Medical Image Analysis 84 (2023) 102698

Contents lists available at ScienceDirect

Medical Image Analysis

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

Orthogonal latent space learning with feature weighting and graph learning for multimodal Alzheimer's disease diagnosis

Zhi Chen, Yongguo Liu *, Yun Zhang, Qiaoqin Li, Alzheimer's Disease Neuroimaging Initiative [1]

Knowledge and Data Engineering Laboratory of Chinese Medicine, School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Latent representation learning
Graph learning
Multimodal fusion

ABSTRACT

Recent studies have shown that multimodal neuroimaging data provide complementary information of the brain and latent space-based methods have achieved promising results in fusing multimodal data for Alzheimer's disease (AD) diagnosis. However, most existing methods treat all features equally and adopt nonorthogonal projections to learn the latent space, which cannot retain enough discriminative information in the latent space. Besides, they usually preserve the relationships among subjects in the latent space based on the similarity graph constructed on original features for performance boosting. However, the noises and redundant features significantly corrupt the graph. To address these limitations, we propose an Orthogonal Latent space learning with Feature weighting and Graph learning (OLFG) model for multimodal AD diagnosis. Specifically, we map multiple modalities into a common latent space by orthogonal constrained projection to capture the discriminative information for AD diagnosis. Then, a feature weighting matrix is utilized to sort the importance of features in AD diagnosis adaptively. Besides, we devise a regularization term with learned graph to preserve the local structure of the data in the latent space and integrate the graph construction into the learning processing for accurately encoding the relationships among samples. Instead of constructing a similarity graph for each modality, we learn a joint graph for multiple modalities to capture the correlations among modalities. Finally, the representations in the latent space are projected into the target space to perform AD diagnosis. An alternating optimization algorithm with proved convergence is developed to solve the optimization objective. Extensive experimental results show the effectiveness of the proposed method.

1. Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disease, is characterized by the accumulation of beta-amyloid plaques and neurofibrillary tangles composed of tau protein leading to brain atrophy (Theofilas et al., 2018; Wang et al., 2018). It is reported that AD has become the fifth leading cause of death in the elderly (Gauthier, 2005) and 115.4 million people are expected to have AD in 2050 (Alberdi et al., 2016). However, no known therapy is able to reverse or prevent the progression of AD. Fortunately, studies have demonstrated that the treatments in its early stage (i.e., mild cognitive impairment (MCI)) are highly desirable to slow down the disease progression (Gauthier, 2005). MCI can be divided into progressive MCI (pMCI) that will progress to AD and stable MCI (sMCI) that remains stable or regains normal cognitive function in a certain period of time. It is reported that MCI subjects progress to AD at a rate of 10%–15% per year (Hänninen et al., 2002). Therefore, early diagnosis of AD and identification of the subtypes of MCI are of great significance for personalized cure and progression delay (Shen et al., 2021; Martinez-Murcia et al., 2016).

As the brain changes related to AD occur prior to any clinical symptom, neuroimaging techniques are more reliable and sensitive than traditional cognitive assessments in AD diagnosis (Qian et al., 2019; Górriz et al., 2019). Therefore, more and more attention has been paid to find effective neuroimaging biomarkers associated with the progression of AD for accurate diagnosis (Jack Jr. et al., 2010; Zhang and Davatzikos, 2013). Commonly used neuroimaging techniques include magnetic resonance imaging (MRI) (Zhang and Davatzikos, 2013), positron emission tomography (PET) (Langbaum et al., 2009), diffusion tensor imaging (DTI) (Kantarci et al., 2017), etc. These neuroimaging

* Corresponding author.
E-mail address: liuyg@uestc.edu.cn (Y. Liu).

[1] Data used in preparation of this article were obtained from the Alzheimer' s Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.media.2022.102698
Received 10 June 2022; Received in revised form 18 October 2022; Accepted 17 November 2022
Available online 21 November 2022

techniques provide quantitative measurements to capture structural and functional brain changes and have been extensively utilized for the diagnosis of AD (Wang et al., 2020). Different modalities can provide different neurological or biological information. For example, structural MRI provides the information related to gray and white matter structures, while PET measures the metabolic rate of glucose. Researches have shown that if the information contained in multimodal neuroimaging data is properly fused, then the final performance of AD diagnosis can be significantly improved (Liu et al., 2020b). Consequently, numerous efforts have been made to mine the intrinsic structures across multiple modalities for automatic AD diagnosis (Shen et al., 2021; Zhang et al., 2021; Shi et al., 2020, 2018; El-Sappagh et al., 2020). For example, Shen et al. (2021) exploited subject's MRI features, PET features and age by a sparse regression model to predict whether MCI will progress to AD. Shi et al. (2020) presented coupled boosting and coupled metric ensemble to capture the complementary information among different neuroimaging modalities for AD diagnosis.

However, the major challenge of multimodal AD diagnosis arises from high dimension, low sample size data. The size of features extracted from multimodal data ranges from hundreds to thousands while the number of available training samples is often only a few hundred or even dozens, which may easily result in overfitting (Ning et al., 2021). To address this issue, previous studies usually adopt feature selection or dimension reduction to reduce data dimensionality and find effective features for AD diagnosis (Zhang et al., 2021; Ning et al., 2021; Zhu et al., 2015; Zhou et al., 2020, 2019b; Hao et al., 2020). For instance, Zhu et al. (2015) combined linear discriminant analysis and locality preserving projection for selecting the important features from multimodal neuroimaging data. To exploit the correlations among different modalities and reduce data dimensionality simultaneously, many latent space-based methods have been explored (Ning et al., 2021; Zhu et al., 2015; Zhou et al., 2020, 2019b), which project multiple modalities into a low dimensional common latent space and then perform the classification in the latent space. For example, Ning et al. (2021) learned low dimensional shared representations for multimodal neuroimaging data and built a classifier on the shared representations for AD diagnosis. Zhou et al. (2020) projected multimodal neuroimaging data into a latent space and then established multiple diversified classifiers in the latent space to improve the diagnosis performance. Although the aforementioned methods have achieved good performance for multimodal AD diagnosis, they still have several limitations:

1. For latent space learning, most existing methods treat different features equally. However, the number of features extracted from multimodal neuroimaging data ranges from hundreds to thousands and different features contribute differently to AD diagnosis (Lei et al., 2017). Some brain regions are known to be preferentially affected by AD. For example, it is reported that the hippocampus is one of the first areas in the brain affected by AD and the changes in the hippocampus represent an early critical event in the progression of AD (Rao et al., 2022). Besides, in the earliest stages of AD, the pathology and volume loss have been widely observed in the entorhinal cortex (Bobinski et al., 1999). Hence, the features related to the important regions should play a major role in AD diagnosis and different features in the multimodal neuroimaging feature set should have different contributions to the latent space learning. Treating different features equally may limit the discriminative power of the learned space. Recent works introduce $\ell_{2,1}$-norm sparsity regularization on the projection matrices to select useful features (Zhu et al., 2015; Zhou et al., 2020, 2019b; Hao et al., 2020). For example, Zhou et al. (2019b) developed a multimodal AD diagnosis effort joints latent representation space and $\ell_{2,1}$-norm sparsity regularization into a unified framework. Hao et al. (2020) proposed a multimodal neuroimaging feature selection method based on a consistent metric constraint, in which $\ell_{2,1}$-norm sparsity regularization is used for feature selection. However, in all aforementioned approaches, a hyperparameter should be set to tune the tradeoff between the loss term and $\ell_{2,1}$-norm sparsity regularization term. How to tactfully determine the hyperparameter of the $\ell_{2,1}$-norm term is a severe challenge. Besides, existing methods usually adopt nonorthogonal projection for learning the latent space, which may lead to imbalanced weights on different projection directions and disturb the Euclidean distance-based similarity of original data (Liu et al., 2017a). As a result, nonorthogonal projection cannot preserve the data structure well and is unsuitable for latent space learning.
2. Previous methods usually exploit the similarity relationships between samples by Laplacian regularization to enhance the diagnosis performance (Ning et al., 2021; Zhou et al., 2019b; Hao et al., 2020; Lei et al., 2017). They assume that if samples are similar to each other according to a predefined similarity graph, then the same or similar relationships are expected to be preserved in the latent space. For example, Zhou et al. (2019b) proposed to learn a latent representation space using samples with missing modalities and design a Laplacian regularization term to ensure that similar inputs have similar latent representations. It is found that the results of Laplacian regularization-based methods are heavily dependent on the quality of the predefined similarity graph (Hao et al., 2020). However, these methods calculate the similarity graph in the original input space. The original neuroimaging feature set inevitably contains unnecessary noises and redundant features, which make the similarity graph unreliable (Wang et al., 2017; Adeli et al., 2018). Therefore, it is unreasonable to directly construct the similarity graph using original features. To alleviate this limitation, it is desirable to learn a good similarity graph from data to deal with the potential noises or corruptions (Wang et al., 2017). For example, Hao et al. (2020) calculated the similarity graph for each modality individually by random forest strategy. Regrettably, the construction of the similarity graph is independent of the classification process, which may degrade classification performance. Besides, existing methods typically construct a graph for each modality individually. However, considering that diverse modalities should share the same underlying structure related to the diagnosis, learning a similarity graph for each modality ignores the correlations of graph structures between multiple modalities and makes the feature fusion unreasonable (Zhou et al., 2019a; Zhan et al., 2018).

In this study, we propose an Orthogonal Latent space learning with Feature weighting and Graph learning (OLFG) approach for multimodal AD diagnosis. OLFG integrates latent representation learning, similarity graph construction, and classifier training, into a unified framework. Specifically, we assume there exists a common latent space for multimodal neuroimaging data and project each modality into the common latent space by a modality-specific projection matrix to capture diagnosis-related information. Meanwhile, a feature weighting matrix is adaptively learned for each modality to rank the feature importance in the latent space learning. Then, to avoid trivial solutions and retain discriminative information in the latent space, orthogonal constraints are imposed on the projection matrices. Moreover, we add a Laplacian regularization term with learned graph to preserve the local structure of the data after projection. Considering that all modalities should admit the same underlying structure related to AD, we integrate different graphs constructed from multiple modalities into an intrinsic joint graph to capture the associations among modalities. Instead of constructing the similarity graph on the original features, our formulation learns the joint graph in a task-oriented manner for accurately encoding the relationships between samples. The joint graph learning and latent space learning are integrated into the unified framework and can be optimized simultaneously. Finally, we project the shared representations into the target space (i.e., label space) to perform AD diagnosis. The experimental results show the effectiveness of OLFG. The main contributions of our work include the following:

1. We integrate latent representation learning, similarity graph construction, and classifier training, into a unified framework and propose a latent space-based approach to address the multimodal AD diagnosis problem.
2. We introduce feature weighting and orthogonal projection to learn the latent space. The weighting matrix is utilized to sort the importance of features in AD diagnosis. Benefiting from the orthogonal constraints on the projection matrices, the proposed method can avoid trivial solutions and preserve enough discriminative information in the latent space.
3. We explore the relationships between training subjects and mine the underlying associations among modalities by a Laplacian regularization with adaptive joint similarity graph. The graph construction is integrated into the learning process to adaptively learn a discriminative and accurate graph structure.
4. An effective iterative algorithm with proved convergence is proposed to solve the objective function. Extensive experimental results on the ADNI-2 and OASIS-3 datasets show that the proposed method performs better than the previous methods in most cases.

The rest of the paper is organized as follows. In Section 2, the formalization of the OLFG model is introduced. In Section 3, an effective algorithm is presented to optimize this approach. In Section 4, experimental results are presented. Section 5 presents the discussion. Finally, Section 6 offers some conclusions.

2. Method

2.1. Notations and definitions

Scalar, vector, and matrix are denoted by normal lowercase letter, bold lowercase letter, and bold uppercase letter, respectively. For matrix $\mathbf{M} \in \mathbb{R}^{d\times n}$, let $\boldsymbol{m}_j$ and m_{ij} represent the jth column and the (i,j)th element of $\mathbf{M}$, respectively. $\|\mathbf{M}\|_F$ denotes the Frobenius norm of $\mathbf{M}$. The transpose and trace of $\mathbf{M}$ are denoted by $\mathbf{M}^\top$ and $\mathrm{tr}(\mathbf{M})$, respectively.

