\mathbf{Xw})^T\mathbf{L}^{ph}(\mathbf{Xw})$ into the objective function to preserve the higher-order structural information of similar data. Therefore, the objective function of our HpMTFS method is defined as follows:

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{Y} - \mathbf{X}_m \mathbf{w}_m\|_2^2 + \mu \|\mathbf{W}\|_{2,1} + \lambda \sum_{m=1}^{M} (\mathbf{X}_m \mathbf{w}_m)^T \mathbf{L}_m{}^{ph} (\mathbf{X}_m \mathbf{w}_m) \tag{9}$$

The first two terms are the same as Eq. (7), $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$, $M = 3$. The third term is the hypergraph p-Laplacian regularization term, λ is its coefficient, and $\mathbf{L}_m{}^{ph}$ is the hypergraph p-Laplacian matrix of the m-th modal data. In addition, to improve the noise immunity of the model, we also introduce a Frobenius norm regularization term in the objective function, which enhances the sparsity of the data and makes each element of $\mathbf{W}$ infinitely close to zero, thus simplifying the model. Therefore, the final objective function is defined as follows:

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{Y} - \mathbf{X}_m \mathbf{w}_m\|_2^2 + \mu \|\mathbf{W}\|_{2,1} + \lambda \sum_{m=1}^{M} (\mathbf{X}_m \mathbf{w}_m)^T \mathbf{L}_m{}^{ph} (\mathbf{X}_m \mathbf{w}_m) + \gamma \|\mathbf{W}\|_F^2 \tag{10}$$

In summary, our HpMTFS method not only extracts common features jointly from multimodal data but also retains higher-order structural information of similar data through hypergraph p-Laplacian regularization term, resulting in more discriminative features.

3.2.2. Optimization algorithm

Since existing sparse learning models are unable to solve objective functions that contain both group-sparsity regularization term, hypergraph p-Laplacian regularization term and Frobenius norm regularization term, we use the accelerated approximate gradient (APG) method [36,37] to solve our objective function. First, separate the objective function in Eq. (10) to a smooth part (Eq. (11)) and a nonsmooth part (Eq. (12)):

$$g(\mathbf{W}) = \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{Y} - \mathbf{X}_m \mathbf{w}_m\|_2^2 + \lambda \sum_{m=1}^{M} (\mathbf{X}_m \mathbf{w}_m)^T \mathbf{L}_m{}^{ph} (\mathbf{X}_m \mathbf{w}_m) + \gamma \|\mathbf{W}\|_F^2 \tag{11}$$

$$h(\mathbf{W}) = \mu \|\mathbf{W}\|_{2,1} \tag{12}$$

The objective function can then be expressed as $f(\mathbf{W}) = g(\mathbf{W}) + h(\mathbf{W})$. Next, constructed the function Ω to approximate $f(\mathbf{W})$:

$$\mathbf{\Omega}_l(\mathbf{W}, \mathbf{W}_i) = g(\mathbf{W}_i) + (\mathbf{W} - \mathbf{W}_i, \nabla g(\mathbf{W}_i)) + \frac{l}{2} \|\mathbf{W} - \mathbf{W}_i\|_F^2 + h(\mathbf{W}) \tag{13}$$

Where $\nabla g(\mathbf{W}_i)$ is the gradient of $g(\mathbf{W})$ at the ith iteration point $\mathbf{W}_i$ and l is the step length. Finally, the APG algorithm is updated by Eq. (14):

$$\begin{aligned} \mathbf{W}_{i+1} &= \arg\min_{\mathbf{W}} \frac{1}{2} \|\mathbf{W} - \mathbf{O}\|_F^2 + \frac{1}{l} h(\mathbf{W}) \\ &= \arg\min_{w_1,\ldots,w_d} \frac{1}{2} \sum_{j=1}^{d} (\|\mathbf{w}_j - \mathbf{o}_j\|_2^2 + \frac{\mu}{l} \|\mathbf{w}_j\|_2) \end{aligned} \tag{14}$$

Where $\mathbf{w}_j$ and $\mathbf{o}_j$ denote the jth row of matrices $\mathbf{W}$ and $\mathbf{O}$ respectively. The step length l is determined by a linear search, and the matrix $\mathbf{O}$ can be calculated using Eq. (15):

$$\mathbf{O} = \mathbf{W}_i - \frac{1}{l} \nabla g(\mathbf{W}_i) \tag{15}$$

Furthermore, following Xi et al. [37] and Liu et al. [38], we decompose the update step of the APG algorithm (Eq. (14)) into d independent subproblems to solve:

$$\mathbf{w}_j^* = \begin{cases} (1 - \frac{\mu}{l\|\mathbf{o}_j\|_2})\mathbf{o}_j, & if\ \|\mathbf{o}_j\|_2 > \frac{\mu}{l} \\ 0, & otherwise \end{cases} \tag{16}$$

And solving Ω by computing the search point $\mathbf{Q}$ according to the method proposed by Liu et al. [38]:

$$\mathbf{Q}_i = \mathbf{W}_i + \eta_i (\mathbf{W}_i - \mathbf{W}_{i-1}) \tag{17}$$

$$\eta_i = \frac{(1 - \psi_{i-1})\psi_i}{\psi_{i-1}} \tag{18}$$

$$\psi_i = \frac{2}{i+3} \tag{19}$$

The specific optimization process is described in Algorithm 1, where K is the maximum number of iterations, σ is the factor controlling the step size l.

Algorithm 1

Input:
Multimodal data matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M]^T$
Class label matrix $\mathbf{Y} = [y_1, y_2, ..., y_N]^T$
Output:
$\mathbf{W}_K, J^*$
Initialization: $\mu \geq 0, \lambda \geq 0, \gamma \geq 0, l_0 \geq 0, \sigma \geq 1, \mathbf{W}_0 = 0, \psi_0 = 1$;
For $i = 1$ to K
1: Compute the search point $\mathbf{Q}_i$ (Equation 17)
2: $l = l_{i-1}$
3: Calculate $\mathbf{W}_{i+1}$ (Equation 14)
4: While $f(\mathbf{W}_{i+1}) > \Omega_l(\mathbf{W}_{i+1}, \mathbf{Q}_i)$, $l = \sigma l$ (Calculate Ω_l from Equation 13)
5: Set $l_i = l$
End
Calculate $J^* = \{j | |\mathbf{w}_j| > 0.01, j = 1, 2, ..., d\}$

3.2.3. Multi-kernel SVM

In this study, we used MKSVM to fuse and classify the final sMRI, FDG PET, and AV-45 PET features. Specifically, generate a corresponding kernel matrix for each modal data at first:

$$k_m(\mathbf{x}_m(i), \mathbf{x}_m(j)) = \phi((\mathbf{x}_m(i))^T \mathbf{x}_m(j)) \tag{20}$$

The M kernel matrices were then linearly fused:

$$k(\mathbf{x}(i), \mathbf{x}(j)) = \sum_{m=1}^{M} \beta_m k_m(\mathbf{x}_m(i), \mathbf{x}_m(j)) \tag{21}$$

Where $\beta m(\beta_m > 0, \beta_1 + \cdots + \beta_M = 1)$ is the kernel combination weight, and the optimal β_m is determined by cross-validation on the training set via grid search. Finally, the optimal β_m and linear kernel are combined to train the MKSVM model. When a new subject data $\mathbf{x}_m$ is input to the trained MKSVM model, the MKSVM model predicts the category of that subject by the following decision function:

$$f(\mathbf{x}) = \mathrm{sign}\left(\sum_{i=1}^{N} y_i \alpha_i \sum_{m=1}^{M} \beta_m k_m(\mathbf{x}_m(i), \mathbf{x}_m) + b\right) \tag{22}$$

In particular, in our experiments, we consider the absolute value of the components in the normal vector of the linear SVM hyperplane (denoted $\mathbf{W}_{SVM}$) to be the weights of the features [5]. Thus, we only need to rank the $\mathbf{W}_{SVM}$ to filter out the brain regions that contribute most to the AD/MCI classification task.

