IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 26, 2018, accepted January 17, 2019, date of publication January 23, 2019, date of current version March 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894530

Sparse Feature Learning With Label Information for Alzheimer's Disease Classification Based on Magnetic Resonance Imaging

LINA XU[1,2], ZHIJUN YAO[3], JING LI[1], CHEN LV[1], HUAXIANG ZHANG[1], AND BIN HU[1,3]

[1]School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China
[2]School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China
[3]Key Laboratory of Wearable Computing of Gansu Provincial, Lanzhou University, Lanzhou 930000, China

Corresponding authors: Huaxiang Zhang (huaxzhang@hotmail.com) and Bin Hu (bh@lzu.edu.cn)

This work was supported in part by the National Basic Research Program of China (973 Program) (No.2014CB744600), in part by the National Natural Science Foundation of China under Grant 61632014, Grant 61210010, Grant 61373081, Grant 61772322, Grant 61572298, Grant 61702310, Grant 61402268, Grant 61401260, and Grant 61601268, in part by the Shandong Provincial Natural Science Foundation under Grant ZR2016FB13, and in part by the Project of Shandong Province Higher Educational Science and Technology Program under Grant J16LN18.

ABSTRACT Neuroimaging techniques have been used for automatic diagnosis and classification of Alzheimer's disease and mild cognitive impairment. How to select discriminant features from these data is the key that will affect the subsequent automatic diagnosis and classification performance. However, in the previous manifold regularized sparse regression models, the local neighborhood structure was constructed directly in the traditional Euclidean distance without fully utilizing the label information of the subjects, which leads to the selection of less discriminative features. In this paper, we propose a novel manifold regularized sparse regression model for learning discriminative features. Specifically, we first adopt $\ell_{2,1}$-norm regularization to jointly select a relevant feature subset among the samples. Then, to select more discriminative features, a novel manifold regularization term is constructed via the relative distance adjusted by the label information, which can simultaneously maintain the compactness of the intra-class samples and the separability of inter-class samples. The proposed feature learning method is further carried out for both the binary classification and the multi-class classification. The experimental results on Alzheimer's Disease Neuroimaging Initiative database demonstrate the effectiveness of the proposed method, which can be utilized for the diagnosis of Alzheimer's disease and mild cognitive impairment.

INDEX TERMS Alzheimer's disease, feature learning, sparse regression, manifold regularization.

I. INTRODUCTION

Alzheimer's disease (AD) is a degenerative, irreversible, incurable dementia and will eventually cause the patients to lose basic living ability. It is estimated that 1 in every 85 persons will be affected by AD by year 2050 [1]. AD patients have been not only a huge economic burden to the society but also a great trouble to the patients as well as their families. Mild cognitive impairment (MCI) is known as a prodromal stage of AD. Previous studies have shown that the normal controls (NC) subjects progress to AD patients at a rate of approximately 1% to 2% per year [2], whereas MCI subjects convert to AD patients at an annualized rate around 10% to 15% [3], [4]. The stage of MCI is a golden period to effectively curb the conversion of MCI subjects into AD patients [5], [6], so the intervention and treatment for MCI subjects can possibly alleviate their pain. Therefore, it is essential to correctly identify AD patients and mine discriminant features for early diagnosis and treatment of AD patients and MCI subjects.

Recently, biomedical signal processing techniques have been widely applied for AD/MCI studies, such as structural Magnetic Resonance Imaging (MRI), positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) [7]–[9]; meanwhile, machine learning methods have presented promising performances in various

The associate editor coordinating the review of this manuscript and approving it for publication was Liangtian Wan.

applications [10]–[14], including biomedical applications. Specifically, structural MRI data provide information about the tissue type of the brain, and could serve as a powerful tool for the analysis of AD patients and MCI subjects due to its clear contrast and high spatial resolution [15], [16]. Many valuable structural features extracted from the structural MRI data have been identified for AD patients and MCI subjects, such as tissue probability maps, cortical thickness and hippocampal volumes [17], [18]. Additionally, network techniques is applied in many fields and a series of research has been studied [19]–[25]. The brain is a complex network system where the interrelationships exist between different brain region, rather than in one single brain region. Therefore, based on the network techniques, using the structural features(e.g. cortical thickness(CT) and local gyrification index (LGI)), the personal morphological brain network can be further constructed and the characteristics of personal network is used for AD/MCI classification. In [15] and [26], personal brain network was constructed using cortical thickness for attaining good classification performance. In [27], personal network was constructed using multiple structural features to improve the accuracy of classifying AD patients and MCI subjects.

In the past decades, many promising performances have been already achieved in AD applications. However, for biomedical signal processing data, even after conducting feature extraction, there is still redundant and irrelevant features, which may lead to poor performance of subsequent classification. It is necessary to conduct feature selection which can remove less discriminant features to obtain an effective feature subset. Feature selection methods can mainly be divided into three categories [28]: the filter models [29], the wrapper models [30] and the Embedded models [31]. The filter(e.g. Minimum Redundancy Maximum Relevance (mRMR) [32]) and the wrapper (e.g. recursive feature elimination algorithm (RFE) [33]) models have been widely applied in AD studies. In [13] and [17], a two-step feature selection method was adopted, including the mRMR filter method and the RFE wrapper method to find an optimal feature subset and gain higher classification performance with the SVM classifier. Yao *et al.* [34] used two filter methods and a RFE wrapper method to select features for AD detection using FDG-PET data. The embedded methods (e.g., the method LARS [35], LASSO [31], Elastic Net [36]) could obtain superior performance over the two selection models mentioned above, and have been successfully used for various application researches including AD studies [37]–[42]. For instance, Zhang *et al.* [38] used Sparse Multi-Task Learning model for feature selection. Zhu *et al.* [44] proposed a Sparse Multi-Task Learning with Subspace Regularization to select features, where subspace regularization was constructed in an unsupervised manner. Jie *et al.* [45] developed a manifold regularized multitask feature learning for multimodality disease classification method. Zu *et al.* [46] presented a Label-aligned manifold regularization for multitask feature selection method. Ye *et al.* [47] introduced a new discriminative regularization term based on intra-class and inter-class Laplacian matrices. Generally, in most recent studies, the local neighborhood relation of the manifold regularization is constructed directly among samples via the traditional Euclidean distance, without considering the label information of all subjects which may lead to selecting redundant features subset. Importantly, the label information could improve the quality of selected neighborhood and may further improve the subsequent classification performance. Therefore, in this paper, based on the label information, we propose a novel manifold regularized sparse feature selection method to select more discriminative features for AD/MCI classification. Specifically, we first introduce a novel label information-based manifold regularization term into sparse regress model, which could better preserve the local structural information and obtain the optimal feature subset. We then adopt the support vector machine (SVM) [48] with radial basis function (RBF) kernel to evaluate the performance of the proposed method. Experimental results show that our method is more effective than several others methods. The main contributions of ourwork can be summarized as follows:

1) We propose a novel manifold regularized sparse feature learning method for MCI/AD Classification based on structural magnetic resonance signal processing data.