A multimodal neuroimaging dataset is denoted by $\mathcal{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\}$, where V is the number of modalities. $\mathbf{X}^v = [\boldsymbol{x}_1^v, \boldsymbol{x}_2^v, \ldots, \boldsymbol{x}_n^v] \in \mathbb{R}^{d_v\times n}$ denotes the original feature matrix in the vth modality, where $\boldsymbol{x}_i^v$ is the ith sample, d_v is the feature dimensionality of the vth modality, and n is the number of samples. $\mathbf{Y} \in \{0,1\}^{c\times n}$ is the corresponding label matrix, where c is the number of classes.

2.2. Orthogonal latent space learning

For multimodal neuroimaging data, we can project the original features from different modalities into the label space by the least square regression model

$$\min_{\mathbf{w}^v,\boldsymbol{b}^v} \sum_{v=1}^{V} \|\mathbf{W}^{v\top}\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{Y}\|_F^2, \tag{1}$$

where $\mathbf{W}^v \in \mathbb{R}^{d_v\times c}$ and $\boldsymbol{b}^v \in \mathbb{R}^c$ are the projection matrix and bias vector for the vth modality, respectively, $\mathbf{1}_n \in \mathbb{R}^n$ is a column vector with all the elements as 1. Eq. (1) treats each modality individually without considering the correlations among different modalities. To exploit the correlations among modalities, we assume that there exists a common latent space which each modality can be projected into. The latent space contains more information extracted from all available modalities for AD diagnosis and fewer noises than the original feature space. Accordingly, we project the multimodal neuroimaging data into the latent space by several modality-specific projections and perform classification in the latent space:

$$\begin{aligned}&\min_{\mathbf{W}^v,\mathbf{H},\mathbf{P},\boldsymbol{b},\boldsymbol{b}^v} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2\\ &+\frac{\alpha}{2}\sum_{v=1}^{V}\|\mathbf{W}^{v\top}\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{H}\|_F^2 + \frac{\beta}{2}\|\mathbf{P}\|_F^2,\end{aligned} \tag{2}$$

where $\mathbf{H} \in \mathbb{R}^{h\times n}$ is the latent feature representation matrix, $\mathbf{W}^v \in \mathbb{R}^{d_v\times h}$ is the projection matrix that projects the vth modality into the latent space, $\mathbf{P} \in \mathbb{R}^{c\times h}$ is the regression matrix that projects the latent representation matrix into the label space, $\boldsymbol{b} \in \mathbb{R}^c$ and $\boldsymbol{b}^v \in \mathbb{R}^h$ are the bias terms, $\|\mathbf{P}\|_F^2$ is the ℓ_2 regularization term, α and β are the regularization parameters, and h is the dimension of the latent space. However, $\mathbf{W}^v$ is a nonorthogonal projection, which may lead to trivial solutions and disturb the Euclidean distance-based similarity of original data (Liu et al., 2017a). To overcome the limitation, we impose orthogonal constrain on $\mathbf{W}^v$:

$$\begin{aligned}&\min_{\mathbf{W}^v,\mathbf{H},\mathbf{P},\boldsymbol{b},\boldsymbol{b}^v} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2 +\\ &\frac{\alpha}{2}\sum_{v=1}^{V}\|\mathbf{W}^{v\top}\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{H}\|_F^2 + \frac{\beta}{2}\|\mathbf{P}\|_F^2,\\ &\text{s.t. } \mathbf{W}^{v\top}\mathbf{W}^v = \mathbf{I}_h\end{aligned} \tag{3}$$

where $\mathbf{I}_h \in \mathbb{R}^{h\times h}$ is an identity matrix.

2.3. Adaptive feature weighting

As can be seen, the model in Eq. (3) equally treats different features when learning the latent space. However, different features typically have different contributions to AD diagnosis. For example, the hippocampus is one of the first regions in the brain affected by AD and low hippocampal volume has been qualified in AD clinical trials (Rao et al., 2022). The features related to the important regions should play a major role when learning the latent space. Moreover, there exist redundant or unrelated features that provide little information for AD diagnosis in the feature set (Lei et al., 2017). These features are not helpful for the latent space learning and may limit the discriminative power of the learned latent space. Therefore, it is unsuitable to treat all features equally. In this paper, we incorporate the feature weighting information into the framework by adaptive feature weighting matrices for showing the rankings of all features and selecting important features:

$$\begin{aligned}&\min_{\mathbf{W}^v,\boldsymbol{\Theta}^v,\mathbf{H},\mathbf{P},\boldsymbol{b},\boldsymbol{b}^v} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2\\ &+\frac{\alpha}{2}\sum_{v=1}^{V}\|\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{H}\|_F^2 + \frac{\beta}{2}\|\mathbf{P}\|_F^2,\\ &\text{s.t. } \mathbf{W}^{v\top}\mathbf{W}^v = \mathbf{I}_h, \boldsymbol{\theta}^{v\top}\mathbf{1}_{d_v} = 1, \boldsymbol{\theta}^v \geq \mathbf{0}\end{aligned} \tag{4}$$

where $\boldsymbol{\Theta}^v \in \mathbb{R}^{d_v\times d_v}$ is a diagonal matrix and $\boldsymbol{\theta}^v \in \mathbb{R}^{d_v}$ is the diagonal of $\boldsymbol{\Theta}^v$, i.e., $\boldsymbol{\theta}^v = \mathrm{diag}(\boldsymbol{\Theta}^v)$. The elements in $\boldsymbol{\theta}^v$ measure the importance of features. A large value in $\boldsymbol{\theta}^v$ means that the corresponding feature plays an important role in the latent space learning while a small one means that the corresponding feature provides little information for learning the latent space.

2.4. Adaptive joint graph regularization

Inspired by the graph Laplacian-regularized approaches (Ning et al., 2021; Zhou et al., 2019b; Hao et al., 2020; Lei et al., 2017), we would also like to preserve the similarity relationships between samples in the latent space for benefiting the diagnosis performance. Previous methods consider that if samples are closed in the original feature space, their corresponding latent representations should be also similar to each other. This can be achieved by the regularization term designed as

$$\sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\boldsymbol{t}_i^v - \boldsymbol{t}_j^v\|_2^2 a_{ij}^v, \tag{5}$$

where $a_{ij}^v = \exp(-\|\boldsymbol{x}_i^v - \boldsymbol{x}_j^v\|_2^2)$ is the (i,j)th element in $\mathbf{A}^v = [a_{ij}^v] \in \mathbb{R}^{n\times n}$ which measures the similarity between two samples on the vth modality, and $\boldsymbol{t}_j^v$ is the jth column of matrix $\mathbf{T}^v = \mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v$. Although Eq. (5) preserves the similarity relationships for each modality in the

latent space, it has the following limitations: (1) It constructs the similarity graph on the original feature space. However, the neuroimaging data usually contain unnecessary noises, which significantly corrupt the similarity graph (Adeli et al., 2018). Therefore, the similarity graph constructed on the original feature space is usually inaccurate. Besides, it has been shown that only a small number of features extracted from neuroimaging data are related to AD (Lei et al., 2017). However, Eq. (5) uses all features to calculate the similarity graph. As a result, the similarity graph constructed on the original feature space only provides little information for AD diagnosis. (2) The construction of the similarity graph is independent of the latent space learning and classifier building. That is, the similarity graph is fixed when learning the latent space and building the classifier. As a result, the predefined graph may not be the most informative for AD diagnosis. (3) It constructs a graph for each modality individually. We project all modalities into a shared latent space and expect that the latent space contains discriminative information for AD diagnosis. Consequently, different modalities should share the same graph structure after projection even though they represent different information for the diagnosis task. Constructing a graph for each modality individually ignores the consistency among the graphs of different modalities.

In this paper, we simultaneously deal with the three issues via an adaptive joint graph learning method. Specifically, we use the same graph for all modalities to capture the underlying associations among multiple modalities. Moreover, to deal with the noises and redundant features in neuroimaging data, we propose to integrate the graph construction into the learning process and adaptively learn the joint graph for graph regularization. Consequently, the graph regularization term for our approach is formulated as

$$\sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\boldsymbol{t}_i^v-\boldsymbol{t}_j^v\|_2^2 s_{ij}+\lambda\|\mathbf{S}-\frac{1}{V}\sum_{v=1}^{V}\mathbf{A}^v\|_F^2, \quad \text{s.t. } \mathbf{S1}=\mathbf{1},\mathbf{S}\geq\mathbf{0} \tag{6}$$

where $\mathbf{S}\in\mathbb{R}^{n\times n}$ is the joint similarity graph. Eq. (6) can be rewritten as

$$\sum_{v=1}^{V}\mathrm{tr}(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{L}_s(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v)^\top)+\lambda\|\mathbf{S}-\frac{1}{V}\sum_{v=1}^{V}\mathbf{A}^v\|_F^2, \quad \text{s.t. } \mathbf{S1}=\mathbf{1},\mathbf{S}\geq\mathbf{0} \tag{7}$$

where $\mathbf{L}_s$ is a Laplacian matrix defined as $\mathbf{L}_s=\mathbf{D}-(\mathbf{S}^\top+\mathbf{S})/2$, $\mathbf{D}$ is a diagonal matrix with the ith diagonal element $d_{ii}=\sum_j(s_{ij}+s_{ji})/2$. In Eq. (7), the first term preserves the sample relationship in the latent space based on the joint similarity graph and the second one regularizes the joint similarity graph toward the sum of graphs of different modalities and makes different modalities agree with each other in the latent space.

2.5. Final objective function

By unifying the objective function in Eqs. (4) and (7), our final objective function is formulated as

$$\begin{aligned}\min_{\mathbf{W}^v,\boldsymbol{\Theta}^v,\mathbf{H},\mathbf{P},\mathbf{S},\boldsymbol{b},\boldsymbol{b}^v}&\frac{1}{2}\|\mathbf{PH}+\boldsymbol{b}\mathbf{1}_n^\top-\mathbf{Y}\|_F^2\\&+\frac{\alpha}{2}\sum_{v=1}^{V}\|\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v+\boldsymbol{b}^v\mathbf{1}_n^\top-\mathbf{H}\|_F^2+\frac{\beta}{2}\|\mathbf{P}\|_F^2\\&+\gamma\sum_{v=1}^{V}\mathrm{tr}(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{L}_s(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v)^\top)+\lambda\|\mathbf{S}-\frac{1}{V}\sum_{v=1}^{V}\mathbf{A}^v\|_F^2.\\&\text{s.t. } \mathbf{W}^{v\top}\mathbf{W}^v=\mathbf{I}_h,\boldsymbol{\theta}^{v\top}\mathbf{1}_{d_v}=1,\boldsymbol{\theta}^v\geq\mathbf{0},\mathbf{S1}=\mathbf{1},\mathbf{S}\geq\mathbf{0}\end{aligned} \tag{8}$$

In Eq. (8), the first two terms show that the multimodal neuroimaging data are projected into a common latent space and the classification is performed in the latent space. Benefitting from the orthogonal constraints on the projection matrices $\{\mathbf{W}^v\}_{v=1}^V$, OLFG can preserve more discriminative information in the latent space than conventional nonorthogonal projection and avoid trivial solutions. Since OLFG assigns an adaptive weight for each feature with the feature weighting matrices $\{\boldsymbol{\Theta}^v\}_{v=1}^V$, the discriminative and nonredundant features can be selected for learning the latent space. The third term is the ℓ_2 regularization term which is adopted to prevent overfitting. The last two terms represent the adaptive joint graph regularization. Different from existing methods that construct the similarity graph in the original feature space before the learning processing, OLFG integrates the similarity graph construction into the optimization framework and learns similarity graph $\mathbf{S}$ in a task-oriented manner. Hence, compared with the graph constructed in the original feature space, the learned graph $\mathbf{S}$ is more robust to noises and more suitable for the diagnosis task.

