HHS Public Access

Author manuscript

*Neuroinformatics*. Author manuscript; available in PMC 2016 July 01.

Published in final edited form as:

*Neuroinformatics*. 2015 July ; 13(3): 277–295. doi:10.1007/s12021-014-9241-6.

# Supervised Discriminative Group Sparse Representation for Mild Cognitive Impairment Diagnosis

**Heung-Il Suk**[1], **Chong-Yaw Wee**[1], **Seong-Whan Lee**[2], and **Dinggang Shen**[1,2,*]

[1]Biomedical Research Imaging Center (BRIC) and Department of Radiology, University of North Carolina at Chapel Hill, North Carolina 27599, USA

[2]Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

## Abstract

Research on an early detection of Mild Cognitive Impairment (MCI), a prodromal stage of Alzheimer's Disease (AD), with resting-state functional Magnetic Resonance Imaging (rs-fMRI) has been of great interest for the last decade. Witnessed by recent studies, functional connectivity is a useful concept in extracting brain network features and finding biomarkers for brain disease diagnosis. However, it still remains challenging for the estimation of functional connectivity from rs-fMRI due to the inevitable high dimensional problem. In order to tackle this problem, we utilize a group sparse representation along with a structural equation model. Unlike the conventional group sparse representation method that does not explicitly consider class-label information, which can help enhance the diagnostic performance, in this paper, we propose a novel supervised discriminative group sparse representation method by penalizing a large within-class variance and a small between-class variance of connectivity coefficients. Thanks to the newly devised penalization terms, we can learn connectivity coefficients that are similar within the same class and distinct between classes, thus helping enhance the diagnostic accuracy. The proposed method also allows the learned common network structure to preserve the network specific and label-related characteristics. In our experiments on the rs-fMRI data of 37 subjects (12 MCI; 25 healthy normal control) with a cross-validation technique, we demonstrated the validity and effectiveness of the proposed method, showing the diagnostic accuracy of 89.19% and the sensitivity of 0.9167.

### Keywords

Alzheimer's Disease (AD); Mild Cognitive Impairment (MCI); resting-state fMRI; functional connectivity; sparse regression learning

*Correspondence: Tel.: +1 919-843-3535, Fax: +1 919-843-2641, dgshen@med.unc.edu.

**Information Sharing Statement**

In this paper, the proposed method was implemented based on SLEP (http://www.public.asu.edu/~jye02/Software/SLEP), which is an open source toolkit for sparse learning. For the group ICA, we used a GIFT toolbox, open to public (http://www.nitrc.org/projects/gift). The ADNI dataset (http://www.adni-info.org) used for evaluating the proposed method in this paper is also freely available.

## 1 Introduction

Alzheimer's Disease (AD), characterized by progressive impairment of cognitive and memory functions, is one of the most prevalent neurodegenerative brain diseases in the elderly subjects. A recent research by Alzheimer's Association reports that AD is the sixth-leading cause of death in the United States, rising significantly every year in terms of the proportion of cause of death [1]. It is also indicated that 10 to 20 percent of people aged 65 or older have Mild Cognitive Impairment (MCI), a prodromal stage of AD [1]. Although it is not clear why some people with MCI progress to AD and some do not, MCI is considered as an early stage of dementia in the particular form and it is estimated that approximately 10% to 15% of individuals with MCI progress to AD in one year. Therefore, it has been of great importance for early detection of MCI and a proper treatment to prevent them from progressing to AD.

Researchers in many scientific fields have devoted their efforts to understand the underlying mechanism of causing the diseases and to identify pathological biomarkers for diagnosis by analyzing different types of imaging modalities such as Magnetic Resonance Imaging (MRI) [17, 28, 33, 34, 84], Positron Emission Tomography (PET) [43], Diffusion Tensor Imaging (DTI) [61, 75], functional MRI (fMRI) [23, 76], etc. However, it is still challenging to diagnose MCI due to its subtlety of the involved cognitive impairment compared to AD.

Among modalities, fMRI has become a successful investigative tool in basic and clinical neuroscience to explore brain functions based on the Blood Oxygenation Level-Dependent (BOLD) signal variation. The fMRI characterizes hemodynamic responses associated with the neural activity. Ever since Biswal et al.'s work [9], resting-state fMRI (rs-fMRI) has been widely considered to explore the intrinsic and spontaneous neuronal activities induced during the resting state [11, 19, 53]. The rs-fMRI has provided effective insights into the inter-connection of structurally segregated and functionally specialized brain regions in normal healthy subjects and also for understanding of the pathological changes that can cause various brain function disorders: AD [23, 29], MCI [53, 57, 70, 82], Schizophrenia [30, 85], depression [3, 15, 22], etc. From a clinical point of view, it is very advantageous to use rs-fMRI to investigate brain activations of the patients, who are not able to perform complicated tasks during scanning, since it does not require a subject to perform any cognitive functions.

Recent studies have witnessed that the functional connectivity, defined as the temporal correlations between spatially distinct brain regions [20], can be a useful tool in finding biomarkers for brain disease diagnosis. A large part of the literature has considered a correlation approach to model the functional connectivity in brain images [68, 74]. However, due to its pairwise computation and full connectedness, it is hard to interpret the resulting connectivity map in terms of group analysis.

Justified by the studies of the numerical characteristics of network connectivity in anatomical brain databases [58,59] and in functional brain images [12,63], it is a reasonable assumption that we could drastically reduce the number of connections, effectively restricting our attention to networks with sparse connectivity. To this end, some groups

utilized independent component analysis to lessen the full-connectivity problems that plague correlation-based method [22,78]. A sparse connectivity can be also constructed via the least absolute shrinkage and selection operator (lasso) [66], which penalizes a linear regression model with $l_1$-norm. While lasso induces sparsity in the regression coefficients, it selects variables in a subject- or task-dependent manner and therefore has a limitation in inducing the group-wise information. Group analysis of brain connectivity has long been another challenging topic, since biomedical research is usually conducted at a group level to extract the population features. Efficient group analysis requires appropriate handling of expected inter-subject variability without destroying inter-group differences. Wee et al. proposed a constrained sparse functional connectivity network [77] via a group sparse representation [80] and earlier Ng and Abugharbieh applied the group sparse representation to fMRI brain decoding [46].

The main goal of the computer-aided AD/MCI diagnosis is to improve the classification performance with high sensitivity and specificity. In order for that, the previous methods in the literature utilized a regression method along with the sparse representation. Interestingly, while discrimination is the main goal of the computer-aided brain disease diagnosis, the optimization is based on a criterion to represent a target vector, without explicitly considering the actual discrimination task. Although these methods have proved their efficacy and validity on normal and AD/MCI classification in their own experiments, we believe that we can further enhance the classification performance by combining the regression model with a discriminative method in a unified framework.

In this work, we present a novel method of classifying AD/MCI and healthy Normal Control (NC) with sparse modeling in a supervised and discriminative manner. More specifically, we combine a group analysis with a class-discriminative feature extraction by extending the group lasso [80] with the introduction of a label-informed regularization term by penalizing a large within-class variance and a small between-class variance of connectivity coefficients. To our best knowledge, there has been no work on brain disease diagnosis and/or medical image analysis with the application of the sparse modeling that explicitly incorporates the discriminative approach into the regression model. We also show that the proposed optimization algorithm finds the group-consistent topological network as the conventional group lasso does and further jointly makes the connectivity coefficients to be similar within a class and distinct between classes. Therefore, we can say that the proposed method allows the learned functional connectivity to preserve the network specific and label-related characteristics.

Fig. 1 illustrates a schematic diagram of the proposed framework for MCI identification with rs-fMRI. Given a set of rs-fMRI images, we preprocess the images and parcellate a brain into ROIs based on an Automatic Anatomical Labeling (AAL) template. From each ROI, we compute a representative time series by averaging the intensity of voxels within a ROI. The ROI feature sequences are then bandpass filtered to remove physiological noise caused by cardiac and respiratory cycles. The bandpass filtered signals are then fed into the proposed supervised discriminative group sparse representation method to identify functional connectivities, whose network structures are consistent across subjects. Utilizing a graph theory, we compute clustering coefficients from the estimated functional

connectivities and perform feature selection with a combination of *t*-test, minimum-Redundancy Maximum-Relevance (mRMR), and Recursive Feature Elimination with Support Vector Machine (SVM-RFE). A linear SVM is considered to classify features for MCI identification.

## 2 Method

### 2.1 Connectivity Construction with a Structural Equation Model

In this study, we assume that the brain activity of a ROI can be represented by a linear combination of the activity of the other ROIs, justified by the investigation that the neural activities in the brain can be thought of firing the activity of the neurons in a voxel based on the neural firings of the other neurons. Mathematically, this can be formulated as a "structural equation model" (SEM) [40], which has been widely used for a study of brain connectivity [48, 86], as follows

$$\mathbf{y}=\mathbf{A}\mathbf{w}+\mathbf{e} \quad (1)$$

where **y** denotes a target response of a ROI, **A** is a predictor matrix composed of the responses of the other ROIs, **w** is a regression coefficient vector, and **e** is a zero-mean random error vector, which is Gaussian distributed and assumed to have independent and identically distributed (*i.i.d*) elements with a common finite variance. It is straightforward that since the trained regression coefficients in **w** indicate the relation between the target ROI and the predictor ROIs, we can consider the coefficients in **w** as a connectional strength or connectivity between the target ROI and the corresponding predictor ROI. Therefore, we can efficiently construct a connectivity map by concatenating the estimated coefficients into a matrix. Hereafter, we use regression coefficients and connectivity coefficients interchangeably.

### 2.2 Identification of class-discriminative brain connectivity model

Due to the curse of dimensionality in modeling with the practical fMRI data sets, it is known that the coefficients estimated by maximum-likelihood from the dataset are neither robust nor stable. However, thanks to the small network characteristics in brain functions [12, 63], we can efficiently restrict our attention to the sparse connectivity. Furthermore, it is also beneficial to investigate the connectional differences between NC and MCI under the same network structure. As studied by Wee et al. [77], the group lasso [80] can be useful as a nice technique to deal with these problems.

Here, we briefly describe the group lasso that can sparsely select subsets of predefined groups with non-zero weights assigned to all predictors for connectivity construction. Formally, it can be written as follows:

$$J(\mathbf{W}_r)=\min_{\mathbf{W}_r}\frac{1}{2}\sum_{n=1}^{N}\left\|\mathbf{y}_r^{(n)}-\mathbf{A}_r^{(n)}\mathbf{w}_r^{(n)}\right\|_2^2+\lambda\|\mathbf{W}_r\|_{2,1} \quad (2)$$

where $r$ and $n$ denote, respectively, indices of an ROI and a subject, $R$ is the number of ROIs, $\mathbf{y}_r^{(n)} \in \mathbb{R}^V$ and $\mathbf{A}_r^{(n)} = \left[\mathbf{y}_1^{(n)} \cdots \mathbf{y}_{r-1}^{(n)} \mathbf{y}_{r+1}^{(n)} \cdots \mathbf{y}_R^{(n)}\right] \in \mathbb{R}^{V \times (R-1)}$ are, respectively, a $V$-length time series of the $r$-th target ROI and a predictor matrix concatenated by the time series of the other $(R-1)$ ROIs, $\mathbf{w}_r^{(n)} \in \mathbb{R}^{(R-1)}$ is a regression coefficient vector, $\mathbf{W}_r = \left[\mathbf{w}_r^{(1)} \cdots \mathbf{w}_r^{(n)} \cdots \mathbf{w}_r^{(N)}\right]$, $N$ is the number of subjects, and $\lambda$ is a regularization parameter. The regularization term is defined as $\|\mathbf{W}_r\|_{2,1} = \sum_g \|\mathbf{w}_{r,g}\|_2$ (Fig. 2(a)), *i.e.*, $l_1$-norm of $\|\mathbf{w}_{r,g}\|_2$, where $\mathbf{w}_{r,g}$ denotes the connectivity coefficients associated with the $g$-th predictor in $\mathbf{A}_r^{(n)}$, for $n = 1, \cdots, N$. The group lasso regularization ensures that all regression models at different subjects/groups share a common set of connections.