2) The selected brain regions by our method can be utilized for the diagnosis of Alzheimer's disease and mild cognitive impairment.

The rest of our paper is organized as follows: Materials and image preprocessing are introduced in Section II. Section III gives the details of the proposed method. The experimental results are in Section IV and discussion is in Section V. Sections VI and VII finally gives the limitations and concludes the paper, respectively.

II. MATERIALS AND IMAGE PREPROCESSING

A. MATERIALS

The data we used in this paper is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). ADNI was launched in 2003 by the National Institute of Biomedical Imaging and Bioengineering, as a non-profit organization led by Principal Investigator Michael W. Weiner, MD. The initial goal of ADNI is to evaluate the progression of early Alzheimer's disease, i.e., MCI, by combining technologies such as magnetic resonance imaging (MRI), fluorodeoxyglucose positron emission tomography (FDG-PET), Cerebrospinal fluid (CSF) and other biological markers. We download baseline MRI data from ADNI database, which includes 165 NC subjects, 142 AD patients and 221 MCI subjects. The MCI Participants were further divided into two categories: cMCI (126) who progressed to AD within 3years from baseline and sMCI (95) who did not progress to AD within the same time period. Table1 lists the detailed demographic information of the participants.

TABLE 1. Demographic information of the participant.

Variables	NC	sMCI	cMCI	AD
Age	76.4±5.4	74.9±7.3	73.4±9.2	76.1±7.5
Male/female	78/87	63/32	73/53	72/70
MMSE	29.2±1.0	27.7±1.7	26.5±1.7	23.2±2.0
CDR	0	0.5	0.5	0.5/1

Note: MMSE: Mini-Mental State Examination. CDR: Clinical Dementia Rating;

B. IMAGE PREPROCESSING AND FEATURE EXTRACTION

Biomedical Signal image pre-processing and feature extraction are performed by the following procedures. First, T1-weighted images were preprocessed using FreeSurfer software (http://surfer.nmr.mgh.harvard.edu). This process involves the following steps: motion correction, non-brain tissue removal, coordinate transformation, grey matter segmentation, and reconstruction of gray/white matter boundaries. In particular, the reconstruction and segmentation errors were visually checked in FreeView and manually corrected. Next, surface inflation and registration were performed [49]–[51]. The cortical thickness was calculated in each vertex based on the distances between the white matter and pial surface. Finally, we obtained the mean cortical thickness of each brain region according to the Automated Anatomical Labeling (AAL) atlas and removed the subcortical tissues [52], resulting in 78 cortical regions (78 ROIs).

III. THE PROPOSED METHOD

In this section, we will propose our method, which is called manifold regularized sparse feature learning with label information. We first briefly introduce the feature selection with sparse regression model, and then present the details of our method, as well as the corresponding iterative optimization algorithm. Finally, we utilize SVM classifier for the three binary classification tasks and one multi-class classification task. An overview of the proposed classification framework is illustrated in Figure1.

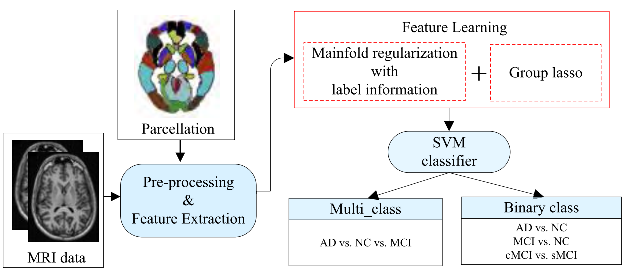

FIGURE 1. Overall workflow of our proposed method.

A. NOTATIONS

We begin with a brief description of some notations used in this paper. For a matrix $X = [x_{i,j}] \in R^{n\times d}$, its transpose, inverse, and trace operator are denoted by X^T, X^{-1} and $tr(X)$, respectively. Its i-th row and i-th column are denoted by x_i and x_i^T, respectively. The $\ell_{2,1}$-norm [53] of a matrix W is the sum of $\ell 2$-norm of the rows of W:$\|W\|_{2,1} = \sum_i \|W\|_2 = \sum_i \sqrt{\sum_j w_{ij}^2}$.

B. FEATURE LEARNING WITH SPARSE REGRESSION MODEL

Let $X \in R^{n\times d}$ denote the feature matrix, where n is the number of training samples, d is the dimension of features. $Y \in R^{n\times c}$ denotes the class label matrix with binary encoding, where c is the number of classes. Consequently, the objective function of sparse regression model with a group lasso [31], [44] is defined as follows:

$$\min_W \frac{1}{2}\|Y - XW\|_F^2 + \lambda\|W\|_{2,1} \tag{1}$$

where $W \in R^{d\times c}$ is the regression coefficient matrix, and λ is a weighting parameter which not only balances the relative importance between the loss term and the regularization term, but also controls the sparsity of elements in W matrix. The $\ell_{2,1}$-norm encourages the sparsity of rows in W, and then common features will be jointly selected among samples corresponding to multi-class label matrix. When $\lambda = 0$, all features are selected, as λ increases, the number of selected features decreases. In other words, a larger λ value means that there are more zero rows in W matrix and fewer features are selected. However, this model only selects common feature subset without considering the local geometric space structure among samples, and the selected features might be less discriminative.

C. MANIFOLD REGULARIZED SPARSE FEATURE LEARNING WITH LABEL INFORMATION

In this section, a joint framework of the sparse regression model is proposed for effectively selecting discriminative features.

