

## Article

# Multi-Modal Feature Selection with Feature Correlation and Feature Structure Fusion for MCI and AD Classification

Zhuqing Jiao <sup>1</sup>, Siwei Chen <sup>1</sup>, Haifeng Shi <sup>2,3,\*</sup> and Jia Xu <sup>4,\*</sup>

<sup>1</sup> School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China; jzq@cczu.edu.cn (Z.J.); chensiwei1223@163.com (S.C.)

<sup>2</sup> Department of Radiology, Changzhou Second People's Hospital, Nanjing Medical University, Changzhou 213003, China

<sup>3</sup> School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213164, China

<sup>4</sup> School of Medicine, Ningbo University, Ningbo 315211, China

\* Correspondence: 18114330152@163.com (H.S.); xujia@nbu.edu.cn (J.X.)

**Abstract:** Feature selection for multiple types of data has been widely applied in mild cognitive impairment (MCI) and Alzheimer's disease (AD) classification research. Combining multi-modal data for classification can better realize the complementarity of valuable information. In order to improve the classification performance of feature selection on multi-modal data, we propose a multi-modal feature selection algorithm using feature correlation and feature structure fusion (FC2FS). First, we construct feature correlation regularization by fusing a similarity matrix between multi-modal feature nodes. Then, based on manifold learning, we employ feature matrix fusion to construct feature structure regularization, and learn the local geometric structure of the feature nodes. Finally, the two regularizations are embedded in a multi-task learning model that introduces low-rank constraint, the multi-modal features are selected, and the final features are linearly fused and input into a support vector machine (SVM) for classification. Different controlled experiments were set to verify the validity of the proposed method, which was applied to MCI and AD classification. The accuracy of normal controls versus Alzheimer's disease, normal controls versus late mild cognitive impairment, normal controls versus early mild cognitive impairment, and early mild cognitive impairment versus late mild cognitive impairment achieve  $91.85 \pm 1.42\%$ ,  $85.33 \pm 2.22\%$ ,  $78.29 \pm 2.20\%$ , and  $77.67 \pm 1.65\%$ , respectively. This method makes up for the shortcomings of the traditional multi-modal feature selection based on subjects and fully considers the relationship between feature nodes and the local geometric structure of feature space. Our study not only enhances the interpretation of feature selection but also improves the classification performance, which has certain reference values for the identification of MCI and AD.

**Keywords:** feature correlation; feature structure fusion; multi-modal; classification; feature selection



**Citation:** Jiao, Z.; Chen, S.; Shi, H.; Xu, J. Multi-Modal Feature Selection with Feature Correlation and Feature Structure Fusion for MCI and AD Classification. *Brain Sci.* **2022**, *12*, 80. <https://doi.org/10.3390/brainsci12010080>

Academic Editors: Elena Tamagno and Luigi De Gennaro

Received: 14 November 2021

Accepted: 29 December 2021

Published: 5 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Alzheimer's Disease (AD) is a neurological disorder associated with memory and mobility impairment and resulting in loss of cognitive function. With the aging of society, more and more elderly people are facing this disease. Studies have shown that the prevalence of AD in developing countries is much higher than that in developed countries [1]. Early mild cognitive impairment (EMCI) and late mild cognitive impairment (LMCI) is an intermediate state between healthy normal people and Alzheimer's patients, and MCI gradually develops into AD with the development of the disease. Thus, determining how to accurately classify MCI and AD is of great significance.

In daily diagnosis, we can obtain massive amounts of medical image data with different structures and types. It helps us to observe the same subject from different perspectives and strengthen the understanding of the disease pathogenic factors. Traditional single-modal data only start from medical image data and observe the subjects from a single

perspective. Obviously, the information complementarity between different modal data is ignored. This is bound to result in the acquired features not being comprehensive enough, affecting the final classification results. By observing the subjects with multi-modal data, we can understand the pathogenic factors of the disease more comprehensively. For example, Zhang et al. [2] combined Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Cerebrospinal Fluid (CSF) data for feature selection. Li et al. [3] adopted two imaging techniques, Arterial Spin Labeling (ASL) and Blood Oxygen Level-Dependent Functional Magnetic Resonance Imaging (BOLD-fMRI), to conduct MCI classification and select features with good characterization. The results show that the classification effect of these two studies is better than that of single-modal data. Structural Magnetic Resonance Imaging (sMRI) and PET have been widely adopted in multi-modal feature selection [4–7]. These two modes can simultaneously obtain the structural and functional features of the brain, which can enhance the ability of feature description and facilitate feature expression.

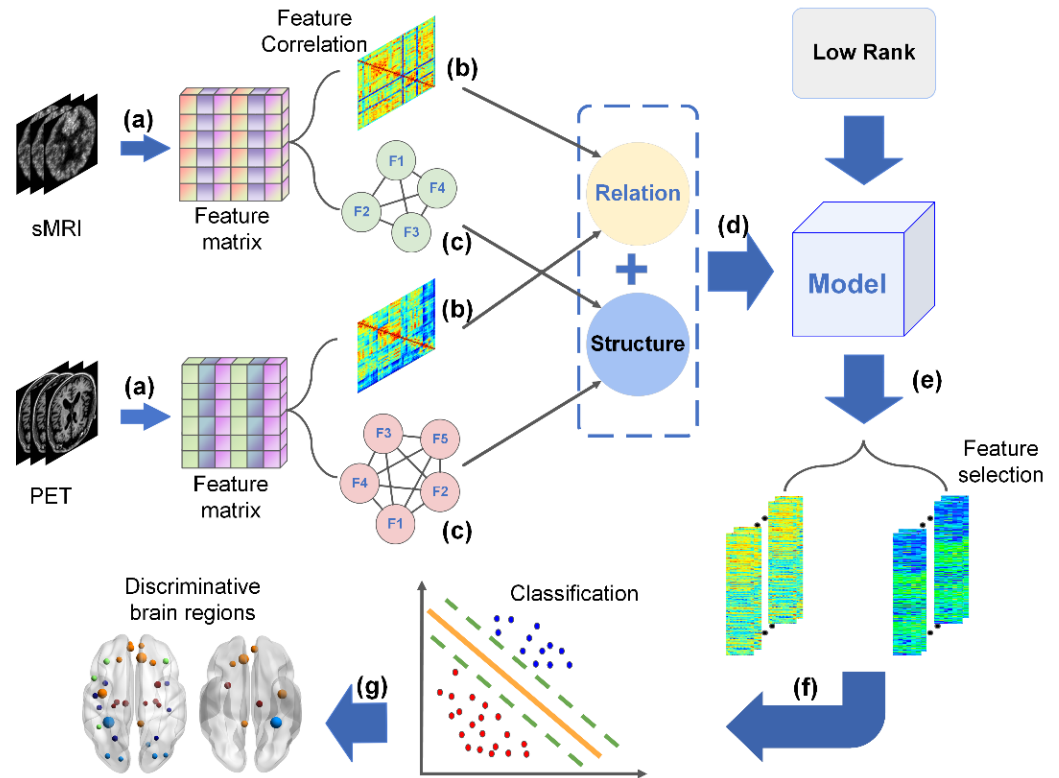
For MCI and AD classification, the most important point is to carry out joint feature selection for the features extracted from multiple modal data. It is essential to screen out the features associated with the disease and improve the classification performance while reducing the feature dimension. In machine learning, feature selection algorithms can be roughly divided into filtering [8], wrapping [9], and embedded [10,11]. Embedded feature selection, which is widely applied, combines the learner with the feature selection process, and automatically completes feature selection when the learner learns. Regularization techniques are often applied to embedded feature selection algorithms. For example, the Lasso algorithm [12] uses  $L_1$ -norm regularizer to achieve feature selection effect with sparse feature weight vectors. Among the existing embedded feature selection algorithms, multi-task learning is often used for feature selection related to disease [13–16]. Its advantages are that it can reveal the potential common characteristics between different tasks, carry out information sharing between tasks, and has good generalization. For example, Jie et al. [17] obtained manifold structures of different modal data by combining manifold learning and multi-task learning, effectively combining information complementarity among multi-modal data; Lei et al. [15] adopted a new regularization to reduce rank relaxation based on multi-task learning, which can better carry out feature selection and reduce redundant features. In recent studies, Shao et al. [18] introduced hypergraph learning derived from multi-task learning and proposed a feature selection algorithm based on hypergraph to reflect the high-order relationship between subjects through the hypergraph Laplacian matrix.

However, the above methods only consider the potential relationship between subjects in different modalities or in the same modality, and do not satisfactorily consider the internal relationship between different modalities or different features in the same modality. Therefore, we propose a multi-modal feature selection with feature correlation and feature structure fusion that applies to MCI and AD classification. First, features are extracted from sMRI and PET data, and the correlation coefficient matrix between different modal features is converted into feature correlation regularization by weighted sum; Then, based on manifold learning, the feature matrix is fused, and feature structure regularization is constructed. A low-rank constraint is added based on the multi-task learning model, and two regularizations are embedded into the improved model to obtain the final feature selection model. The multi-modal features are selected by the proposed model, and the selected features are linearly fused into a support vector machine (SVM) for classification, and the final classification results are obtained. Then, the effects of different feature correlation calculation methods, the fusion coefficients of the feature matrix, and different regularized weight coefficients on classification performance are discussed. Finally, the brain regions corresponding to the selected features are analyzed to find the discriminative brain regions affected by MCI and AD diseases respectively.

## 2. Materials and Methods

### 2.1. Research Framework

Our research framework which is shown in Figure 1 mainly includes the following steps.



**Figure 1.** The research framework. (a) Original sMRI and PET images were preprocessed, then regions of interest were extracted by AAL template as sMRI and PET features, and the corresponding feature matrices of sMRI and PET were obtained, respectively. (b) Correlation coefficients between feature nodes of each modal data were calculated to obtain the feature correlation matrix, and the feature correlation regularization was obtained by linear fusion. (c) Feature matrix was weighted fusion, then the adjacent nodes were calculated to obtain the adjacency matrix, and the feature graph Laplacian matrix was constructed according to the cosine distance method to obtain the feature structure regularization. (d) The two regularizations were embedded into the multi-task model with low-rank constraint for feature selection. (e) Feature vectors with good characterization were selected by the proposed model, standardized respectively, and the features extracted from multi-modal data were linearly fused to obtain a new fused feature matrix. (f) The test set and training set were divided from the fused feature matrix, the training set was trained by the 10-fold cross-validation method and SVM to obtain the classification model, and the classification performance of the model was verified by the test set; (g) The corresponding discriminative brain regions of the selected feature nodes were visualized to analyze the discriminative brain regions affected by MCI and AD respectively.