The justification for our assumption on the same connectional structure between diseased patients and healthy controls is that although, between-group analysis may identify disease induced network structure changes, it is limited at an individual level, particularly for the purpose of disease classification. In this study, rather than investigating group specific connection networks, which are, of course, unlikely to be the same for patients and healthy controls, we focus on constructing a common network structure, in which it is effective to distinguish patients from healthy controls. That is, by imposing the functional connectivity of subjects to have the same network structure, but different connectivity strengths, we can efficiently diagnose subjects at an individual level.

Although the main goal in brain disease diagnosis is to enhance the classification performance between normal subjects and patients, the group lasso is, however, a regression method based on the penalization that does not explicitly include the discrimination task in the framework. To this end, we propose a novel method of jointly learning common functional brain networks across subjects via group sparse representation and class-discriminative coefficients with label-informed regularization terms.

In order to incorporate the categorical or class-label information, we utilize a well-known discriminative information of "Within-Class-Variance" (WCV) $f_W(\mathbf{w}_{r,g})$ and "Between-Class-Variance" (BCV) $f_B(\mathbf{w}_{r,g})$ [21] defined as follows:

$$f_W(\mathbf{w}_{r,g}) = \frac{1}{N^+} \sum_{n \text{ s.t. } l(n)='+'} (\mathbf{w}_{r,g}(n) - \hat{w}_{r,g}^+)^2 + \frac{1}{N^-} \sum_{n \text{ s.t. } l(n)='-'} (\mathbf{w}_{r,g}(n) - \hat{w}_{r,g}^-)^2 \quad (3)$$

$$f_B(\mathbf{w}_{r,g}) = (\hat{w}_{r,g}^+ - \hat{w}_{r,g}^-)^2 \quad (4)$$

where $\mathbf{w}_{r,g}(n)$ denotes the $n$-th element of a vector $\mathbf{w}_{r,g}$, $l(n)$ is a class-label of the $n$-th subject, $\hat{w}_{r,g}^+ = \frac{1}{N^+} \sum_{n \text{ s.t. } l(n)='+'} \mathbf{w}_{r,g}(n)$ and $\hat{w}_{r,g}^- = \frac{1}{N^-} \sum_{n \text{ s.t. } l(n)='-'} \mathbf{w}_{r,g}(n)$. The functions $f_W(\mathbf{w}_{r,g})$ and $f_B(\mathbf{w}_{r,g})$ compute, respectively, WCV and BCV of the coefficients with regard to the $g$-th predictor in the estimation of the $r$-th ROI. The idea of exploiting WCV and BCV to extract class-discriminative features is similar to the Linear Discriminant Analysis (LDA) [38] that considers the Fisher criterion defined as the ratio of the above functions. Unlike LDA, in this paper, we take the difference of WCV and BCV for computational efficiency.

By injecting the functions of $f_W(\mathbf{w}_{r,g})$ and $f_B(\mathbf{w}_{r,g})$ into Eq. (2), we devise a new objective function formulated as follows:

$$J(\mathbf{W}_r)=\min_{\mathbf{W}_r} L(\mathbf{W}_r)+\lambda_1\|\mathbf{W}_r\|_{2,1}+\lambda_2\left(\sum_g\|f_W(\mathbf{w}_{r,g})\|_2-\sum_g\|f_B(\mathbf{w}_{r,g})\|_2\right) \quad (5)$$

where $L(\mathbf{W}_r)=\frac{1}{2}\sum_n\left\|\mathbf{y}_r^{(n)}-\mathbf{A}_r^{(n)}\mathbf{w}_r^{(n)}\right\|_2^2$, and $\lambda_1$ and $\lambda_2$ are the regularization parameters. Through the two regularization terms in the parentheses, we penalize the high WCV and the low BCV. With the introduction of those two penalty terms, the connectivity coefficients for the subjects within a class are imposed to be similar to each other while those between classes to be distinct (Fig. 2(b)). We call this novel label-informed sparse model as '*Supervised Discriminative Group Lasso*' (SDGL).

Fortunately, we can simplify the variance related terms in Eq. (5) with appropriate algebraic operations as follows:

$$\sum_g\|f_W(\mathbf{w}_{r,g})\|_2=\|\mathbf{W}_rD_1\|_{2,1}^2 \quad (6)$$

$$\sum_g\|f_B(\mathbf{w}_{r,g})\|_2=\|\mathbf{W}_rD_2\|_{2,1}^2 \quad (7)$$

where $D_1$ and $D_2$ denote definitive matrices to compute WCV and BCV of the coefficients in $\mathbf{W}_r$, respectively. Specifically, $D_1$ is a composite matrix that computes the sum of the differences between the coefficients and their mean in each class, and $D_2$ is a matrix that computes the difference between the mean of the coefficients of two different classes. Refer to Appendix 4 for deduction of these matrices.

Interestingly, from Eq. (6) and Eq. (7), we can see that WCV and BCV are represented as a square of the coefficients in the original group lasso multiplied by a definitive matrix. It is clear that the proposed SDGL is equal to the conventional group lasso by setting $\lambda_2$ to zero. Replacing the two functions with matrix operations, we can rewrite the objective function in Eq. (5) as the following optimization problem:

$$\min_{\mathbf{W}_r} L(\mathbf{W}_r)+\lambda_1\|\mathbf{W}_r\|_{2,1}+\lambda_2\left(\|\mathbf{W}_rD_1\|_{2,1}^2-\|\mathbf{W}_rD_2\|_{2,1}^2\right). \quad (8)$$

### 2.3 Optimization Algorithm

Due to the non-smooth terms in the SDGL, it is challenging to solve the optimization problem directly. In order to tackle this optimization problem, we use the Accelerated Gradient Method (AGM) [45], in which the computation of the proximal operator with the composite of non-smooth penalties is one of the key steps. Here, we should note that the notations and explanations in the following paragraphs are based on Zhou et al.'s work [83] and from here on, we omit the subscript *r* for clarity.

In our case, the proximal operator can be defined as

$$\pi(\mathbf{V})=\underset{\mathbf{W}}{\operatorname{argmin}}\frac{1}{2}\|\mathbf{W}-\mathbf{V}\|_2^2+\lambda_1\|\mathbf{W}\|_{2,1}+\lambda_2\left(\|\mathbf{W}D_1\|_{2,1}^2-\|\mathbf{W}D_2\|_{2,1}^2\right) \quad (9)$$

where **W** and **V** denote, respectively, the approximate solution and the search point during optimization in AGM [31]. It is clear that we can decompose Eq. (9) into a row-wise operation, and then what we need is to solve the optimization problem defined as

$$\pi(\mathbf{v}_g)=\underset{\mathbf{w}_g}{\operatorname{argmin}}\frac{1}{2}\|\mathbf{w}_g-\mathbf{v}_g\|_2^2+\lambda_1\|\mathbf{w}_g\|_2+\lambda_2\left(\|\mathbf{w}_gD_1\|_2^2-\|\mathbf{w}_gD_2\|_2^2\right) \quad (10)$$

where $\mathbf{v}_g$ and $\mathbf{w}_g$ denote $g$-th rows of **V** and **W**, respectively.

Employing the decompositional property exhibited in [83], we can efficiently compute the proximal operator of Eq. (10) in two stages. The first stage is to estimate the coefficients in group lasso, which ensure that the brain networks across subjects to be consistent. In the second stage, we impose the group-constrained coefficients to be similar within a class and distinct between classes. That is, given an proximal operator defined in Eq. (10), we first solve the following optimization problem with the conventional group lasso

$$\pi_1(\mathbf{v}_g)=\underset{\mathbf{w}_g}{\operatorname{argmin}}\frac{1}{2}\|\mathbf{w}_g-\mathbf{v}_g\|_2^2+\lambda_1\|\mathbf{w}_g\|_2 \quad (11)$$

and then we tackle another proximal operator by taking the coefficients learned from the group lasso as the search point for class-discriminative learning as follows

$$\pi_2(\pi_1(\mathbf{v}_g))=\underset{\mathbf{w}_g}{\operatorname{argmin}}\frac{1}{2}\|\mathbf{w}_g-\pi_1(\mathbf{v}_g)\|_2^2+\lambda_2\left(\|\mathbf{w}_gD_1\|_2^2-\|\mathbf{w}_gD_2\|_2^2\right). \quad (12)$$

Refer to Appendix 4 for proof. The complete and concise algorithm for the proximal operator in Eq. (10) is given in Algorithm 1. In our experiments, we used a SLEP toolbox[1] [32] to implement the algorithm.

We should note that the proposed method finds the brain networks that are consistent across subjects as the conventional group lasso does, and meanwhile it makes the coefficients to be similar within a class and distinct between classes. That is, it allows the learned functional connectivity to preserve the network specific and label-related characteristics simultaneously.

**Algorithm 1**

Proximal operator associated with the proposed Supervised Discriminative Group Lasso (SDGL)

**Input:** $\mathbf{V}\in\mathbb{R}^{(R-1)\times N}$, $D_1\in\mathbb{R}^{N\times N}$, $D_2\in\mathbb{R}^{N\times N}$, $\lambda_1$, $\lambda_2$

[1]Available at 'http://www.public.asu.edu/~jye02/Software/SLEP/index.htm'

**Output:** Coefficient matrix **W**

**for** *g to* $(R-1)$ **do**

$$\mathbf{u}_g = \text{argmin}_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}-\mathbf{v}_g\|_2^2 + \lambda_1\|\mathbf{w}\|_2$$

$$\mathbf{w}_g = \text{argmin}_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}-\mathbf{u}_g\|_2^2 + \lambda_2(\|\mathbf{w}D_1\|_2^2 - \|\mathbf{w}D_2\|_2^2)$$

**end**

### 2.4 Functional Connectivity, Feature Selection, and Classifier Learning

The functional connectivity that represents inter-regional correlations in neuronal variability is estimated from the trained sparse connectivity coefficients $\bar{\mathbf{W}} = [\bar{\mathbf{w}}_i]_{i=1,\cdots,R} \in \mathbb{R}^{R\times R}$, where $\bar{\mathbf{w}}_i = [\mathbf{w}_i(1), \cdots, \mathbf{w}_i(i-1), 0, \mathbf{w}_i(i+1), \cdots, \mathbf{w}_i(R)]^T$ and $R$ denotes the number of ROIs. In order to obtain a functional connectivity representation, we take the average of the coefficient matrix and its transposed one, $\mathbf{C} = (\bar{\mathbf{W}} + \bar{\mathbf{W}}^T)/2$. Fisher's $z$ transformation, $\mathbf{Z}_{ij} = [ln(1+\mathbf{C}_{ij}) - ln(1-\mathbf{C}_{ij})]/2$, where $\mathbf{C}_{ij}$ denotes the $(i, j)$-th entry in $\mathbf{C}$, is then performed to improve the normality of correlation coefficients. The functional connectivity is then represented by a $z$-map. Fig. 3 shows the sample functional connectivities of a NC subject and a MCI patient in five non-overlapping frequency bands.