4. Experimentation and analysis

4.1. Experimental settings

To validate the effectiveness of our HpMTFS method, we conducted four sets of binary classification experiments on the ADNI-2 dataset: (1) AD vs. NC, (2) pMCI vs. sMCI, (3) MCI vs. NC, and (4) AD vs. NC. For each set of binary classification experiments, we used a 10-fold cross-validation strategy to avoid the impact of bias on classification performance caused by randomly dividing the dataset. In addition, we used four specific metrics to assess classification performance: classification accuracy (ACC=$(TP+TN)/(TP+TN+FP+FN)$), sensitivity (SEN=$TP/(TP+FN)$), specificity (SPE=$TN/(TN+FP)$), and area under the subject operating characteristic curve (AUC), where TP, TN, FP, and FN indicate true positive, true negative, false positive and false negative, respectively.

Notably, our HpMTFS method involves four hyperparameters, namely the number of neighbors k, the group-sparsity regularization coefficient μ, the hypergraph p-Laplace regularization coefficient λ, and the Frobenius norm regularization coefficient γ. All four hyperparameters were determined by grid search with 10-fold cross-validation on the training set. Specifically, in the 10-fold cross-validation used to calculate classification performance, we perform another 10-fold cross-validation to find the optimal μ, λ and γ before training on each fold and then use the optimal hyperparameters found to train on that fold. Next, different k values are set to perform 10-fold cross-validation, which is used to calculate the classification performance and determine the optimal k value. The range of values of k is $\{3, 5, 7, 10, 20, 30, 40\}$ and the range of values of μ, λ and γ are $\{0.001, 0.01, 0.1, 1, 10, 100\}$. In addition, in the MKSVM with a linear kernel, the value of C is 1, the kernel combination coefficients β_{MRI} and $\beta_{FDG-PET}$ take values in the range $\{0, 0.3, 0.6, 0.9\}$, and $\beta_{AV45-PET} = 1 - \beta_{MRI} - \beta_{FDG-PET}$.

4.2. Classification results

Table 2 shows the classification results for AD vs. NC, pMCI vs. sMCI, MCI vs. NC, and AD vs. MCI (number of nearest neighbor k = 7). To highlight the superiority of the proposed method, we used a single modal image (sMRI or FDG PET or AV-45 PET) and different combinations of multimodal images (sMRI+FDG PET or FDG PET+AV-45 PET or sMRI+AV-45 PET or sMRI+FDG PET+AV-45 PET (All)) to perform each group of classification tasks. As can be seen from Table 2: (1) classification performance using multimodal images is mostly better than that using only single modal images for all classification tasks, as multimodal image features achieve complementary information and provide more valid information. (2) Classification results using three modal images outperform those using two modal images because three image features contain more valid information than two image features. (3) The classification accuracy of our proposed method was 98.78% and 86.47% for the AD and NC classification task and the AD and MCI classification task, respectively. It indicates that our proposed method can effectively distinguish AD patients from NC subjects or MCI patients. (4) Most of the classification accuracies of existing research methods for pMCI and sMCI classification task are below 90% [8,39], while our proposed method achieves 92.62% classification accuracy, 84.29% sensitivity, and 96.98% specificity using three modal images. It indicates that our proposed method has good classification performance in distinguishing different subtypes of MCI.

Furthermore, as can be seen in Table 2, the classification performance of our proposed method for MCI and NC is not very satisfactory, with an accuracy of 78.15%, a sensitivity of 70%, and a specificity of 87.31%. Therefore, we further classified AD, pMCI, sMCI, and NC. Table 3 shows the classification results for AD vs. pMCI, AD vs. sMCI,

Table 2
Classification results using different modalities images in AD vs. NC, pMCI vs. sMCI, MCI vs. NC, and AD vs. MCI.

Data	AD vs. NC (%)			pMCI vs. sMCI (%)			MCI vs. NC (%)			AD vs. MCI (%)		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
sMRI	95.09	93.05	96.70	84.71	74.46	90.11	74.48	70.45	78.98	80.44	70.90	87.17
FDG PET	73.91	78.71	70.18	60.62	42.86	70.55	58.01	57.14	59.04	64.51	62.52	65.95
AV-45 PET	52.35	41.76	60.64	47.26	29.46	56.87	48.15	51.19	44.71	49.85	37.91	58.12
sMRI+FDG PET	97.85	96.52	98.89	86.74	73.21	93.90	77.38	69.07	86.78	86.18	76.52	93.10
sMRI+AV-45 PET	96.00	93.71	97.87	88.69	81.61	92.47	76.08	66.55	86.81	78.71	61.95	90.67
FDG PET+AV-45 PET	96.61	92.33	1.00	88.19	78.75	93.13	76.35	67.10	86.73	83.83	71.52	92.60
All	**98.78**	97.24	1.00	**92.62**	84.29	96.98	**78.15**	70.00	87.30	**86.47**	76.00	94.07

Table 3
Classification results using different modalities images in AD vs. pMCI , AD vs. sMCI , pMCI vs. NC, and sMCI vs. NC.

Data	AD vs. pMCI (%)			AD vs. sMCI (%)			pMCI vs. NC (%)			sMCI vs. NC (%)		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
sMRI	77.66	90.05	51.96	89.54	91.10	87.86	92.42	84.29	95.56	73.83	53.08	88.95
FDG PET	62.79	75.05	38.57	72.51	76.43	68.24	74.18	55.00	81.75	54.39	37.20	66.75
AV-45 PET	41.66	41.71	41.59	60.50	76.81	26.96	61.91	38.04	71.29	52.65	44.78	58.45
sMRI+FDG PET	81.39	84.67	74.64	92.78	90.33	95.44	96.43	88.75	99.47	74.47	53.19	90.00
sMRI+AV-45 PET	71.60	73.71	67.68	90.31	85.48	95.49	96.02	87.32	99.44	75.11	55.44	89.50
FDG PET+AV-45 PET	79.98	80.48	78.93	92.03	92.38	91.65	96.02	90.00	98.36	75.77	54.78	91.14
ALL	**83.22**	84.67	80.54	**94.61**	92.43	96.98	**98.40**	94.29	1.00	**77.36**	56.32	92.81