### 2.2. Data Acquisition and Preprocessing

The data were collected from Alzheimer’s Disease Neuroimaging Initiative (ADNI), which focuses on the prediction and diagnosis of AD. ADNI was approved by the Institutional Review Boards (IRBs), and all subjects were reviewed and approved by the IRBs within the ADNI study, meeting all ethical standards for data collection. Our study included structural magnetic resonance imaging (sMRI) and positron emission tomography (PET) imaging data of 73 normal controls (NCs), 53 EMCI subjects, 49 LMCI subjects, and 69 AD subjects. The specific information of subjects is shown in Table 1.

**Table 1.** The subject's information.

Characteristic	Normal	EMCI	LMCI	AD
Number	73	53	49	69
Male/Female	39/34	23/30	22/27	36/33
Age (mean $\pm$ SD)	75.9 $\pm$ 6.79	72.9 $\pm$ 7.6	73.1 $\pm$ 8.23	72.54 $\pm$ 6.12
MMSE (mean $\pm$ SD)	28.14 $\pm$ 1.21	27.1 $\pm$ 1.62	26.3 $\pm$ 2.1	22.54 $\pm$ 2.16

There are different inclusion and exclusion criteria for the four categories of subjects: For normal controls, cognition must be normal, and without memory impairment, the Mini-Mental State Exam (MMSE) score should be between 24 and 30, Clinical Dementia Rating (CDR) and Memory Box Score should be 0, while any normal controls with significant neurologic disease must be excluded. EMCI subjects must have subjective memory problems, MMSE score between 24 and 30, CDR and Memory Box score must both be 0.5; Subjects with any significant neurological disease other than suspected early Alzheimer's disease need to be excluded. For LMCI subjects, the inclusion criteria were consistent with EMCI, and the criteria to distinguish EMCI and LMCI were determined by Wechsler Memory Scale. Exclusion criteria are consistent with EMCI. Patients with AD must have subjective memory problems, MMSE score between 20 and 26, and CDR must be 0.5 or 1. Probable AD needs to meet the NINCDS/ADRDA criteria [19]. Subjects with other neurological disorders besides Alzheimer's had to be excluded.

SPM12 software [20] was used to preprocess sMRI and PET original images with voxel-based morphometric (VBM) analysis methods. For sMRI data, spatial standardization was first carried out. The MNI152 standard brain template was used to map the same region of each original image to the template region one by one, which helps to eliminate the brain differences caused by individual factors. Then, the image was segmented into gray matter, white matter, and cerebrospinal fluid, and the noise was eliminated by a smoothing operation. Finally, the AAL template [21] was used to extract the average gray matter density of the brain region of interest (ROI) as the sMRI data features. For PET data, the realignment of the images was carried out first. The images were coregistered onto the MNI152 brain space [21,22] for normalization and smoothing operation, and the width of 8 mm of a Gaussian filter was adopted. Finally, the glucose metabolism of the cerebral regions of interest was extracted using AAL template as PET data features.

### 2.3. Joint Feature Learning with Low-Rank Constraint

The multi-task learning model has been widely applied into multi-modal feature selection. Its main advantage is that it can mine deep common data features among different tasks and realize information sharing among multiple modal data [23].  $L_{2,1}$ -norm regularizer can minimize the loss function while making weight vectors as sparse as possible, while selecting feature vectors with representation. Previous studies have shown that low-rank constraint can also find shared information well [24] and can measure the similarity between matrix row vectors. Therefore, low-rank constraint is introduced to capture the potential relationship between different task features. It promises to improve the information sharing between different tasks in the multi-task model and improve the model generalization performance. The following model is established:

$$\min_{w_1, w_2, \dots, w_m} \sum_{i=1}^m \|Y_i - X_i w_i\|_2^2 + \gamma \text{rank}(W) \quad (1)$$

where  $X_i = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^{N \times p}$  is the feature matrix of the  $i$ -th modality,  $N$  represents the number of subjects,  $p$  represents the number of features, which is also the number of regions of interest;  $Y_i = [y_1, y_2, \dots, y_N]^T \in \mathbb{R}^{N \times 1}$  represents the number of subjects labels in the  $i$ -th modality,  $W = [w_1, w_2, \dots, w_m] \in \mathbb{R}^{p \times m}$  is the feature weight matrix, and each element in  $w_i$  represents the corresponding feature weight value in the  $i$ -th modality,  $m$  is

the number of the modality;  $rank(\cdot)$  represents low-rank constraint, and  $\gamma$  is the low-rank constraint coefficient.

In fact, the low-rank constraint for a matrix is nonconvex and it is a typical NP-hard problem. It has been proved that trace norm can be used to approximate low-rank constraint [25,26]. Finally, the loss function of multi-modal feature learning based on low-rank constraint is obtained, as shown in Equation (2):

$$\min_{w_1, w_2, \dots, w_m} \sum_{i=1}^m \|Y_i - X_i w_i\|_2^2 + \gamma \|W\|_* \quad (2)$$

where  $\|\cdot\|_*$  represents trace norm of the matrix,  $\|W\|_* = \sum_i \lambda_i$  is the sum of all singular values of the matrix  $W$ .

#### 2.4. Feature Correlation and Feature Structure Regularization

In multi-modal data, features are often related to each other [27]. Feature selection is to select a feature from several highly correlated features, when one feature is restricted, which will inevitably lead to the selection of highly correlated features [28]. Therefore, we consider the correlation of features between different modalities, the weighted average of the feature correlation matrix of various modalities. Finally, we propose a new feature correlation regularization, as shown in Equation (3):

$$tr(W^T \sum_{i=1}^m R_i W) \quad (3)$$

where  $R_i$  is the correlation coefficient matrix of the  $i$ -th modality, and  $tr(\cdot)$  represents the trace of the matrix.

The common calculation methods of correlation coefficient include the Pearson correlation coefficient, the Spearman correlation coefficient, and the Kendall correlation coefficient. The Pearson correlation coefficient can measure the linear correlation of two variables and its value lies between  $-1$  and  $1$ . The Spearman correlation coefficient and Kendall correlation coefficient, compared with the Pearson correlation coefficient, have more relaxed requirements for data and wider application scope [29,30].

Furthermore, when the distance between two feature vectors is close in space, the distance between their corresponding weight vectors should also be close. Inspired by manifold learning and feature fusion [17,31–33], we use the weighted fusion multi-modal feature matrix to construct the Laplacian matrix to preserve the local geometric structure of features, so we have the following feature structure regularization:

$$\begin{aligned} & \frac{1}{2} \sum_{j,k} h_{jk} \|W_{j\cdot} - W_{k\cdot}\|_2^2 \\ & = tr(W^T (S - H) W) \\ & = tr(W^T L_F W) \end{aligned} \quad (4)$$

where  $W_{j\cdot}$  and  $W_{k\cdot}$  represent the  $j$ -th row and  $k$ -th row vectors of the weight matrix respectively,  $H \in \mathbb{R}^{p \times p}$  represents an adjacency matrix of the features.  $S \in \mathbb{R}^{p \times p}$  is a degree and diagonal matrix, the principal diagonal element is the degree of each feature node in the adjacency matrix  $H$ , and the calculation equation is  $S_{ii} = \sum_{i=1}^p H_{i\cdot}$ .  $L_F \in \mathbb{R}^{p \times p}$  represents the Laplacian matrix computed after the fusion of the feature matrices, and  $L_F = S - H$ .

For the adjacency matrix  $\mathbf{H}$ , there are three construction methods: 0–1 weighting, heat-kernel function, and cosine distance [34]. We adopt the cosine distance method to construct, and its calculation equation is as follows:

$$h_{ij} = \begin{cases} \frac{\mathbf{X}_i^T \mathbf{X}_j}{\|\mathbf{X}_i\| \|\mathbf{X}_j\|}, & \text{if } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $h_{ij}$  is the  $i$ -th row and  $j$ -th column element of the adjacency matrix  $\mathbf{H}$ , which is used to measure the similarity between  $i$ -th and  $j$ -th columns of the feature vectors in the feature matrix  $\mathbf{X}$ .  $\mathbf{X}_i$  and  $\mathbf{X}_j$  represent the feature vectors of the  $i$ -th and  $j$ -th columns of the feature matrix  $\mathbf{X}$ , respectively.

### 2.5. Multi-Modal Feature Selection

In this work, an improved feature selection algorithm is proposed. Based on the multi-task learning model, trace norm is introduced to improve information sharing between different modalities, and the feature correlation regularization and feature structure regularization proposed above are introduced. The potential correlation between features is learned and the local geometric structure of features is preserved while minimizing the loss function, in order to improve the generalization ability of the model. Finally, we obtain the final established loss function:

$$\min_{w_1, w_2, \dots, w_m} \sum_{i=1}^m \|\mathbf{Y}_i - \mathbf{X}_i \mathbf{w}_i\|_2^2 + \alpha \text{tr}(\mathbf{W}^T \sum_{i=1}^m \mathbf{R}_i \mathbf{W}) + \beta \text{tr}(\mathbf{W}^T \mathbf{L}_F \mathbf{W}) + \gamma \|\mathbf{W}\|_* \quad (6)$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are regularization parameters, and they are all real numbers greater than zero. The loss function is divided into four parts, the first is empirical error, the second is feature correlation regularization, the third is feature structure regularization, and the fourth is trace norm.