First, we introduce a graph regularization term (Locality Preserving Projections (LPP)) as follows:

$$G = \sum_{i,j}^{n} s_{i,j}\left\|x_i W - x_j W\right\|_2^2 = tr[W^T X^T LXW] \tag{2}$$

where $L = D - S$ is a positive semi-definite symmetric matrix and $S = [s_{i,j}] \in R^{n\times n}$ is the similarity matrix with elements $s_{i,j}$ which represent the similar relationship between each pair of samples x_i and x_j. $D = [d_{i,j}] \in R^{n\times n}$ is a diagonal matrix with $d_{i,i} = \sum_{j=1}^n s_{ij}$. $s_{i,j}$ in Eq. (2) is defined as:

$$s_{i,j} = \begin{cases} 1, & x_i \in N_k(x_j)\ or\ x_j \in N_k(x_i),\ i \neq j \\ 0, & otherwise \end{cases} \tag{3}$$

where $N_k(x_i)$ denotes the k-neighbors of subject x_i. How to choose the optimal k-neighbors among the samples of different classes is the key issue in constructing the neighbor relationship. Unlike the traditional methods in [44], [46], and [47], $s_{i,j}$ is directly constructed based on the Euclidean distance, where the label information of all the subjects is not fully utilized. To solve this problem, we first introduced an adjustable parameter with a label information to control the absolute Euclidean distance between samples, and then the relative distance can be defined as follows:

$$rd_{ij} = \begin{cases} \delta \left\| x_i - x_j \right\| & \text{if } x_i \text{ and } x_j \text{ belong to the same class}, \\ & 0 < \delta \leq 1 \\ \frac{1}{\delta} \left\| x_i - x_j \right\| & \text{otherwise} \end{cases} \tag{4}$$

For each subject x_i, its k-neighbors $N_k(x_i)$ are calculated via the relative distance rd_{ij} which is controlled by the adjustable parameter δ. Notably, if two samples x_i and x_j belong to the same class, the relative distance rd_{ij} between them is reduced; contrarily, if two samples x_i and x_j belong to different classes, the relative distance rd_{ij} between them is extended. Therefore, intra-class samples are closer, while inter-class samples are far away. As a result, the optimal neighbors with the most similar features are easily selected via the relative distance. By constructing the k-neighbors graph, the regularization term in Eq. (2) could subtly preserve the relationship among the most similar neighbors ($k \in [5, 10]$). Interestingly, when parameter δ is equal to 1, the novel subspace regularization term in Eq. (2) will degenerate into the ordinary subspace regularization constructed in an unsupervised manner. This indicates that the neighbor relationship could be conducted by our proposed method in both unsupervised and supervised manner.

Substituting Eq. (2) into Eq. (1), we can obtain our proposed model as follows:

$$\min_W \frac{1}{2} \|Y - XW\|_F^2 + \lambda_1 \sum_{i,j}^{n} s_{i,j} \left\| x_i W - x_j W \right\|_2^2 + \lambda_2 \|W\|_{2,1} \tag{5}$$

where λ_1 and λ_2 are two balancing parameters. Our proposed method could preserve the relationship among the optimal neighbors and select more discriminative features to achieve superior performance in subsequent classification tasks. More importantly, feature selection is still guaranteed to perform in the original space which is easy to interpret or investigate the selected features.

D. OPTIMIZATION ALGORITHM

We use an iterative algorithm to solve the optimization problem in Eq. (5). Specifically, we first separate our objective function into the smooth part and non-smooth part. The smooth part is shown as follows:

$$\frac{1}{2} \|Y - XW\|_F^2 + \lambda_1 \sum_{i,j}^{n} s_{i,j} \left\| x_i W - x_j W \right\|_2^2 \tag{6}$$

While non-smooth part is shown as follows:

$$\lambda_2 \|W\|_{2,1} \tag{7}$$

According to [54], we can define $\phi(x) = \sqrt{x^2 + \varepsilon}$, ε is a smoothing term, which is usually set to a small value, so that the next iterative procedures can be guaranteed to converge. $\|W\|_{2,1}$ can then be rewritten as $\sum_{i=1}^{d} \sqrt{\|w_i\|_2^2 + \varepsilon}$ and can be optimized in a half-quadratic way [55], [56]. Therefore, we can replace $\|W\|_{2,1}$ as:

$$\|W\|_{2,1} = Tr(W^T R W) \tag{8}$$

where $R = Diag(r)$, r is an auxiliary vector of the $\ell_{2,1}$-norm. r_i can be computed as follows:

$$r_i = \frac{1}{\sqrt{\|w_i\|_2^2 + \varepsilon}} \tag{9}$$

From the above formula, our objective functions can be approximately expressed as follows:

$$\begin{aligned} &\min_W \frac{1}{2} \|Y - XW\|_F^2 + \lambda_1 \sum_{i,j}^{n} s_{i,j} \left\| x_i W - x_j W \right\|_2^2 + \lambda_2 \|W\|_{2,1} \\ &= \min_W \frac{1}{2} \|Y - XW\|_F^2 + \lambda_1 tr[W^T X^T L X W] \\ &\quad + \lambda_2 tr(W^T R W) \end{aligned} \tag{10}$$

Taking derivative of Eq. (10) with respect to W and set it to zero, we can get:

$$\begin{aligned} \frac{df(W)}{dW} &= X^T X W - X^T Y + \lambda_1 (X^T L X + X^T L^T X) W \\ &\quad + \lambda_2 (R W + R^T W) = 0 \end{aligned} \tag{11}$$

We can then get the solution for W as:

$$\begin{aligned} W = (X^T X - X^T + \lambda_1 (X^T L X + X^T L^T X) \\ + \lambda_2 (R + R^T))^{-1} X^T Y \end{aligned} \tag{12}$$

The main optimization procedure of our proposed method is summarized in Algorithm1.

Algorithm 1 Our Proposed Method

Input: Samples X and the label matrix Y, parameters λ_1 and λ_2
Output: The regression coefficient matrix W
1 Construct the Laplacian matrix L
2 Repeat
3 Compute $\mathrm{r_i}$ according to Eq. (9)
4 Update W according to Eq. (12)
5 Until Convergence

E. CLASSIFICATION

The optimal solution W obtained from Eq. (5) can be considered as a new representation of the samples X. Due to the $\ell_{2,1}$-norm imposed on W, many rows in W are approximately equal to zero. This indicates that the features corresponding to these approximate zero rows are uninformative in representing the class labels. Therefore, we ignore them and then select the top ranked rows as the results of feature selection [57]. We then use the Support Vector Machines (SVM) with linear

TABLE 2. Classification performance of different methods for three binary classification.