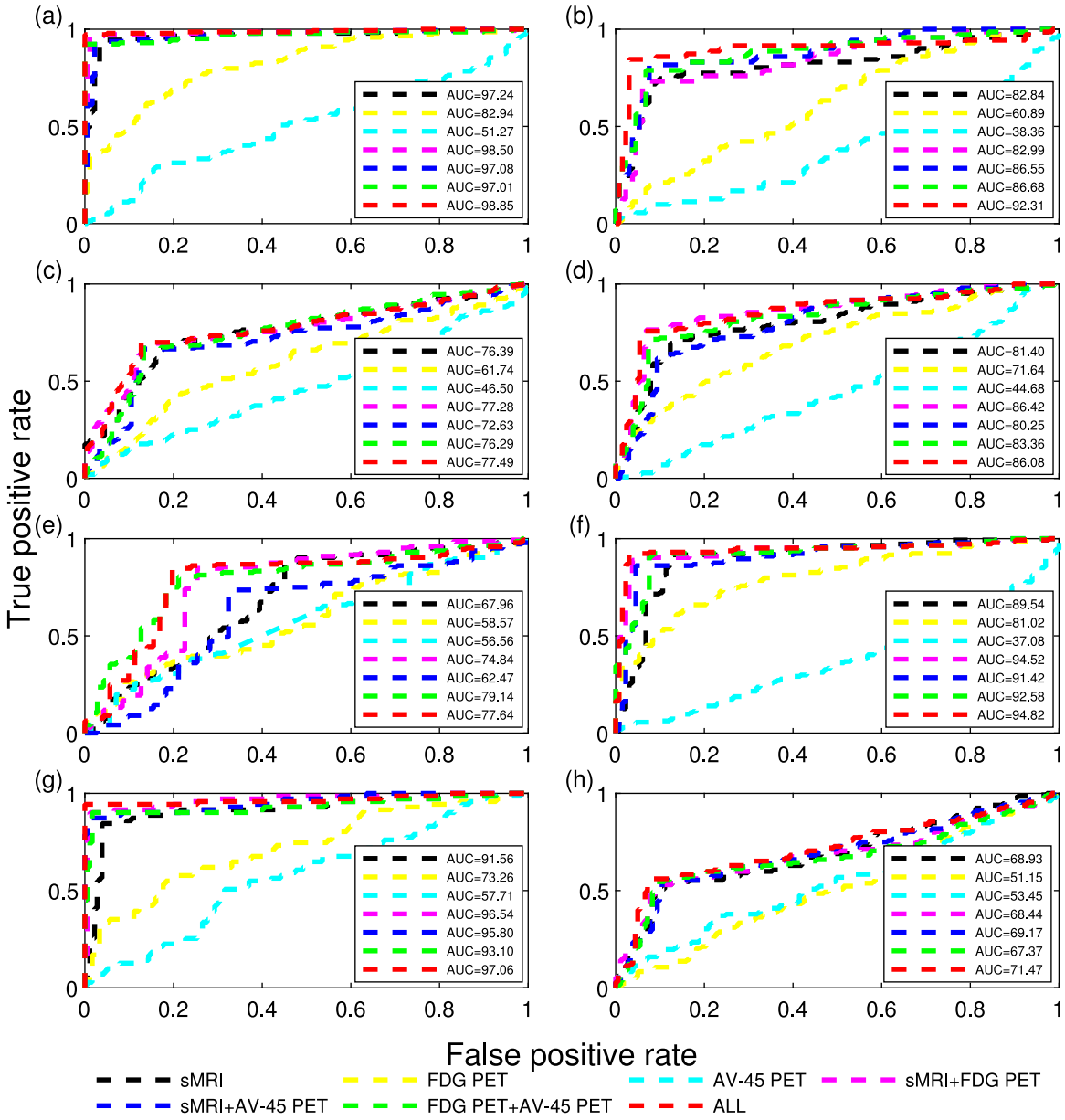


Fig. 3. ROC curves obtained for different classification tasks using different methods. (a) AD vs. NC, (b) pMCI vs. sMCI, (c) MCI vs. NC, (d) AD vs. MCI, (e) AD vs. pMCI, (f) AD vs. sMCI, (g) pMCI vs. NC, and (h) sMCI vs. NC.

pMCI vs. NC, and sMCI vs. NC, respectively. From Table 3, we can also see that most of the classification performances using multimodal images are better than those using only single modal images. Interestingly, the classification accuracy of our proposed method was greater than 94.61% for pMCI vs. NC and AD vs. sMCI, compared to 83.22% and 77.36% for AD vs. pMCI and sMCI vs. NC, respectively. We speculate that this may be because pMCI subjects eventually develop AD, and therefore their features extracted from sMRI images, FDG PET images, and AV-45 PET images may be similar to those of AD subjects, whereas sMCI subjects do not develop AD, and therefore their features extracted from sMRI images, FDG PET images, and AV-45 PET images may be similar to those of NC subjects. Therefore, the final AD and pMCI

Table 4
The AAL atlas names of the top 10 contributing brain regions in different classification tasks.

Classification tasks	Top 10 brain regions
AD vs. NC	Hippocampus_L (No. 37), Temporal_Mid_R (No. 86), Putamen_L (No. 73), Fusiform_R (No. 56), Amygdala_L (No. 41), Amygdala_R (No. 42), Temporal_Inf_R (No. 90), Calcarine_R (No. 44), Angular_R (No. 66), Hippocampus_R (No. 38)
pMCI vs. sMCI	Temporal_Mid_R (No. 86), Cingulum_Mid_R (No. 34), Frontal_Mid_L (No. 7), Caudate_R(No. 72), Cingulum_Ant_R (No. 32), Occipital_Mid_R (No. 52), Temporal_Inf_R (No. 90), Temporal_Inf_L (No. 89), Temporal_Pole_Sup_L (No. 83), Fusiform_R (No. 56)
MCI vs. NC	Hippocampus_R (No. 38), Frontal_Mid_R (No. 8), Cingulum_Mid_L (No. 33), Hippocampus_L (No. 37), Fusiform_R (No. 56), Fusiform_L (No. 55), Frontal_Sup_Orb_L (No. 5), Temporal_Mid_R (No. 86), Caudate_R (No. 72), Supp_Motor_Area_R (No. 20)
AD vs. MCI	Temporal_Inf_R (No. 90), Frontal_Mid_Orb_R (No. 26), Frontal_Mid_L (No. 7), Insula_R (No. 30), Angular_R (No. 66), Temporal_Inf_L (No. 89), Cingulum_Post_R (No. 36), Calcarine_L (No. 43), Occipital_Mid_L (No. 51), Temporal_Mid_R (No. 86)
AD vs. pMCI	Thalamus_R (No. 78), Frontal_Sup_Orb_R (No. 6), Thalamus_L (No. 77), Caudate_R (No. 72), Occipital_Sup_L (No. 49), Angular_R (No. 66), Cingulum_Post_R (No. 36), Hippocampus_R (No. 38), Temporal_Inf_L (No. 89), Frontal_Inf_Tri_R (No. 14)
AD vs. sMCI	Frontal_Mid_Orb_R (No. 26), Rectus_R (No. 28), Frontal_Sup_Orb_R (No. 6), Olfactory_R (No. 22), Fusiform_L (No. 55), Temporal_Mid_R (No. 86), Supp_Motor_Area_R (No. 20), Frontal_Mid_Orb_L (No. 25), Putamen_R (No. 74), Cingulum_Mid_R (No. 34)
pMCI vs. NC	Hippocampus_L (No. 37), Frontal_Sup_Orb_R (No. 6), Temporal_Mid_R (No. 86), Cingulum_Mid_R (No. 34), Frontal_Mid_L (No. 7), Thalamus_L (No. 77), Cingulum_Mid_L (No. 33), Insula_L (No. 29), Precuneus_R (No. 68), Fusiform_R (No. 56)
sMCI vs. NC	Frontal_Inf_Oper_R (No. 12), Cingulum_Ant_L (N0. 31), Hippocampus_L (No. 37), Frontal_Mid_Orb_L (No. 25), Rectus_L (No. 27), Cingulum_Post_L (No. 35), Frontal_Inf_Tri_R (No. 14), Cingulum_Mid_L (No. 33), Frontal_Mid_Orb_R (No. 26), Frontal_Sup_Orb_R (No. 6)

classification and sMCI and NC classification gave poor classification results.