It is natural to convert the $z$-map into a graph by considering each ROI as a node and the signed correlation coefficient as an edge. We extract features utilizing a graph theory. Specifically, the weighted Local Clustering Coefficient (*wLCC*), which quantifies the cliqueness of the nodes in a graph, is considered.

$$wLCC(r) = 2 \times \frac{\sum_{q\in\mathbb{S}_r} E_{r,q}}{C_r(C_r-1)} \quad (13)$$

where $C_r$ is the number of ROIs connected to the $r$-th ROI in $\mathbf{Z}$, $\mathbb{S}_r$ is a sub-network composed of nodes directly connected to the $r$-th ROI, and $E_{r,q}$ is the connection coefficient between the $r$-th and the $q$-th ROIs. A feature vector is then constructed by stacking the set of clustering coefficients over all ROIs, *i.e.*, $F = [wLCC(1), \cdots, wLCC(R)]^T \in \mathbb{R}^R$. In our multi-spectrum approach, for which we divide the frequency band of interest into multiple non-overlapping sub-bands, we have multiple sets of the clustering coefficients, one for each frequency band. In that case, we further concatenate the feature vector of each frequency band into a single large one. For example, in our experiment, we decomposed a frequency band into five non-overlapping sub-bands. So the feature vector was defined as $\mathbf{F} = [F_1^T, F_2^T, F_3^T, F_4^T, F_5^T]^T \in \mathbb{R}^{5R}$, where $F_i$ denotes a feature vector constructed from the $i$-th sub-band.

Thanks to its efficacy and robustness proved in many fields, SVM is considered for classification [26, 49, 62, 76, 81]. The feature selection and the optimal parameter setting in classifier learning is another important issue to enhance classification performance. In this

study, a nested cross-validation technique is applied for the optimal parameter setting as well as performance evaluation.

Given training samples from $N$ subjects, in our case we have one sample from each subject, we first leave one subject out for test, and consider the samples from the remaining $N - 1$ subjects for feature selection and parameter setting for the optimal classifier learning. Since we employ a linear SVM for classification, there is one parameter that controls the relative importance of maximizing the margin and minimizing the amount of slack. From the $N - 1$ training samples, we further leave out another sample from the remaining $N - 1$ for validation. We select features by applying three methods sequentially, *i.e.*, *t*-test, mRMR [47], and SVM-RFE [50]. In order to find an optimal parameter for SVM, we adopt a grid search algorithm. The combination of features and a parameter that gives the best performance constructs the optimal SVM model to classify the test data.

## 3 Experimental Results and Discussion

### 3.1 Participants and diagnosis

The experimental protocols were approved by the institutional ethics board at Duke University Medical Center in compliance with the Health Insurance Portability and Accountability Act. In this study, we recruited 37 participants (12 MCI patients and 25 socio-demographically matched normal healthy controls), who submitted a written consent on the experiment, by the Duke-UNC Brain Imaging and Analysis Center (BIAC), Durham, North Carolina, USA. The diagnosis of the participants was performed by the expert consensus panels at the Joseph and Kathleen Bryan Alzheimer's Disease Research Center (Bryan ADRC) and the Department of Psychiatry at Duke University Medical Center. The diagnosis confirmation was made by consensus with the ultimate decision by a board-certified neurologist in concert with data available from a battery of general neurological examination, neuropsychological assessment evaluation, collateral subject symptom and functional capacity reports. The considered neuropsychological battery was a revised Consortium to Establish a Registry for Alzheimer's Disease (CERAD) [42] including 1) Mini-Mental State Examination (MMSE) [18]; 2) immediate and delayed verbal memory (Logical Memory subtest of the Wechsler Memory Scale-Revised [73]); 3) visual immediate memory (Benton Visual Retention Test [7]); 4) verbal initiation/lexical fluency (Controlled Oral Word Association Test from the Multilingual Aphasia Examination [8]); 5) attentional/executive functions (Trail Making Test [51], Symbol Digit Modality Test [56], Digit Span sub-test of the Wechsler Adult Intelligence Scale-Revised [72], and a separate ascending Digit Span task modeled after the Digit Ordering Test [14]); 6) premorbid verbal ability (Shipley Vocabulary Test [55]); 7) Finger Oscillation [52] and Grooved Pegboard [39] Tests; and 8) Self Rating of Memory Function [60].

We conformed a subject as MCI if he or she met the following criteria: 1) age > 55 years and any race; 2) recent worsening of cognition, but still functioning independently; 3) MMSE score between 24 and 30; 4a) score ≤ −1.5 SD on at least two Bryan ADRC cognitive battery memory tests for single-domain amnestic MCI; or 4b) score ≤ −1.5 SD on at least one of the formal memory tests and score ≤ −1.5 SD on at least one other cognitive domain task (e.g., language, visuospatial-processing, or judgment/executive function) for

multi-domain MCI; 5) 4 or lower for baseline Hachinski score; 6) does not meet the NINCDS-ADRDA [41] or DSM-IV-TR [2] criteria for dementia; 7) no psychological symptoms or history of depression; and 8) capacity to give informed consent and follow study procedures.

Similarly, the subjects meeting the following criteria were considered as normal healthy controls: 1) age > 55 years and any race; 2) adequate visual and auditory acuity to properly complete neuropsychological testing; 3) no self-report of neurological or depressive illness; 4) shows no evidence of depression based on the Diagnostic Interview Schedule port of the Duke Depression Evaluation Schedule; 5) normal score on a non-focal neurological examination; 6) a score > −1 SD on any formal memory tests and a score > −1 SD on any formal executive function or other cognitive test; and 7) demonstrates a capacity to give informed consent and follow study procedures.

In order for safety purposes and minimizing biases, subjects were excluded from the study if they have: 1) any of the traditional MRI contraindications, such as foreign metallic implants or pacemakers; 2) a past head injury or neurological disorder associated with MRI abnormalities, including dementia, brain tumors, epilepsy, Parkinson's disease, demyelinating diseases, *etc.*; 3) any physical or intellectual disability affecting completion of assessments; 4) documentation of other Axis I psychiatric disorders; and 5) any prescription medication (or nonprescription drugs) with known neurological effects. It is noteworthy that the diagnosis of all cases were made on clinical grounds without reference to MRI. Demographic and clinical information of the participants is summarized in Table 1.

### 3.2 Data Acquisition

Our data were acquired on a 3.0T GE scanner (Signa EXCITE, GE Healthcare) using a SENSE inverse-spiral pulse sequence with the following parameters: Repetition Time (TR)=2,000 ms; Echo Time (TE)=32 ms; flip angle=77°; 64×64 acquisition matrix with a rectangular Field Of View (FOV) of 256×265 $mm^2$; 34 axial slices parallel to the anterior commissure-posterior commissure plane, with voxel size of 4×4×4$mm^3$. In total, 150 rs-fMRI volumes were acquired per scan for each subject. The T1-weighted anatomical MRI images were also acquired on the same scanner with the following parameters: TE=2.976 ms; TR=7.460 ms; flip angle=12°; 256×224 acquisition matrix with a rectangular FOV of 256×256 $mm^2$; slice thickness of 1mm. In total, 216 slices were acquired using the FSPGR ASSET sequence.

During scanning, all the subjects were asked to keep their eyes open and to fixate on a crosshair in the middle of the screen. The experiment lasted for 5 minutes with no change of the stimulus, which allows the neural excitation related to the stimulus to vanish quickly. This procedure also prevents a subject from falling in sleep and unavoidable saccade eyes' movement that may occur with eyes closed.

### 3.3 Preprocessing

Since there were not dummy scans at the beginning, we discarded the first 10 fMRI image volumes of each subject for magnetization equilibrium. In order to remove extraneous

sources of variation and to isolate the fMRI signals, the remaining 140 fMRI images were preprocessed by applying the typical procedures of slice timing, motion correction, and spatial normalization using SPM8[2]. In this study, we realigned images with TR/2 as a reference time point to minimize the relative errors across TRs. In the head motion correction step, we realigned images to the first volume across the subjects.

In this study, the global signal regression was not considered since it is still controversial in the fields if regressing out the global signals is meaningful [19, 35, 44, 63]. However, in order to reduce the effects of CerebroSpinal Fluid (CSF), ventricles, and White Matter (WM), and to focus on the signals of Gray Matter (GM), we regressed out the nuisance signals caused from these regions along with those of the six head-motion profiles. Then we considered only the signals in GM for further processing by minimizing the physiological noises caused by cardiac and respiratory cycles from WM and/or CSF [69].

In the spatial normalization, the fMRI images of each subject were coregistered to their own T1-weighted structure image to avoid the unfavorable smoothness effect that can limit the ability of the normalization procedures to match precisely the corresponding anatomical regions across subjects [5]. The fMRI brain space was then parcellated into 116 ROIs based on the AAL template [67]. In order for the parcellation, we warped the AAL template into the subject naive space with the deformation fields estimated by HAMMER [54]. The representative mean time series of each ROI was computed by averaging the intensity of all voxels in a ROI. Therefore, we had a set of time series $\mathbf{X} \in \{X^{(n)} \in \mathbb{R}^{R\times V}\}_{n=1}^{N}$ where $N$ is the number of subjects, and $R$ and $V$ denote, respectively, the number of ROIs (=116) and the number of volumes (=140).

It is well investigated that the Low Frequency Fluctuation (LFF) in rs-fMRI is a dominant characteristic observed in the resting state brain signals [9]. In order to utilize the LFF features in rs-fMRI, we performed a temporal band-pass filtering with a frequency interval of $0.025 \leq f \leq 0.100$ Hz on $\mathbf{X}$. It has been shown that frequency range between 0.025 and 0.06 or 0.07 is reliable for test-retest experiment [36]. Based on Wee et al.'s work [74], we further decomposed this frequency interval into five equally spaced non-overlapping frequency bands (0.025–0.03929 Hz, 0.03929–0.05357 Hz, 0.05357–0.06786 Hz, 0.06786–0.08214 Hz, 0.08214–0.100 Hz). We can perform frequency-specific analysis of brain features with the frequency-decomposed signals. Finally, the bandpass-filtered regional fMRI time series were used to learn the coefficient matrix $\mathbf{W}_r$ in Eq. (5) over all ROIs, *i.e.*, $r \in \{1, \cdots, R\}$, for each subject.

### 3.4 Finding Functional Connectivity

In terms of learning a classifier, there are in general two methods: supervised and semi-supervised learning. In supervised learning, we construct a classification model from the labeled training samples, each of which consists of an input feature vector and the corresponding target label. Meanwhile, semi-supervised learning takes both labeled and unlabeled samples into account in building a model. That is, the semi-supervised method

[2]Available at 'http://www.fil.ion.ucl.ac.uk/spm/software/spm8/'

first extracts general information from the samples available, regardless of the availability of label information, and then builds a classifier with only the labeled samples. In the real world, there exist a lot of unlabeled data available and from a practical point of view, it can be informative to utilize the unlabeled data to extract meaningful information and thus to build a robust classifier.

In our case of MCI diagnosis, we exploit the concept of the semi-supervised learning and benefit from the brain functional information of the test subject in finding functional connectivity. That is, in order to obtain a robust network structure from a larger number of samples, we use both the training and test samples in optimization of the proposed SDGL. Here, we should emphasize that the label information of the test samples is never used. This approach can also be classified into online learning since it utilizes the information of a new sample in learning a model as time goes by. Although it is inevitable for the computational burden due to the increasing number of samples, we believe that it is still acceptable to spend more time to get a more accurate diagnostic result in the clinic.