Algorithm 1 GPI method

Input: the symmetric matrix $\mathbf{C}^v\in\mathbb{R}^{d_v\times d_v}$ and matrix $\mathbf{B}^v\in\mathbb{R}^{d_v\times h}$;

```
Initialize W^v with random values;
Initialize α such as O = αI_{d_v} − C^v is a positive definite matrix;
while not converged do
    Update M = 2OW^v + 2B^v;
    Calculate USV^⊤ = M via the compact SVD method;
    Update W^v = UV^⊤;
end while
return W^v;
```

3. Optimization

3.1. Optimization algorithm

In this section, we develop an alternating optimization algorithm to solve the proposed problem. The optimization of Eq. (8) involves $\{\mathbf{W}^v\}_{v=1}^V,\{\boldsymbol{\Theta}^v\}_{v=1}^V,\mathbf{H},\mathbf{P},\mathbf{S}$, and bias vectors (i.e., $\boldsymbol{b}$ and $\boldsymbol{b}^v$). In each iteration, we minimize the objective function w.r.t. one variable while fixing the other variables. The main optimizing steps are summarized as follows.

(1) Optimize $\mathbf{W}^v$: When $\{\boldsymbol{\Theta}^v\}_{v=1}^V,\mathbf{H},\mathbf{P},\mathbf{S}$, and bias vectors are fixed, the optimization problem in Eq. (8) can be separated into V independent subproblems

$$\begin{aligned}\min_{\mathbf{W}^v}&\frac{\alpha}{2}\|\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v+\boldsymbol{b}^v\mathbf{1}_n^\top-\mathbf{H}\|_F^2\\&+\gamma\,\mathrm{tr}(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{L}_s(\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v)^\top).\\&\text{s. t. } \mathbf{W}^{v\top}\mathbf{W}^v=\mathbf{I}_h,\end{aligned} \tag{9}$$

Eq. (9) is equivalent to the following problem

$$\begin{aligned}\min_{\mathbf{W}^v}&\mathrm{tr}(\mathbf{W}^{v\top}\mathbf{C}^v\mathbf{W}^v)-2\,\mathrm{tr}(\mathbf{W}^{v\top}\mathbf{B}^v),\\&\text{s.t. } \mathbf{W}^{v\top}\mathbf{W}^v=\mathbf{I}_h,\end{aligned} \tag{10}$$

where $\mathbf{C}^v=\boldsymbol{\Theta}^v\mathbf{X}^v(\mathbf{I}_n+\frac{2\gamma}{\alpha}\mathbf{L}_s)\mathbf{X}^{v\top}\boldsymbol{\Theta}^{v\top},\mathbf{B}^v=\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{H}^{v\top}$, and $\mathbf{H}^v=\mathbf{H}-\mathbf{b}^v\mathbf{1}_n^\top$. Eq. (10) has the same form as the quadratic problem on the Stiefel manifold (QPSM). We use the generalized power iteration (GPI) algorithm (Nie et al., 2017) to solve Eq. (10). Algorithm 1 describes the GPI method.

(2) Optimize $\boldsymbol{\Theta}^v$: By fixing other variables except for $\boldsymbol{\Theta}^v$, Eq. (8) is equivalent to the following problem

$$\begin{aligned}\min_{\boldsymbol{\Theta}^v}&\mathrm{tr}(\boldsymbol{\Theta}^v\mathbf{X}^v(\mathbf{I}+\frac{2\gamma}{\alpha}\mathbf{L}_s)\mathbf{X}^{v\top}\boldsymbol{\Theta}^{v\top}\mathbf{W}^v\mathbf{W}^{v\top})\\&-\mathrm{tr}(2\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{H}^{v\top}\mathbf{W}^{v\top}).\\&\text{s.t. } \boldsymbol{\theta}^{v\top}\mathbf{1}_{d_v}=1,\boldsymbol{\theta}^v\geq\mathbf{0}\end{aligned} \tag{11}$$

We provide the following lemma to solve Eq. (11).

Lemma 1. *If* $\mathbf{U}$ *is a diagonal matrix and* $\mathbf{K}$ *is a symmetric matrix, then* $\mathrm{tr}(\mathbf{UKUL}) = \boldsymbol{u}^\top(\mathbf{K}^\top \circ \mathbf{L})\boldsymbol{u}$*, where* $\boldsymbol{u} = \mathrm{diag}(\mathbf{U})$.

Proof.

$$\begin{aligned}\mathrm{tr}(\mathbf{UKUL}) &= \boldsymbol{u}^T \mathrm{diag}(\mathbf{KUL}) \\ &= \boldsymbol{u}^T \mathrm{vec}\{\boldsymbol{k}_i^\top \mathbf{U} \boldsymbol{l}_i\} \\ &= \boldsymbol{u}^T \mathrm{vec}\{(\boldsymbol{k}_i \circ \boldsymbol{l}_i)^\top \boldsymbol{u}\} \\ &= \boldsymbol{u}^T (\mathbf{K}^\top \circ \mathbf{L})^\top \boldsymbol{u} \\ &= \boldsymbol{u}^T (\mathbf{K}^\top \circ \mathbf{L}) \boldsymbol{u},\end{aligned} \tag{12}$$

where $\circ$ is the Hadamard product. □

By using Lemma 1, Eq. (11) becomes

$$\begin{aligned}&\min_{\boldsymbol{\theta}^v} \boldsymbol{\theta}^{v\top} \mathbf{Q} \boldsymbol{\theta}^v - \boldsymbol{\theta}^{v\top} \boldsymbol{r}, \\ &\text{s.t. } \boldsymbol{\theta}^{v\top} \mathbf{1}_{d_v} = 1, \boldsymbol{\theta}^v \geq \mathbf{0}\end{aligned} \tag{13}$$

where $\mathbf{Q} = \mathbf{X}^v(\mathbf{I} + \frac{2\gamma}{\alpha}\mathbf{L}_s)\mathbf{X}^{v\top} \circ \mathbf{W}^v\mathbf{W}^{v\top}$ and $\boldsymbol{r} = \mathrm{diag}(2\mathbf{X}^v\mathbf{H}^{v\top}\mathbf{W}^{v\top})$, respectively. We employ the augmented Lagrangian multiplier (ALM) method to solve Eq. (13). To solve Eq. (13) is equivalent to minimize the augmented Lagrange function $L_{\boldsymbol{\theta}^v}$ defined as

$$\begin{aligned}L_{\boldsymbol{\theta}^v} = \boldsymbol{\theta}^{v\top} \mathbf{Q} \boldsymbol{\theta}^v - \boldsymbol{\theta}^{v\top} \boldsymbol{r} + \frac{\mu}{2}\|\boldsymbol{\theta}^v - \boldsymbol{g} + \frac{1}{\mu}\boldsymbol{\delta}_1\|_2^2 + \frac{\mu}{2}(\boldsymbol{\theta}^T \mathbf{1}_{d_v} - 1 + \frac{1}{\mu}\delta_2)^2, \\ \text{s.t. } \boldsymbol{g} \geq \mathbf{0}\end{aligned} \tag{14}$$

where $\boldsymbol{\delta}_1 \in \mathbb{R}^{d_v}$ and $\delta_2 \in \mathbb{R}$ are Lagrange multipliers, μ is a penalty factor. Then, we solve the following problems iteratively:

step 1: By fixing other variables, $\boldsymbol{\theta}^v$ can be updated by solving the following optimization problem

$$\min_{\boldsymbol{\theta}^v} \frac{1}{2}\boldsymbol{\theta}^{v\top}\mathbf{E}\boldsymbol{\theta}^v - \boldsymbol{\theta}^{v\top}\boldsymbol{f}, \tag{15}$$

where $\mathbf{E} = 2\mathbf{Q} + \mu\mathbf{I}_{d_v} + \mu\mathbf{1}_{d_v}\mathbf{1}_{d_v}^\top$ and $\boldsymbol{f} = \mu\boldsymbol{g} + \mu\mathbf{1}_{d_v} - \delta_2\mathbf{1}_{d_v} - \boldsymbol{\delta}_1 + \boldsymbol{r}$, respectively. By setting the derivative of Eq. (15) w.r.t. $\boldsymbol{\theta}^v$ to zero, we can obtain its optimal solution

$$\boldsymbol{\theta}^v = \mathbf{E}^{-1}\boldsymbol{f}. \tag{16}$$

Step 2: By fixing other variables, we can obtain $\boldsymbol{g}$ by solving the following problem:

$$\begin{aligned}&\min_{\boldsymbol{g}} \|\boldsymbol{g} - (\boldsymbol{\theta}^v + \frac{1}{\mu}\boldsymbol{\delta}_1)\|_2^2. \\ &\quad \text{s.t. } \boldsymbol{g} \geq \mathbf{0}\end{aligned} \tag{17}$$

We can obtain its optimal solution

$$\boldsymbol{g} = (\boldsymbol{\theta}^v + \frac{1}{\mu}\boldsymbol{\delta}_1)_+. \tag{18}$$

Step 3: Update the Lagrange multipliers and the penalty factor by the following equations

$$\begin{cases}\boldsymbol{\delta}_1 = \boldsymbol{\delta}_1 + \mu(\boldsymbol{\theta}^v - \boldsymbol{g}) \\ \delta_2 = \delta_2 + \mu(\boldsymbol{\theta}^{v\top}\mathbf{1}_{d_v} - 1), \\ \mu = \epsilon\mu\end{cases} \tag{19}$$

where the parameter $\epsilon > 1$ is manually set. Algorithm 2 summarizes the entire algorithm for solving Eq. (13) in detail.

(3) Optimize $\boldsymbol{b}^v$: By fixing other variables except for $\boldsymbol{b}^v$, Eq. (8) can be rewritten as

$$\min_{\boldsymbol{b}^v} \frac{1}{2}\|\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{H}\|_F^2. \tag{20}$$

By setting the derivative of Eq. (20) w.r.t. $\boldsymbol{b}^v$ to zero, we can obtain its optimal solution

$$\boldsymbol{b}^v = \frac{1}{n}(\mathbf{H}\mathbf{1}_n - \mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v\mathbf{1}_n). \tag{21}$$

(4) Optimize $\mathbf{H}$: With fixed $\{\mathbf{W}^v\}_{v=1}^V, \{\boldsymbol{\Theta}^v\}_{v=1}^V, \mathbf{P}, \mathbf{S}$, and bias vectors, the optimization is equivalent to solving the following problem

$$\min_{\mathbf{H}} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2 + \frac{\alpha}{2}\sum_{v=1}^V \|\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top - \mathbf{H}\|_F^2. \tag{22}$$

Algorithm 2 Algorithm to solve Eq. (13)

Input: Matrix $\mathbf{Q}$, vector $\boldsymbol{r}$, and parameter μ;

1: Initialize $\boldsymbol{\theta}^v = \left(1/d_v, 1/d_v, \cdots, 1/d_v\right)^\top, \boldsymbol{\delta}_1 = \mathbf{0}, \delta_2 = 0, \boldsymbol{g} = \boldsymbol{\theta}^v$;
2: **while** not converged **do**
3: Update $\boldsymbol{\theta}^v$ by Eq. (16);
4: Update $\boldsymbol{g}$ by Eq. (18);
5: Update Lagrange multipliers $\boldsymbol{\delta}_1, \delta_2$ and penalty factor μ by Eq. (19).
6: **end while**
7: **return** $\boldsymbol{\theta}^v$;

Taking the derivative of Eq. (22) w.r.t. $\mathbf{H}$ equal zero, we obtain

$$\mathbf{H} = (\alpha V\mathbf{I} + \mathbf{P}^\top\mathbf{P})^{-1}(\alpha\sum_{v=1}^V \mathbf{Z}_v + \mathbf{P}^\top\overline{\mathbf{Y}}), \tag{23}$$

where $\mathbf{Z}_v = \mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v + \boldsymbol{b}^v\mathbf{1}_n^\top$ and $\overline{\mathbf{Y}} = \mathbf{Y} - \boldsymbol{b}\mathbf{1}_n^\top$, respectively.