The solution of the objective loss function of Equation (6) is a convex optimization problem. Combining with the existing optimization algorithm [35,36], an optimization algorithm is proposed to solve this problem. First, the loss function is divided into a convex and non-convex function. The trace norm is a non-convex regularization term, and the remainder is convex terms. Let  $\varphi(\mathbf{W}) = \eta(\mathbf{W}) + \gamma \|\mathbf{W}\|_*$ , where  $\eta(\mathbf{W})$  is the differentiable part, and the original loss function can be rewritten as:

$$\min_{w_1, w_2, \dots, w_m} \eta(\mathbf{W}) + \gamma \|\mathbf{W}\|_* \quad (7)$$

For any given  $\mathbf{W}_{k-1}$ , consider the second-order approximate form of  $\varphi(\mathbf{W})$  at  $\mathbf{W}_{k-1}$ , and we obtain:

$$\begin{aligned} \varphi(\mathbf{W}) &\approx Q(\mathbf{W}, \mathbf{W}_{k-1}) \\ &= \eta(\mathbf{W}_{k-1}) + \langle \mathbf{W} - \mathbf{W}_{k-1}, \nabla \eta(\mathbf{W}_{k-1}) \rangle + \frac{s}{2} \|\mathbf{W} - \mathbf{W}_{k-1}\|_F^2 + \gamma \|\mathbf{W}\|_* \end{aligned} \quad (8)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product,  $\|\cdot\|_F$  is the Frobenius norm of the matrix,  $\nabla \eta(\mathbf{W}_{k-1})$  represents the derivative of the differentiable function  $\eta(\cdot)$  at  $\mathbf{W}_{k-1}$ , and the iterative updating equation of the weight matrix  $\mathbf{W}$  is further obtained:

$$\mathbf{W}_k = \text{prox}_{\gamma \|\cdot\|_*} \left( \mathbf{W}_{k-1} - \frac{1}{s} \nabla \eta(\mathbf{W}_{k-1}) \right) \quad (9)$$

where  $s$  is the step length, and the calculation of the proximal operator  $\text{prox}_{\gamma \|\cdot\|_*}(\cdot)$  is shown in Equation (10):

$$\text{prox}_{\gamma \|\cdot\|_*}(\mathbf{Z}) = \frac{1}{2} \|\mathbf{W}_k - \mathbf{Z}\|_F^2 + \gamma \|\mathbf{W}\|_* \quad (10)$$



For the solution of Equation (10), according to the conclusions of existing studies [37,38], it can be computed by singular value decomposition (SVD) of  $\mathbf{W}_{k-1} - \frac{1}{s}\nabla\eta(\mathbf{W}_{k-1})$ , as shown in Equation (11):

$$\text{prox}_{\gamma\|\cdot\|_*}(\mathbf{Z}) = \mathbf{U}\mathbf{D}_\gamma\mathbf{V}^T \quad (11)$$

where  $\mathbf{U}\mathbf{D}_\gamma\mathbf{V}^T$  is the SVD of  $\mathbf{Z}$ ,  $\mathbf{D}_\gamma$  is a diagonal matrix, and the diagonal element of  $\mathbf{D}_\gamma$  is  $(\mathbf{D}_\gamma)_{ii} = \max\{\mathbf{D}_{ii} - \gamma, 0\}$ . Equation (11) can make the weight matrix  $\mathbf{W}$  become low-rank while shrinking the singular values.

In the above optimization algorithm, the time complexity of this algorithm can still be achieved  $O(1/M)$ , where  $M$  is the maximum number of iterations of the algorithm, despite the existence of a non-differentiable trace norm approximate low-rank constraint. Meanwhile, we summarize the optimization algorithm flow of loss function to show the above iterative update process more clearly, as shown in Table 2.

**Table 2.** Optimization algorithm of loss function.

Line No.	Optimization Algorithm of Loss Function
1	Input: $\mathbf{X}_i \in \mathbb{R}^{N \times p}$ represents the feature matrix of the $i$ -th modality; $\mathbf{Y}_i \in \mathbb{R}^{N \times 1}$ represents the label corresponding to the $i$ -th modality subjects.
2	Output: $\mathbf{W} \in \mathbb{R}^{p \times m}$ represents the weight matrix of the feature.
3	Normalize feature matrix $\mathbf{X}_i$ and initialize $\mathbf{W}_0, s$ ;
4	Compute feature correlation matrix of $i$ -th modality, and weighted average;
5	Weighted fusion of the feature matrix, and compute feature Laplacian matrix $\mathbf{L}_F$ ;
6	Do
7	Compute SVD of $\mathbf{W}_{k-1} - \frac{1}{s}\nabla\eta(\mathbf{W}_{k-1})$ ;
8	Update $\mathbf{Z}_{k-1} = \mathbf{W}_{k-1} - \frac{1}{s}\nabla\eta(\mathbf{W}_{k-1})$ ;
9	Update $\mathbf{W}_k = \text{prox}_{\gamma\ \cdot\ _*}(\mathbf{Z}_{k-1})$ ;
10	Update step length $s$ ;
11	While it reaches the maximum number of iterations or converges

## 2.6. Classification and Evaluation Measures

SVM is suitable for binary classification with small subjects. It has good generalization ability, and can avoid dimensional disasters, and is often applied to disease classification [39–41]. In our study, the loss function of Equation (6) is used for feature selection, and the multi-modal features obtained are linearly fused. Then the features are input into the SVM to classify MCI and AD, and the performance of the model is estimated from different classification indexes.

This work mainly includes six indicators to evaluate the classification performance. The first four common classification indicators are accuracy (ACC), area under curve (AUC), sensitivity (SEN), specificity (SPE). Meanwhile, the geometric mean (GMean) and F1 Score (F1) are used to further measure classification performance to overcome the influence of different proportions of positive and negative subjects on the classification results.

Each indicator is defined as follows:

$$\text{ACC} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (12)$$

$$\text{AUC} = P(P_- < P_+) \quad (13)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (15)$$

$$\text{GMean} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}} \quad (16)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (17)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative respectively. Accuracy (ACC) represents the proportion of correctly classified subjects to all subjects, sensitivity (SEN) and specificity (SPE) describe the proportion of positive and negative subjects that are correctly classified, respectively, while area under the curve (AUC) describes the area of the ROC curve.

### 3. Results

#### 3.1. Classification Performance

In the experiment, four methods were selected for comparison and respectively applied to MCI and AD classification. The classification performance of each method is shown in Table 3. The baseline method is the widely used Lasso feature selection method [12], CMTL [42] is a multi-task learning method based on clustering, MTFS [2] is a multi-task feature selection method with  $L_{2,1}$ -norm regularizer and applied to AD, and HMTFS [18] is a multi-modal feature selection method based on MTFS that introduces hypergraph. In addition to being less sensitive in NC vs. LMCI classification, the proposed method has a better classification performance for NC vs. AD, NC vs. EMCI and EMCI vs. LMCI than previous methods. Compared to the four comparison methods, feature correlation regularization can better reflect the potential relationship between multiple features, while feature structure regularization can preserve the local geometric structure of features.

**Table 3.** Classification performance of different methods.

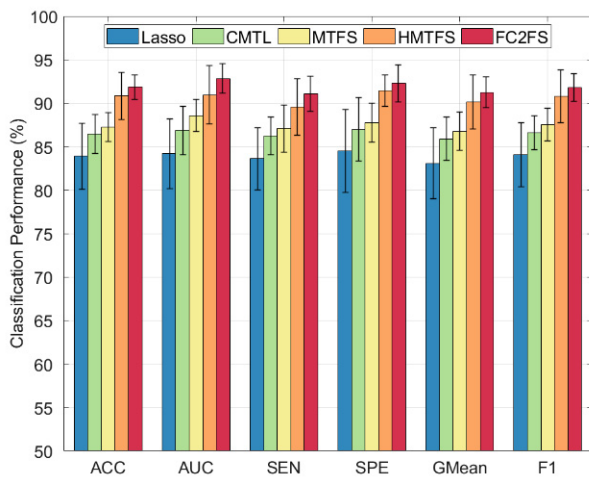
Task	Methods	ACC (%) ±STD	AUC (%) ±STD	SEN (%) ±STD	SPE (%) ±STD	GMean (%) ±STD	F1 (%) ±STD
NC vs. AD	Lasso	83.89 ± 3.80	84.18 ± 4.00	83.60 ± 3.55	84.50 ± 4.76	83.09 ± 4.06	84.08 ± 3.67
	CMTL	86.44 ± 2.26	86.86 ± 2.79	86.24 ± 2.14	86.99 ± 3.62	85.91 ± 2.51	86.59 ± 1.98
	MTFS	87.28 ± 1.66	88.58 ± 1.87	87.09 ± 2.70	87.76 ± 2.26	86.75 ± 2.21	87.52 ± 1.88
	HMTFS	90.83 ± 2.72	90.96 ± 3.38	89.54 ± 3.24	91.45 ± 1.80	90.14 ± 3.12	90.78 ± 3.01
	FC2FS	91.85 ± 1.42	92.84 ± 1.69	91.07 ± 2.02	92.27 ± 2.12	91.23 ± 1.77	91.81 ± 1.59
NC vs. LMCI	Lasso	78.81 ± 3.69	78.66 ± 5.09	80.24 ± 3.75	76.41 ± 4.71	76.85 ± 4.28	82.36 ± 2.97
	CMTL	80.45 ± 2.72	80.96 ± 3.09	82.38 ± 2.40	77.96 ± 4.87	78.75 ± 3.68	83.09 ± 2.53
	MTFS	81.10 ± 2.55	82.20 ± 3.84	83.14 ± 2.18	78.65 ± 3.89	79.52 ± 3.51	83.96 ± 2.49
	HMTFS	84.39 ± 2.30	84.96 ± 3.52	85.65 ± 2.07	82.42 ± 3.36	83.00 ± 3.02	87.09 ± 2.17
	FC2FS	85.33 ± 2.22	85.11 ± 1.86	85.38 ± 2.13	84.85 ± 2.73	83.91 ± 2.55	87.45 ± 1.59
NC vs. EMCI	Lasso	70.67 ± 2.47	70.81 ± 4.54	74.24 ± 1.91	66.47 ± 3.89	68.90 ± 2.96	74.18 ± 2.10
	CMTL	71.52 ± 2.48	70.06 ± 3.34	74.44 ± 3.07	68.44 ± 4.21	69.82 ± 3.56	75.30 ± 2.56
	MTFS	72.56 ± 3.55	71.86 ± 2.86	75.10 ± 3.11	68.89 ± 3.83	69.69 ± 3.30	75.80 ± 2.89
	HMTFS	74.90 ± 2.41	74.41 ± 1.68	79.77 ± 2.72	69.86 ± 3.11	73.37 ± 2.61	77.56 ± 2.32
	FC2FS	78.29 ± 2.20	78.03 ± 2.41	82.02 ± 2.13	74.73 ± 3.04	77.18 ± 2.58	81.00 ± 1.93
EMCI vs. LMCI	Lasso	71.58 ± 2.43	70.07 ± 3.62	72.41 ± 3.06	70.65 ± 2.11	70.09 ± 2.20	71.77 ± 1.76
	CMTL	73.08 ± 2.37	70.89 ± 3.81	73.71 ± 2.31	72.15 ± 3.32	70.95 ± 3.12	72.91 ± 2.23
	MTFS	73.54 ± 2.22	74.14 ± 4.84	75.45 ± 2.98	72.58 ± 3.29	72.48 ± 2.62	73.16 ± 2.49
	HMTFS	75.46 ± 3.12	74.18 ± 3.97	75.77 ± 3.36	74.74 ± 3.53	74.08 ± 3.78	75.46 ± 2.98
	FC2FS	77.67 ± 1.65	77.63 ± 3.17	77.72 ± 2.54	78.94 ± 2.47	76.83 ± 2.35	78.35 ± 1.75