Mathematically, since we do not have the label information for the test samples, the composite matrices of $D_1$ and $D_2$ cannot be defined, and thus the optimization problem in Eq. (9) cannot be solved in its current form. To this end, we defined composite matrices $\hat{D_1}$ and $\hat{D_2}$ by concatenating zero-vectors to $D_1$ and $D_2$ in Eqs. (6) and (7) as follows:

$$\hat{D}_i = \begin{bmatrix} D_i & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \in \mathbb{R}^{(K+L)\times(K+L)} \quad (14)$$

where $i \in \{1, 2\}$, and $K$ and $L$ denote, respectively, the total number of training and test samples. Here, it is assumed that the last $L$ samples are for test without loss of generality. By setting the row and column vectors zero, which corresponds to the test samples, and solving the optimization problem of Eq. (9) with the replacement of $D_1$ and $D_2$ with $\hat{D_1}$ and $\hat{D_2}$, we can find the network structures consistent across the training and test samples, and the connectivity coefficients to be similar within a class and distinct between classes. Note that we used the label information of only the training samples. We should further note that while both the training and test samples are utilized in learning the connectivity coefficients, the feature selection and classifier learning were performed with only the samples of the training subjects.

### 3.5 Classification Performance

Since the group Independent Component Analysis (ICA) is one of the most widely used method for rs-fMRI analysis [13,28,65], we considered group ICA-based feature extraction, linked with a linear SVM, as a baseline method. Specifically, we applied a temporal concatenation approach, so called spatial group ICA [13], which decomposed fMRI data into spatial maps of neuronal activation and their respective time courses. For the group ICA, we applied a fast ICA algorithm [25] in a GIFT toolbox[3]. Regarding the number of independent components, we set it to 30 by following Li et al.'s work [28]. After performing

[3]Available at 'http://www.nitrc.org/projects/gift'.

group ICA, for each subject, we computed the correlation coefficients of every pair of time courses and then used them as features to train a linear SVM. A feature selection described in Section 2.4 is also applied for dimension reduction before the SVM learning.

In order for obtaining the optimal regularization parameters in Eq. (5), we considered $\lambda_1 \in \{0.05,, 0.1, 0.15, 0.2, 0.5\}$ and $\lambda_2 \in \{0.05,, 0.1, 0.15, 0.2, 0.5\}$, and further performed a grid search. For the *t*-test in the feature selection step, *p*-values of {0.001, 0.005, 0.01, 0.05} were considered.

Let TP, TN, FP, and FN denote, respectively, true positive, true negative, false positive, and false negative. In this work, we consider the following quantitative measurements to validate the effectiveness of the proposed method by comparing with the competing group lasso [80].

- Accuracy (ACC) = (TP+TN) / (TP+TN+FP+FN)
- Sensitivity (SEN) = TP / (TP+FN)
- Specificity (SPEC) = TN / (TN+FP)
- Receiver operating characteristic (ROC) curve
- Area under the ROC curve (AUC)
- F-Score = 2 × SPEC × SEN / (SPEC+SEN)
- Youden's index (YI) = SEN − (1−SPEC)
- Balanced accuracy (BAC) = (SEN+SPEC) / 2

The most direct comparison between two methods can be the accuracy, which counts the number of correctly classified samples in a test set. Table 2 presents that the proposed method outperforms both the simple group ICA and the conventional group lasso in both single- and multi-spectrum approaches with the performances of 86.49% and 89.19% in single- and multi-spectrum, respectively. It is noteworthy that the proposed method with a single-spectrum approach outperformed the conventional group lasso with a multi-spectrum approach. That is, the proposed method can produce a higher accuracy with a smaller computational cost. Here, we should also note that the accuracy of the group lasso in multi-spectrum is lower than the one reported by Wee et al. [77]. The main reason for that comes from the difference in preprocessing. In this work, we regressed out the nuisance signals from the regions of CSF and WM as done in [77], but we further considered the six head-motion profiles in our regression. From a signal processing point of view, the regression step allows us to acquire more noise-free signals to be analyzed and the application of the six head-motion profiles to regression is helpful for robustness to noises. Henceforth, we believe that the result from our experiment is more faithful. For reference, the performances obtained from the data, which followed the same preprocessing in Wee et al.'s work, *i.e.*, without using head-motion profiles in regression, are also presented in Table 2.

Regarding the sensitivity and specificity, the higher the sensitivity, the lower the chance of mis-diagnosing MCI patients, and the higher the specificity, the lower the chance of mis-diagnosing normal subjects to MCI. Although the specificity of the proposed method is similar or slightly better than the other methods, the proposed SDGL in multi-spectrum

overwhelms the competing methods, reporting a sensitivity of 0.9167. Clinically, it is much more beneficial to have a high sensitivity, *i.e.*, correct identification of MCI patients.

One of the most effective methods of evaluating the performance of diagnostic tests in brain disease as well as other medical areas is the Receiver Operating Characteristic (ROC) curve, which is defined as a plot of test true positive rate versus its false positive rate. We illustrated the ROC curve along with the partest graph of the multi-spectrum SDGL in Fig. 4, which produced the highest accuracy. In accordance with the ROC, the Area Under the ROC Curve (AUC), a combined measure of sensitivity and specificity, is also widely considered for performance measurement. The AUC can be thought as a measure of the overall performance of a diagnostic test. The larger the AUC, the better the overall performance of the diagnostic test. The AUC of the multi-spectrum SDGL is 0.9567, which also outperforms the other methods.

Youden's Index (YI) [79], which equally weights the performance of a method on positive and negative samples, evaluates the ability of a method to avoid failure in diagnosis. The higher the value of YI, the better the ability to avoid failure in diagnosis. Based on the YIs in Table 2, we can say that the diagnosis result from the proposed method along with multi-spectrum, whose YI is 0.7967, is much more reliable than the competing methods.

Based on the measurements mentioned above, it is obvious that the proposed SDGL method clearly outperforms both the simple group ICA and the conventional group lasso. With respect to the spectrum analysis, the multi-spectrum approach showed better performance in both the proposed SDGL and the conventional group lasso. However, the proposed method showed an improvement of 10% in both single- and multi-spectrum approaches. We believe that since the main goal of the application of the group lasso to rs-fMRI is to find out the underlying functional network itself, not discrimination between classes, it fails to extract class-discriminative features from the bandpass-filtered signals. Meanwhile, since the proposed SDGL considers both the group-constrained representation and the class-discriminative characteristics simultaneously, it succeeded in correctly identifying MCI patients in both single- and multi-spectrum approaches.

### 3.6 Discussion

In order to see which ROIs are discriminative for MCI identification, we presented the distribution of the ROIs selected by the proposed SDGL in Fig. 5. It should be noted that each frequency band in the multi-spectrum approach shows different distributions. Here, we define the Most Discriminant ROIs (MDRs) and the Second most Discriminant ROIs (SDRs) based on the following rules:

- MDRs = $\{r : F_i(r) > \mu_i + 2\sigma_i, \forall_i\}$
- SDRs = $\{r : \mu_i + \sigma_i < F_i(r) < \mu_i + 2\sigma_i, \forall_i\}$

where $F_i(r)$ is the frequency of the $r$-th ROI to be selected in the $i$-th frequency band, $\mu_i = \frac{1}{R}\sum_{r=1}^{R} F_i(r)$ and $\sigma_i = \left[\frac{1}{R}\sum_{r=1}^{R}(F_i(r) - \mu_i)^2\right]^{\frac{1}{2}}$ denote, respectively, the mean and the standard deviation of the frequencies.

To sum up, the selected ROIs coincide with the studies on the MCI and/or AD in the literature. The MDRs are marked in Fig. 6(a): Left Posterior Cingulate Gyrus (35) [4, 63], Left Postcentral Gyrus (57) [71], Left Putamen (73) [24], Left Lobule IV, V of Cerebellar Hemisphere (97), Left Lobule VI of Cerebellar Hemisphere (99), and Lobule VI of Vermis (112) [64]. The SDRs are marked in Fig. 6(b): Right Anterior Cingulate Gyrus (32) [16], Right Amygdala (42) [6], Left Fusiform Gyrus (55) [10], Left Pallidum (75) [63], Left Crus I of Cerebellar Hemisphere (91), Right Lobule IV, V of Cerebellar Hemisphere (98), Right Lobule IX of Cerebellar Hemisphere (106), and Lobule VII Vermis (113) [64]. The number in a parenthesis denotes an index of an ROI in an AAL template.

The functional connectivity maps estimated from the multi-bandpass filtered signals with the proposed SDGL and the conventional group lasso are presented in Fig. 7. Although they are not exactly the same, there exists the similarity between the methods in terms of the network structure, which means that the proposed SDGL holds the group constraint characteristic across subjects as the group lasso does. Meanwhile, it is clear that they show different connectivity coefficients in each frequency range. Based on the performances obtained in our experiments, we can say that the coefficient differences resulted from the proposed label-informed regularization terms in Eq. (5) help enhance the classification accuracy.

For a detailed representation of the functional connectivity in terms of brain areas, lobes, and hemispheres, we also created a connectogram using Circos [27] in Fig. 8. We should note that since the connection weights between ROIs vary across subjects regardless of the classes of NC and MCI, we cannot directly visualize them in a single graph. We circumvented this issue as follows. First, we count the number of valid connections between two ROIs over all subjects, in which any connection was considered as valid if and only if its absolute value of the connectivity coefficient is larger than the mean of the absolute coefficients in a subject. We accumulated the counts over experiments, *i.e.*, cross-validation, and then normalized it.

From the figure, we can see that the connectivity varies across decomposed frequency bands (Fig. 8(b) – Fig. 8(f)). Interestingly, the connections are the densest in the frequency band of [0.025–0.03929], which means that a huge amount of the network connections occur in the low frequency range. There is a tendency for the connections to concentrate on a small number of ROIs in the higher frequency ranges: Posterior Cingulate Gyrus Left (35) and Heschl Gyrus Left (79) in [0.03929–0.05357], Vermis 9 (115) in [0.05357–0.06786], Posterior Cingulate Gyrus Left (35) and Pallidum Left (75) in [0.06786–0.08214], and Temporal Pole: Superior Temporal Gyrus Right (84) in [0.08214–0.1].

Meanwhile, although it is sparser, the functional connectivity estimated from a single-spectrum approach is similar to that of the lowest frequency band in a multi-spectrum approach. Based on the similarity of those functional connectivities as well as the similar classification accuracy of the proposed method with single- and multi-spectrum approaches in Table 2, we hypothesize that the discriminative information between NC and MCI is mostly involved in the connectivities commonly observed in Fig. 8(a) and Fig. 8(b). The functional connectivities shown in Figs. 8(c)–8(f) are supplementary to further improve the accuracy of 2.7%. Therefore, we believe that it would be interesting to further analyze the

functional connectivities commonly observed in Fig. 8(a) and Fig. 8(b), which we would like to leave as our forthcoming research issue.

Note that since we applied leave-one-out cross-validation, the functional connectivity of a subject estimated in different cross-validation trials could be different. In this regard, we measured the variability of the networks by computing the standard deviation of the functional connectivities by different cross-validation trials and further visualized the respective connectivity variation maps in Fig. 9. In the map, each element denotes the standard deviation of functional connectivity between the corresponding ROIs. From the maps, we can see that the maximal standard deviation is less than 0.5, implying that the network structures, *i.e.*, functional connectivity, obtained with the proposed method are stable.