(5) Optimize $\mathbf{P}$: When $\{\mathbf{W}^v\}_{v=1}^V, \{\boldsymbol{\Theta}^v\}_{v=1}^V, \mathbf{H}, \mathbf{S}$, and bias vectors are fixed in Eq. (8), the optimization of $\mathbf{P}$ is equal to solving the following problem

$$\min_{\mathbf{P}} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2 + \frac{\beta}{2}\|\mathbf{P}\|_F^2. \tag{24}$$

By setting the first-order partial derivative of the objective function in Eq. (24) w.r.t. $\mathbf{P}$ as 0, we have the following closed-form solution

$$\mathbf{P} = \overline{\mathbf{Y}}\mathbf{H}^\top(\mathbf{H}\mathbf{H}^\top + \beta\mathbf{I})^{-1}. \tag{25}$$

(6) Optimize $\boldsymbol{b}$: By fixing other variables except for $\boldsymbol{b}$, Eq. (8) can be rewritten as

$$\min_{\boldsymbol{b}} \frac{1}{2}\|\mathbf{PH} + \boldsymbol{b}\mathbf{1}_n^\top - \mathbf{Y}\|_F^2. \tag{26}$$

By setting the derivative of Eq. (8) w.r.t. $\boldsymbol{b}$ to zero, we can get the closed-form solution of Eq. (26):

$$\boldsymbol{b} = \frac{1}{n}(\mathbf{Y}\mathbf{1}_n - \mathbf{P}\mathbf{H}\mathbf{1}_n). \tag{27}$$

(7) Optimize $\mathbf{S}$: By fixing all other variables except $\mathbf{S}$ in Eq. (8) and denoting $1/V\sum_{v=1}^V \mathbf{A}^v$ by $\mathbf{A}$ and $\mathbf{W}^{v\top}\boldsymbol{\Theta}^v\mathbf{X}^v$ by $\mathbf{T}^v$, we have the following subproblem

$$\begin{aligned}&\min_{\mathbf{S}} \frac{\gamma}{\lambda}\sum_{v=1}^V \mathrm{tr}(\mathbf{T}^v\mathbf{L}_s\mathbf{T}^{v\top}) + \|\mathbf{S} - \mathbf{A}\|_F^2. \\ &\quad \text{s.t. } \mathbf{S}\mathbf{1} = \mathbf{1}, \mathbf{S} \geq \mathbf{0}\end{aligned} \tag{28}$$

We can further reformulate Eq. (28) as

$$\min_{\boldsymbol{s}_i^\top\mathbf{1}=1, \boldsymbol{s}_i \geq \mathbf{0}} \sum_{i=1}^n\sum_{j=1}^n (s_{ij} - a_{ij})^2 + \frac{\gamma}{2\lambda}\sum_{i=1}^n\sum_{j=1}^n s_{ij}\sum_{v=1}^V \|\boldsymbol{t}_i^v - \boldsymbol{t}_j^v\|_2^2. \tag{29}$$

Since Eq. (29) is independent of different i, it can be decomposed into a series of independent minimization problems w.r.t. $\{\boldsymbol{s}_i\}_{i=1}^n$:

$$\begin{aligned}&\min_{\boldsymbol{s}_i} \|\boldsymbol{s}_i - (\boldsymbol{a}_i - \frac{\gamma}{4\lambda}\boldsymbol{e}_i)\|_2^2, \\ &\quad \boldsymbol{s}_i^\top\mathbf{1} = 1, \boldsymbol{s}_i \geq \mathbf{0}\end{aligned} \tag{30}$$

where $\boldsymbol{e}_i$ is a vector with the jth element $e_{ij} = \sum_{v=1}^V \|\boldsymbol{t}_i^v - \boldsymbol{t}_j^v\|_2^2$. The Lagrangian function of Eq. (30) can be written as

$$L(\boldsymbol{s}_i, \eta, \boldsymbol{\beta}_i) = \frac{1}{2}\|\boldsymbol{s}_i - (\boldsymbol{a}_i - \frac{\gamma}{4\lambda}\boldsymbol{e}_i)\|_2^2 - \eta(\boldsymbol{s}_i^\top\mathbf{1} - 1) - \boldsymbol{\beta}_i^\top\boldsymbol{s}_i, \tag{31}$$

where η and $\boldsymbol{\beta}_i$ represent the Lagrangian multipliers. According to the Karush–Kuhn–Tucker condition, we can get the optimal solution

$$\boldsymbol{s}_i = (\boldsymbol{a}_i - \frac{\gamma}{4\lambda}\boldsymbol{e}_i + \eta)_+. \tag{32}$$

We summarize the detailed algorithm to solve Eq. (8) in Algorithm 3.

Algorithm 3 Algorithm to solve Eq. (8)

Input: Multimodal neuroimaging data $\mathcal{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\}$, label matrix $\mathbf{Y}$, similarity matrix $\mathbf{A}^v$, and model parameters $\alpha, \beta, \gamma, \lambda, h$.;

while not converged **do**
  for $v = 1 : V$ **do**
    Update $\mathbf{W}^v$ by Algorithm 1;
  end for
  for $v = 1 : V$ **do**
    Update $\mathbf{\Theta}^v$ by Algorithm 2;
  end for
  Update $\boldsymbol{b}^v$ by Eq. (21);
  Update $\mathbf{H}$ by Eq. (23);
  Update $\mathbf{P}$ by Eq. (25);
  Update $\boldsymbol{b}$ by Eq. (27);
  for $i = 1 : n$ **do**
    Update $\boldsymbol{s}_i$ by Eq. (32);
  end for
end while

3.2. Complexity analysis and model prediction

We analyze the computational complexity of Algorithm 3. The computational cost of the proposed algorithm mainly lies on updating variables $\{\mathbf{W}^v\}_{v=1}^V$, $\{\mathbf{\Theta}^v\}_{v=1}^V$, $\mathbf{H}$, $\mathbf{P}$, and $\mathbf{S}$. Algorithm 3 employs Algorithm 1 to update variables $\{\mathbf{W}^v\}_{v=1}^V$, which takes the computational complexity of $\mathcal{O}(\sum_v (d_v^2 h t_1 + d_v^2 n + d_v n h + d_v n^2))$, where t_1 is the iteration number in Algorithm 1. We use Algorithm 2 to update $\{\mathbf{\Theta}^v\}_{v=1}^V$ and its computational complexity is $\mathcal{O}(\sum_v (d_v^3 t_2))$, where t_2 is the iteration number in Algorithm 2. Since $c \ll n, h$, and $h \ll d_v$ in practice, the computational complexity values for updating variable $\mathbf{H}$ and $\mathbf{P}$ are $\mathcal{O}(nh^2 + \sum_v h d_v^2 + \sum_v n d_v^2)$ and $\mathcal{O}(nh^2 + h^3)$, respectively. For updating variable $\mathbf{S}$, its computational complexity $\mathcal{O}(\sum_v n^2 d_v)$. Therefore, the complexity of the algorithm is summarized as $\mathcal{O}(t_3 \sum_v (d_v^2 h t_1 + d_v^2 n + d_v n^2 + d_v^3 t_2))$, where t_3 is the iteration number in Algorithm 3.

For a given testing sample $\{\boldsymbol{x}_{te}^v\}_{v=1}^V$, we obtain the latent representation by averaging the latent representation from each modality, i.e., $\boldsymbol{h}_{te} = (\sum_{v=1}^V \mathbf{W}^{v\top} \mathbf{\Theta}^v \boldsymbol{x}_{te}^v + \boldsymbol{b}^v)/V$. Then, the label for this testing sample is given as $\boldsymbol{y}_{te} = \mathbf{P}\boldsymbol{h}_{te} + \boldsymbol{b}$.

3.3. Convergence analysis

The convergence of the proposed optimization algorithm is proved in this section. The functions for optimizing $\mathbf{H}, \mathbf{P}, \boldsymbol{b}$ and $\boldsymbol{b}^v$ are convex; Since Eq. (30) is a constrained quadratic minimization, updating $\mathbf{S}$ is a convex problem. The convergence of the ALM algorithm for updating $\mathbf{\Theta}^v$ has been generally proven in Bertsekas (2014). The GPI algorithm is used to optimize $\mathbf{W}^v$ and the convergence of the GPI algorithm has been proved in Nie et al. (2017). As a result, the objective function value in Eq. (8) decreases monotonically in each iteration and Algorithm 3 will converge to its local optima. Fig. 1 shows the convergence results of objective value with the number of iterations on the ADNI-2 dataset. It is seen that the objective value drops to a stable value in a few iterations, which indicates that the proposed optimization algorithm has good convergence properties.

4. Experiments

In this section, we carry out experiments to evaluate the effectiveness of the proposed OLFG method. We first introduce the datasets used in our experiments and then present the baseline methods and evaluation settings. Afterward, we compare the performance of OLFG with related approaches. Finally, we perform a comprehensive study to show the effects of different components of OLFG on the diagnosis performance.

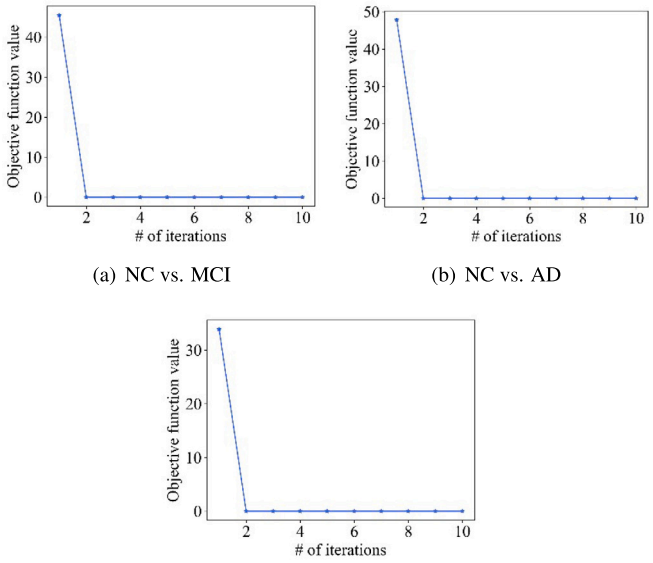


Fig. 1. Convergence of Algorithm 3 on the ADNI-2 dataset.

4.1. Experimental datasets

The neuroimaging data in this study are from the ADNI data repository (http://adni.loni.usc.edu) (Weiner et al., 2010). The ADNI project was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations with a 60 million, 5-year public–private partnership. The main purpose of this project is to verify whether MRI, PET, other biomarkers, and neuropsychological assessments can be combined to measure the progression of AD and its early stage. As the preprocessed neuroimaging features of the ADNI-2 dataset are provided on the public ADNI website, we conduct experiments on the ADNI-2 dataset to provide a high degree of reproducibility. The ADNI-2 dataset includes 757 patients (144 ADs, 330 MCIs, and 283 normal controls (NCs)). We further divide MCI subjects into pMCI subjects that will progress to AD within a predefined period and sMCI subjects that remain stable or convert to normal. A longer follow-up period allows to earlier possible AD detection and therapeutic intervention. However, the use of long-term follow-up data will also reduce the available training sample size. Considering that almost all subjects in the ADNI-2 dataset have complete follow-up data in 3 years and most related works (Moradi et al., 2015; Young et al., 2013; Wee et al., 2013) aim to predict the conversion from MCI to AD within 3 years, we set the follow-up period as 3 years. As 11 MCI patients only have the diagnostic information at baseline, they could not be classified as sMCI or pMCI. As a result, there are 234 sMCI subjects and 85 pMCI subjects in the ADNI-2 dataset. The demographic information is summarized in Table 1. To further validate the generalization performance of OLFG, we also conduct performance evaluation on an independent external cohort, i.e., the OASIS-3 dataset (LaMontagne et al., 2019), containing 29 AD patients and 329 normal controls.

We adopt the features of MRI and PET that can be downloaded from the public ADNI website to provide a high degree of reproducibility. For MRI, we use the features processed by a team from University of California at San Francisco. They performed cortical reconstruction and volumetric segmentation with the FreeSurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/). Preprocessed 3T T1 weighted image data (gradient warping, scaling, B1 correction and N3 inhomogeneity correction) were run with FreeSurfer. The cortical volume (CV), surface

Table 1
Demographic information of subjects.