In NC vs. AD classification, ACC, AUC, SEN, SPE, GMean and F1 reach  $91.85 \pm 1.42\%$ ,  $92.84 \pm 1.69\%$ ,  $91.07 \pm 2.02\%$ ,  $92.27 \pm 2.12\%$ ,  $91.23 \pm 1.77\%$ , and  $91.81 \pm 1.59\%$ , respectively. Compared with the other four methods, the classification performance is improved. It is worth noting that the method improved greatly in NC vs. EMCI, and the six classification indexes achieve  $78.29 \pm 2.20\%$ ,  $78.03 \pm 2.41\%$ ,  $82.02 \pm 2.13\%$ ,  $74.73 \pm 3.04\%$ ,  $77.18 \pm 2.58\%$ , and  $81.00 \pm 1.93\%$ , respectively. MTFS has a better classification performance than Lasso and CMTL, indicating that the introduction of  $L_{2,1}$ -norm regularizer can effectively sparse multi-modal features and capture effective features, which is consistent

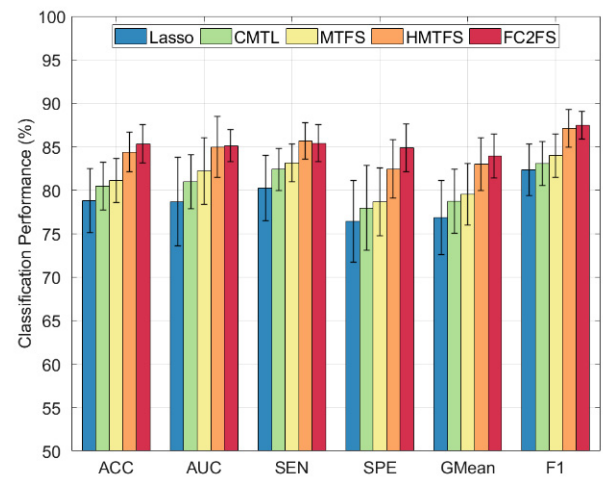


with the research results of Zhang et al. [2]. In addition, HMTFS has better classification performance than MTFS, indicating that introducing hypergraph regularization can indeed discover high-order relations between subjects, which is consistent with the research results of Shao et al. [18]. Besides, FC2FS has better classification performance than HMTFS, indicating that feature correlation and feature structure regularization can effectively discover more potential features and improve classification performance, which proves the effectiveness of the method.

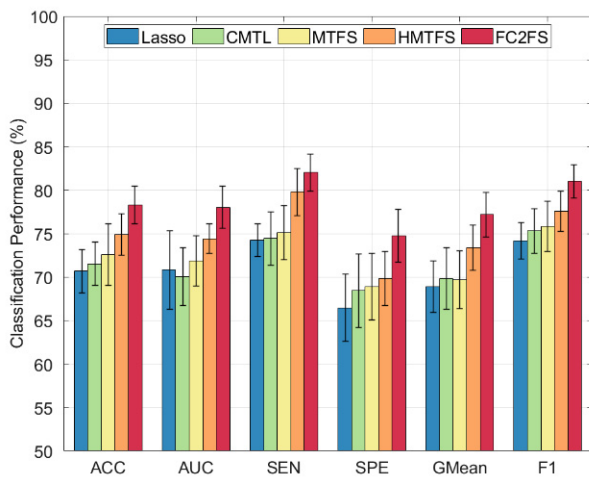
We used bar charts to represent the classification performance of the five methods more vividly on MCI and AD, aiming at better display of the experiment results of different methods, as shown in Figure 2.



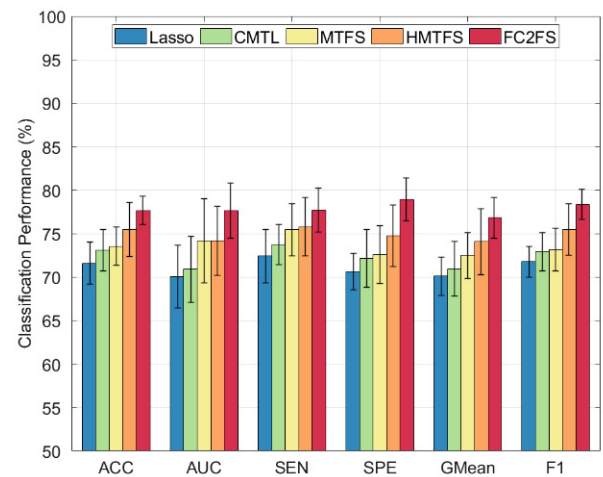
(a)



(b)



(c)



(d)

**Figure 2.** Classification performance of different methods. (a) NC vs. AD, (b) NC vs. LMCI, (c) NC vs. EMCI, (d) EMCI vs. LMCI.

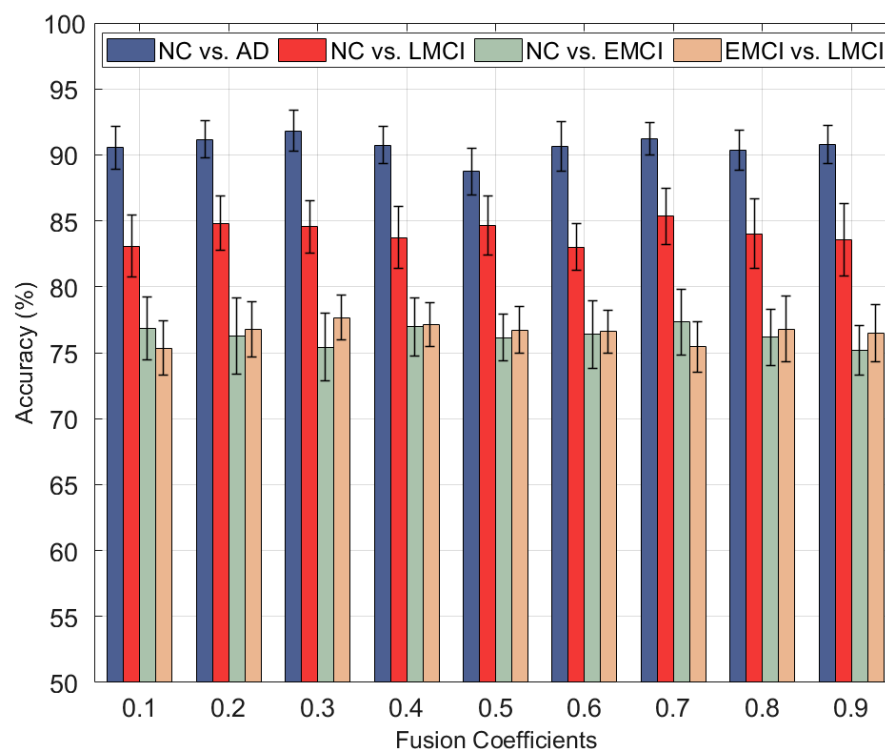
### 3.2. Parameter Sensitivity and Correlation Analysis

The selection of different parameters has different effects on the experiment results, and the parameter selection directly affects the performance of the method for MCI and AD classification. This section mainly analyzes the various influences of the main parameters

involved in the experiment and calculates the correlation coefficient on the experimental results, while exploring the optimal parameter selection of the experiment. The influence of the weighted fusion coefficient of the feature matrix, two kinds of regularization parameters, and the common calculation methods of the feature correlation coefficient on the classification performance are analyzed. Ten-fold cross-validation was adopted to make the experiment results credible [43]. According to the results of each experiment, the mean value of the ten experiment results was calculated randomly as the value of the classification performance index.

### 3.2.1. The Influence of Fusion Coefficient on Classification Accuracy

First,  $\alpha$ ,  $\beta$ , and  $\gamma$  values were fixed,  $\alpha$  and  $\beta$  were set to  $2^{-2}$  and  $\gamma$  to  $2^{-1}$ , and the Pearson correlation was used to calculate the correlation coefficient matrix. Then the feature matrix fusion coefficient  $\tau$  of sMRI data was set, with the range of change of 0 to 1 and step size of 0.1 decimal, then giving the feature matrix fusion coefficient of PET data as  $1 - \tau$ . Further, the influence of fusion with different fusion coefficients on classification performance was explored, and experimental results were obtained as shown in Figure 3. Among them, the optimal fusion coefficient of NC vs. AD and EMCI vs. LMCI is 0.3, while the optimal fusion coefficient of NC vs. LMCI and NC vs. EMCI is 0.7. Using different fusion coefficient combinations to fuse the feature matrix directly affects the classification accuracy. In addition, the analysis of the four experiment results shows that in NC vs. AD and EMCI vs. LMCI classification, the sMRI fusion coefficient is 0.3, while the PET fusion coefficient is 0.7, indicating that sMRI data contributes more to classification performance than PET data. In NC vs. LMCI and NC vs. EMCI, the sMRI fusion coefficient was 0.7, while the PET fusion coefficient was 0.3, indicating that sMRI data had more influence on classification performance than PET data.