Method	AD vs. NC				MCI vs. NC				cMCI vs. sMCI			
	ACC(%)	SEN(%)	SPE(%)	AUC	ACC(%)	SEN(%)	SPE(%)	AUC	ACC(%)	SEN(%)	SPE(%)	AUC
Raw	88.10	91.421	84.65	0.9465	69.87	57.58	79.32	0.7838	56.23	63.27	21.62	0.5947
mRMR	89.08	92.20	85.66	0.9530	70.26	59.91	78.60	0.7899	61.58	79.79	38.91	0.6789
RFE	89.27	92.00	86.15	0.9524	70.44	58.58	79.73	0.7781	62.18	77.98	44.39	0.6818
SR	89.87	93.56	85.78	0.9534	70.78	61.58	78.15	0.7885	62.47	77.17	44.25	0.6802
ED	90.40	92.36	88.67	0.9536	70.78	61.31	78.34	0.7899	63.46	78.27	45.36	0.6783
Proposed	90.40	92.36	88.67	0.9536	70.89	61.39	78.42	0.7902	63.69	78.56	45.39	0.6787

kernel [48] for classification. We build three binary classifiers, i.e., AD vs. NC, MCI vs. NC and cMCI vs. sMCI respectively, and one multi-class classifier i.e., AD vs. MCI vs. NC. Here, we chose to use one-versus-one approach for multi-class classification. To evaluate the performance of our method, we apply a 10-fold cross-validation method, where all samples are randomly divided into 10 parts, and each part is left out in turn as test set, while the rest are used for training sets that undergo the nested feature selection, the optimal parameter values (i.e., λ_1, λ_2, δ and k) as mentioned above. We repeat this process 10 times independently to avoid the possible bias caused by randomly partitioning the dataset and the average results are reported. For binary classification tasks, we adopt four measures i.e. accuracy (ACC), sensitivity (SEN), specificity (SPE) and Area Under Curve (AUC) to quantify the performance of different methods, whereas only accuracy is calculated to evaluate the multi-class classification performance. In addition, for binary classification, we applied a grid search algorithm and 3-fold cross-validation to select the optimal parameter C of SVM (range from 2^{-8} to 2^8) on the training data. For multi-class classification, to short computing time, the parameter C is set to the default value on the training data.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETTINGS

We test the performance of our proposed method using cortical thickness of MRI data obtained from ADNI database. For each ROI, we compute the mean cortical thickness of the brain region, and thus each subject contains 78 features.

We compare our method with several state-of-the-art feature selection methods including mRMR and RFE. mRMR is a filter feature selection method that maximizes the correlation between each feature and class label variables, while minimizes the redundancy between each feature pair simultaneously [32]. RFE is a wrapper feature selection method, which removes the minimum discriminant features from the feature set to find an optimal feature subset [33]. To reveal the validity of feature selection, we also conduct the classification task using all features without feature selection (denoted as 'Raw'). Since our method combines the novel manifold regularization method and a sparse regression model with an $\ell_{2,1}$-norm method in a unified framework. When $\lambda_1 = 0$, our proposed model will be sparse regression model with a group lasso (denoted as SR) without the manifold regularization term. When $\lambda_2 = 0$, all features without feature selection are used for classification(denoted as 'Raw'). It's necessary that we should compare our method with SR method to justify the rationale of our method where adds the novel manifold regularization into SR method. In addition, to further verify the validity of relative distance term adjusted by the label information, we also make comparison with the relative distance in Eq. (5) with $0 < \delta \leq 1$ (our method) and the traditional Euclidean distance in Eq. (5) with $\delta = 1$ (denoted as ED). For fairness comparisons, all methods are conducted on the same training and test samples.

B. CLASSIFICATION PERFORMANCE

The comparison performance of three binary classification tasks are presented in Table2. We can see that our method achieves better performance than the other methods. Specifically, for AD vs. NC classification ($\lambda_1 = 1$, $\lambda_2 = 8$, $k = 9$, $\delta = 1$), our method improves the classification accuracy by 1.32% (mRMR), and 1.13% (RFE), respectively. For MCI vs. NC classification ($\lambda_1 = 0.01$, $\lambda_2 = 2$, $k = 8$, $\delta = 0.9$), our method improves the classification accuracy by 0.63% (mRMR)and 0.45% (RFE), respectively. For cMCI vs. sMCI classification ($\lambda_1 = 0.01$, $\lambda_2 = 5$, $k = 7$, $\delta = 0.1$), our proposed method achieves a classification accuracy of 63.65%, which is 2.11% and 1.51% higher than that of mRMR and RFE method, respectively. We also can see that the classification accuracy of our method is slightly higher than SR, and ED (only except for AD vs. NC classification, our method has the same classification performance as the ED method.). For further validation, we perform the significance test by using the standard paired t-test on the classification accuracy between SR, ED, our method and other methods. The p-values are shown in Table3. The smaller the p-value is, the more significant the difference is. If p-value is less than 0.05, it indicates a significant difference between two sets of data. We can see that our method achieves smaller p-values than SR and ED methods. For AD vs. NC classification, our proposed method is significantly better than all the other method. For cMCI vs. sMCI classification, our method is significantly better than mRMR. Only, for MCI vs. NC classification, there are no significant differences

TABLE 3. Significance test for three binary classification between SR, ED, our proposed method and other methods.

Method	*p*-value		
	AD vs. NC	MCI vs. NC	cMCI vs. sMCI
Proposed			
Raw	0.001	0.012	<0.001
mRMR	0.012	0.326	0.039
RFE	0.028	0.595	0.173
SR			
Raw	0.1015	0.2126	<0.001
MRMR	0.7339	0.4124	0.2140
RFE	1	0.7083	0.7153
ED			
Raw	0.001	0.1957	<0.001
MRMR	0.012	0.4545	0.0565
RFE	0.028	0.6999	0.2602

in classification performance between our method and the other methods, so do SR and ED. It is worth mentioning that, for MCI classification, we can see that there are no significant differences in performance between SR, ED and the other methods(except Raw), however, there are significant differences in performance between our method and the other methods(except RFE).

We also demonstrate the Receiver Operating Characteristic (ROC) curves for different methods in Figure2. Our method achieves the Area Under the ROC curve of 0.9536, 0.7089 and 0.6365 for AD vs. NC, MCI vs. NC and cMCI vs. sMCI classifications, respectively, indicating the effective classification performance of our method over the other methods in most case.

TABLE 4. Comparison of classification performance for 3-class i.e., cMCI vs. sMCI classification.

Method	ACC (%)
Raw	57.43
mRMR	57.88
RFE	57.91
SR	58.73
ED	58.97
Proposed	59.16

In addition, the classification accuracy for 3-class i.e., AD vs. NC vs. MCI (($\lambda_1 = 0.1$, $\lambda_2 = 3$, $k = 9$, $\delta = 0.4$)) is presented in Table 4. Our proposed method achieves a classification accuracy of 59.16 %, while the other methods are 57.43%,57.88%,57.91%,58.73 and 58.97% respectively. The significance test are shown in Table5. The *p*-values further confirmed the effectiveness of our method over mRMR and RFE. We also can see that our method achieves smaller *p*-values than SR and ED. There are significant differences in performance between our method, ED and the other methods, however, there are no significant differences in performance between SR and RFE.