	ADNI-2					OASIS-3	
	NC	MCI	AD	sMCI	pMCI	NC	AD
Gender(Male/Female)	130/153	183/147	84/60	130/104	47/38	140/189	12/17
Age (mean±std)	72.2±9.5	70.9±10.0	73.0±13.3	70.4±10.8	70.2±14.9	71.7±8.4	76.1±6.6
MMSE (mean±std)	29.0±1.2	28.0±1.8	23.1±2.1	28.3±1.6	27.2±1.8	29.1±1.2	24.5±4.0
Education (mean±std)	16.6±2.5	16.4±2.6	15.7±2.7	16.4±2.7	16.2±2.5	–	–

area (SA), cortical thickness average (TA), and standard deviation of thickness (TS) of cortical regions and subcortical regions were extracted as features. Meanwhile, total intracranial volume (ICV) and left and right hemisphere SA were also calculated. For PET, we use the features processed by a team from University of California at Berkeley. Each florbetapir image (AV45 Coreg, Avg, Std Img and Vox Siz, Uniform Resolution) was coregistered using SPM8 to that subject's MRI image that is closest in time. Then, they performed skull-strip, segment, and delineate cortical and subcortical regions in all MRI scans using FreeSurfer. The standardized uptake value ratios (SUVRs) for frontal, anterior/posterior cingulate, lateral parietal, lateral temporal regions were extracted. The data and details of the processing procedure are available at the public ADNI website (http://adni.loni.usc.edu/). As several regions do not satisfy the quality control requirements, there exist features with missing values in the dataset. We remove the features with missing values and a total of 311 MRI features and 115 PET features are used for the ADNI-2 dataset. As 12 PET features in the ADNI-2 dataset are not inclued in the OASIS-3 dataset, only 103 PET features are used for the OASIS-3 dataset and the preprocessed features can be downloaded from https://www.oasis-brains.org.

4.2. Compared methods

We compare OLFG with eight baselines:

- RMSR (Ning et al., 2021): It learns a bi-directional mapping between original space and shared space to exploit the underlying associations inherent in multimodal data, and then projects the shared representations into the target space for AD diagnosis.
- CMLS (Zhou et al., 2020): It projects multimodal data into a common latent space to capture the cross-modality correlations. Then, it ensembles multiple diversified classifiers to obtain a robust classification result.
- LRL (Zhou et al., 2019b): It learns a latent representation space with multimodal data and then projects the latent representation to the label space for AD diagnosis.
- SRSML (Zhu et al., 2015): It integrates LDA and LPP into a sparse feature selection framework to incorporate global and local information jointly for selecting informative features from the original feature set.
- R2DLSR (Lei et al., 2017): It is a discriminative sparse learning framework that incorporates the intrinsic relationships among features and subjects for AD diagnosis.
- L2PSC (Zhu et al., 2017): It embeds three kinds of relationships, i.e., feature–feature relation, sample–sample relation, and response–response relation, into a sparse learning framework for AD diagnosis.
- PTDCN (Gao et al., 2021): It contains a dense convolutional network for each modality to learn the deep features for AD classification. Moreover, three pathwise transfer blocks are built to communicate information across different modalities for performance promotion.
- MMDL (Zhang et al., 2019): It builds two independent convolutional neural networks for extracting features from the MRI and PET images. Then, the correlation analysis is performed to evaluate the consistency of the output of the two networks and the results are combined with the clinical neuropsychological tests for AD diagnosis.

4.3. Experimental settings

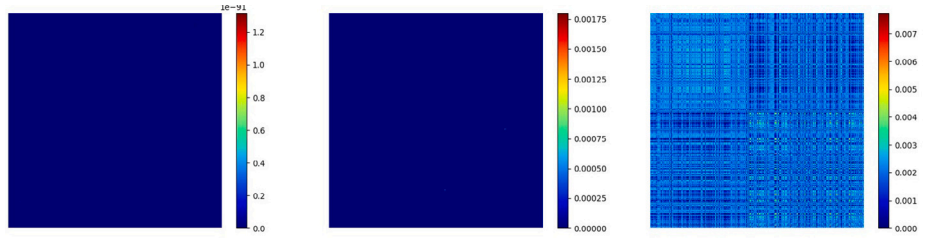
Following related works (Zhu et al., 2017; Ning et al., 2021), we evaluate all comparison methods using 10-fold cross-validation strategy and the final results are computed by averaging the repeated experiments. We use accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the curve (AUC) as our evaluation criteria. All experiments are repeatedly conducted 10 times for recording the average performance. The proposed method has five user-defined hyperparameters: $\alpha, \beta, \gamma, \lambda$, and h. Parameters $\alpha, \beta, \gamma, \lambda$ are determined by grid search in the range of $\{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$ and parameter h are determined by grid search in the range of $\{2, 3, 5, 10, 20, 30, 40, 50\}$. Three classification tasks are performed, including NC vs. MCI, NC vs. AD, and sMCI vs. pMCI. Our experiments are implemented on a Windows 10 desktop computer with a 3.6 GHz Intel Core i3-8100 CPU and 8 GB RAM. Our code has been released at https://github.com/chenz96/OLFG.

4.4. Performance comparison

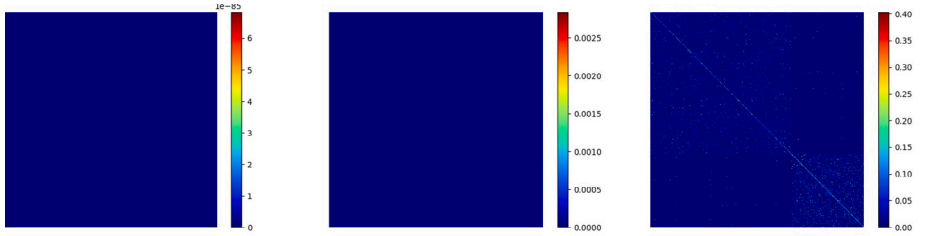
In this section, we compare our proposed OLFG to several state-of-the-art baselines. Table 2 summarizes the mean ACC, SEN, SPE, and AUC values for all the compared methods. It can be seen that OLFG performs the best among all evaluated methods in most cases, which demonstrates the superiority of OLFG for the multimodal AD diagnosis tasks. Specifically, OLFG outperforms other latent space-based methods (i.e., RMSR, CMLS, and LRL). This implies that OLFG can learn a more discriminative latent space than other latent space-based methods. The reason is that OLFG focuses on useful features through adaptive feature weighting and imposes the orthogonal constraint on the projection matrices to retain discriminative information in the latent space. Moreover, OLFG obtains better performance when compared with other Laplacian regularization-based methods (i.e., RMSR, LRL, R2DLSR, L2PSC). There are two potential reasons for this phenomenon. First, other Laplacian regularization-based methods calculate the similarity graph on the original input space and the graph construction is independent of the following processes. In contrast, graph learning and latent representation learning are integrated into a unified framework and can be optimized jointly in OLFG, which can help to build an accurate graph in a task-driven manner. Second, other Laplacian regularization-based methods construct a similarity graph for each modality, ignoring the fact that different modalities should share a common underlying graph structure which is vital for the diagnosis. OLFG constructs a joint graph for multiple modalities, which can accurately encode the relationships among different modalities. In addition, the results reported for NC vs. MCI and sMCI vs. pMCI are lower than the performance for NC vs. AD for all models. The main reason is that the brain differences between normal controls and MCI patients, as well as between sMCI and pMCI patients, are much lower than those between normal controls and AD patients. As a result, it is easier for diagnosis models to extract information that can be used to distinguish AD patients from normal controls. Finally, the performance achieved by OLFG is comparable to those of deep learning methods, i.e., PT-DCN and CNN. As shown in Table 3, in terms of ACC, the proposed method outperforms PT-DCN by a margin of 2.8% and MMDL by a margin of 3.5% for NC vs. AD classification. The improvements achieved by OLFG can be attributed to the capability in modeling the contributions of different features for AD diagnosis. In contrast, PTDCN

Table 2
Performance comparison of different methods on the ADNI-2 dataset for NC vs. MCI, NC vs. AD, and sMCI vs. pMCI classifications.

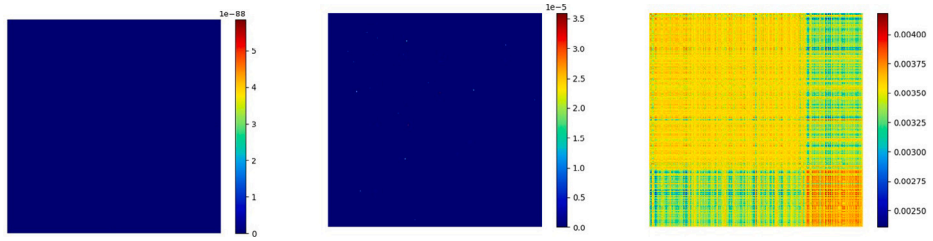
Method	NC vs. MCI				NC vs. AD				sMCI vs. pMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
RMSR	0.658	0.594	0.732	0.707	0.918	0.944	0.905	0.967	0.730	0.849	0.689	0.811
CMLS	0.656	0.570	0.756	0.688	0.876	0.942	0.841	0.953	0.749	**0.858**	0.710	**0.828**
LRL	0.617	0.536	0.710	0.667	0.909	0.950	0.887	0.964	0.737	0.383	0.868	0.800
SRSML	0.640	0.558	0.735	0.669	0.921	0.861	0.931	0.958	0.784	0.506	0.885	0.801
R2DLSR	0.658	0.564	0.767	0.702	0.878	0.953	0.841	0.955	0.721	0.815	0.689	0.800
L2PSC	0.664	0.564	**0.781**	0.702	0.895	**0.951**	0.866	0.958	0.749	0.849	0.714	0.805
PTDCN	0.638	0.673	0.598	0.708	0.900	0.814	0.944	0.947	0.772	0.336	0.929	0.806
MMDL	0.661	0.673	0.647	0.712	0.907	0.813	0.953	0.943	0.781	0.366	0.946	0.771
OLFG	**0.671**	**0.697**	0.640	**0.719**	**0.947**	0.890	**0.982**	**0.970**	**0.802**	0.425	**0.953**	0.814



(a) NC vs. MCI



(b) NC vs. AD



(c) sMCI vs. pMCI

Fig. 2. Visualization of the similarity graphs on the ADNI-2 dataset. The first two columns correspond to the input single-view similarity graphs respectively. The third column corresponds to the similarity graphs learned by OLFG.

and MMDL take whole brain images as input, which is insufficient for discovering important brain regions.

Moreover, we further perform an extra group of experiments on the OASIS-3 dataset. In this group of experiments, the proposed OLFG model and competing methods are trained using the samples in the ADNI-2 dataset and tested on the independent OASIS-3 dataset. As there is no MCI patient in the OASIS-3 dataset, we only show the results for the NC vs. AD classification task in this group of experiments. The results achieved by different methods are reported in Table 3. OLFG yields the best performance for most evaluation metrics on the OASIS-3 dataset. Specifically, the ACC and AUC values of OLFG are 0.949 and 0.973, which are better than those of the best competing method, i.e., CMLS. This set of results also demonstrates the effectiveness of OLFG for AD diagnosis.

4.5. Quantitative component-wise evaluation

In this section, we conduct ablation studies to evaluate the different components in the proposed model. Five variants of the proposed method are implemented:

- OLFG_P: A variant of OLFG that only employs PET data for training.
- OLFG_M: A variant of OLFG that only employs MRI data for training.
- OLFG_A: A variant of OLFG without updating the joint similarity graph.
- OLFG_G: A variant of OLFG by neglecting the graph Laplacian regularization.

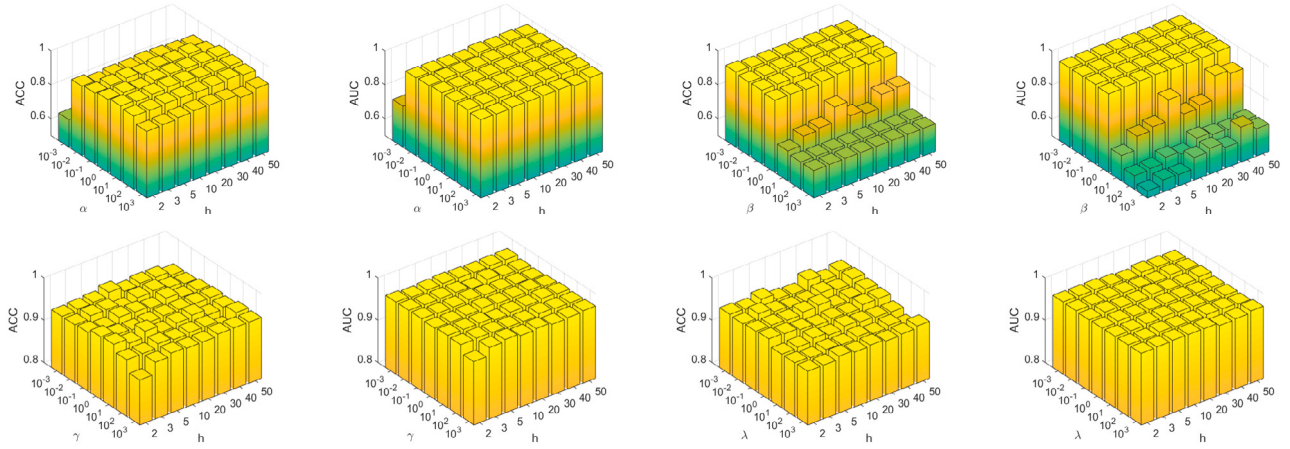


Fig. 3. The diagnosis performance of OLGF versus the variations of parameters on the ADNI-2 dataset.