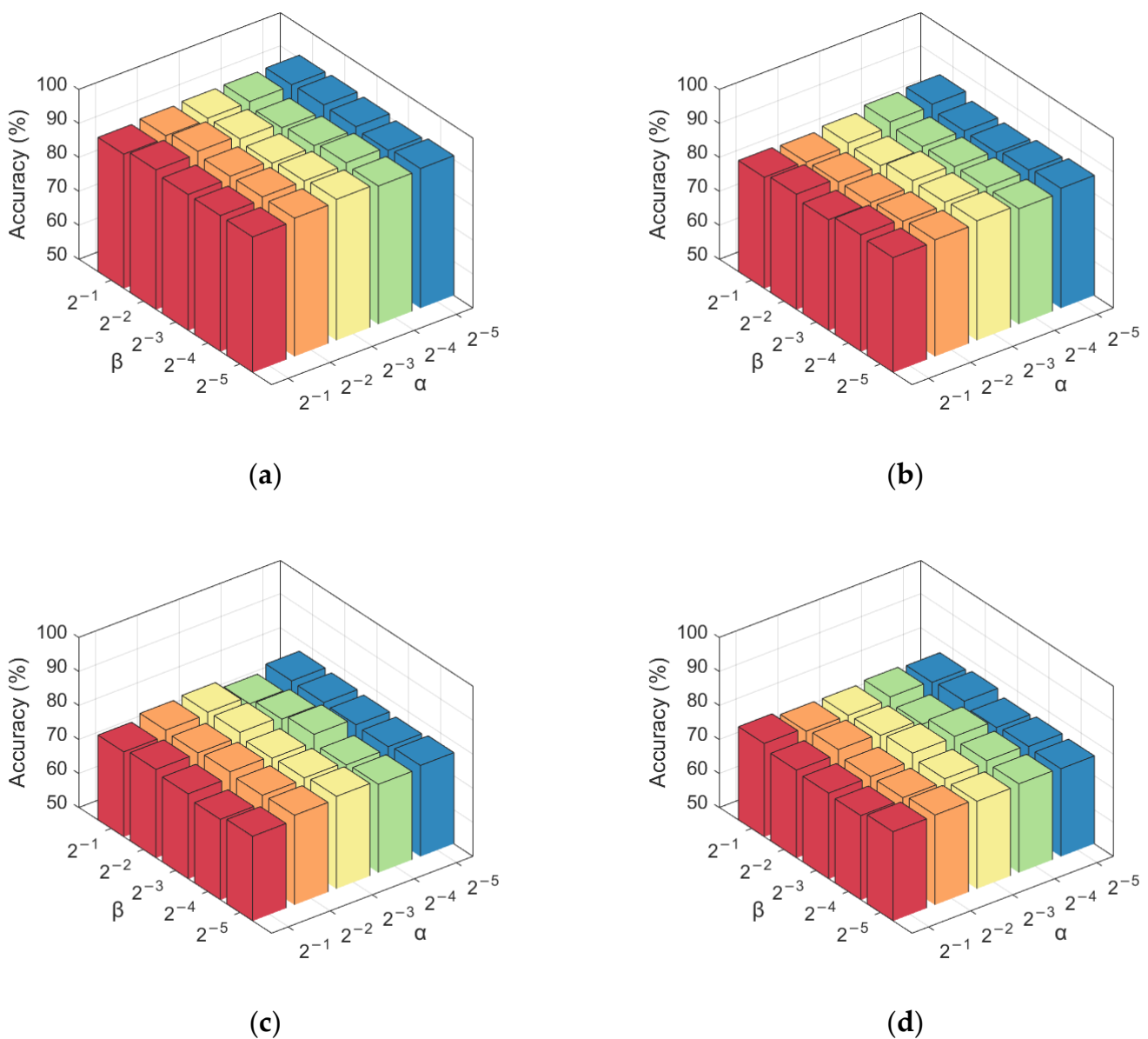


**Figure 3.** Influence of different fusion coefficients on classification accuracy.

### 3.2.2. Effects of Regularization Parameters on Classification Performance

In the established feature selection model, there are three regularization parameters, namely  $\alpha$ ,  $\beta$ , and  $\gamma$ . In the experiment, the appropriate  $\gamma$  value was first selected, and

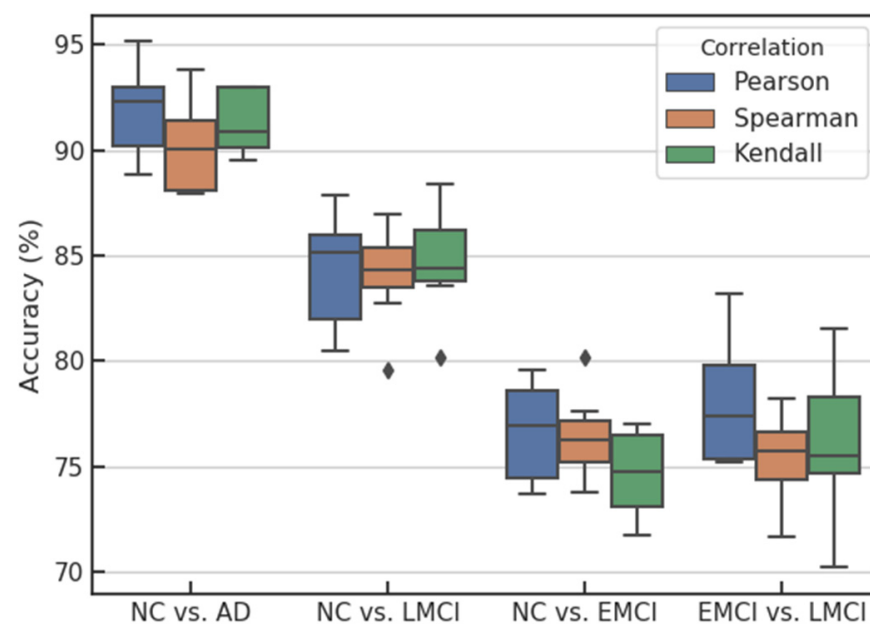
the value range of  $\alpha$  and  $\beta$  was set as  $\{2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}\}$  to explore the influence of different regularization parameter combinations on classification accuracy and reduce the time complexity of the model. Then the values of each group of  $\alpha$  and  $\beta$  were fixed. Finally, the classification accuracy of each group of values was calculated using the 10-fold cross-validation method, and the results were obtained as shown in Figure 4. Through analysis, it can be found that the classification accuracy does not fluctuate greatly under different regularization parameter combinations, which indicates that the method has a certain stability. In most cases, for each fixed  $\alpha$  ( $\beta$ ), as the value of  $\beta$  ( $\alpha$ ) decreases, the classification accuracy generally shows a trend of increasing first and then decreasing. The reason may be that with the decrease of the regularization parameter value, the weight of the feature correlation regularization and feature structure regularization decreases. The feature selection model's ability to capture the correlation between features is weakened, which results in partial effective features being ignored and reduced accuracy.



**Figure 4.** Influence of regularization parameters on classification accuracy. (a) NC vs. AD, (b) NC vs. LMCI, (c) NC vs. EMCI, (d) EMCI vs. LMCI.

### 3.2.3. The Influence of Correlation Coefficient Calculation Methods on Classification Performance

The influence of the proposed method on classification accuracy was analyzed when different methods were applied to calculate the feature correlation coefficients. The Pearson correlation coefficient, Spearman correlation coefficient and Kendall correlation coefficient were used to calculate the influence of correlation between features on classification accuracy under the three conditions, as shown in Figure 5. The results show that when the Pearson correlation coefficient is used to average the feature correlation matrix, the median accuracy obtained is always higher than that calculated by using the other two correlation coefficients, and the Pearson correlation has a larger fluctuation range than the other two correlation coefficients. The main reason may be that the Pearson correlation coefficient is sensitive to outliers. When more outliers of the feature correlation coefficient are generated, the Pearson correlation coefficient is greatly affected, while the Spearman correlation coefficient and the Kendall correlation coefficient are correlation coefficients based on matrix rank, so they are robust to outliers.



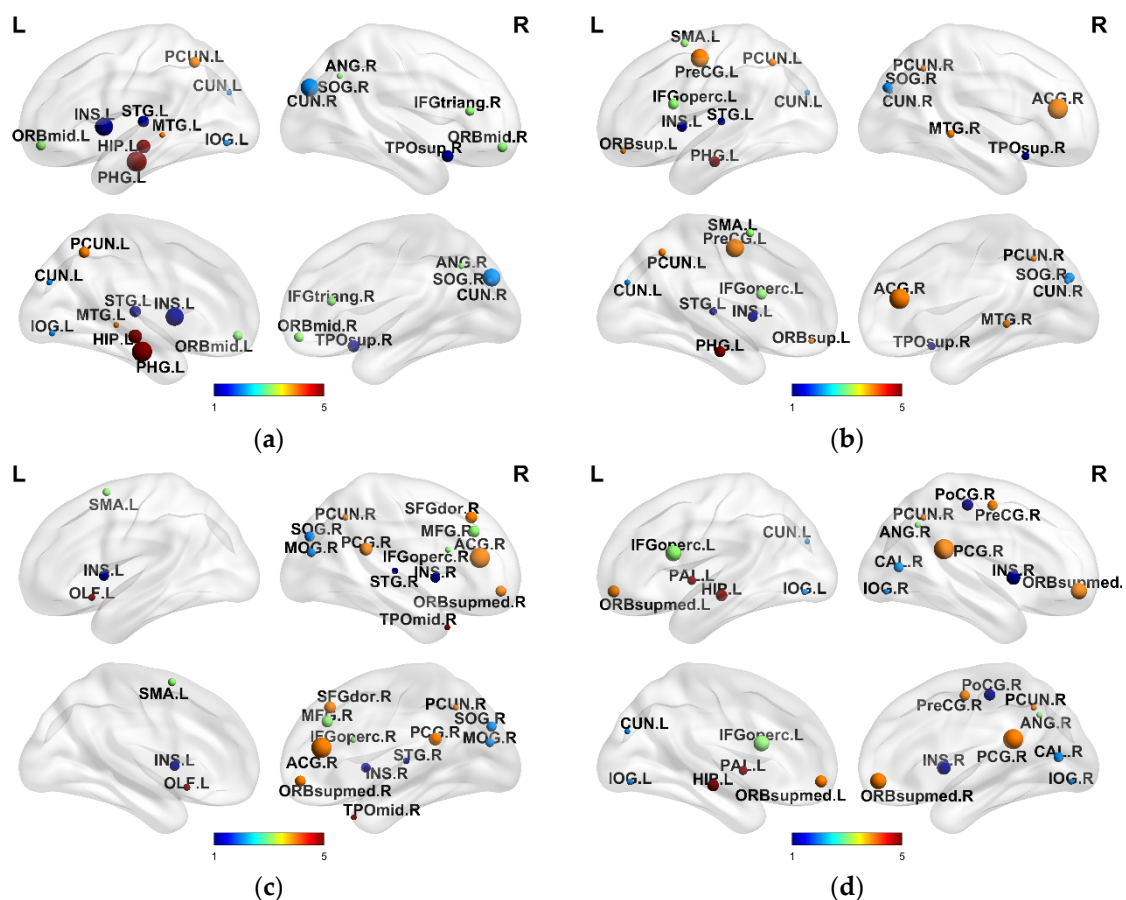
**Figure 5.** Influence of correlation coefficient calculation on classification accuracy.