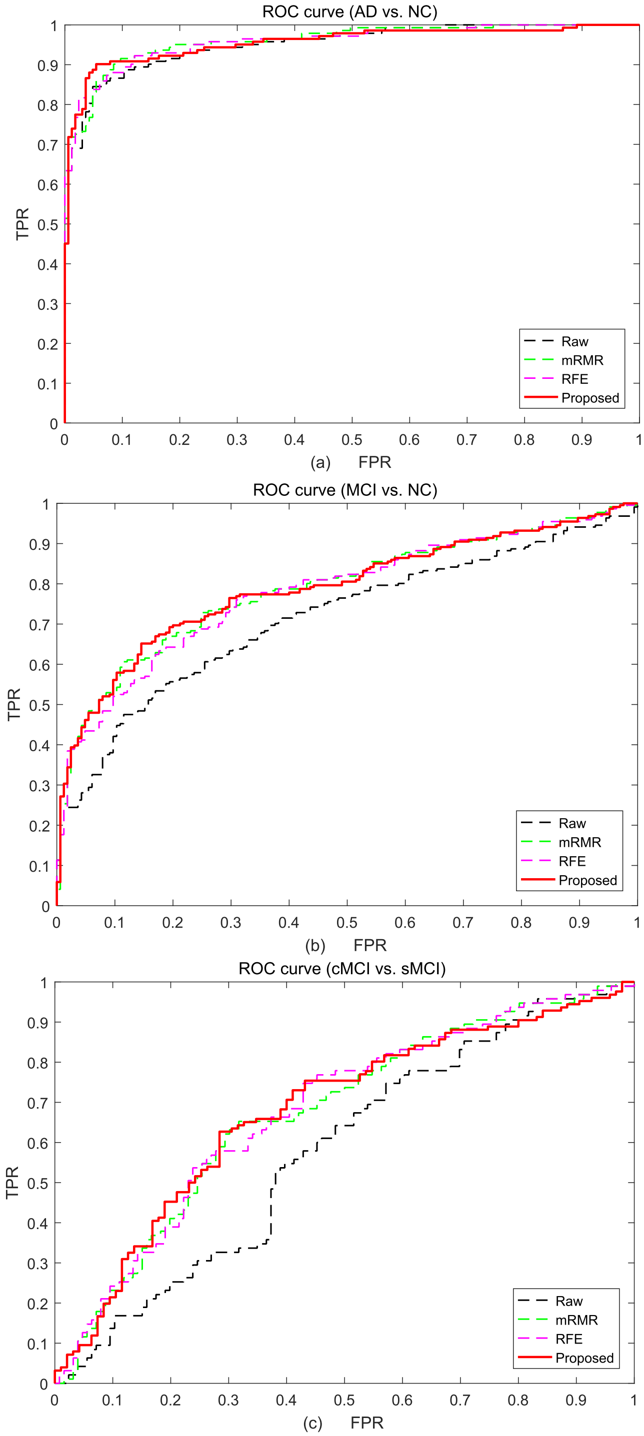

FIGURE 2. ROC curve of the classification performance of different methods. (a), (b) and (c) show the ROC curve of Classification of AD vs. NC, MCI vs. NC and cMCI vs. sMCI, respectively(TPR:true positive rate;FPR:false positive rate).

We investigated the top 10 selected brain regions by our method for MCI classification and 3-calss classification. Specifically, the most frequently selected brain regions in each cross-validation were defined as the top 10 ROIs. The results are shown in Table6 and Table7. We also visualized the top 10 ROIs in Figure3 and in Figure4.

TABLE 5. Significance test for 3-class between SR, ED, our proposed method and other methods.

Method	p-value
	AD vs. NC vs. MCI
Proposed	
Raw	<0.001
mRMR	<0.001
RFE	0.008
SR	
Raw	<0.001
MRMR	0.019
RFE	0.082
ED	
Raw	<0.001
MRMR	0.004
RFE	0.029

TABLE 6. Top 10 selected ROIs for MCI classification.

Rank	Index	Selected ROIs
1	65	Angular gyrus left
2	31	Anterior cingulate and paracingulate gyri left
3	32	Anterior cingulate and paracingulate gyri right
4	40	Parahippocampal gyrus right
5	70	Paracentral lobule right
6	34	Median cingulate and paracingulate gyri right
7	33	Median cingulate and paracingulate gyri left
8	83	Temporal pole: superior temporal gyrus left
9	45	Cuneus left
10	29	Insula left

TABLE 7. Top 10 selected ROIs for 3-class classification.

Rank	Index	Selected ROIs
1	84	Temporal pole: superior temporal gyrus right
2	87	Temporal pole: middle temporal gyrus left
3	39	Parahippocampal gyrus left
4	83	Temporal pole: superior temporal gyrus left
5	40	Parahippocampal gyrus right
6	82	Superior temporal gyrus right
7	17	Rolandic operculum left
8	69	Paracentral lobule left
9	10	Middle frontal gyrus, orbital part right
10	24	Superior frontal gyrus, medial right

Note: Index: The number of the brain region in AAL;

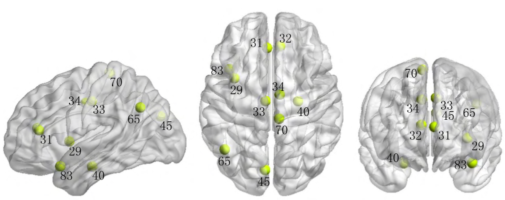

FIGURE 3. Top 10 selected ROIs by our method for MCI classification.

For MCI classification(Table6 and Figure3), the most discriminative regions are Angular gyrus, Cingulate, Parahippocampal gyrus, etc., which have been reported in AD/MCI studies [44], [58]–[62] and also shown to be highly related to AD/MCI diagnosis [44], [59], [63], [64].

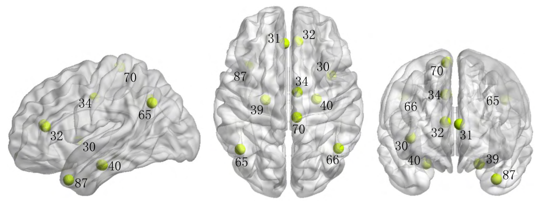

FIGURE 4. Top 10 ROIs selected by our method for 3-class classification.

Also, we can see from Table7 and Figure4, for 3-class, the top five-ranked discriminative regions are known to be highly related to AD/MCI which have been reported in the previous studies [46], [44], [58], [59], [62].

V. DISCUSSION

A. PERFORMANCE COMPARISON

The classification results listed in Table2 and Table4 demonstrate that our method generally outperforms the other feature section methods. Specifically, SR also obtains higher accuracies for all classification tasks, compared with mRMR and RFE. This indicates that the embedding feature section method is more effective than the filter and wrapper method in AD/MCI data. Notably, by introducing the manifold regularization into the sparse regression model, our proposed method achieves better classification performance than SR. Furthermore, from Table 3 and Table 5, We can see that our method both achieves smaller p-values than SR and ED methods.And there are significant differences in performance between our method and the other methods in most case. This indicates that combining the manifold regularization and SR in a unified framework can help enhance the classification performance, and relative distance term in Eq. (5) adjusted by the label information of our method is also effective for improving classification performance. This verifies the conclusion that the manifold regularization with label information could positively detect the most discriminative features and remove irrelevant features for improving the classification performance. Besides, all feature selection methods outperform the Raw, which implies the necessity of feature selection for AD/MCI classification.