## 4 Conclusion

Since it is reported that a large portion of MCI patients progresses to AD [1], there has been great interest in many scientific fields for early detection of MCI and a proper treatment. In this paper, we proposed a novel method of identifying MCI with group sparse representation in a supervised and discriminative manner. Specifically, in order to reflect the categorical or class-label information in the model, we utilized a well-known discriminative information of the within-class-variance and the between-class-variance [21] for penalization. We also devised an efficient algorithm to optimize our objective function using the accelerated gradient method. It is noteworthy that the proposed method jointly learns the coherent brain network structures across subjects regardless of the classes, while imposing similar connectional coefficients within a class and distinct coefficients between classes, but still maintaining individual network characteristics.

Our experimental results on rs-fMRI data validated the effectiveness of the proposed method with the classification accuracy of 89.19% and the sensitivity of 0.9167 in a multi-spectrum approach. The class-discriminative ROIs selected in our framework coincide with those reported in the studies on MCI and AD in the literature. It is also observed that the functional connectivities vary across the frequency ranges, showing the densest connectivities in the low frequency range of [0.025–0.03929].

Some studies have shown that sub-regions within larger rs-fMRI networks can be different in different disease or behavioral states [37]. From this perspective, artificially limiting the connectome to the coarse AAL template can be limitation of this study. It would be an interesting issue to verify the performance of the proposed method on a finer brain parcellation scale with higher resolution images and more number of ROIs. Finally, although we demonstrated and validated the superiority of the proposed method in our experiments, the sample size of the rs-fMRI images was very limited and unbalanced.

## Acknowledgements

This work was supported in part by NIH grants EB006733, EB008374, EB009634, AG041721, MH100217, and AG042599, and also by ICT R&D program of MSIP/IITP (14-824-09-014, Basic Software Research in Human-level Lifelong Machine Learning) funded by the Korean government.

## References


1. Alzheimer's Association. 2012 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2012; 8(2):131–168.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition - Text Revision (DSMIV-TR). American Psychiatric Association. 2000
3. Anand A, Li Y, Wang Y, Wu J, Gao S, Bukhari L, Mathews VP, Kalnin A, Lowe MJ. Activity and connectivity of brain mood regulating circuit in depression: A functional magnetic resonance study. Biological Psychiatry. 2005; 57(10):1079–1088. [PubMed: 15866546]
4. Bai F, Watson DR, Yu H, Shi Y, Yuan Y, Zhang Z. Abnormal resting-state functional connectivity of posterior cingulate cortex in amnestic type mild cognitive impairment. Brain Research. 2009; 1302:167–174. [PubMed: 19765560]
5. Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M, Peterson BS. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. PLoS ONE. 2012; 7(12):e50,698.
6. Basso M, Yang J, Warren L, MacAvoy M, Varma P, Bronen R, Dyck C. Volumetry of amygdala and hippocampus and memory performance in alzheimer's disease. Psychiatry Research: Neuroimaging. 2006; 146(3):251–261.
7. Benton AL. The visual retention test as a constructional praxis task. Confinia Neurologica. 1962; 22:141–155. [PubMed: 13967555]
8. Benton, AL.; Hamsher, K. Multilingual Aphasia Examination manual. Iowa City: University of Iowa; 1976.
9. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine. 1995; 34(4):537–541. [PubMed: 8524021]
10. Bokde ALW, Lopez-Bayo P, Meindl T, Pechler S, Born C, Faltraco F, Teipel SJ, Mller HJ, Hampel H. Functional connectivity of the fusiform gyrus during a face-matching task in subjects with mild cognitive impairment. Brain. 2006; 129(5):1113–1124. [PubMed: 16520329]
11. Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network. Annals of the New York Academy of Sciences. 2008; 1124:1–38. [PubMed: 18400922]
12. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience. 2009; 312:186–198.
13. Calhoun V, Adali T, Pearlson G, Pekar J. A method for making group inferences from functional MRI data using independent component analysis. Human Brain Mapping. 2001; 14(3):140–151. [PubMed: 11559959]
14. Cooper JA, Sagar HJ, Jordan N, Harvey NS, Sullivan EV. Cognitive impairment in early, untreated parkinsons disease and its relationship to motor function. Brain. 1991; 114(5):2095–2122. [PubMed: 1933236]
15. Craddock RC III, PE H, Hu XP, Mayberg HS. Disease state prediction from resting state functional connectivity. Magnetic Resonance in Medicine. 2009; 62:1619–1628. [PubMed: 19859933]
16. Dai W, Lopez O, Carmichael O, Becker J, Kuller L, Gach H. Mild cognitive impairment and alzheimer disease: patterns of altered cerebral blood flow at mr imaging. Radiology. 2009; 250(3): 856–866. [PubMed: 19164119]
17. Fan Y, Rao H, Hurt H, Giannetta J, Korczykowski M, Shera D, Avants BB, Gee JC, Wang J, Shen D. Multivariate examination of brain abnormality using both structural and functional MRI. NeuroImage. 2007; 36(4):1189–1199. [PubMed: 17512218]
18. Folstein MF, Folstein SE, McHugh PR. "mini-mental state". A practical method for grading the cognitive state of patient for the clinician. Journal of Psychiatric Research. 1975; 12(3):189–198. [PubMed: 1202204]
19. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Essen DCV, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(27):9673–9678. [PubMed: 15976020]

20. Friston KJ, Frith CD, Liddle PF, Frackowiak RS. Functional connectivity: The principal-component analysis of large (PET) data sets. Journal of Cerebral Blood Flow and Metabolism. 1993; 13:5–14. [PubMed: 8417010]
21. Fukunaga, K. Introduction to Statistical Pattern Recognition, 2 edn. Academic Press Professional; 1990.
22. Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, Reiss AL, Schatzberg AF. Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus. Biological Psychiatry. 2007; 62(5):429–437. [PubMed: 17210143]
23. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimers disease from healthy aging: Evidence from functional MRI. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(13):4637–4642. [PubMed: 15070770]
24. Han SD, Arfanakis K, Fleischman DA, Leurgans SE, Tuminello ER, Edmonds EC, Bennett DA. Functional connectivity variations in mild cognitive impairment: Associations with cognitive function. Journal of the International Neuropsychological Society. 2012; 18:39–48. [PubMed: 22005016]
25. Hyv A. Independent component analysis: algorithms and applications. Neural Networks. 2000; 13(4–5):411–430. DOI http://dx.doi.org/10.1016/S0893-6080(00)00026-5. [PubMed: 10946390]
26. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nature Neuroscience. 2005; 8(5):679–685.
27. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Research. 2009; 19(9):1639–1645. [PubMed: 19541911]
28. Li S, Eloyan A, Joel S, Mostofsky S, Pekar J, Bassett SS, Caffo B. Analysis of group ICA-based connectivity measures from fMRI: Application to Alzheimer's disease. PLoS ONE. 2012; 7(11):e49,340.
29. Li SJ, Li Z, Wu G, Zhang MJ, Franczak M, Antuono PG. Alzheimer disease: Evaluation of a functional MR imaging index as a marker. Radiology. 2002; 225:253–259. [PubMed: 12355013]
30. Liang M, Zhou Y, Jiang T, Liu Z, Tian L, Liu H, Hao Y. Widespread functional disconnectivity in Schizophrenia with resting-state functional magnetic resonance imaging. Neuroreport. 2006; 17:209–213. [PubMed: 16407773]
31. Liu, J.; Ji, S.; Ye, J. Multi-task feature learning via efficient l2,1-norm minimization; Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence; 2009. p. 339-348.
32. Liu, J.; Ji, S.; Ye, J. SLEP: Sparse Learning with Efficient Projections. Arizona State University; 2009.
33. Liu M, Zhang D, Shen D. Ensemble sparse classification of Alzheimer's disease. NeuroImage. 2012; 60(2):1106–1116. [PubMed: 22270352]
34. Liu M, Zhang D, Shen D. the Alzheimer's Disease Neuroimaging Initiative. Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. Human Brain Mapping. 2013; 35(4):1305–1319. [PubMed: 23417832]
35. Lynall ME, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, Bullmore ET. Functional connectivity and brain networks in schizophrenia. Journal of Neuroscience. 2010; 30:9477–9487. [PubMed: 20631176]
36. Malinen S, Vartiainen N, Hlushchuk Y, Koskinen M, Ramkumar P, Forss N, Kalso E, Hari R. Aberrant temporal and spatial brain activity during rest in patients with chronic pain. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:6493–6497. [PubMed: 20308545]
37. Marrelec G, Fransson P. Assessing the influence of different roi selection strategies on functional connectivity analyses of fmri data acquired during steady-state conditions. PLoS ONE. 2011; 6(4):e14,788.
38. Martinez AM, Mart'inez AM, Kak AC. Pca versus lda. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 23:228–233.

39. Matthews, CG.; Klove, H. Instruction Manual for the Adult Neuropsychology Test Battery. Madison, WI: University of Wisconsin Medical School; 1964.
40. Mcintosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitzl B. Network analysis of cortical visual pathways mapped with pet. Journal of Neuroscience. 1994; 14:655–666. [PubMed: 8301356]
41. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984; 34(7):939–944. [PubMed: 6610841]
42. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology. 1989; 39(9):1159–1165. [PubMed: 2771064]
43. Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, Reiman EM, Holthoff V, Kalbe E, Sorbi S, Diehl-Schmid J, Perneczky R, Clerici F, Caselli R, Beuthien-Baumann B, Kurz A, Minoshima S, de Leon MJ. Multicenter standardized 18f-fdg pet diagnosis of mild cognitive impairment, alzheimer's disease, and other dementias. The Journal of Nuclear Medicine. 2008; 49(3):390–398.
44. Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? NeuroImage. 2009; 44(3):893–905. [PubMed: 18976716]
45. Nesterov, Y. Introductory Lectures on Convex Optimization: A Basic Course (Applied Optimization). 1 edn.. Netherlands: Springer; 2009.
46. Ng B, Abugharbieh R. Generalized sparse regularization with application to fmri brain decoding. Proceedings of the 22nd international conference on Information processing in medical imaging, IPMI'11. 2011:612–623.
47. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005; 27(8):1226–1285. [PubMed: 16119262]
48. Penny W, Stephan K, Mechelli A, Friston K. Modelling functional integration: a comparison of structural equation and dynamic causal models. NeuroImage. 2004; 23 Supplement 1(0):S264–S274. [PubMed: 15501096]
49. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: A tutorial overview. NeuroImage. 2009; 45:199–209.
50. Rakotomamonjy A. Variable selection using SVM based criteria. Journal of Machine Learning Research. 2003; 3:1357–1370.
51. Reitan RM. Validity of the trail making test as an indicator of organic brain damage. Perceptual and Motor Skills. 1958; 8:271–276.
52. Reitan, RM.; Wolfson, D. Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. Tucson, AZ: Neuropsychological Press; 1993.
53. Rombouts SARB, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: An fMRI study. Human Brain Mapping. 2005; 26(4):231–239. [PubMed: 15954139]
54. Shen D, Davatzikos C. HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging. 2002; 21(11):1421–1439. [PubMed: 12575879]
55. Shipley, WC. Western Psychological Services. Los Angeles: Calif; 1946. Institute of Living Scale.
56. Smith A. The symbol-digit modalities test: a neuropsychologic test of learning and other cerebral disorders. Learning Disorders. 1968; 3:83–91.
57. Sorg C, Riedl V, Mhlau M, Calhoun VD, Ler TEL, Drzezga A, Kurz HFA, Zimmer C, Wohlschlger AM. Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104(47):18,760–18,765.