Table 3
Performance of different methods that trained on the ADNI-2 dataset and tested on the OASIS-3 dataset for NC vs. AD classification.

Method	NC vs. AD			
	ACC	SEN	SPE	AUC
RMSR	0.846	0.931	0.839	0.950
CMLS	0.935	0.828	0.945	0.957
LRL	0.827	0.793	0.830	0.883
SRSML	0.901	0.724	0.916	0.904
R2DLSR	0.807	**0.966**	0.793	0.969
L2PSC	0.810	**0.966**	0.796	0.959
PTDCN	0.921	0.897	0.924	0.963
MMDL	0.914	0.886	0.918	0.954
OLFG	**0.949**	0.862	**0.957**	**0.973**

- OLFG_F: A variant of OLFG that does not employ the feature weighting matrices for latent space learning.

Tables 4 and 5 show the results of OLFG and its five variants. Obviously, the metrics output by OLFG are higher than those by its variants in most cases, indicating the effectiveness of the components in the diagnosis tasks. Specifically, by comparing OLFG_P and OLFG_M with OLFG, we find that OLFG achieves better diagnosis performance, showing the effectiveness of OLFG for fusing multimodal data. Besides, it is observed that the performance of OLFG drops when the feature weighting matrices are removed by comparing OLFG with OLFG_F. For example, the ACC and AUC values of OLFG for the NC vs. MCI task are 0.671 and 0.719, respectively, while those of OLFG_F are 0.654 and 0.692, respectively. That is because OLFG has the ability to recognize the important features from the original features with the feature weighting matrices. As a result, the latent space learned by OLFG contains more information related to the diagnosis than OLFG_F. From the results among OLFG_A, OLFG_G and OLFG, it can be found that the metrics provided by OLFG are higher than those by OLFG_A and OLFG_G. This is because OLFG learns an integrated global graph from multimodal data, which captures the correlations among different modalities and reduces the influence of noises. To further show the effectiveness of the graph learning strategy in OLFG, we show the similarity graphs constructed on the original features of different modalities and the similarity graphs learned by OLFG on the ADNI-2 dataset. The results are shown in Fig. 2. Obviously, due to the influence of noises and redundant features in the original space, the block diagonalization of the graphs constructed in the original feature space is distributed. In contrast, the graphs generated by OLFG are block diagonalization and have more structural information than the original graphs. The results demonstrate that the graphs learned by OLFG are helpful for improving the diagnosis performance.

5. Discussion

In this section, we first analyze the parameter sensitivity and show the discriminative features identified by OLFG. Then, we compare our results with those of several state-of-the-art methods. Finally, we analyze the limitations and possible future directions of our work.

5.1. Parameter evaluation

In this part, we discuss the behaviors of the major hyperparameters for a better understanding of our model. There are five parameters (the dimension of latent representation h, regularizer parameters α, β, γ, and λ) that need to be set in OLFG. Hyperparameter h is selected from $\{2, 3, 5, 10, 20, 30, 40, 50\}$ and hyperparameters $\alpha, \beta, \gamma, \lambda$ are searched from $\{10^{-3}, 10^{-2}, \ldots, 10^{2}, 10^{3}\}$. For simplicity, the NC vs. AD task on the ADNI-2 dataset is used to study the influence of parameters, as shown in Fig. 3. It can be observed that OLFG is sensitive to the changes of α and β. When α is large and β is small, our method can obtain a stable and relatively good performance. However, when $\alpha < 10^{-2}$ or $\beta > 10^{0}$, the performance drops dramatically. When α is small, the impact of classification loss decreases and the classification error increases. When β is large, $\mathbf{P}$ is excessively sparse and provides little information for the classification. We also find that the variation of γ and λ has few impacts on ACC and AUC as shown in Figs. 3(c) and 3(d), which demonstrates that OLFG is robust to γ and λ. Moreover, it is observed that OLFG still obtains relatively good ACC and AUC when h is very small, showing that projecting the high dimensional multimodal data into a low dimensional space is effective for improving the diagnosis performance.

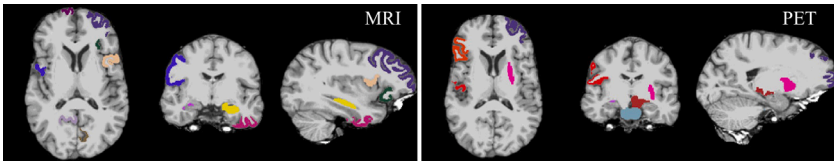
5.2. Qualitative results

In this section, we investigate the top features identified by OLFG. The feature weighting values in $\{\boldsymbol{\theta}_v\}_{v=1}^{V}$ can be used to evaluate the importance of each feature. Small or even zero weighting values will be assigned to uninformative or redundant features, while large weighting values will be assigned to informative features. We sort all features according to their absolute feature weighting values and report the top 10 features frequently selected in cross-validation in Table 6 and visualize the brain regions related to the top 10 features in Fig. 4.

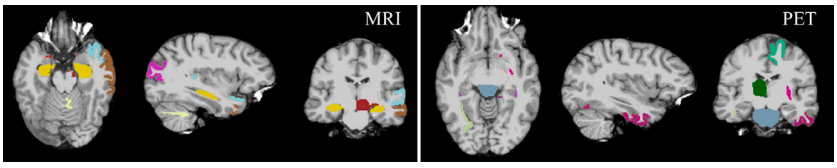
As can be seen, some important brain regions are selected by OLFG for all tasks, such as the hippocampus, entorhinal, inferior temporal, and inferior lateral ventricle. These identified brain regions have been recognized to be highly related to AD and its early stage diagnosis (Eichenbaum, 2004; Astillero-Lopez et al., 2022; Scheff et al., 2011; Scelsi et al., 2020). For example, Henneman et al. (2009) investigated the hippocampal measures in patients with AD, MCI, and normal

Table 4
Performance of OLFG and its variants on the ADNI-2 dataset for NC vs. MCI, NC vs. AD, and sMCI vs. pMCI classifications.

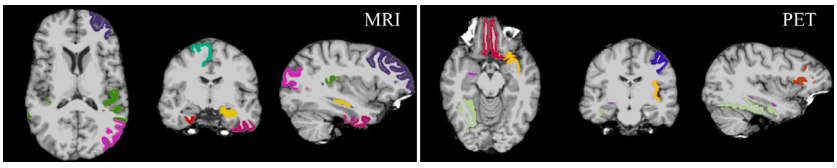
Method	NC vs. MCI				NC vs. AD				sMCI vs. pMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
OLFG_P	0.669	0.667	0.672	0.715	0.897	0.832	0.929	0.948	0.787	0.392	0.932	0.784
OLFG_M	0.594	0.652	0.526	0.634	0.906	0.800	0.961	0.947	0.774	0.408	0.906	0.741
OLFG_A	0.653	0.673	0.629	0.716	0.939	0.860	0.980	**0.971**	0.777	0.356	0.932	0.810
OLFG_G	0.664	0.687	0.625	0.712	0.936	0.854	**0.984**	0.966	0.759	0.318	0.919	0.806
OLFG_F	0.654	0.671	0.620	0.692	0.892	0.833	0.922	0.944	0.784	0.371	0.936	0.804
OLFG	**0.671**	**0.697**	**0.640**	**0.719**	**0.947**	**0.890**	0.982	0.970	**0.802**	**0.425**	**0.953**	**0.814**



(a) NC vs. MCI



(b) NC vs. AD



(c) sMCI vs. pMCI

Fig. 4. Visualization of the regions selected by OLFG for the ADNI-2 dataset: (a) NC vs. MCI, (b) NC vs. AD, and (c) sMCI vs. pMCI.

Table 5
Performance of OLFG and its variants that trained on the ADNI-2 dataset and tested on the OASIS-3 dataset for NC vs. AD classifications.

Method	NC vs. MCI			
	ACC	SEN	SPE	AUC
OLFG_P	0.913	**0.897**	0.915	0.937
OLFG_M	0.930	0.655	0.954	0.946
OLFG_A	0.946	0.690	**0.973**	0.958
OLFG_G	0.948	0.764	0.966	0.969
OLFG_F	0.933	0.799	0.942	0.967
OLFG	**0.949**	0.862	0.957	**0.973**

controls. The results show that the hippocampal atrophy rate best discriminates MCI from normal controls and regional measures of hippocampal atrophy are the strongest predictors of progression to AD. The entorhinal is one of the earliest areas involved in AD and the volume reduction of the entorhinal has been widely reported in AD patients (Astillero-Lopez et al., 2022). Scheff et al. (2011) analyzed the impact of the inferior temporal gyrus and found that it is affected during the early clinical stages of AD and related to some of the early AD-related symptoms. The putamen is selected by OLFG for the tasks of NC vs. AD and NC vs. MCI classifications. It is reported that the decreased volumes of the putamen correlate independently to poor cognitive test performance (de Jong et al., 2008). For the tasks of NC vs. AD and sMCI vs. pMCI classifications, the superior frontal, isthmuscingulater, and fusiform are selected by OLFG. These areas are thought to be highly correlated with cognitive decline and dementia (Li et al., 2013; Ma et al., 2020; Wei et al., 2018). The rostral middle frontal and parstriangularis are selected for the tasks of NC vs. MCI and sMCI vs. pMCI classifications. Zhao et al. Zhao et al. (2015) found that the thickness of the rostral middle frontal is correlated with cognitive impairment. The parstriangularis has been shown to have a role in cognitive control of memory (Yang et al., 2019). These results suggest that our method is able to identify the relevant brain regions for AD diagnosis.

5.3. Comparison to the related prior works

In this part, we compare the performance of OLFG with those of existing methods for AD diagnosis using multimodal data from ADNI datasets, including the traditional machine learning methods and the deep learning methods. For compared methods, the results reported in the literature are presented in Table 7. It should be noted that due to the differences in dataset selection, preprocessing framework, and dataset partition, the results of different methods are actually incomparable. As shown in Table 7, we can have several observations as follows. First, the results on the ADNI-2 dataset are usually lower than those on the ADNI-1 dataset. For example, for the method proposed by Hao et al. (2020), the AUC on the ADNI-1 dataset (0.980) is higher than that on the ADNI-2 dataset (0.950). The possible reason is that the ADNI-2 dataset uses a different study protocol than the ADNI-1 dataset. Second, the performance of OLFG is higher than those of several deep learning methods (i.e., Suk et al., 2016, Huang et al., 2019, Liu et al., 2020a, and Gao et al., 2021) for most evaluation metrics. The improvements achieved by OLFG can be attributed to the capability in modeling the contributions of different features for AD diagnosis. In contrast, deep learning methods usually take whole brain images as input, which is insufficient for discovering important brain regions. Moreover, compared

Table 6
The features selected by OLFG for different classification tasks on the ADNI-2 dataset.