B. EFFECT OF PARAMETERS

In our method, there are three main parameters including an adjustable parameter δ and two regularization parameters (i.e., λ_1 and λ_2). To evaluate their effects on the final prediction performances, we first study the parameter sensitivity of λ_1 and λ_2 by fixing δ. Specifically, λ_1 is tested from 10^{-3} to 10^3 and λ_1 is tested from 1 to 10. As shown in Figure5, we can see that λ_1 and λ_2 only have minor effects on the prediction accuracy of our method.

We also study different values of δ by fixing λ_1 and λ_2, δ is tested from 0 to 1. As shown in Figure 6, the classification performance varies with the parameter δ, which means that parameter δ is valid in our method and the selection of parameter δ is very important for final classification performance.

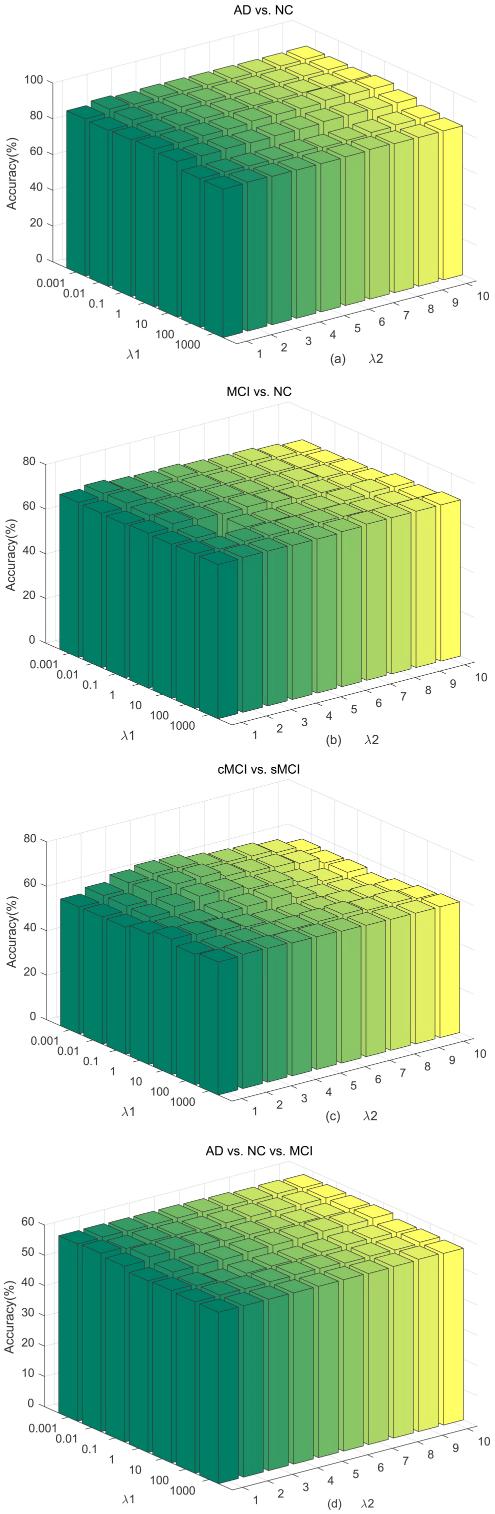

FIGURE 5. **The classification accuracy with regularization parameters λ_1 and λ_2.**

Specifically, when $\delta = 1$, for AD vs. NC, our method achieves the best classification accuracy of 90.40% which indicates that it is relatively easy to distinguish AD from NC, and thus there is not need to adjust the samples distance. For MCI vs. NC, when $\delta = 0.9$, our method achieves the best

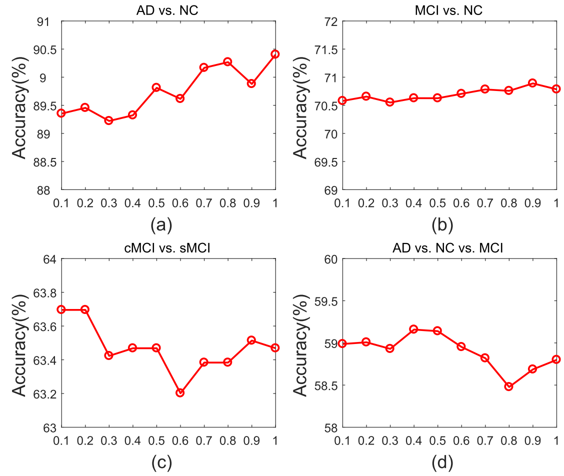

FIGURE 6. **The classification accuracy with the adjustable parameter δ. (a) Adjustable parameter δ. (b) Adjustable parameter δ. (c) Adjustable parameter δ. (d) Adjustable parameter δ.**

classification accuracy of 70.89% and 62.23%, respectively, which denotes that it is not very easy to distinguish MCI and NC, so there is need to slightly adjust the samples distance. For AD vs. NC vs. MCI, the best classification accuracy of 59.16% is achieved when $\delta = 0.4$. Since it is relatively difficult to distinguish MCI, NC and AD, there is need to relatively largely adjust the samples distance. However, for cMCI vs. sMCI, the best classification accuracy of 63.69% is achieved when $\delta = 0.1$, which denotes that it is difficult to distinguish cMCI from sMCI, and it's necessary that the distance of two samples within the same class is reduced on a larger scale and vice versa. This verified the conclusion that, for different classification tasks, the optimal neighbors are effectively selected via calculating the relative distance adjusted by label information, which contribute to the selection of more discriminative features for further improving the subsequent classification.

VI. LIMITATIONS

There are still some limitations in our study that should be considered in future studies. First, the optimal parameters are data dependent. How to obtain the optimal parameter automatically is still an open problem. In the future, the method of parameter optimization should be exploited to set the optimal parameter. Second, we only use cortical thickness for AD/MCI classification. However, there exist other structural features (e.g., volume and area) may also contain commentary information that can help to improve the classification performance. Finally, the personal network should be further constructed, and the characteristics of network may describe structural changes in a more accurate way.