58. Sporns O, Toning G, Edelman G. Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. Cerebral Cortex. 2000; 10(2):127–141. [PubMed: 10667981]
59. Sporns O, Zwi JD. The small world of the cerebral cortex. Neuroinformatics. 2004; 2:145–161. [PubMed: 15319512]
60. Squire LR, Zouzounis JA. Self-ratings of memory dysfunction: different findings in depression and amnesia. Journal of Clinical and Experimental Neuropsychology. 1988; 10(6):727–738. [PubMed: 3235647]
61. Stebbins G, Murphy C. Diffusion tensor imaging in alzheimer's disease and mild cognitive impairment. Behavioural Neurology. 2009; 21(1):39–49. [PubMed: 19847044]
62. Suk HI, Lee SW. A novel bayesian framework for discriminative feature extraction in brain-computer interfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(2):286–299. [PubMed: 22431526]
63. Supekar K, Menon V, Rubin D, Musen M, Greicius MD. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Computational Biology. 2008; 4:e1000,100.
64. Thomann PA, Schlfer C, Seidl U, Santos VD, Essig M, Schrder J. The cerebellum in mild cognitive impairment and alzheimers disease a structural mri study. Journal of Psychiatric Research. 2008; 42(14):1198–1202. [PubMed: 18215400]
65. Tian L, Kong Y, Ren J, Varoquaux G, Zang Y, Smith SM. Spatial vs. temporal features in ICA of resting-state fMRI - a quantitative and qualitative investigation in the context of response inhibition. PLoS ONE. 2013; 8(6):e66,572.
66. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. 1996; 58(1):267–288.
67. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002; 15(1):273–289. [PubMed: 11771995]
68. Uddin LQ, Kelly AC, Biswal BB, Castellanos FX, Milham MP. Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. Human Brain Mapping. 2009; 30(2):625–637. [PubMed: 18219617]
69. Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic functional connectivity as a tool for human connectomics: Theory, properties and optimization. Journal of Neurophysiology. 2010; 103:297–321. [PubMed: 19889849]
70. Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T. Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. Human Brain Mapping. 2007; 28(10):967–978. [PubMed: 17133390]
71. Wang Z, Nie B, Li D, Zhao Z, Han Y. Effect of acupuncture in mild cognitive impairment and alzheimer disease: A functional mri study. PLoS ONE. 2012; 7(8):e42,730.
72. Wechsler, D. Manual for the Wechsler Adult Intelligence Scale - Revised. New York: Psychological Corporation; 1981.
73. Wechsler, D. WMS-R: Wechsler Memory Scale-Revised Manual. The Psychological Corporation; 1987.
74. Wee CY, Yap PT, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PLoS ONE. 2012; 7(5):e37,828.
75. Wee CY, Yap PT, Li W, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Enriched white matter connectivity networks for accurate identification of MCI patients. Neuroimage. 2011; 54(3):1812–1822. [PubMed: 20970508]
76. Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Identification of mci individuals using structural and functional connectivity networks. Neuroimage. 2012; 59(3):2045–2056. [PubMed: 22019883]
77. Wee, CY.; Yap, PT.; Zhang, D.; Wang, L.; Shen, D. Constrained sparse functional connectivity networks for mci classification. In: Ayache, N.; Delingette, H.; Golland, P.; Mori, K., editors.

Medical Image Computing and Computer-Assisted Intervention, MICCAI 2012, Lecture Notes in Computer Science. Vol. 7511. Berlin Heidelberg: Springer; 2012. p. 212-219.

78. Wu L, Eichele T, Calhoun VD. Reactivity of hemodynamic responses and functional connectivity to different states of alpha synchrony: A concurrent eeg-fmri study. NeuroImage. 2010; 52(4): 1252–1260. [PubMed: 20510374]
79. Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3(1):32–35. [PubMed: 15405679]
80. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B. 2006; 68(1):49–67.
81. Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. NeuroImage. 2012; 59(2):895–907. [PubMed: 21992749]
82. Zhang D, Shen D. Alzheimer's Disease Neuroimaging, I. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS ONE. 2012; 7(3):e33,182.
83. Zhou, J.; Liu, J.; Narayan, VA.; Ye, J. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012. Modeling disease progression via fused sparse group lasso; p. 1095-1103.
84. Zhou L, Wang Y, Li Y, Yap PT, Shen D. the Alzheimer's Disease Neuroimaging, I. Hierarchical anatomical brain networks for MCI prediction: Revisiting volumetric measures. PLoS ONE. 2011; 6(7):e21,935.
85. Zhou Y, Liang M, Tian L, Wang K, Hao Y, Liu H, Liu Z, Jiang T. Functional disintegration in paranoid Schizophrenia using resting-state fMRI. Schizophrenia Research. 2007; 97(1–3):194–205. [PubMed: 17628434]
86. Zhuang J, Peltier S, He S, LaConte S, Hu X. Mapping the connectivity with structural equation modeling in an fmri study of shape-from-motion task. NeuroImage. 2008; 42(2):799–806. [PubMed: 18599316]

# Appendix A

## Derivation of the Definitive Matrices

$$\begin{aligned}
&\frac{1}{N^c}\sum_{ns.t.l(n)=c}\left(\mathbf{w}_{r,g}(n)-\hat{w}^c_{r,g}\right)^2\\
&=\frac{1}{N^c}\sum_{ns.t.l(n)=c}\left(\mathbf{w}_{r,g}\mathbf{e}^c_n-\mathbf{w}_{r,g}\mathbf{m}^c\right)^2\\
&=\frac{1}{N^c}\sum_{ns.t.l(n)=c}\{\mathbf{w}_{r,g}(\mathbf{e}^c_n-\mathbf{m}^c)\}\{\mathbf{w}_{r,g}(\mathbf{e}^c_n-\mathbf{m}^c)\}^T\\
&=\frac{1}{N^c}\sum_{ns.t.l(n)=c}\mathbf{w}_{r,g}(\mathbf{e}^c_n\\
&-\mathbf{m}^c)(\mathbf{e}^c_n-\mathbf{m}^c)^T\mathbf{w}^T_{r,g}=\frac{1}{N^c}\mathbf{w}_{r,g}\left(\sum_{ns.t.l(n)=c}\mathbf{e}^c_n\mathbf{e}^{cT}_n\right.\\
&\left.-\mathbf{e}^c_n\mathbf{m}^{cT}-\mathbf{m}^c\mathbf{e}^{cT}_n+\mathbf{m}^c\mathbf{m}^{cT}\right)\mathbf{w}^T_{r,g}\\
&=\mathbf{w}_{r,g}\frac{1}{N^c}(\mathbf{I}_c\\
&-\mathbf{I}^c\mathbf{M}^{cT}-\mathbf{M}^c\mathbf{I}^{cT}+\mathbf{M}^c\mathbf{M}^{cT})\mathbf{w}^T_{r,g}\\
&=\mathbf{w}_{r,g}K^c\mathbf{w}^T_{r,g}
\end{aligned}\tag{15}$$

where $N^c$ is the number of training samples of the class $c \in \{+, -\}$, $\mathbf{w}_{r,g}(n)$ denotes the $n$-th element of a vector $\mathbf{w}_{r,g}$, $\hat{w}^c_{r,g}=\frac{1}{N^c}\sum_{ns.t.l(n)=c}\mathbf{w}_{r,g}(n)$, $\mathbf{e}^c_n$ is an $N$-dimensional unit vector with the $n$-th element 1 if $l(n) = c$, 0 otherwise, $\mathbf{I}$ is a square diagonal matrix with $\mathbf{I}^c_{nn}=1$ if $l(n) = c$, 0 otherwise, $\mathbf{M}^c$ is a square matrix with the columns set to $\mathbf{m}^c=[m^c_1,\cdots,m^c_n,\cdots,m^c_N]^T$, and $m^c_n=\frac{1}{N^c}$ if $l(n) = c$, otherwise 0.

$K^c=\frac{1}{N^c}(\mathbf{I}^c-\mathbf{I}^c\mathbf{M}^{cT}-\mathbf{M}^c\mathbf{I}^{cT}+\mathbf{M}^c\mathbf{M}^{cT})$.

$$
\begin{aligned}
& f_W(\mathbf{w}_{r,g}) \\
& = \frac{1}{N^+} \sum_{n \text{s.t.} l(n)=`+'} (\mathbf{w}_{r,g}(n) - \hat{w}^+_{r,g})^2 \\
& + \frac{1}{N^-} \sum_{n \text{s.t.} l(n)=`-'} (\mathbf{w}_{r,g}(n) - \hat{w}^-_{r,g})^2 \\
& = \mathbf{w}_{r,g} K^+ \mathbf{w}^T_{r,g} \\
& + \mathbf{w}_{r,g} K^- \mathbf{w}^T_{r,g} \\
& = \mathbf{w}_{r,g} (K^+ \\
& \quad + K^-) \mathbf{w}^T_{r,g} = \mathbf{w}_{r,g} \hat{K} \mathbf{w}^T_{r,g} = \mathbf{w}_{r,g} D_1 D_1^T \mathbf{w}^T_{r,g} = (\mathbf{w}_{r,g} D_1)(\mathbf{w}_{r,g} D_1)^T \\
& = \|\mathbf{w}_{r,g} D_1\|_2^2 (\text{matrix decomposition})
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
& f_B(\mathbf{w}_{r,g}) \\
& \quad = (\hat{w}^+_{r,g} - \hat{w}^-_{r,g})^2 \\
& \quad = (\mathbf{w}_{r,g} \mathbf{m}^+ - \mathbf{w}_{r,g} \mathbf{m}^-)^2 \\
& \quad = \{\mathbf{w}_{r,g}(\mathbf{m}^+ - \mathbf{m}^-)\}\{\mathbf{w}_{r,g}(\mathbf{m}^+ - \mathbf{m}^-)\}^T = \mathbf{w}_{r,g}(\mathbf{m}^+ \\
& \quad \quad - \mathbf{m}^-)(\mathbf{m}^+ - \mathbf{m}^-)^T \mathbf{w}^T_{r,g} \\
& \quad = (\mathbf{w}_{r,g} D_2)(\mathbf{w}_{r,g} D_2)^T \\
& \quad = \|\mathbf{w}_{r,g} D_2\|_2^2
\end{aligned}
\tag{17}
$$

where $D_2 = (\mathbf{m}^+ - \mathbf{m}^-)$.