	#	Selected MRI feature	Selected PET feature
NC vs. AD	1	Subcortical Volume (aseg.stats) of RightHippocampus	Left_putamen_suvr
	2	Thickness Average (aparc.stats) of LeftEntorhinal	Right_inf_lat_vent_suvr
	3	Subcortical Volume (aseg.stats) of LeftHippocampus	Right_thalamus_proper_suvr
	4	Cortical Volume (aparc.stats) of LeftEntorhinal	Ctx_rh_superiorfrontal_suvr
	5	Thickness Average (aparc.stats) of LeftSuperiorTemporal	Summarysuvr_composite_refnorm
	6	Subcortical Volume (aseg.stats) of LeftVentralDC	left_inf_lat_vent_suv
	7	Subcortical Volume (aseg.stats) of RightCerebellumWM	Ctx_lh_inferiortemporal_suvr
	8	Thickness Average (aparc.stats) of RightEntorhinal	Ctx_rh_fusiform_suvr
	9	Thickness Average (aparc.stats) of LeftInferiorParietal	Brainstem_suvr
	10	Thickness Average (aparc.stats) of LeftMiddleTemporal	Ctx_lh_isthmuscingulate_suvr
NC vs. MCI	1	Subcortical Volume (aseg.stats) of LeftHippocampus	Left_putamen_suvr
	2	Subcortical Volume (aseg.stats) of RightInferiorLateralVentricle	Right_inf_lat_vent_suvr
	3	Surface Area (aparc.stats) of LeftRostralMiddleFrontal	Brainstem_suvr
	4	Cortical Volume (aparc.stats) of LeftParsOpercularis	Ctx_rh_postcentral_suvr
	5	Surface Area (aparc.stats) of RightPrecuneus	Ctx_rh_parstriangularis_suvr
	6	Cortical Volume (aparc.stats) of LeftLateralOrbitofrontal	Summarysuvr_wholecerebnorm
	7	Cortical Volume (aparc.stats) of LeftInferiorTemporal	Ctx_lh_rostralmiddlefrontal_suvr
	8	Cortical Volume (aparc.stats) of RightFrontalPole	Right_ventraldc_suvr
	9	Surface Area (aparc.stats) of RightPrecentral	Summarysuvr_composite_refnorm
	10	Surface Area (aparc.stats) of LeftPericalcarine	Cingulate_suvr
sMCI vs. pMCI	1	Cortical Volume (aparc.stats) of RightSuperiorFrontal	Summarysuvr_wholecerebnorm
	2	Subcortical Volume (aseg.stats) of LeftHippocampus	Ctx_lh_isthmuscingulate_suvr
	3	Thickness Average (aparc.stats) of LeftInferiorTemporal	Ctx_lh_medialorbitofrontal_suvr
	4	Cortical Volume (aparc.stats) of RightEntorhinal	Ctx_rh_paracentral_suvr
	5	Thickness Average (aparc.stats) of LeftRostralMiddleFrontal	Ctx_rh_parstriangularis_suvr
	6	Thickness Average (aparc.stats) of LeftSupramarginal	Ctx_rh_fusiform_suvr
	7	Cortical Volume (aparc.stats) of LeftTemporalPole	Ctx_lh_insula_suvr
	8	Cortical Volume (aparc.stats) of Icv	Ctx_rh_medialorbitofrontal_suvr
	9	Thickness Average (aparc.stats) of Rightsupramarginal	right_inf_lat_vent_suvr
	10	Surface Area (aparc.stats) of Leftinferiorparietal	ctx_lh_precentral_suvr

Table 7
Comparison of the multimodal classification performances reported in the literature.

Method	Subject	Dataset	NC vs. MCI				NC vs. AD				sMCI vs. pMCI			
			ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Suk et al. (2016)	51 AD, 52 NC, 43 pMCI, 56 sMCI	ADNI-1	0.788	0.908	0.560	–	0.951	0.920	0.980	–	0.730	0.530	0.890	–
Liu et al. (2017b)	186 AD, 226 NC, 395 MCI, 169 pMCI and 226 sMCI	ADNI-1	0.800	0.862	0.688	0.805	0.931	0.900	0.957	0.948	0.790	0.608	0.925	0.797
Shi et al. (2018)	51 AD, 99 MCI, 52 NC, 43 pMCI and 56 sMCI	ADNI-1	0.870	0.942	0.713	–	0.969	0.950	0.984	–	0.765	0.625	0.863	–
Huang et al. (2019)	465 AD, 567 MCI, 480 CN	–	–	–	–	–	0.901	0.909	0.892	0.908	0.722	0.734	0.713	0.775
Zhou et al. (2019b)	171 AD, 204 NC, 157 pMCI, 205 sMCI,	ADNI-1	–	–	–	–	–	–	–	–	0.743	–	–	0.755
Hao et al. (2020)	51 AD, 52 NC, 99 MCI, 43 pMCI and 56 sMCI	ADNI-1	0.845	0.940	0.662	0.810	0.976	0.984	0.967	0.980	0.778	0.674	0.855	0.769
	211 NC, 160 AD	ADNI-2					0.937	0.952	0.918	0.950				
Shi et al. (2020)	51 AD, 99 MCI, and 52 NC	ADNI-1	0.799	0.846	0.710	0.766	0.946	0.954	0.944	0.937	–	–	–	–
Liu et al. (2020a)	97 AD, 233 MCI, 119 NC	ADNI-1	0.762	0.795	0.698	0.775	0.889	0.866	0.908	0.925	–	–	–	–
Gao et al. (2021)	352 AD, 427 NC, 234 pMCI, 342sMCI	ADNI-1 for training and ADNI-2 for testing	–	–	–	–	0.920	0.891	0.940	0.956	0.753	0.773	0.741	0.786
OLFG	144 AD, 330 MCI, 283 NC, 234 sMCI, 85 pMCI	ADNI-2	0.671	0.697	0.640	0.719	0.947	0.890	0.982	0.970	0.802	0.425	0.953	0.814

with deep learning methods, our proposed method has the advantage of interpretability in discovering useful neuroimaging biomarkers. Third, OLFG achieves comparable performance in the three AD diagnosis tasks compared with other methods. Moreover, OLFG has better overall performance in sMCI vs. pMCI classification. For instance, OLFG achieves a 1.2% improvement in terms of ACC compared with other methods in sMCI vs. pMCI classification. These results imply the effectiveness of OLFG for capturing the discriminative information from multimodal neuroimaging data for AD diagnosis.

5.4. Limitations and future directions

There are several challenges for OLFG to be solved in the future. First, we validate the performance of OLFG on the samples with complete MRI and PET data. However, the missing-modality problem commonly exists in neuroimaging datasets due to patient dropout or poor data quality. Two terms in our objective function involve the multimodal input data, i.e., the latent space learning term and the Laplacian regularization term. In our future work, we will modify the two terms to make our method applicable to incomplete multimodal input data and validate its effectiveness. Second, as the sample size in the experiments might not be able to fully validate the generalizability of OLFG, it would be highly desirable to make use of the data from multiple datasets, such as ADNI, AIBL, and OASIS, for improving the sample size. However, the major challenges of using the data from multiple datasets arise from inconsistent disease labels and different neuroimaging protocols: 1) The disease labels and patient groups included in different datasets are usually diverse. As a consequence, if we want to combine the two datasets for improving the sample size, the label-inconsistent samples in the OASIS dataset have to be discarded. We will deal with this problem by borrowing ideas from semi-supervised learning (Liu et al., 2020c). Specifically, perturbed

samples are created by adding Gaussian noises into a label-inconsistent sample. Then, we encourage the model prediction results on these samples to be similar. In this way, the model outputs under perturbations are similar and the model robustness can be improved (Liu et al., 2022). (2) The neuroimaging protocols used for different datasets are usually different. For example, the amyloid imaging tracer for the ADNI-2 dataset is Florbetapir (AV45) while the AIBL dataset does not provide the AV45 PET scans (The AIBL dataset provides the PIB PET scans for most subjects). Since the scans with different protocols are usually not comparable (Pan et al., 2021; Bourgeat et al., 2022), the two datasets cannot be used together. To address this issue, we will design a generative adversarial network to learn a map from PIB PET scans to AV45 PET scans. Then, the AV45 PET scans for the AIBL dataset are synthesized using the existing PIB PET scans. Consequently, the synthetic PET scans in the AIBL dataset can be used with the PET scans in the ADNI-2 dataset to increase the sample size.

6. Conclusion

In this paper, we present an orthogonal latent space learning with feature weighting and graph learning approach named OLFG for multimodal AD diagnosis. We project multimodal neuroimaging data into a common latent space to capture the correlations among different modalities. Meanwhile, an adaptive feature weighting matrix is learned for each modality to measure the feature importance in the latent space learning. To retain enough statistical and structural information, OLFG utilizes orthogonal projection to learn the latent space. Moreover, we develop a graph Laplacian regularization term with learned graph to preserve the relationships among samples in the latent space. Different from most existing approaches which calculate the similarity graph on the original feature space, we integrate the similarity graph construction into the learning processing for accurately encoding the relationships among samples and improving the robustness to noises. Finally, the learned latent feature representation is projected to the label space for AD diagnosis. Experimental results on the ADNI-2 and OASIS-3 datasets demonstrate the effectiveness of OLFG.

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 authors do not have permission to share data. However, data can be downloaded from the websites of ADNI and OASIS. The link to the code has been shared in the paper.

Acknowledgments

This research was supported in part by the National Key R&D Program of China under grant 2019YFC1710300, and the Sichuan Science and Technology Program under grants 2019YFS0019, 2020YFS0283, and 2021YJ0184, and the rehabilitation technology innovation center by joint collaboration of ministry of education and Fujian province under grant X2022016-Collaboration.

Data collection and sharing for ADNI was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

The OASIS-3 data were provided in part by OASIS Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

References

Adeli, E., Thung, K.-H., An, L., Wu, G., Shi, F., Wang, T., Shen, D., 2018. Semi-supervised discriminative classification robust to sample-outliers and feature-noises. IEEE Trans. Pattern Anal. Mach. Intell. 41 (2), 515–522. http://dx.doi.org/10.1109/TPAMI.2018.2794470.

Alberdi, A., Aztiria, A., Basarab, A., 2016. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. Artif. Intell. Med. 71, 1–29. http://dx.doi.org/10.1016/j.artmed.2016.06.003.

Astillero-Lopez, V., Gonzalez-Rodriguez, M., Villar-Conde, S., Flores-Cuadrado, A., Martinez-Marcos, A., Ubeda-Banon, I., Saiz-Sanchez, D., 2022. Neurodegeneration and astrogliosis in the entorhinal cortex in Alzheimer's disease: Stereological layer-specific assessment and proteomic analysis. Alzheimer's & Dementia http://dx.doi.org/10.1002/alz.12580.

Bertsekas, D.P., 2014. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, http://dx.doi.org/10.1016/C2013-0-10366-2.

Bobinski, M., De Leon, M.J., Convit, A., De Santi, S., Wegiel, J., Tarshish, C.Y., Saint Louis, L., Wisniewski, H.M., 1999. MRI of entorhinal cortex in mild Alzheimer's disease. Lancet 353 (9146), 38–40. http://dx.doi.org/10.1016/S0140-6736(05)74869-8.

Bourgeat, P., Doré, V., Burnham, S.C., Benzinger, T., Tosun, D., Li, S., Goyal, M., LaMontagne, P., Jin, L., Rowe, C.C., et al., 2022. β-Amyloid PET harmonisation across longitudinal studies: Application to AIBL, ADNI and OASIS3. NeuroImage 262, 119527.

Eichenbaum, H., 2004. Hippocampus: cognitive processes and neural representations that underlie declarative memory. Neuron 44 (1), 109–120. http://dx.doi.org/10.1016/j.neuron.2004.08.028.

El-Sappagh, S., Abuhmed, T., Islam, S.R., Kwak, K.S., 2020. Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data. Neurocomputing 412, 197–215. http://dx.doi.org/10.1016/j.neucom.2020.05.087.

Gao, X., Shi, F., Shen, D., Liu, M., 2021. Task-induced pyramid and attention GAN for multimodal brain image imputation and classification in Alzheimers disease. IEEE J. Biomed. Health Inf..

Gauthier, S., 2005. Alzheimer's disease: the benefits of early treatment. Euro. J. Neurol. 12, 11–16. http://dx.doi.org/10.1111/j.1468-1331.2005.01322.x.

Górriz, J.M., Ramirez, J., Suckling, J., Consortium, M.A., et al., 2019. On the computation of distribution-free performance bounds: Application to small sample sizes in neuroimaging. Pattern Recognit. 93, 1–13. http://dx.doi.org/10.1016/j.patcog.2019.03.032.