In addition, to further evaluate the robustness of the classification model, we plot ROC plots of eight groups of binary classification tasks using different modality images. From Fig. 3(a)–(h), it can be seen that (1) in all classification tasks, the AUC values using three modal image classification (0.998, 0.923, 0.775, 0.861, 0.776, 0.948, 0.971, 0.715) are greater than those using single modal image classification, and most of the AUC values using three modal image classification are greater than those using two modal images. It indicates that the classification model obtained by training with the features extracted by the proposed method has good robustness. (2) The AUC values for the MCI vs. NC classification are smaller than those for AD vs. NC, pMCI vs. sMCI, and AD vs. MCI, which is consistent with the results for the first four groups (Table 2). The AUC values for AD vs. pMCI were lower than those for AD vs. sMCI, and the AUC values for sMCI vs. NC were lower than those for pMCI vs. NC, similar to the classification results for the latter four groups (Table 3).

4.3. Brain region analysis

To identify the brain regions that contributed most to the AD classification tasks, we rank the contributions of all brain regions by the $\mathbf{W}_{SVM}$ mentioned in Section 3.2.3. Table 4 lists the AAL atlas names of the top 10 contributing brain regions in the eight classification tasks of AD vs. NC, pMCI vs. sMCI, MCI vs. NC, AD vs. MCI, AD vs. pMCI, AD vs. sMCI, pMCI vs. NC, and sMCI vs. NC, with their specific visualizations are shown in Fig. 4(a)–(h). Since the first four classification tasks overlapped with the last four classification tasks (e.g. AD vs. MCI and AD vs. pMCI (or AD vs. sMCI)), we only analyzed the top 10 brain regions of the first four classification tasks in more detail. For AD and NC classification, our approach selected the hippocampus (37/38), amygdala (41/42), temporal pole (86/90), and other brain regions; For pMCI and sMCI classification, our approach selected brain regions in the para-cingulate gyrus (32/34), temporal pole (83/86/89/90) and caudate nucleus (72); For MCI and NC classification, our approach selected brain regions in the hippocampus (37/38), pallidum (55/56) and caudate nucleus (72); For AD and MCI classification, our approach selected brain regions such as the temporal pole (86/89/90), posterior cingulate gyrus (36) and middle frontal gyrus (7). Many studies have confirmed the link between these brain areas and AD/MCI [40–43]. For example, Ban et al. [41] found significant changes in hippocampus and thalamus in AD and MCI patients compared to NC controls; Persson et al. [42] found that the hippocampus was smaller and the caudate nucleus was larger in AD/MCI patients. In addition, as early as 1994, the temporal pole was identified as the more severely damaged brain region in AD patients [43]. Overall, our approach achieves classification by working with AD-sensitive brain regions, suggesting that our method is applicable to AD classification tasks.

5. Discussion

5.1. Effect of different regularization terms on classification performance

To investigate the effect of the group-sparsity regularization term, the hypergraph p-Laplace regularization term, and the Frobenius norm regularization term on classification performance, we set one (or two or three) regularization coefficients to 0, and observed the change in classification performance for each of the eight classification tasks. Fig. 5 shows the changes in accuracy, sensitivity, and specificity after the introduction of the different regularization terms, respectively. As can be seen in Fig. 5, the accuracy, sensitivity, and specificity of each group of classification tasks improved to varying degrees with the introduction of different regularization terms. Specifically, when $\mu = \lambda = \gamma = 0$, i.e. when no regularization term is introduced, our method degenerates to the MKSVM method. When $\lambda = \gamma = 0$, i.e. when group-sparsity regularization is introduced, our method degenerates to the MTFS method, where the classification accuracy, sensitivity, and specificity are higher than in the previous case. It suggests that group-sparsity regularization terms help to extract fewer relevant features, thus improving the accuracy of classification. When $\gamma = 0$, i.e. when both group-sparsity regularization and hypergraph p-Laplacian regularization are introduced, our method is similar to the M2TFS method, with higher classification accuracy, sensitivity, and specificity than the first two cases. It is good evidence that retaining the higher-order structural information of the modal data can improve the classification

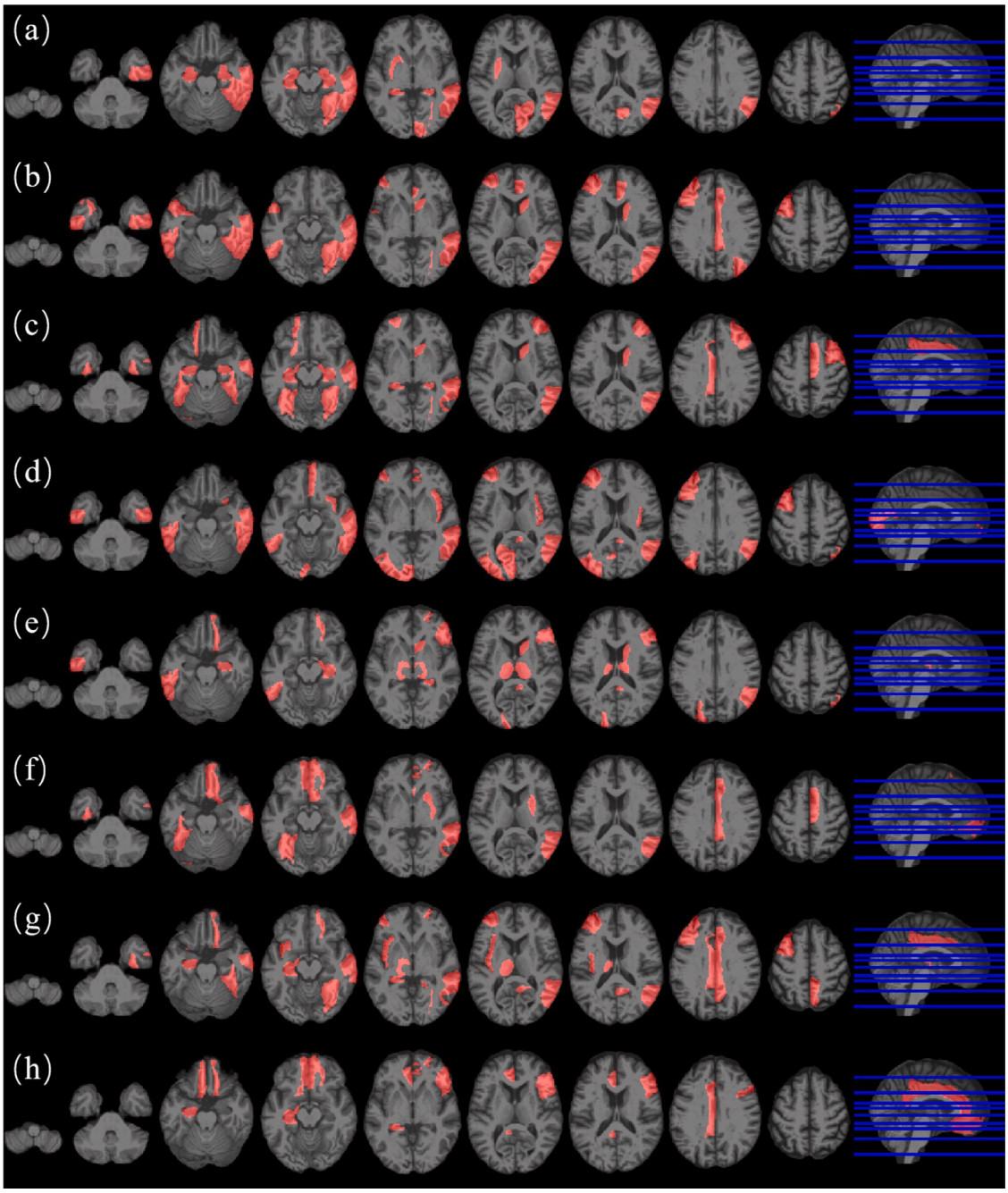
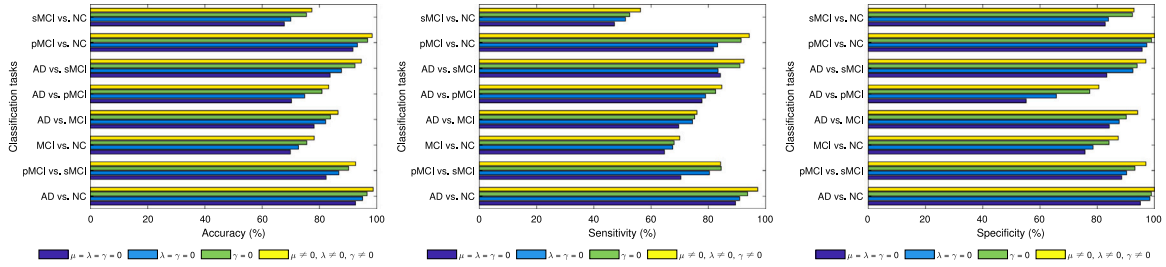