### 3.3. Discriminative Brain Regions

The optimal regularized parameters determined by the 10-fold cross-validation method were selected in the experiment to find the most discriminant biologic features in MCI and AD classification. Further statistics obtain the brain regions corresponding to the top 15 feature vectors with different classification results. These brain areas are called discriminative brain regions, as shown in Table 4. The influence degree of MCI and AD on brain regions was discussed, and the BrainNet Viewer toolbox [44] was used to visually display the selected discriminative brain regions, as shown in Figure 6. As can be seen from the obtained results, most of the selected discriminative brain regions in NC vs. AD and NC vs. LMCI classification were confirmed, while only a small part of the selected discriminative brain regions in NC vs. EMCI and EMCI vs. LMCI classification was confirmed by previous studies. This phenomenon explains the lower performance of the latter classification.

**Table 4.** Discriminative brain regions.

NC vs. AD				NC vs. LMCI			
ID	Regions	Abbreviation	References	ID	Regions	Abbreviation	References
29	Insula_L	INS.L	Jon et al. [45] Bi et al. [46]	11	Frontal_Inf_Oper_L	IFGoperc.L	Liu et al. [47]
81	Temporal_Sup_L	STG.L	Liu et al. [48]	39	ParaHippocampal_L	PHG.L	Lee et al. [49]
45	Cuneus_L	CUN.L	Le et al. [50]	29	Insula_R	INS.R	Li et al. [51]
10	Frontal_Mid_Orb_R	ORBmid.R		32	Cingulum_Ant_R	ACG.R	Li et al. [51]
67	Precuneus_L	PCUN.L	Bailly et al. [52]	1	Precentral_L	PreCG.L	Cai et al. [53]
85	Temporal_Mid_L	MTG.L	Liu et al. [48]	5	Frontal_Sup_Orb_L	ORBsup.L	Li et al. [51]
37	Hippocampus_L	HIPL	Salvatore et al. [54]	19	Supp_Motor_Area_L	SMA.L	
53	Occipital_Inf_L	IOG.L		45	Cuneus_L	CUN.L	Le et al. [50]
50	Occipital_Sup_R	SOG.R		46	Cuneus_R	CUN.R	Le et al. [50]
39	ParaHippocampal_L	PHG.L	Katsel et al. [55]	50	Occipital_Sup_R	SOG.R	
84	Temporal_Pole_Sup_R	TPOsup.R	Salvatore et al. [54]	67	Precuneus_L	PCUN.L	Bailly et al. [52]
66	Angular_R	ANG.R		68	Precuneus_R	PCUN.R	Bailly et al. [52]
46	Cuneus_R	CUN.R	Le et al. [50]	81	Temporal_Sup_L	STG.L	
14	Frontal_Inf_Tri_R	IFGtriang.R	Salvatore et al. [54]	84	Temporal_Pole_Sup_R	TPOsup.R	Salvatore et al. [54]
9	Frontal_Mid_Orb_L	ORBmid.L	Zhang et al. [56]	86	Temporal_Mid_R	MTG.R	
NC vs. EMCI				EMCI vs. LMCI			
ID	Regions	Abbreviation	References	ID	Regions	Abbreviation	References
88	Temporal_Pole_Mid_R	TPOmid.R		68	Precuneus_R	PCUN.R	Lee et al. [49]
12	Frontal_Inf_Oper_R	IFGoperc.R	Chen et al. [57]	75	Pallidum_L	PAL.L	
82	Temporal_Sup_R	STG.R	Lee et al. [49]	37	Hippocampus_L	HIPL	Wu et al. [58]
29	Insula_L	INS.L	Anna et al. [59]	66	Angular_R	ANG.R	Lee et al. [49]
32	Cingulum_Ant_R	ACG.R		2	Precentral_R	PreCG.R	
50	Occipital_Sup_R	SOG.R		11	Frontal_Inf_Oper_L	IFGoperc.L	
68	Precuneus_R	PCUN.R	Lee et al. [49]	25	Frontal_Mid_Orb_L	ORBsupmed.L	
4	Frontal_Sup_R	SFGdor.R		30	Insula_R	INS.R	Bi et al. [46]
8	Frontal_Mid_R	MFG.R	Lee et al. [49]	36	Cingulum_Post_R	PCG.R	
19	Supp_Motor_Area_L	SMA.L		44	Calcarine_R	CAL.R	
21	Olfactory_L	OLF.L	Vasavada et al. [60]	53	Occipital_Inf_L	IOG.L	
26	Frontal_Mid_Orb_R	ORBsupmed.R		54	Occipital_Inf_R	IOG.R	
30	Insula_R	INS.R	Anna et al. [59]	26	Frontal_Mid_Orb_R	ORBsupmed.R	
36	Cingulum_Post_R	PCG.R		45	Cuneus_L	CUN.L	
52	Occipital_Mid_R	MOG.R		58	Postcentral_R	PoCG.R	

**Figure 6.** Visualization of discriminative brain regions. (a) NC vs. AD, (b) NC vs. LMCI, (c) NC vs. EMCI, (d) EMCI vs. LMCI.

By analyzing the brain regions obtained by NC vs. AD and NC vs. LMCI classification, we can find that among the selected discriminative brain regions, the ones belonging to temporal lobe, prefrontal lobe and occipital lobe account for a large proportion of the first 15 discriminative brain regions. NC vs. EMCI and EMCI vs. LMCI classification results showed that the discriminative brain regions belonging to the prefrontal lobe and occipital lobe accounted for a large proportion. The temporal lobe region mainly includes five discriminative brain regions: the left superior temporal gyrus (STG.L), the left middle temporal gyrus (MTG.L), the left hippocampus (HIP.L), the left parahippocampal gyrus (PHG.L), and the right temporal pole superior temporal gyrus (TPOsup.r). The temporal lobe is closely related to language and memory, among which damage to the left superior temporal gyrus (STG.L) will cause sensory aphasia, while damage to the left hippocampus (HIP.L) and left parahippocampus (PHG.L), one of the important organs in the brain involved in learning and memory storage, will lead to atrophy and memory impairment. Relevant studies have confirmed that the volume and morphology of the AD hippocampus will change compared with normal subjects [61,62]. The prefrontal lobe has the function of managing cognition, emotion, and behavior, which is mainly related to motor and higher mental function, while occipital lobe lesions will not only lead to visual impairment but are also accompanied by memory and motor defects. The selected discriminative brain regions belonging to the prefrontal lobe and occipital lobe, mainly include the right middle frontal gyrus (MFG.R), left middle frontal gyrus, orbital part (ORBmid.L), left superior orbital cortex (ORBsup.L), right inferior frontal gyrus, triangular part (IFGtriang.R), left opercular part of inferior frontal gyrus (IFGoperc.L), right anterior cingulate gyrus (ACG.R), left insula (INS.L), left cuneus (CUN.L), and right cuneus (CUN.R).

Notably, the right posterior cingulate gyrus (PCG.r) and the left precuneus (PCUN.L) were selected in NC vs. EMCI classification. These two discriminative brain regions are associated with the process of memory formation, indicating that compared with normal subjects, memory has been changed during the EMCI stage. The right angular gyrus (ANG.R) was identified in the EMCI vs. LMCI classification and is an important biological feature that distinguishes the first two [54]. However, there are still a small number of brain regions that have not been confirmed by previous studies among the selected discriminative brain regions. This may be caused by the fact that some of these brain regions do indeed have a strong impact on MCI and AD classification, but the existing relevant studies have not proved it. In addition, there may still be a few redundant features in the feature selection, leading to the selection of brain regions weakly related to the disease.

#### 4. Discussion

There are few studies on the relationship and structure between feature nodes among the existing multi-modal feature selection methods for diseases. Most of the methods focusing on the relationship between the same modality or different modality subjects, do not consider the influence of the relationship between feature nodes and structure on the model, and lack interpretation. For example, Jie et al. [17] used manifold learning to measure the distance between different subjects to maintain the adjacent structure between subjects, and applied it to MCI classification, achieving a good classification performance and verifying the effectiveness of the method. However, this method ignores the similarity relation between feature nodes and local geometry structure and lacks explanation for feature selection.

The results of this experiment show that this feature relation cannot be ignored in MCI and AD classification and has a positive influence on feature selection. It is worth mentioning that in the study of Lei et al. [63], four feature relations were regularized and the  $L_{2,1}$ -norm regularizer was introduced to sparse feature weight vectors, and it was finally applied to the classification of Parkinson's disease, achieving good classification performance and good interpretability. Yet, this method has some disadvantages, the obvious one is that there are too many model parameters, and the time complexity of the method increases in practical application. In our study, the multi-modal feature selection method with feature



correlation and feature structure fusion fully considers the internal connection between feature nodes, solves the problem of too many parameters in the loss function, reduces the time complexity of feature selection, and brings better classification performance.

Additionally, it is worth noting that we extract feature vectors directly from the original sMRI and PET images to obtain the feature matrix, and each feature vector represents a different brain region in the AAL template. When learning feature weight, the loss function essentially learns the weight value of the brain region through the training set and selects the corresponding feature vectors of the brain regions that are helpful to improve the classification performance. Therefore, the proposed method improves the interpretability of the model.