## Appendix B

## Proof of Two Stage Proximal Operator

Given a target proximal operator of

$$
\pi(\mathbf{v}) = \underset{\mathbf{w}}{\text{argmin}} \frac{1}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2 + \lambda_2 \left( \|\mathbf{w} D_1\|_2^2 - \|\mathbf{w} D_2\|_2^2 \right). \tag{18}
$$

we can decompose it into two proximal operators as follows:

$$
\pi_1(\mathbf{v}) = \underset{\mathbf{w}}{\text{argmin}} \frac{1}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2 \tag{19}
$$

$$
\pi_2(\mathbf{v}) = \underset{\mathbf{w}}{\text{argmin}} \frac{1}{2} \|\mathbf{w} - \mathbf{v}\|_2^2 + \lambda_2 \left( \|\mathbf{w} D_1\|_2^2 - \|\mathbf{w} D_2\|_2^2 \right). \tag{20}
$$

Then it holds that

$$
\pi(\mathbf{v}) = \pi_2(\pi_1(\mathbf{v})). \tag{21}
$$

The necessary and sufficient optimality conditions for Eq. (18), Eq. (19), and Eq. (20) can be written as

$$0 \in \pi(\mathbf{v}) - \mathbf{v} + \lambda_1 \partial g(\pi(\mathbf{v})) + \lambda_2 \partial h(\pi(\mathbf{v})) \quad (22)$$

$$0 \in \pi_1(\mathbf{v}) - \mathbf{v} + \lambda_1 \partial g(\pi(\mathbf{v})) \quad (23)$$

$$0 \in \pi_2(\pi_1(\mathbf{v})) - \pi_1(\mathbf{v}) + \lambda_2 \partial h(\pi_2(\pi_1(\mathbf{v}))) \quad (24)$$

where the partial derivatives are defined as

$$\partial g(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|_2} & \text{if} \mathbf{x} \neq \mathbf{0} \\ \{\mathbf{y} : \|\mathbf{y}\|_2 \leq 1\} & \text{otherwise} \end{cases} \quad (25)$$

and

$$\partial h(\mathbf{x}) = 2\mathbf{x}(D_1 D_1^T - D_2 D_2^T). \quad (26)$$

It follows from Eq. (24) and Eq. (26) that if $\pi_1(\mathbf{v}) = 0$ then $\pi_2(\pi_1(\mathbf{v})) = 0$. That is, the group sparsity $\pi_1(\mathbf{v})$ via the group lasso still holds for $\pi_2(\pi_1(\mathbf{v}))$. Therefore, we have

$$0 \in \pi_2(\pi_1(\mathbf{v})) - \pi_1(\mathbf{v}) + \lambda_2 \pi_2(\pi_1(\mathbf{v}))(D_1 D_1^T - D_2 D_2^T). \quad (27)$$

Since Eq. (18) has a unique solution, we can get Eq. (21) from Eq. (22) and Eq. (27). Note that thanks to the matrix multiplication of $D_1 D_1^T$ in the partial derivative of Eq. (26), there is no need to explicitly decompose the matrix $\hat{K}$ in Eq. (16).

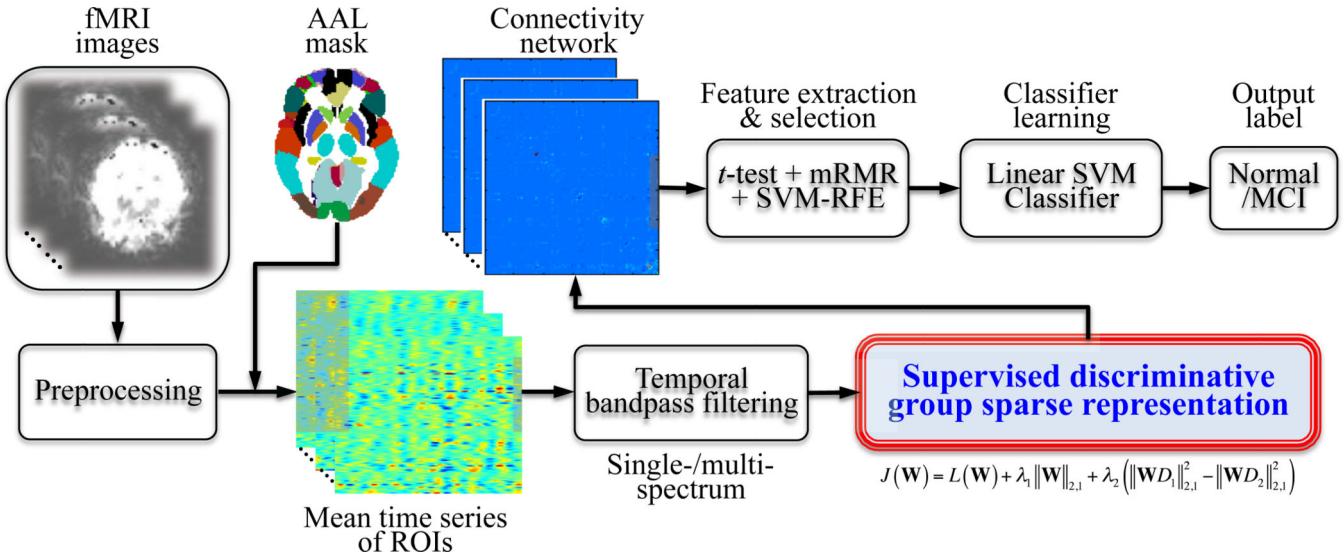



**Fig. 1.**

A schematic diagram of the proposed framework for MCI classification. (AAL: Automated Anatomical Labeling, ROI: Region Of Interest, mRMR: minimum-Redundancy Maximum-Relevance, SVM: Support Vector Machine, RFE: Recursive Feature Elimination).

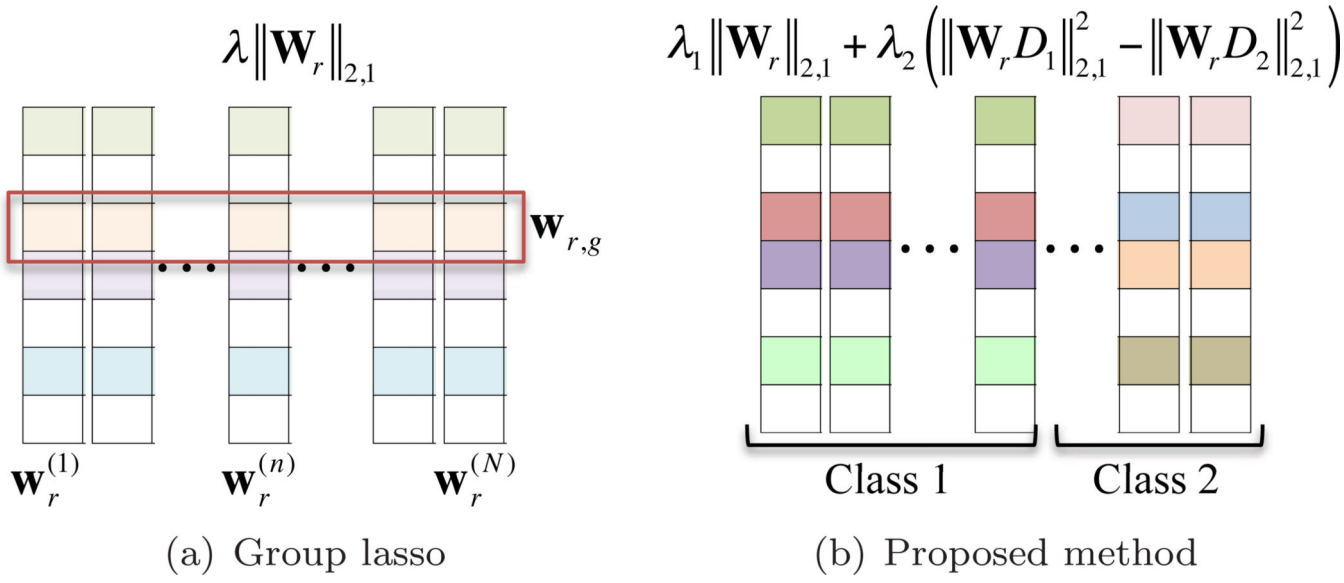



**Fig. 2.**

Pictorial illustration of the concept and the regularization term(s) in an objective function of the conventional group lasso and the proposed supervised discriminative group lasso.

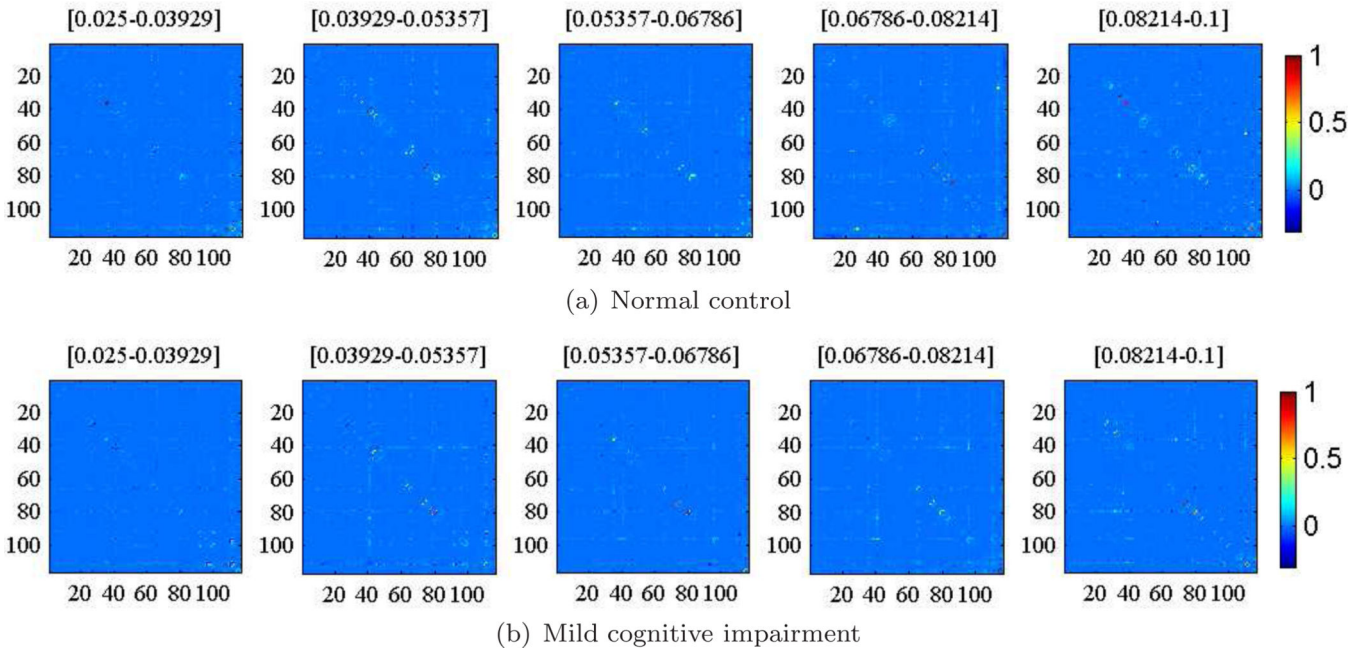



**Fig. 3.**

Sample functional connectivities of normal control and mild cognitive impairment estimated by the proposed method.