Fig. 4. The brain regions corresponding to the top 10 weight values for different classification tasks (the blue lines on the right-sided sagittal image correspond to axial image slices). (a) AD vs. NC. (b) pMCI vs. sMCI. (c) MCI vs. NC. (d) AD vs. MCI. (e) AD vs. pMCI. (f) AD vs. sMCI. (g) pMCI vs. NC. (h) sMCI vs. NC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

performance of the model. When $\mu \neq 0$, $\lambda \neq 0$, and $\gamma \neq 0$, i.e. when all three regularization terms are introduced, the overall classification performance is better than in the first three cases, indicating that the Frobenius norm regularization is more resistant to interference.

5.2. Effect of different hyperparameters on classification performance

In this study, we used four hyperparameters, including a nearest neighbor number k and three regularization parameters (the group-sparsity regularization coefficient μ, the hypergraph p-Laplacian regularization coefficient λ, and the Frobenius norm regularization coefficient γ), which need to be adjusted. Figs. 6 and 7 show the effect of different values of μ, λ, and γ on each group of classification tasks when the value of k is 7. In Fig. 6(a)–(h), the x-axis represents the range of μ, the y-axis represents the classification accuracy, and the different colored curves represent the different values of λ. When the value of γ is fixed at 1, the λ curves for all classification tasks do not fluctuate much as μ increases to 10, indicating that our method has good stability over an appropriate range of μ values. When μ is greater than 10, all λ curves tend to decrease. This may be because when μ is too large, there is too much sparsity resulting in too few common features being extracted. In Fig. 7(a)–(h), the x-axis represents the range of values of γ and the rest is similar to Fig. 6. When μ is fixed at 10, all λ curves show large fluctuations, indicating that the higher-order structural features of the data have a greater impact on the classification results.



(a) (b) (c)

Fig. 5. The changes in accuracy, sensitivity, and specificity after the introduction of the different regularization terms, respectively. (a) Accuracy under different regularization conditions. (b) Sensitivity under different regularization conditions. (c) Specificity under different regularization conditions. Among them, $\mu = \lambda = \gamma = 0$ indicates no regularization term is introduced; $\lambda = \gamma = 0$ indicates group-sparsity regularization is introduced; $\gamma = 0$ indicates both group-sparsity regularization and hypergraph p-Laplacian regularization are introduced; $\mu \neq 0$, $\lambda \neq 0$, $\gamma \neq 0$ indicates group-sparsity regularization, hypergraph p-Laplacian regularization, and Frobenius norm regularization are introduced simultaneously.

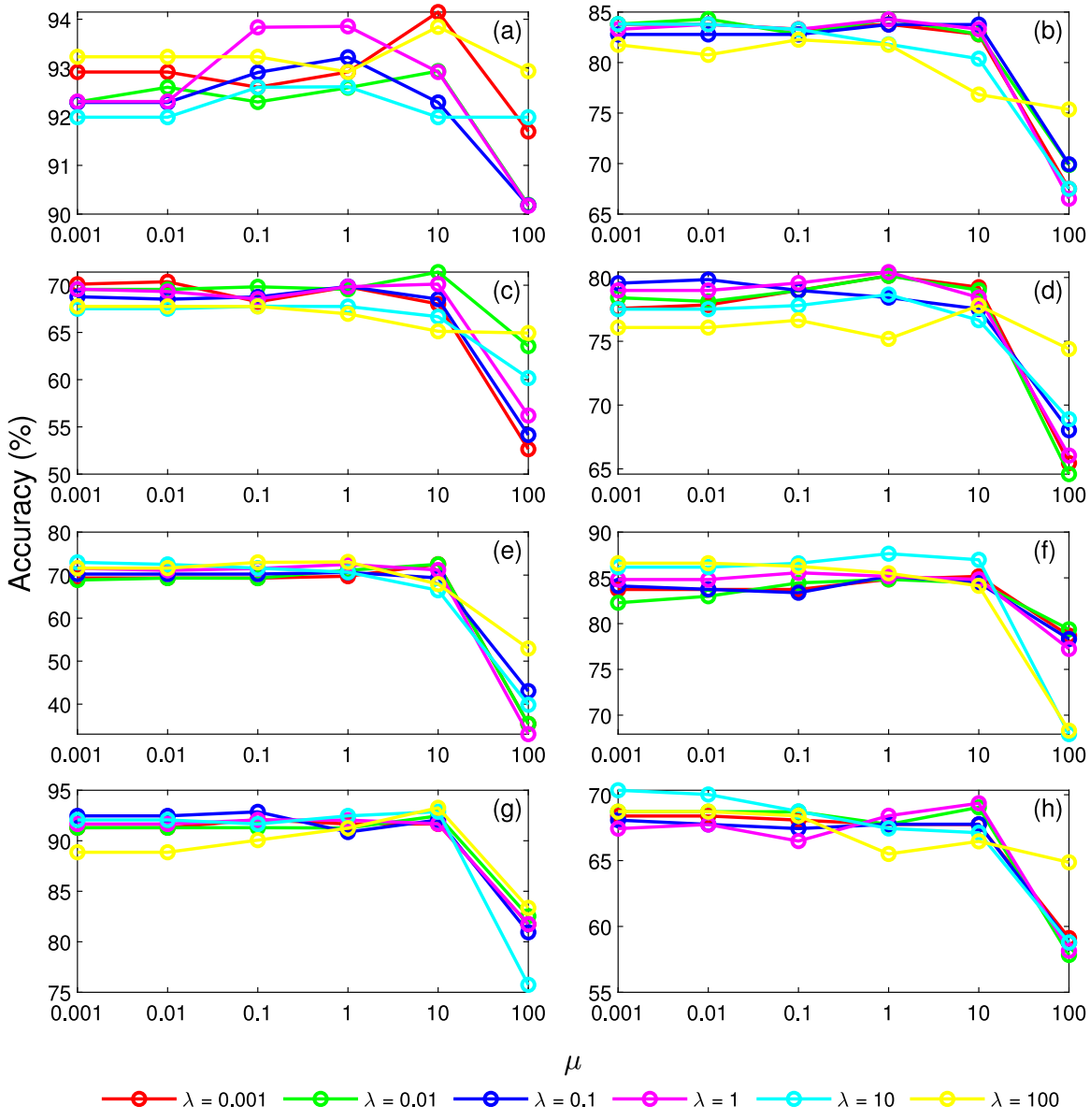


Fig. 6. The effect of different μ and λ values on the classification task when the γ value is fixed (γ=1). (a) AD vs. NC, (b) pMCI vs. sMCI, (c) MCI vs. NC, (d) AD vs. MCI, (e) AD vs. pMCI, (f) AD vs. sMCI, (g) pMCI vs. NC, and (h) sMCI vs. NC. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