In summary, our work has potential clinical application value in clinical diagnosis. On the one hand, since the scale is subjective in clinical use [64] and the determination of patients with cognitive impairment is also personal, applying this method in the clinic will reduce human intervention, assist clinical diagnosis, and make diagnostic results more objective. On the other hand, the experimental results showed that the sensitivity and specificity of the method were significantly improved in the classification of NC and EMCI, which is clinically significant [65], reducing the risk of misdiagnosis of normal controls as early cognitive impairment patients require timely drug intervention. At the same time, the experiment proved that the method is more suitable for accurately capturing and identifying patients with subtle changes in brain regions. This property is better suited for diagnosing more difficult cognitive impairment associated with certain diseases, such as End-Stage Renal Disease (ESRD) combined with cognitive impairment [66], the exact neuropathological mechanism of which is still unclear. Cognitive impairment is a comorbidity of ESRD, and treatment of ESRD may also change brain function and structure [67], making it more challenging to identify MCI. In the future, based on the proposed model, we will further explore the identification of ESRD patients with cognitive impairment.

## 5. Conclusions

In this study, a multi-modal feature selection algorithm with feature correlation and feature structure fusion is proposed and applied to MCI and AD classification. In this method, low-rank constraint is introduced based on multi-task learning, moreover, feature correlation and feature structure regularization are adopted considering feature node relations. Finally, feature learning is carried out according to the constructed loss function. Experimental results showed that the proposed method performed better than the comparison methods in classification performance.

Nevertheless, our work has some limitations. When constructing the feature correlation coefficient matrix, only the relatively common calculation method of the correlation coefficient was considered, and the method that can better measure the correlation between two features or even multiple feature nodes remains to be discussed. Moreover, only the linear fusion of multi-modal features was input into the SVM classifier during classification. In the future, the integration model [68] deserves to be discussed to combine several weak classifiers into a strong classifier, and the classification performances of MCI and AD need to be further improved.

**Author Contributions:** Conceptualization, Z.J. and H.S.; methodology, Z.J. and S.C.; software, S.C.; formal analysis, S.C.; writing—original draft preparation, Z.J. and S.C.; writing—review and editing, H.S. and J.X.; supervision, H.S.; project administration, J.X.; funding acquisition, Z.J. and J.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 51877013) awarded to Z.J., the National Natural Science Foundation of China (Grant No. 52007087) awarded to J.X., the Jiangsu Provincial Key Research and Development Program (Grant No. BE2021636) awarded to Z.J., the Science and Technology Project of Changzhou City (Grant No. CE20205056) awarded to Z.J., the Natural Science Foundation of Ningbo City (Grant No. 202003N4116) awarded to J.X., and the Fund from the Educational Commission of Zhejiang Province (Grant No. Y202044047) awarded to J.X.; This work was also sponsored by Qing Lan Project of Jiangsu Province.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study, due to research data can be used by public and open database. Moreover, the researchers recorded the information in such a way that subjects would not be directly identified or indirectly identified through relevant identifiers.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://loni.ucla.edu/ADNI>). 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: [www.loni.ucla.edu/ADNI/Collaboration/ADNI\\_Authorship\\_list.pdf](http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf).

**Conflicts of 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.

## References

1. Kalaria, R.N.; Maestre, G.E.; Arizaga, R.; Friedland, R.P.; Group TWFND. Alzheimer's disease and vascular dementia in developing countries: Prevalence, management, and risk factors. *Lancet Neurol.* **2008**, *7*, 812–826. [[CrossRef](#)]
2. Zhang, D.Q.; Shen, D.G. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage* **2012**, *59*, 895–907. [[CrossRef](#)]
3. Li, Y.; Liu, J.Y.; Gao, X.Q.; Jie, B.; Minjeong, K.; Yap, P.T.; Wee, C.Y.; Shen, D.G. Multimodal hyper-connectivity of functional networks using functionally-weighted LASSO for MCI classification. *Med. Image Anal.* **2019**, *52*, 80–96. [[CrossRef](#)]
4. Zhang, Y.D. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Front. Comput. Neurosci.* **2015**, *9*, 66. [[CrossRef](#)]
5. Dukart, J.; Kherif, F.; Mueller, K.; Adaszewski, S.; Schroeter, M.L.; Frackowiak, R.S.; Draganski, B.; Alzheimer's Disease Neuroimaging Initiative. Generative FDG-PET and MRI model of aging and disease progression in Alzheimer's disease. *PLoS Comput. Biol.* **2013**, *9*, e1002987. [[CrossRef](#)] [[PubMed](#)]
6. Wang, S.H. Detection of Alzheimer's disease by three-dimensional displacement field estimation in structural magnetic resonance imaging. *J. Alzheimer's Dis.* **2016**, *50*, 233–248. [[CrossRef](#)] [[PubMed](#)]
7. Zhang, Y.D. Detection of Alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC. *Biomed. Signal Process. Control.* **2015**, *21*, 58–73. [[CrossRef](#)]
8. Andrea, B.; Sun, X.D.; Bernd, B.; Jörg, R.; Michel, L. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* **2020**, *143*, 106839.
9. Majdi, M.; Seyedali, M. Whale optimization approaches for wrapper feature selection. *Appl. Soft Comput.* **2018**, *62*, 441–453.
10. Ji, Y.X.; Zhang, Y.T.; Shi, H.F.; Jiao, Z.Q.; Wang, S.H.; Wang, C. Constructing dynamic brain functional networks via hyper-graph manifold regularization for mild cognitive impairment classification. *Front. Neurosci.* **2021**, *15*, 358. [[CrossRef](#)] [[PubMed](#)]
11. Wang, S.H. Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimed. Tools Appl.* **2018**, *77*, 10393–10417. [[CrossRef](#)]
12. Robert, T. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288.
13. Wang, S.H. ADVIAN: Alzheimer's disease VGG-inspired attention network based on convolutional block attention module and multiple way data augmentation. *Front. Aging Neurosci.* **2021**, *13*, 687456. [[CrossRef](#)]
14. Saba, E.; Anya, M.; Wei, X. Prognosis and Diagnosis of Parkinson's Disease Using Multi-Task Learning. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1457–1466.
15. Lei, B.Y.; Cheng, N.N.; Alejandro, F.F.; Tan, E.L.; Cao, J.W.; Yang, P.; Ahmed, E.; Du, J.; Xu, Y.W.; Wang, T.F. Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease. *Med. Image Anal.* **2020**, *61*, 101652. [[CrossRef](#)] [[PubMed](#)]
16. Ma, Q.M.; Zhang, T.H.; Marcus, V.Z.; Shen, H.; Theodore, D.S.; Daniel, H.W.; Raquel, E.G.; Fan, Y.; Hu, D.W.; Geraldo, F.B. Classification of multi-site MR images in the presence of heterogeneity using multi-task learning. *NeuroImage Clin.* **2018**, *19*, 476–486. [[CrossRef](#)] [[PubMed](#)]

17. Jie, B.; Zhang, D.Q.; Cheng, B.; Shen, D.G. Manifold Regularized Multi-Task Feature Selection for Multi-Modality Classification in Alzheimer's Disease. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 275–283.
18. Shao, W.; Peng, Y.; Zu, C.; Wang, M.L.; Zhang, D.Q.; Alzheimer's Disease Neuroimaging Initiative. Hypergraph based multi-task feature selection for multimodal classification of Alzheimer's disease. *Comput. Med Imaging Graph.* **2020**, *80*, 101663. [[CrossRef](#)] [[PubMed](#)]
19. Dubois, B.; Feldman, H.H.; Jacova, C.; Dekosky, S.T.; Barberger-Gateau, P.; Cummings, J.; Delacourte, A.; Galasko, D.; Gauthier, S.; Jicha, G. Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol.* **2007**, *6*, 734–746. [[CrossRef](#)]
20. John, A.; Barnes, G.; Chen, C.C.; Jean, D.; Guillaume, F.; Karl, F.; Stefan, K.; James, K.; Vladimir, L.; Rosalyn, M. *SPM12 Manual*; Wellcome Trust Centre for Neuroimaging: London, UK, 2014; p. 2464.
21. Nathalie, T.M.; Brigitte, L.; Dimitri, P.; Fabrice, C.; Olivier, E.; Nicolas, D.; Bernard, M.; Marc, J. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **2002**, *15*, 273–289.
22. Alan, C.E.; Andrew, L.J.; Louis, C.; Sylvain, B. Brain templates and atlases. *NeuroImage* **2012**, *62*, 911–922.
23. Zhang, Y.; Yang, Q. A survey on multi-task learning. *IEEE Trans. Knowl. Data Eng.* **2021**, *1*, 1. [[CrossRef](#)]
24. Wang, S.; Chang, X.J.; Li, X.; Sheng, Q.Z.; Chen, W.T. Multi-task support vector machines for feature selection with shared knowledge discovery. *Signal Process.* **2016**, *120*, 746–753. [[CrossRef](#)]
25. Carlo, C.; Dimitris, S.; Massimiliano, P. Reexamining low rank matrix factorization for trace norm regularization. *arXiv* **2017**, arXiv:1706.08934.
26. Nie, F.P.; Huang, H.; Ding, C. Low-Rank Matrix Recovery via Efficient Schatten P-Norm Minimization. In Proceedings of the Twenty-sixth AAAI Conference on Artificial Intelligence, Toronto, ON, Canada, 22–26 July 2012.
27. Girish, C.; Ferat, S. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28.
28. Huang, L.L.; Tang, J.; Chen, S.-B.; Ding, C.; Luo, B. An Efficient Algorithm for Feature Selection with Feature Correlation. In Proceedings of the International Conference on Intelligent Science and Intelligent Data Engineering, Nanjing, China, 15–17 October 2012; pp. 639–646.
29. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. *Pearson Correlation Coefficient, Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
30. Hauke, J.; Kossowski, T. *Comparison of Values of Pearson's and Spearman's Correlation Coefficient on The Same Sets of Data*; Wydział Nauk Geograficznych i Geologicznych Uniwersytetu im. Adama Mickiewicza: Poznań, Poland, 2011.
31. Zhao, Y.; You, X.G.; Yu, S.J.; Xu, C.; Yuan, W.; Jing, X.Y.; Zhang, T.P.; Tao, D.C. Multi-view manifold learning with locality alignment. *Pattern Recognit.* **2018**, *78*, 154–166. [[CrossRef](#)]
32. Wang, S.H. Alzheimer's Disease Detection by Pseudo Zernike Moment and Linear Regression Classification. *CNS Neurol. Disord. Drug Targets* **2017**, *16*, 11–15. [[CrossRef](#)]
33. Jiao, Z.Q.; Xia, Z.W.; Ming, X.L.; Cheng, C.; Wang, S.H. Multi-scale feature combination of brain functional network for eMCI classification. *IEEE Access* **2019**, *7*, 74263–74273. [[CrossRef](#)]
34. Belkin, M.; Niyogi, P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Adv. Neural Inf. Process. Syst.* **2001**, *14*, 585–591.
35. Zhang, Y.D. Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. *J. Alzheimer's Dis.* **2018**, *65*, 855–869. [[CrossRef](#)]
36. Lorenzo, R.; Silvia, V.; Bang Công, V. Convergence of stochastic proximal gradient algorithm. *arXiv* **2014**, arXiv:1403.5074.
37. Bamdev, M.; Gilles, M.; Francis, B.; Rodolphe, S. Low-rank optimization with trace norm penalty. *SIAM J. Optim.* **2013**, *23*, 2124–2149.
38. Pong, T.K.; Tseng, P.; Ji, S.W.; Ye, J.P. Trace norm regularization: Reformulations, algorithms, and multi-task learning. *SIAM J. Optim.* **2010**, *20*, 3465–3489. [[CrossRef](#)]
39. Zhang, Y.D. Classification of Alzheimer Disease Based on Structural Magnetic Resonance Imaging by Kernel Support Vector Machine Decision Tree. *Prog. Electromagn. Res. Pier* **2014**, *144*, 185–191. [[CrossRef](#)]
40. Jiao, Z.Q.; Ji, Y.X.; Gao, P.; Wang, S.H. Extraction and analysis of brain functional statuses for early mild cognitive impairment using variational auto-encoder. *J. Ambient. Intell. Humaniz. Comput.* **2020**. preprint. [[CrossRef](#)]
41. Jiao, Z.Q.; Ji, Y.X.; Jiao, T.X.; Wang, S.H. Extracting sub-networks from brain functional network using graph regularized nonnegative matrix factorization. *Comput. Model. Eng. Sci.* **2020**, *123*, 845–871. [[CrossRef](#)]
42. Zhou, J.; Chen, J.; Ye, J. Clustered Multi-Task Learning Via Alternating Structure Optimization. *Adv. Neural Inf. Process. Syst.* **2011**, *2011*, 702. [[PubMed](#)]
43. Qiao, L.S.; Chen, S.C.; Tan, X.Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2009**, *43*, 331–341. [[CrossRef](#)]
44. Xia, M.R.; Wang, J.H.; He, Y. BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS ONE* **2013**, *8*, e68910.
45. Jon, D.; Hojjat, A. Graph theory and brain connectivity in Alzheimer's disease. *Neuroscientist* **2017**, *23*, 616–626.