Hänninen, T., Hallikainen, M., Tuomainen, S., Vanhanen, M., Soininen, H., 2002. Prevalence of mild cognitive impairment: a population-based study in elderly subjects. Acta Neurol. Scand 106 (3), 148–154.

Hao, X., Bao, Y., Guo, Y., Yu, M., Zhang, D., Risacher, S.L., Saykin, A.J., Yao, X., Shen, L., Alzheimer's Disease Neuroimaging Initiative, et al., 2020. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease. Med. Image Anal. 60, 101625. http://dx.doi.org/10.1016/j.media.2019.101625.

Henneman, W., Sluimer, J., Barnes, J., Van Der Flier, W., Sluimer, I., Fox, N., Scheltens, P., Vrenken, H., Barkhof, F., 2009. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. Neurology 72 (11), 999–1007. http://dx.doi.org/10.1212/01.wnl.0000344568.09360.31.

Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X., (ADNI), A.D.N.I., 2019. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. Front. Neurosci. 13, 509.

Jack Jr., C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9 (1), 119–128. http://dx.doi.org/10.1016/S1474-4422(09)70299-6.

de Jong, L.W., van der Hiele, K., Veer, I.M., Houwing, J., Westendorp, R., Bollen, E., de Bruin, P.W., Middelkoop, H., van Buchem, M.A., van der Grond, J., 2008. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain 131 (12), 3277–3285.

Kantarci, K., Murray, M.E., Schwarz, C.G., Reid, R.I., Przybelski, S.A., Lesnick, T., Zuk, S.M., Raman, M.R., Senjem, M.L., Gunter, J.L., et al., 2017. White-matter integrity on DTI and the pathologic staging of Alzheimer's disease. Neurobiol. Aging 56, 172–179. http://dx.doi.org/10.1016/j.neurobiolaging.2017.04.024.

LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al., 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. MedRxiv.

Langbaum, J.B., Chen, K., Lee, W., Reschke, C., Bandy, D., Fleisher, A.S., Alexander, G.E., Foster, N.L., Weiner, M.W., Koeppe, R.A., et al., 2009. Categorical and correlational analyses of baseline fluorodeoxyglucose positron emission tomography images from the Alzheimer's disease neuroimaging initiative (ADNI). Neuroimage 45 (4), 1107–1116. http://dx.doi.org/10.1016/j.neuroimage.2008.12.072.

Lei, B., Yang, P., Wang, T., Chen, S., Ni, D., 2017. Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. IEEE Trans. Cybern. 47 (4), 1102–1113. http://dx.doi.org/10.1109/TCYB.2016.2644718.

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., Yu, C., 2013. Subregions of the human superior frontal gyrus and their connections. Neuroimage 78, 46–58.

Liu, H., Han, J., Nie, F., 2017a. Semi-supervised orthogonal graph embedding with recursive projections.. In: IJCAI. pp. 2308–2314. http://dx.doi.org/10.24963/ijcai.2017/321.

Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Shen, L., Xu, M., Initiative, A.D.N., et al., 2020a. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. Neuroimage 208, 116459.

Liu, J., Pan, Y., Wu, F.-X., Wang, J., 2020b. Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification. Neurocomputing 400, 322–332. http://dx.doi.org/10.1016/j.neucom.2020.03.006.

Liu, F., Tian, Y., Chen, Y., Liu, Y., Belagiannis, V., Carneiro, G., 2022. ACPL: Anti-curriculum pseudo-labelling for semi-supervised medical image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20697–20706.

Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A., 2020c. Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Trans. Med. Imaging 39 (11), 3429–3440.

Liu, M., Zhang, J., Yap, P.-T., Shen, D., 2017b. View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. Med. Image Anal. 36, 123–134.

Ma, D., Fetahu, I.S., Wang, M., Fang, R., Li, J., Liu, H., Gramyk, T., Iwanicki, I., Gu, S., Xu, W., et al., 2020. The fusiform gyrus exhibits an epigenetic signature for Alzheimer's disease. Clin. Epigenetics 12 (1), 1–16. http://dx.doi.org/10.1186/s13148-020-00916-3.

Martinez-Murcia, F.J., Górriz, J.M., Ramírez, J., Ortiz, A., 2016. A structural parametrization of the brain using hidden Markov models-based paths in Alzheimer's disease. Int. J. Neural Syst. 26 (07), 1650024. http://dx.doi.org/10.1142/S0129065716500246.

Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Alzheimer's Disease Neuroimaging Initiative, et al., 2015. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage 104, 398–412.

Nie, F., Zhang, R., Li, X., 2017. A generalized power iteration method for solving quadratic problem on the stiefel manifold. Sci. China Inf. Sci. 60 (11), 1–10. http://dx.doi.org/10.1007/s11432-016-9021-9.

Ning, Z., Xiao, Q., Feng, Q., Chen, W., Zhang, Y., 2021. Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis. IEEE Trans. Med. Imaging 40 (6), 1632–1645. http://dx.doi.org/10.1109/TMI.2021.3063150.

Pan, Y., Liu, M., Xia, Y., Shen, D., 2021. Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. IEEE Trans. Pattern Anal. Mach. Intell..

Qian, P., Chen, Y., Kuo, J.-W., Zhang, Y.-D., Jiang, Y., Zhao, K., Al Helo, R., Friel, H., Baydoun, A., Zhou, F., et al., 2019. Mdixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification. IEEE Trans. Med. Imaging 39 (4), 819–832. http://dx.doi.org/10.1109/TMI.2019.2935916.

Rao, Y.L., Ganaraja, B., Murlimanju, B., Joy, T., Krishnamurthy, A., Agrawal, A., 2022. Hippocampus and its involvement in Alzheimer's disease: a review. 3 Biotech 12 (2), 1–10. http://dx.doi.org/10.1007/s13205-022-03123-4.

Scelsi, C., Rahim, T., Morris, J., Kramer, G., Gilbert, B., Forseen, S., 2020. The lateral ventricles: A detailed review of anatomy, development, and anatomic variations. Am. J. Neuroradiol. 41 (4), 566–572.

Scheff, S.W., Price, D.A., Schmitt, F.A., Scheff, M.A., Mufson, E.J., 2011. Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease. J. Alzheimer's Dis. 24 (3), 547–557. http://dx.doi.org/10.3233/JAD-2011-101782.

Shen, H.T., Zhu, X., Zhang, Z., Wang, S.-H., Chen, Y., Xu, X., Shao, J., 2021. Heterogeneous data fusion for predicting mild cognitive impairment conversion. Inf. Fusion 66, 54–63. http://dx.doi.org/10.1016/j.inffus.2020.08.023.

Shi, Y., Suk, H.-I., Gao, Y., Lee, S.-W., Shen, D., 2020. Leveraging coupled interaction for multimodal Alzheimer's disease diagnosis. IEEE Trans. Neural Netw. Learn. Syst. 31 (1), 186–200. http://dx.doi.org/10.1109/TNNLS.2019.2900077.

Shi, J., Zheng, X., Li, Y., Zhang, Q., Ying, S., 2018. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J. Biomed. Health Inf. 22 (1), 173–183. http://dx.doi.org/10.1109/JBHI.2017.2655720.

Suk, H.-I., Lee, S.-W., Shen, D., 2016. Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. Brain Struct. Funct. 221 (5), 2569–2587.

Theofilas, P., Ehrenberg, A.J., Nguy, A., Thackrey, J.M., Dunlop, S., Mejia, M.B., Alho, A.T., Leite, R.E.P., Rodriguez, R.D., Suemoto, C.K., et al., 2018. Probing the correlation of neuronal loss, neurofibrillary tangles, and cell death markers across the Alzheimer's disease Braak stages: a quantitative study in humans. Neurobiol. Aging 61, 1–12. http://dx.doi.org/10.1016/j.neurobiolaging.2017.09.007.

Wang, X., Huang, W., Su, L., Xing, Y., Jessen, F., Sun, Y., Shu, N., Han, Y., 2020. Neuroimaging advances regarding subjective cognitive decline in preclinical Alzheimer's disease. Mol. Neurodegen. 15 (1), 1–27. http://dx.doi.org/10.1186/s13024-020-00395-3.

Wang, C., Saar, V., Leung, K.L., Chen, L., Wong, G., 2018. Human amyloid β peptide and tau co-expression impairs behavior and causes specific gene expression changes in Caenorhabditis elegans. Neurobiol. Dis. 109, 88–101. http://dx.doi.org/10.1016/j.nbd.2017.10.003.

Wang, Y., Zhang, L., Tong, X., Nie, F., Huang, H., Mei, J., 2017. LRAGE: Learning latent relationships with adaptive graph embedding for aerial scene classification. IEEE Trans. Geosci. Remote Sens. 56 (2), 621–634. http://dx.doi.org/10.1109/TGRS.2017.2752217.

Wee, C.-Y., Yap, P.-T., Shen, D., Initiative, A.D.N., 2013. Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum. Brain Mapping 34 (12), 3411–3425.

Wei, H., Kong, M., Zhang, C., Guan, L., Ba, M., Alzheimer's Disease Neuroimaging Initiative, et al., 2018. The structural MRI markers and cognitive decline in prodromal Alzheimer's disease: a 2-year longitudinal study. Quant. Imaging Med. Surg. 8 (10), 1004.

Weiner, M.W., Aisen, P.S., Jack Jr., C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., et al., 2010. The Alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimer's & Dementia 6 (3), 202–211. http://dx.doi.org/10.1016/j.jalz.2010.03.007.

Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., Wu, Y., Li, W., Yang, C., Li, X., et al., 2019. Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. Gen. Psych. 32 (2).

Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S., Initiative, A.D.N., et al., 2013. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. NeuroImage Clin. 2, 735–745.

Zhan, K., Niu, C., Chen, C., Nie, F., Zhang, C., Yang, Y., 2018. Graph structure fusion for multiview clustering. IEEE Trans. Knowl. Data Eng. 31 (10), 1984–1993. http://dx.doi.org/10.1109/TKDE.2018.2872061.

Zhang, T., Davatzikos, C., 2013. Optimally-discriminative voxel-based morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease. Neuroimage 79, 94–110. http://dx.doi.org/10.1016/j.neuroimage.2013.04.063.

Zhang, F., Li, Z., Zhang, B., Du, H., Wang, B., Zhang, X., 2019. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. Neurocomputing 361, 185–195.

Zhang, Y., Wang, S., Xia, K., Jiang, Y., Qian, P., Initiative, A.D.N., et al., 2021. Alzheimer's disease multiclass diagnosis via multimodal neuroimaging embedding feature selection and fusion. Inf. Fusion 66, 170–183. http://dx.doi.org/10.1016/j.inffus.2020.09.002.

Zhao, H., Li, X., Wu, W., Li, Z., Qian, L., Li, S., Zhang, B., Xu, Y., 2015. Atrophic patterns of the frontal-subcortical circuits in patients with mild cognitive impairment and Alzheimer's disease. PLoS One 10 (6), e0130017.

Zhou, R., Chang, X., Shi, L., Shen, Y.-D., Yang, Y., Nie, F., 2019a. Person reidentification via multi-feature fusion with adaptive graph learning. IEEE Trans. Neural Netw. Learn. Syst. 31 (5), 1592–1601. http://dx.doi.org/10.1109/TNNLS.2019.2920905.

Zhou, T., Liu, M., Thung, K.-H., Shen, D., 2019b. Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. IEEE Trans. Med. Imaging 38 (10), 2411–2422. http://dx.doi.org/10.1109/TMI.2019.2913158.

Zhou, T., Thung, K.-H., Liu, M., Shi, F., Zhang, C., Shen, D., 2020. Multi-modal latent space inducing ensemble SVM classifier for early dementia diagnosis with neuroimaging data. Med. Image Anal. 60, 101630. http://dx.doi.org/10.1016/j.media.2019.101630.

Zhu, X., Suk, H.-I., Lee, S.-W., Shen, D., 2015. Subspace regularized sparse multi-task learning for multiclass neurodegenerative disease identification. IEEE Trans. Biomed. Eng. 63 (3), 607–618. http://dx.doi.org/10.1109/TBME.2015.2466616.

Zhu, X., Suk, H.-I., Wang, L., Lee, S.-W., Shen, D., Initiative, A.D.N., et al., 2017. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. Med. Image Anal. 38, 205–214. http://dx.doi.org/10.1016/j.media.2015.10.008.