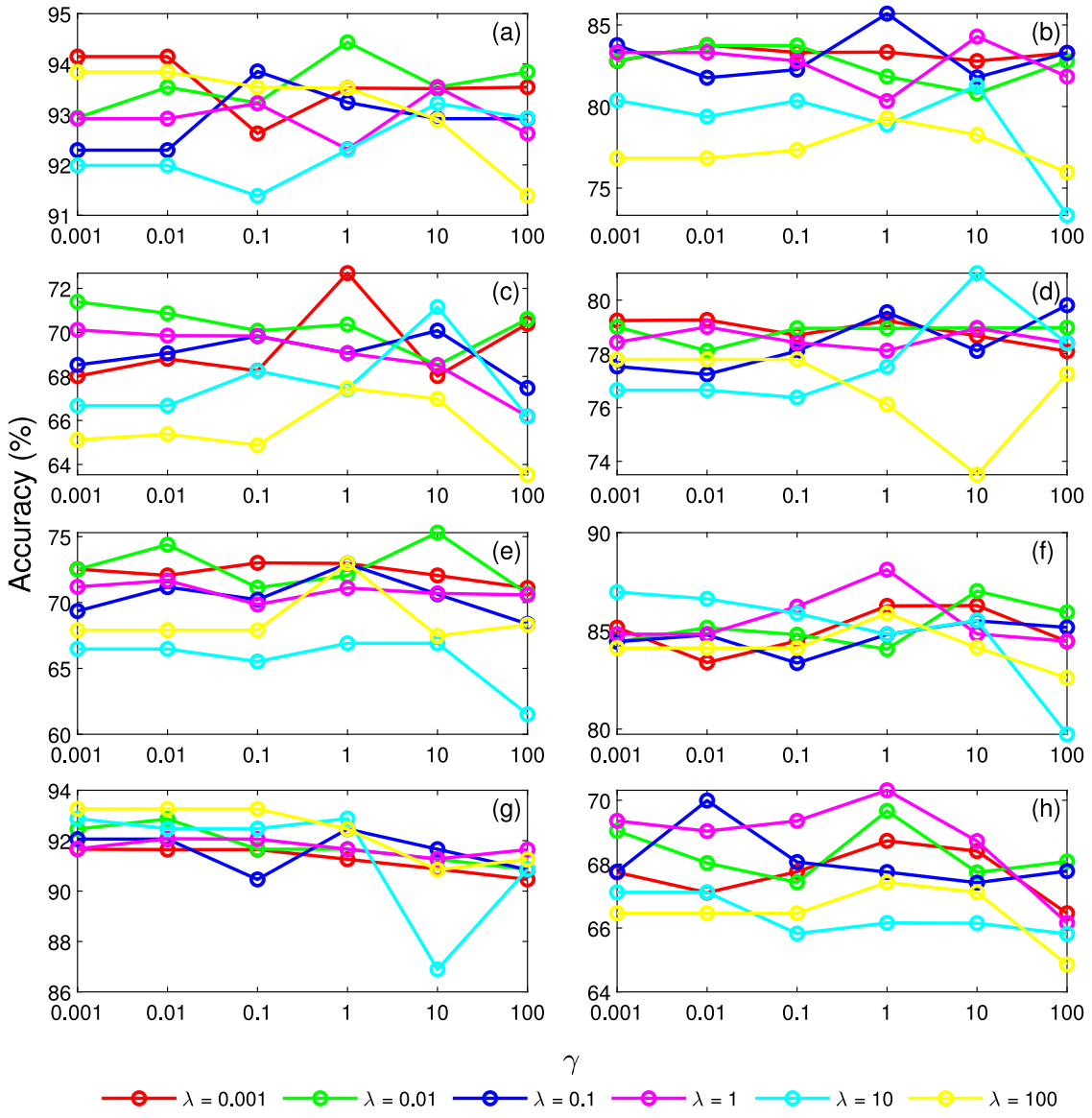


Fig. 7. The effect of different γ and λ values on the classification task when the μ value is fixed (μ=10). (a) AD vs. NC, (b) pMCI vs. sMCI, (c) MCI vs. NC, (d) AD vs. MCI, (e) AD vs. pMCI, (f) AD vs. sMCI, (g) pMCI vs. NC, and (h) sMCI vs. NC.

From the above analysis, it can be seen that the optimal μ, λ, and γ values for different groups of classification tasks are different. Based on the optimal μ, λ, and γ, we further analyzed the effect of the number of nearest neighbors on the classification results. Fig. 8 shows the effect of different k values on each group of classification tasks, where the x-axis indicates the number of nearest neighbors, the y-axis indicates the classification accuracy, and the different colored curves represent different classification tasks. As can be seen from Fig. 8, the curves for each color do not fluctuate much as the value of k increases, meaning that the number of nearest neighbors has no significant effect on the classification performance. Therefore, it can be assumed that the higher-order features extracted by the hypergraph method can reflect the structure of the data very well.

5.3. Comparison with existing methods

In Table 5, we compare the classification results of the proposed HpMTFS method with those obtained from (1) studies combining single modal feature selection and traditional machine learning for AD classification [6–10] and (2) studies combining multimodal feature selection and traditional machine learning for AD classification [12,15,16,19]. It is worth noting that for comparison purposes, all studies used for comparison used data from the ADNI dataset, with the difference being that the subjects used in the studies were different. Although the subject image data used for the studies varied, the ADNI study team performed quality control and pre-processing on all image data. Therefore, we can make a crude comparison of our method with these existing methods to validate the effectiveness of our proposed method. As can be seen from Table 5, our proposed method outperforms most existing methods in terms of AD and NC classification, AD and MCI classification, and pMCI and sMCI classification. Specifically, our method obtained a classification accuracy of 98.78% in AD vs. NC, a 3.75% improvement over the second best performance (95.03%); Our method obtained a classification accuracy of 86.47% in AD vs. MCI, a 24.4% improvement over the existing research (62.07%); Our method achieved a classification accuracy of 92.62% in pMCI vs. sMCI, an improvement of 13.37% over the existing method (79.25%). In addition, our method obtained an accuracy of 78.15% in MCI and NC classification, which is only 1.12% away from the highest accuracy (79.27%). In general, our method has good performance in the classification and diagnosis of AD.

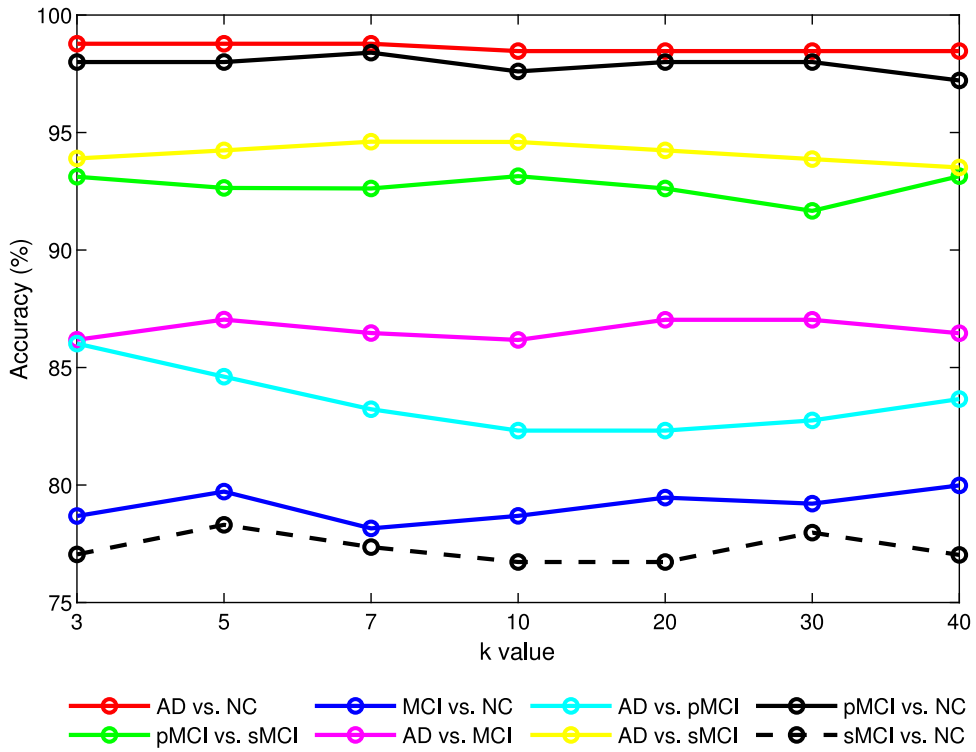


Fig. 8. The classification results on different neighbor size of hyperedges. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Comparison of the classification accuracy of our method with the existing methods.