VII. CONCLUSION

In this paper, we propose a novel manifold regularized sparse feature learning method for MCI/AD Classification based on structural magnetic resonance imaging. We first used a $\ell_{2,1}$-norm regularization to jointly select common features among the samples of different classes. We then constructed a new manifold regularization with class label information to preserve the relationship among the optimal neighbors. Experimental results demonstrate that our proposed method can achieve preferable classification performance compared with several state-of-the-art methods for AD/MCI classification.

ACKNOWLEDGMENT

For this project, data collection and sharing were provided 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).

REFERENCES

[1] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's Dementia*, vol. 3, no. 3, pp. 186–191, 2007.

[2] J. Bischkopf, A. Busse, and M. C. Angermeyer, "Mild cognitive impairment—A review of prevalence, incidence and outcome according to current approaches," *Acta Psychiatrica Scandinavica*, vol. 106, no. 6, pp. 403–414, 2010.

[3] C. Misra, Y. Fan, and C. Davatzikos, "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI," *NeuroImage*, vol. 44, no. 4, pp. 1415–1422, 2009.

[4] M. Grundman *et al.*, "Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials," *Arch. Neurol.*, vol. 61, no. 1, pp. 59–66, 2004.

[5] J. C. Morris and J. Cummings, "Mild cognitive impairment (MCI) represents early-stage Alzheimer's disease," *J. Alzheimers Disease*, vol. 7, no. 3, pp. 235–239, 2005.

[6] R. C. Petersen *et al.*, "Current concepts in mild cognitive impairment," *Arch. Neurol.*, vol. 58, no. 12, pp. 1985–1992, 2001.

[7] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *Neuroimage*, vol. 39, no. 4, pp. 1731–1743, 2008.

[8] L. Fang *et al.*, "Topological organization of metabolic brain networks in pre-chemotherapy cancer with depression: A resting-state PET study," *PLoS ONE*, vol. 11, nos. 1–11, p. e0166049, 2016.

[9] Z. Yao *et al.*, "Resting-state time-varying analysis reveals aberrant variations of functional connectivity in autism," *Frontiers Hum. Neurosci.*, vol. 10, no. 13, p. 463, 2016.

[10] L. Wan, G. Han, L. Shu, and N. Feng, "The critical patients localization algorithm using sparse representation for mixed signals in emergency healthcare system," *IEEE Syst. J.*, vol. 12, no. 1, pp. 52–63, Mar. 2018.

[11] Z. Ning, X. Wang, J. Rodrigues, and F. Xia, "Joint computation offloading,power allocation, and channel assignment for 5G-enabled traffic management systems," *IEEE Trans. Ind. Informat.*. doi: 10.1109/TII.2019.2892767.

[12] L. Wan, G. Han, L. Shu, S. Chan, and N. Feng, "PD source diagnosis and localization in industrial high-voltage insulation system via multimodal joint sparse representation," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2506–2516, Apr. 2016.

[13] L. Wan, G. Han, L. Shu, S. Chan, and T. Zhu, "The application of DOA estimation approach in patient tracking systems with high patient density," *IEEE Trans. Ind. Informat.*, vol. 12, no. 6, pp. 2353–2364, Dec. 2016.

[14] L. Wan, X. Kong, and F. Xia, "Joint range-Doppler-angle estimation for intelligent tracking of moving aerial targets," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1625–1636, Jun. 2017, doi: 10.1109/JIOT.2017.2787785.

[15] C.-Y. Wee, P.-T. Yap, and D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," *Hum. Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, 2013.

[16] Y. Zhang, S. Wang, P. Phillips, Z. Dong, G. Ji, and J. Yang, "Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOT-VAC," *Biomed. Signal Process. Control*, vol. 21, pp. 58–73, Aug. 2015.

[17] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, May 2011.

[18] E. Gerardin *et al.*, "Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging," *NeuroImage*, vol. 47, no. 4, pp. 1476–1486, 2009.

[19] J. Zhang *et al.*, "Joint resource allocation for latency-sensitive services over mobile edge computing networks with caching," *IEEE Internet Things J.*, to be published.

[20] X. Wang *et al.*, "Optimizing content dissemination for real-time traffic management in large-scale Internet of vehicle systems," *IEEE Trans.Veh. Technol.*, to be published. doi: 10.1109/TVT.2018.2886010.

[21] Z. Ning, X. Wang, and J. Huang, "Non-orthogonal multiple access for mobile edge computing enabled vehicular networks," *IEEE Veh. Technol. Mag.*, to be published. doi: 10.1109/MVT.2018.2882873.

[22] X. Wang *et al.*, "Privacy-preserving content dissemination for vehicular social networks: Challenges and solutions," *IEEE Commun. Surveys Tuts.*, to be publishe. doi: 10.1109/COMST.2018.2882064.

[23] X. Wang *et al.*, "A city-wide real-time traffic management system: Enabling crowdsensing in social Internet of vehicles," *IEEE Commun. Mag.*, 56, vol. 9, pp. 19–25, Sep. 2018.

[24] Z. Ning, P. Dong, X. Kong, and F. Xia, "A cooperative partial computation off loading scheme for mobile edge computing enabled Internet of Things," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2868616.

[25] X. Wang *et al.*, "A privacy-preserving message forwarding framework for opportunistic cloud of things," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5281–5295, Dec. 2018, doi: 10.1109/JIOT.2018.2864782.

[26] W. Zheng, Z. Yao, B. Hu, X. Gao, H. Cai, and P. Moore, "Novel cortical thickness pattern for accurate detection of Alzheimer's disease," *J. Alzheimers Disease*, vol. 48, no. 4, pp. 995–1008, 2015.

[27] W. Zheng, Z. Yao, Y. Xie, J. Fan, and B. Hu, "Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features," *Biol. Psychiatry, Cogn. Neurosci. Neuroimaging*, vol. 3, no. 10, pp. 887–897, 2018.

[28] X. Zhu, Z. Huang, H. T. Shen, J. Cheng, and C. Xu, "Dimensionality reduction by mixed kernel canonical correlation analysis," *Pattern Recognit.*, vol. 45, no. 8, pp. 3003–3016, 2012.

[29] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proc. AAAI*, 1992, pp. 129–134.

[30] H. Liu and R. Setiono, "Feature selection and classification—A probabilistic wrapper approach," in *Proc. 9th Int. Conf. Ind. Eng. Appl. AI ES*, 1997, pp. 419–424.

[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B, Methodol.*, vol. 73, no. 3, pp. 267–288, 1996.

[32] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[33] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.

[34] Z. Yao, B. Hu, H. Nan, W. Zheng, and Y. Xie, "Individual metabolic network for the accurate detection of Alzheimer's disease based on FDGPET imaging," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 1328–1335.

[35] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–451, 2004.

[36] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc.*, vol. 67, no. 2, pp. 301–320, 2005.