46. Bi, X.A.; Xu, Q.; Luo, X.H.; Sun, Q.; Wang, Z.G. Analysis of progression toward Alzheimer's disease based on evolutionary weighted random support vector machine cluster. *Front. Neurosci.* **2018**, *12*, 716. [[CrossRef](#)] [[PubMed](#)]
47. Liu, J.; Pan, Y.; Wu, F.X.; Wang, J.X. Enhancing the feature representation of multi-modal MRI data by combining multi-view information for MCI classification. *Neurocomputing* **2020**, *400*, 322–332. [[CrossRef](#)]
48. Liu, X.Z.; Chen, W.; Tu, Y.H.; Hou, H.T.; Huang, X.Y.; Chen, X.L.; Guo, Z.W.; Bai, G.H.; Chen, W. The Abnormal functional connectivity between the hypothalamus and the temporal gyrus underlying depression in Alzheimer's disease patients. *Front. Aging Neurosci.* **2018**, *10*, 37. [[CrossRef](#)] [[PubMed](#)]
49. Lee, E.S.; Yoo, K.; Lee, Y.B.; Chung, J.; Lim, J.E.; Yoon, B.; Jeong, Y. Default Mode Network Functional Connectivity in Early and Late Mild Cognitive Impairment: Results from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer Dis. Assoc. Disord.* **2016**, *30*, 289–296. [[CrossRef](#)]
50. Le, G.; Kristian, S.F.; Otto, M.H.; Lan, L.; Lasse, A.; Birgitte, B.A.; Eva, B.; Anne-Mette, H.; Peter, H.; Steen, G.H. A visual rating scale for cingulate island sign on 18F-FDG-PET to differentiate dementia with Lewy bodies and Alzheimer's disease. *J. Neurol. Sci.* **2020**, *410*, 116645.
51. Li, C.X.; Li, Y.J.; Zheng, L.; Zhu, X.Q.; Shao, B.X.; Fan, G.; Liu, T.; Wang, J.; Alzheimer's Disease Neuroimaging Initiative. Abnormal brain network connectivity in a triple-network model of Alzheimer's disease. *J. Alzheimer's Dis.* **2019**, *69*, 237–252. [[CrossRef](#)]
52. Bailly, M.; Destrieux, C.; Hommet, C.; Mondon, K.; Cottier, J.P.; Beaufils, E.; Vierron, E.; Vercoullie, J.; Ibazizene, M.; Voisin, T. Precuneus and cingulate cortex atrophy and hypometabolism in patients with Alzheimer's disease and mild cognitive impairment: MRI and 18F-FDG PET quantitative analysis using FreeSurfer. *BioMed Res. Int.* **2015**, *2015*, 583931. [[CrossRef](#)] [[PubMed](#)]
53. Cai, S.P.; Huang, L.Y.; Zou, J.; Jing, L.L.; Zhai, B.Z.; Ji, G.J.; Von Deneen, K.M.; Ren, J.C.; Ren, A.; Alzheimer's Disease Neuroimaging Initiative. Changes in thalamic connectivity in the early and late stages of amnesic mild cognitive impairment: A resting-state functional magnetic resonance study from ADNI. *PLoS ONE* **2015**, *10*, e0115573. [[CrossRef](#)] [[PubMed](#)]
54. Salvatore, C.; Cerasa, A.; Castiglioni, I. MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis. *Front. Aging Neurosci.* **2018**, *10*, 135. [[CrossRef](#)] [[PubMed](#)]
55. Katsel, P.; Roussos, P.; Beerli, M.; Gama-Sosa, M.; Gandy, S.; Khan, S.; Haroutunian, V. Parahippocampal gyrus expression of endothelial and insulin receptor signaling pathway genes is modulated by Alzheimer's disease and normalized by treatment with anti-diabetic agents. *PLoS ONE* **2018**, *13*, e0206547. [[CrossRef](#)]
56. Zhang, Y.; Zhang, H.; Chen, X.B.; Liu, M.X.; Zhu, X.F.; Lee, S.W.; Shen, D.G. Strength and similarity guided group-level brain functional network construction for MCI diagnosis. *Pattern Recognit.* **2019**, *88*, 421–430. [[CrossRef](#)]
57. Chen, X.B.; Zhang, H.; Lee, S.W.; Shen, D.G. Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification. *Neuroinformatics* **2017**, *15*, 271–284. [[CrossRef](#)] [[PubMed](#)]
58. Wu, L.Y.; Rowley, J.; Mohades, S.; Leuzy, A.; Dauar, M.T.; Shin, M.; Fonov, V.; Jia, J.P.; Gauthier, S.; Rosa-Neto, P. Dissociation between brain amyloid deposition and metabolism in early mild cognitive impairment. *PLoS ONE* **2012**, *7*, e47905. [[CrossRef](#)]
59. Anna, L.; Islem, R. Joint pairing and structured mapping of convolutional brain morphological multiplexes for early dementia diagnosis. *Brain Connect.* **2019**, *9*, 22–36.
60. Vasavada, M.M.; Wang, J.L.; Eslinger, P.J.; Gill, D.J.; Sun, X.Y.; Karunanayaka, P.; Yang, Q.X. Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment. *J. Alzheimer's Dis.* **2015**, *45*, 947–958. [[CrossRef](#)] [[PubMed](#)]
61. Pennanen, C.; Kivipelto, M.; Tuomainen, S.; Hartikainen, P.; Hänninen, T.; Laakso, M.P.; Hallikainen, M.; Vanhanen, M.; Nissinen, A.; Helkala, E.L. Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiol. Aging* **2004**, *25*, 303–310. [[CrossRef](#)]
62. Killiany, R.; Hyman, B.; Gomez-Isla, T.; Moss, M.; Kikinis, R.; Jolesz, F.; Tanzi, R.; Jones, K.; Albert, M. MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology* **2002**, *58*, 1188–1196. [[CrossRef](#)] [[PubMed](#)]
63. Lei, H.J.; Huang, Z.W.; Zhou, F.; Elazab, A.; Tan, E.L.; Li, H.C.; Qin, J.; Lei, B.Y. Parkinson's disease diagnosis via joint learning from multiple modalities and relations. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 1437–1449. [[CrossRef](#)]
64. Crowe, M.; Andel, R.; Wadley, V.; Cook, S.; Unverzagt, F.; Marsiske, M. Subjective cognitive function and decline among older adults with psychometrically defined amnesic MCI. *Int. J. Geriatr. Psychiatry A J. Psychiatry Late Life Allied Sci.* **2006**, *21*, 1187–1192. [[CrossRef](#)]
65. Fraser, K.C.; Lundholm Fors, K.; Eckerström, M.; Themistocleous, C.; Kokkinakis, D. Improving the Sensitivity and Specificity of MCI Screening with Linguistic Information. In Proceedings of the LREC Workshop: RaPID-2, Miyazaki, Japan, 8 May 2018.
66. Yue, Z.; Wang, P.; Li, X.; Ren, J.; Wu, B. Abnormal brain functional networks in end-stage renal disease patients with cognitive impairment. *Brain Behav.* **2021**, *11*, e02076. [[CrossRef](#)]
67. Lu, H.; Gu, Z.; Xing, W.; Han, S.; Wu, J.; Zhou, H.; Ding, J.; Zhang, J. Alterations of default mode functional connectivity in individuals with end-stage renal disease and mild cognitive impairment. *BMC Nephrol.* **2019**, *20*, 1–8. [[CrossRef](#)]
68. Zhou, Z.H. *Ensemble Learning, Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 181–210.