Type	Method	Data(Number)	Classification results (%)			
			AD vs. NC	pMCI vs. sMCI	MCI vs. NC	AD vs. MCI
Single-modal	Chu et al. [6]	sMRI (ADNI:580)	84.30	–	67.30	–
	Ahmed et al. [7]	sMRI (ADNI:509)	83.77	–	69.45	62.07
	Liu et al. [8]	sMRI (ADNI:459)	93.06	79.25	–	–
	Illán et al. [9]	FDG PET (ADNI:401)	88.24	–	68.09	–
	Padilla et al. [10]	FDG PET (ADNI:105)	86.59	–	–	–
Multi-modal	Liu et al. [12]	sMRI+ FDG PET (ADNI:202)	94.37	–	78.80	–
	Gray et al. [15]	sMRI+ FDG PET +CSF +Genetic(ADNI:147)	89.00	–	74.6	–
	Zhang et al. [16]	sMRI+ FDG PET +CSF (ADNI:202)	93.20	–	76.4	–
	Jie et al. [19]	sMRI+ FDG PET (ADNI:202)	95.03	–	**79.27**	–
	Proposed method	sMRI+FDG PET+ AV-45 PET(528)	**98.78**	**92.62**	78.15	**86.47**

6. Limitations

In this study, we propose a hypergraph p-Laplacian regularized multi-task feature learning method to jointly extract common features from multimodal image data to improve the accuracy of AD classification. The method consists of four steps, namely data pre-processing, hypergraph construction, multi-task feature selection with hypergraph p-Laplacian regularization, and multi-kernel classification. To validate our method, eight sets of classification experiments were performed using baseline sMRI images, FDG PET images, and AV-45 PET images of 528 subjects in the ADNI-2 dataset. The results show that the proposed method not only allows for the joint extraction of common features of multimodal data but also helps to identify biomarkers associated with AD. However, there are limitations to our study. First, the ADNI dataset collects demographic, neuropsychological, imaging, genetic, cerebrospinal fluid, and blood data from individual subjects according to a uniform standard, whereas our method uses only imaging data for classification. Secondly, the ADNI dataset was scanned longitudinally for all recruited subjects, i.e. MRI scans and PET scans were performed on subjects at different time points (including baseline, 6 months, 12 months, 18 months, 24 months, 36 months, and 48 months). However, in this study, we only used sMRI images, FDG PET images, and AV-45 PET images acquired at a single time point (baseline time) to classify AD. Finally, we have only examined the problem of binary classification and not the performance of multivariate classification, which is of greater clinical relevance. In the future, we will combine these three aspects to further refine our study and extending our method to other applications such as diagnosis of cancer, parkinson's, coronary heart disease and other diseases.

7. Conclusion

In this study, we propose an AD classification method called the hypergraph p-Laplacian regularized multi-task feature learning method. The method captures the intrinsic correlation between different tasks

through multi-task learning to jointly extract common features of multimodal data. In particular, we introduced two regularization terms, namely (1) a hypergraph p-Laplacian regularization term preserves higher-order structural information of similar data and thus obtain more discriminative brain region features; and (2) a Frobenius norm regularization term improves the noise immunity of the model, which further improves the classification accuracy of AD. Finally, we validated the performance of the proposed method on the ADNI-2 dataset, showing that the proposed method not only captures the correlation information between multimodal data well, but also helps to identify biomarkers associated with AD.

CRediT authorship contribution statement

Yanjiao Ban: Conceptualization, Methodology, Software, Writing – original Draft. **Huan Lao:** Data curation, Writing – original draft, Writing – review & editing. **Bin Li:** Conceptualization, Investigation. **Wenjun Su:** Conceptualization, Data curation. **Xuejun Zhang:** Supervision, Writing – review & editing.

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

The datasets generated during and/or analyzed during the current study are available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) repository (www.adni.loni.usc.edu).

Acknowledgments

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), United States. As such, the investigators at the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in the analysis or writing of this report.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104326.

References

[1] N.C. Berchtold, C.W Cotman, Evolution in the conceptualization of dementia and [alzheimer]'s disease: Greco-Roman period to the 1960s, Neurobiol. Aging 19 (3) (1998) 173–189, http://dx.doi.org/10.1016/S0197-4580(98)00052-9.

[2] L Zhao, 2020 Alzheimer's disease facts and figures, Alzheimer's and Dementia 16 (3) (2020) 391–460, http://dx.doi.org/10.1002/alz.12068.

[3] Alzheimer's Association, 2019 Alzheimer's disease facts and figures, Alzheimer's & Dementia 15 (3) (2019) 321–387, http://dx.doi.org/10.1016/j.jalz.2019.01.010.

[4] Alzheimer's Association, 2021 Alzheimer's disease facts and figures, Alzheimer's & Dementia 17 (3) (2021) 327–406, http://dx.doi.org/10.1002/alz.12328.

[5] S. Klöppel, C.M. Stonnington, C. Chu, B. Draganski, R.I. Scahill, J.D. Rohrer, N.C. Fox, C.R. Jack, J. Ashburner, R.S. Frackowiak, Automatic classification of MR scans in Alzheimer's disease, Brain 131 (3) (2008) 681–689, http://dx.doi.org/10.1093/brain/awm319.

[6] C. Chu, A.L. Hsu, K.H. Chou, P. Bandettini, C. Lin, Alzheimer's Disease Neuroimaging Initiative, Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images, Neuroimage 60 (1) (2012) 59–70, http://dx.doi.org/10.1016/j.neuroimage.2011.11.066.

[7] O.B. Ahmed, M. Mizotin, J. Benois-Pineau, M. Allard, G. Catheline, C.B. Amar, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex, Comput. Med. Imaging Graph. 44 (2015) 13–25, http://dx.doi.org/10.1016/j.compmedimag.2015.04.007.

[8] M.X. Liu, D.Q. Zhang, D.G. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474, http://dx.doi.org/10.1109/TMI.2016.2515021.

[9] A.I. Illán, M.J. Górriz, J. Ramírez, D. Salas-Gonzalez, M.M. López, 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis, Inform. Sci. 181 (4) (2011) 903–916, http://dx.doi.org/10.1016/j.ins.2010.10.027.

[10] P. Padilla, M. Lopez, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, I. Alvarez, NMF-svm based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease, IEEE Trans. Med. Imaging 31 (2) (2012) 207, http://dx.doi.org/10.1109/TMI.2011.2167628.

[11] J.H. Jiang, X.H. Shu, X. Liu, Z.M. Huang, Alzheimer's Disease Neuroimaging Initiative, A computed aided diagnosis tool for Alzheimer's disease based on 11C-PiB PET imaging technique, in: 2015 IEEE International Conference on Information and Automation, 2015, http://dx.doi.org/10.1109/ICInfA.2015.7279610.