[37] J. Huo, Y. Gao, Y. Shi, W. Yang, and H. Yin, "Ensemble of sparse cross-modal metrics for heterogeneous face recognition," in *Proc. 24th ACM Int. Conf Multimedia* , 2016, pp. 1405–1414.

[38] J. Krapac, M. Allan, J. Verbeek, and F. Juried, "Improving Web image search results using query-relative classifiers," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1094–1101.

[39] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized Multiview Analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2160–2167.

[40] H. Zhang and J. Lu, "Creating ensembles of classifiers via fuzzy clustering and deflection," *Fuzzy Sets Syst.*, vol. 161, no. 13, pp. 1790–1802, 2010.

[41] X. Zhu, Z. Huang, H. Cheng, J. Cui, and H. T. Shen, "Sparse hashing for fast multimedia search," *ACM Trans. Inf. Syst.*, vol. 31, no. 2, 2013, Art. no. 9.
[42] M. Liu, D. Zhang, E. Adeli-Mosabbeb, and D. Shen, "Relationship induced multi-atlas learning for Alzheimer's disease diagnosis," in *Medical Computer Vision: Algorithms for Big Data*. Cham, Switzerland: Springer, 2015.
[43] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.
[44] X. Zhu, H.-I. Suk, and D. Shen, "Sparse discriminative feature selection for multi-class Alzheimer's disease classification," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, 2014.
[45] B. Jie, D. Zhang, B. Cheng, and D. Shen, "Manifold regularized multitask feature learning for multimodality disease classification," *Hum. Brain Mapping*, vol. 36, no. 2, pp. 489–507, 2015.
[46] C. Zu, B. Jie, M. Liu, S. Chen, D. Shen, and D. Zhang, "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," *Brain Imag. Behav.*, vol. 10, no. 4, pp. 1148–1159, 2016.
[47] T. Ye, Z. Chen, B. Jie, D. Shen, and D. Zhang, "Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease," *Brain Imag. Behav.*, vol. 10, no. 3, pp. 739–749, 2015.
[48] V. N. Vapnik, "The nature of statistical learning theory," *Technometrics*, vol. 38, no. 4, p. 409, 2000.
[49] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis: I. Segmentation and surface reconstruction," *NeuroImage*, vol. 9, no. 2, pp. 179–194, 1999.
[50] B. Fischl, A. Liu, and A. M. Dale, "Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 70–80, Jan. 2001.
[51] B. Fischl, M. I. Sereno, and A. M. Dale, "Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system," *NeuroImage*, vol. 9, no. 2, pp. 195–207, 1999.
[52] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.
[53] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., B, Statist. Methodol.*, vol. 68, no. 1, pp. 49–67, 2006.
[54] W. S. Zheng, L. Wang, T. Tan, and R. He, "$\ell_{2,1}$ regularized correntropy for robust feature selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, vol. 157, no. 10, pp. 2504–2511.
[55] M. Nikolova, and M. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, 2005.
[56] K. Wang, R. He, L. Wang, W. Wang, and T. Tan, "Joint feature selection and subspace learning for cross-modal retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2010–2023, Oct. 2016.
[57] X. Zhu, X. Wu, W. Ding, and S. Zhang, "Feature selection by joint graph sparse coding," in *Proc. SIAM Int. Conf. Data Mining*, 2013, pp. 803–811.
[58] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017.
[59] A. Convit, J. De Asis, M. J. de Leon, C. Y. Tarshish, S. De Santi, and H. Rusinek, "Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease," *Neurobiol. Aging*, vol. 21, no. 1, pp. 19–26, 2000.
[60] S. Derflinger *et al.*, "Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized," *J. Alzheimer's Disease*, vol. 25, no. 2, pp. 347–357, 2011.
[61] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of clinical scores in Alzheimer's disease," in *Multimodal Brain Image Analysis*. Berlin, Germany: Springer, 2011.
[62] X. Zhu, H.-I. Suk, and D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *NeuroImage*, vol. 100, pp. 91–105, Oct. 2014.
[63] G. Chételat *et al.*, "FDG-PET measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment," *Neurocase*, vol. 11, no. 1, pp. 14–25, 2005.
[64] N. C. Fox and J. M. Schott, "Imaging cerebral atrophy: Normal ageing to Alzheimer's disease," *Lancet*, vol. 363, no. 9406, pp. 392–394, Jan. 2004.

LINA XU received the B.E. degree in computer science and technology from the Shandong University of Technology, in 2000, and the M.E. degree in computer application technology from Northwest Minzu University, in 2006. She is currently pursuing the Ph.D. degree with Shandong Normal University. She is also a Teacher with the School of Computer Science and Technology, Shandong Jianzhu University. Her research interests include machine learning, pattern recognition, and biomedical engineering.

ZHIJUN YAO received the B.Sc. degree from the Electronic Information Science Department, Lanzhou University, in 1999, and the M.Sc. and Ph.D. degrees from the School of Information Science and Engineering, Lanzhou University, in 2002 and 2011, respectively, where he is currently a Professor and also a Master Instructor. His research interests include biomedical engineering, cognitive neuroscience, and neuroimage analysis.

JING LI received the B.E. degree in electronic information engineering and technology, and the M.E. degree in communication and information systems both from Shandong University, China. She is currently a Teacher with the School of Shandong Management University, China. Her research papers have been published in several international journals, international conferences, and national journals. Her research interests include machine learning and multimedia information processing.

CHEN LV received the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2015. His research interests include bioinformatics, data mining, and software engineering. He is currently a Teacher with the School of Information Science and Engineering, Shandong Normal University, China.

HUAXIANG ZHANG received the Ph.D. degree from Shanghai Jiao Tong University, in 2004. He was an Associate Professor with the Department of Computer Science, Shandong Normal University, China, from 2004 to 2005. He is currently a Professor with the School of Information Science and Engineering, Institute of Data Science and Technology, Shandong Normal University. He has authored or co-authored over 170 journal and conference papers and holds ten invention patents. His current research interests include machine learning, pattern recognition, evolutionary computation, cross-media retrieval, web information processing, and bioinformatics.

BIN HU received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science. He is currently a Professor, and also the Dean of the School of Information Science and Engineering, Lanzhou University, China. His research interests include computational psychophysiology, pervasive computing, and mental health care. He is an IET Fellow, the Member-at-Large of the ACM China, the Vice Chair of the International Society for Social Neuroscience (China Committee), and a member of the IET Healthcare Technology Network. He is the Chair of the IEEE SMC TC on Computational Psychophysiology, the Chair Professor of the National Recruitment Program of Global Experts, the Chief Scientist of the National Fundamental Research Program of China (973 Program), and a member of the Division of Computer Science Review Panel, Natural Science Foundation China.