[12] S. Liu, S. Liu, W. Cai, H. Che, M.J. Fulham, Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1132–1140, http://dx.doi.org/10.1109/TBME.2014.2372011.

[13] H.I. Suk, S.W. Lee, D. Shen, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, NeuroImage 101 (2014) 569–582, http://dx.doi.org/10.1016/j.neuroimage.2014.06.077.

[14] X. Zhu, H.I. Suk, D. Shen, A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis, NeuroImage 100 (2014) 91–105, http://dx.doi.org/10.1016/j.neuroimage.2014.05.078.

[15] K.R. Gray, P. Aljabar, R.A. Heckemann, A. Hammers, D. Rueckert, Random forest-based similarity measures for multi-modal classification of Alzheimer's disease, Neuroimage 65 (2013) 167–175, http://dx.doi.org/10.1016/j.neuroimage.2012.09.065.

[16] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, Multimodal classification of Alzheimer's disease and mild cognitive impairment, Neuroimage 55 (3) (2011) 856–867, http://dx.doi.org/10.1016/j.neuroimage.2011.01.008.

[17] D. Zhang, D. Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, Neuroimage 59 (2) (2012) 895–907, http://dx.doi.org/10.1016/j.neuroimage.2011.09.069.

[18] F. Liu, C.Y. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification, Neuroimage 84 (2014) 466–475, http://dx.doi.org/10.1016/j.neuroimage.2013.09.015.

[19] B. Jie, D. Zhang, B. Cheng, D. Shen, Manifold regularized multitask feature learning for multimodality disease classification, Hum. Brain Mapp. 36 (2) (2014) 489–507, http://dx.doi.org/10.1002/hbm.22642.

[20] Y. Gao, Z. Zhang, H. Lin, X. Zhao, C. Zou, Hypergraph learning: Methods and practices, IEEE Trans. Pattern Anal. Mach. Intell. 44 (5) (2022) 2548–2566, http://dx.doi.org/10.1109/TPAMI.2020.3039374.

[21] Y. Gao, M. Wang, D. Tao, Q. Ji, 3-D object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303, http://dx.doi.org/10.1109/TIP.2012.2199502.

[22] Y. Huang, Q. Liu, Video object segmentation by hypergraph cut, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, http://dx.doi.org/10.1109/CVPR.2009.5206795.

[23] X. Zheng, W. Zhu, C. Tang, M. Wang, Gene selection for microarray data classification via adaptive hypergraph embedded dictionary learning, Gene 706 (2019) 188–200, http://dx.doi.org/10.1016/j.gene.2019.04.060.

[24] Z. Tian, T. Hwang, R. Kuang, A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge, Bioinformatics 25 (21) (2009) 2831–2838, http://dx.doi.org/10.1093/bioinformatics/btp467.

[25] Y. Gao, C.Y. Wee, M. Kim, P. Giannakopoulos, M.L. Montandon, S. Haller, D. Shen, MCI identification by joint learning on multiple mri data, vol. 9350, Springer International Publishing, 2015, pp. 78–85, http://dx.doi.org/10.1007/978-3-319-24571-3_10.

[26] W. Shao, Y. Peng, C. Zu, M. Wang, TADN Initiative, Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease, Comput. Med. Imaging Graph. 80 (2) (2019) 101663, http://dx.doi.org/10.1016/j.compmedimag.2019.101663.

[27] C. Gaser, Dahnke R., CAT-a computational anatomy toolbox for the analysis of structural MRI data, 2022, http://dx.doi.org/10.1101/2022.06.11.495736, bioRxiv.

[28] H. Lao, X.J. Zhang, Regression and classification of Alzheimer's disease diagnosis using NMF-TDNet features from 3D brain MR image, IEEE J. Biomed. Health Inf. 26 (3) (2022) 1103–1115, http://dx.doi.org/10.1109/JBHI.2021.3113668.

[29] M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, S.M. Smith, Bayesian analysis of neuroimaging data in FSL, Neuroimage 45 (1) (2009) S173–S186, http://dx.doi.org/10.1016/j.neuroimage.2008.10.055.

[30] Y. Shi, C. Zu, M. Hong, L. Zhou, L. Wang, X. Wu, J. Zhou, D. Zhang, Y. Wang, ASMFS: Adaptive-similarity-based multi-modality feature selection for classification of Alzheimer's disease, Pattern Recognit. 126 (2020) http://dx.doi.org/10.1016/j.patcog.2022.108566.

[31] C Berge, Graphes Et Hypergraphes, Dunod, Paris, 1970.

[32] X. Ma, W. Liu, S. Li, Y. Zhou, Hypergraph p-Laplacian regularization for remote sensing image recognition, IEEE Trans. Geosci. Remote Sens. 57 (3) (2018) 1585–1595, http://dx.doi.org/10.1109/TGRS.2018.2867570.

[33] L.H. Tran, L.H. Tran, Un-normalized hypergraph p-Laplacian based semi-supervised learning methods, 2018, http://dx.doi.org/10.48550/arXiv.1811.02986, arXiv e-prints..

[34] S. Saito, D. Mandic, Suzuki H., Hypergraph p-Laplacian: A differential geometry view, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018, http://dx.doi.org/10.1609/aaai.v32i1.11823.

[35] A. Argyriou, T. Evgeniou, M. Pontil, Convex multi-task feature learning, Mach. Learn. 73 (3) (2008) 243–272, http://dx.doi.org/10.1007/s10994-007-5040-8.

[36] Y. Nesterov, Convex optimization, in: S.I. Gass, M.C. Fu (Eds.), Encyclopedia of Operations Research and Management Science, Springer, Boston, MA, 2013, http://dx.doi.org/10.1007/978-1-4419-1153-7_1171.

[37] C. Xi, W. Pan, J.T. Kwok, J.G. Carbonell, Accelerated gradient method for multi-task sparse learning problem, in: 2009 Ninth IEEE International Conference on Data Mining, 2019, http://dx.doi.org/10.1109/ICDM.2009.128.

[38] J. Liu, J. Ye, Efficient L1/Lq norm regularization, 2010, http://dx.doi.org/10.48550/arXiv.1009.4766, arXiv preprint.

[39] H. Lao, X.J. Zhang, Alzheimer's disease by combining 3D discrete wavelet transform and 3D moment invariants, IET Image Process. (2022) http://dx.doi.org/10.1049/ipr2.12605.

[40] H. Lao, X.J. Zhang, Y.Y. Tang, C. Liang, Alzheimer's disease diagnosis based on the visual attention model and equal-distance ring shape context features, IET Image Process. 15 (10) (2021) 2351–2362, http://dx.doi.org/10.1049/ipr2.12218.

[41] Y.J. Ban, X.J. Zhang, H. Lao, Diagnosis of Alzheimer's disease using structure highlighting key slice stacking and transfer learning, Med. Phys. 49 (9) (2022) 5855–5869, http://dx.doi.org/10.1002/mp.158883D.

[42] K. Persson, V.D. Bohbot, N. Bogdanovic, G. Selbaek, A. Braekhus, K. Engedal, Finding of increased caudate nucleus in patients with alzheimer's disease, Acta Neurol. Scand. 137 (2018) 224–232, http://dx.doi.org/10.1111/ane.12800.

[43] S.E. Arnold, B.T. Hyman, G.V. Hoesen, Neuropathologic changes of the temporal pole in Alzheimer's disease and Pick's disease, Arch. Neurol. 51 (2) (1994) 145–150, http://dx.doi.org/10.1001/archneur.1994.00540140051014.