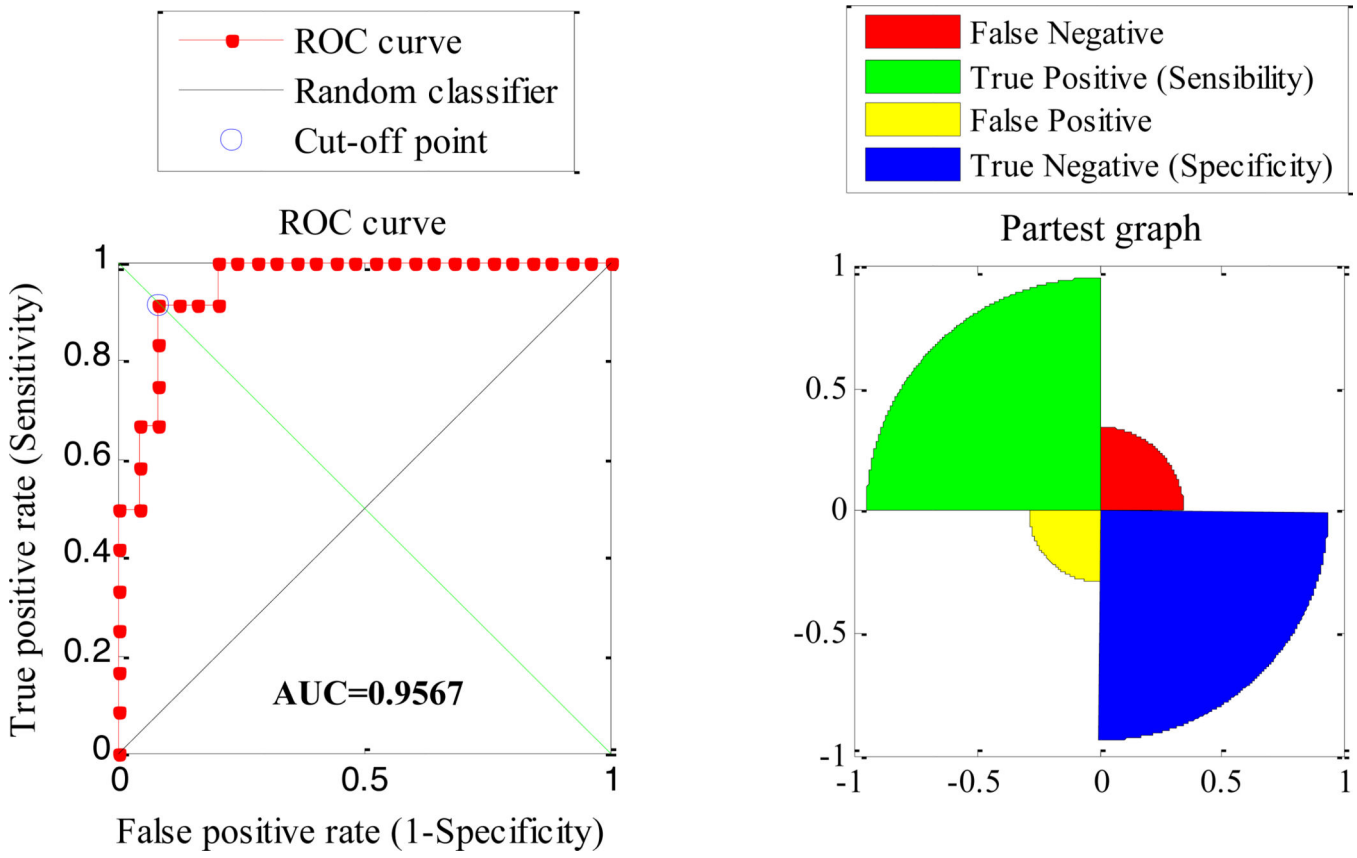



**Fig. 4.**

A receiver operating characteristic (ROC) curve and a partest graph of the proposed supervised discriminative group lasso with a multi-spectrum approach.

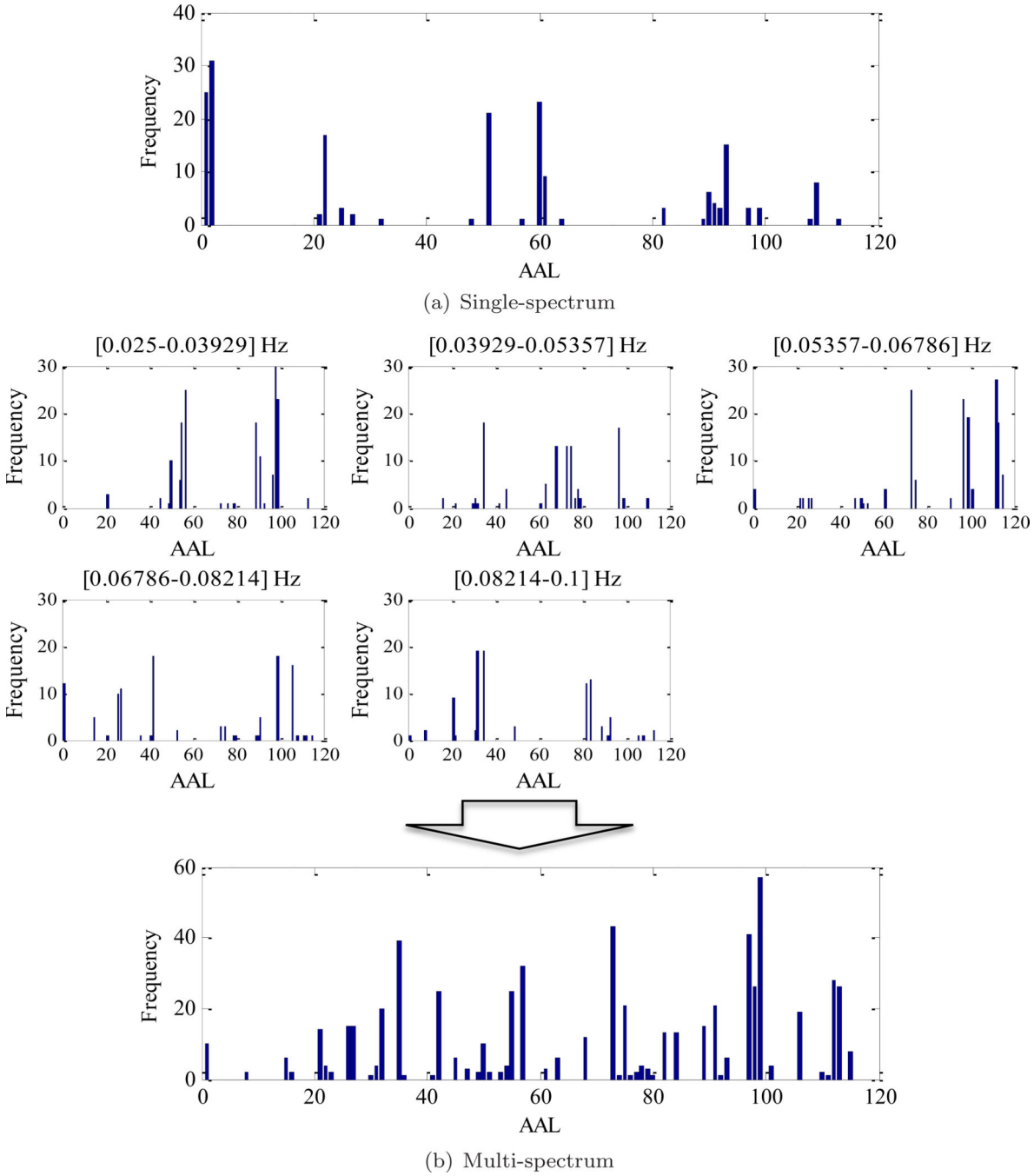
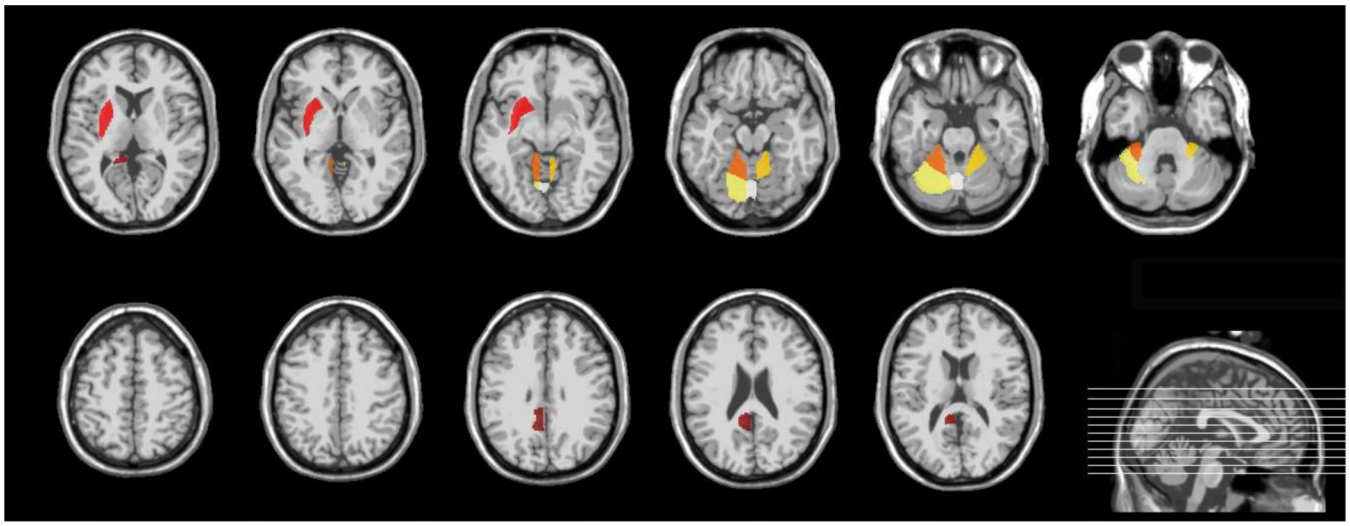
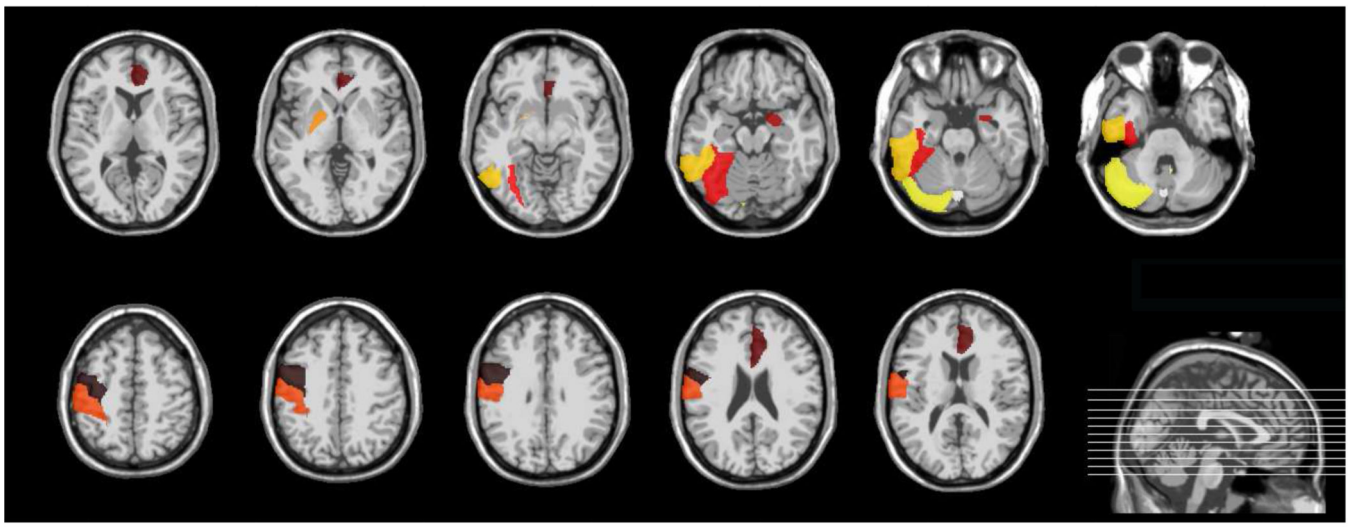



**Fig. 5.**

Distributions of the selected ROIs in the proposed supervised discriminative group lasso. The y-axis denotes a frequency of a ROI being selected in classification. For the multi-spectrum case, the upper five small graphs are from each of the decomposed frequency bands.



(a) Most discriminant ROIs: ≥(Mean+2SD)



(b) Second most discriminant ROIs: ≥(Mean+SD)

**Fig. 6.**

The most and second most discriminative ROIs selected by the proposed supervised discriminative group lasso. Refer to the text for the meaning of the most and second most discriminative ROIs. (SD: Standard Deviation)

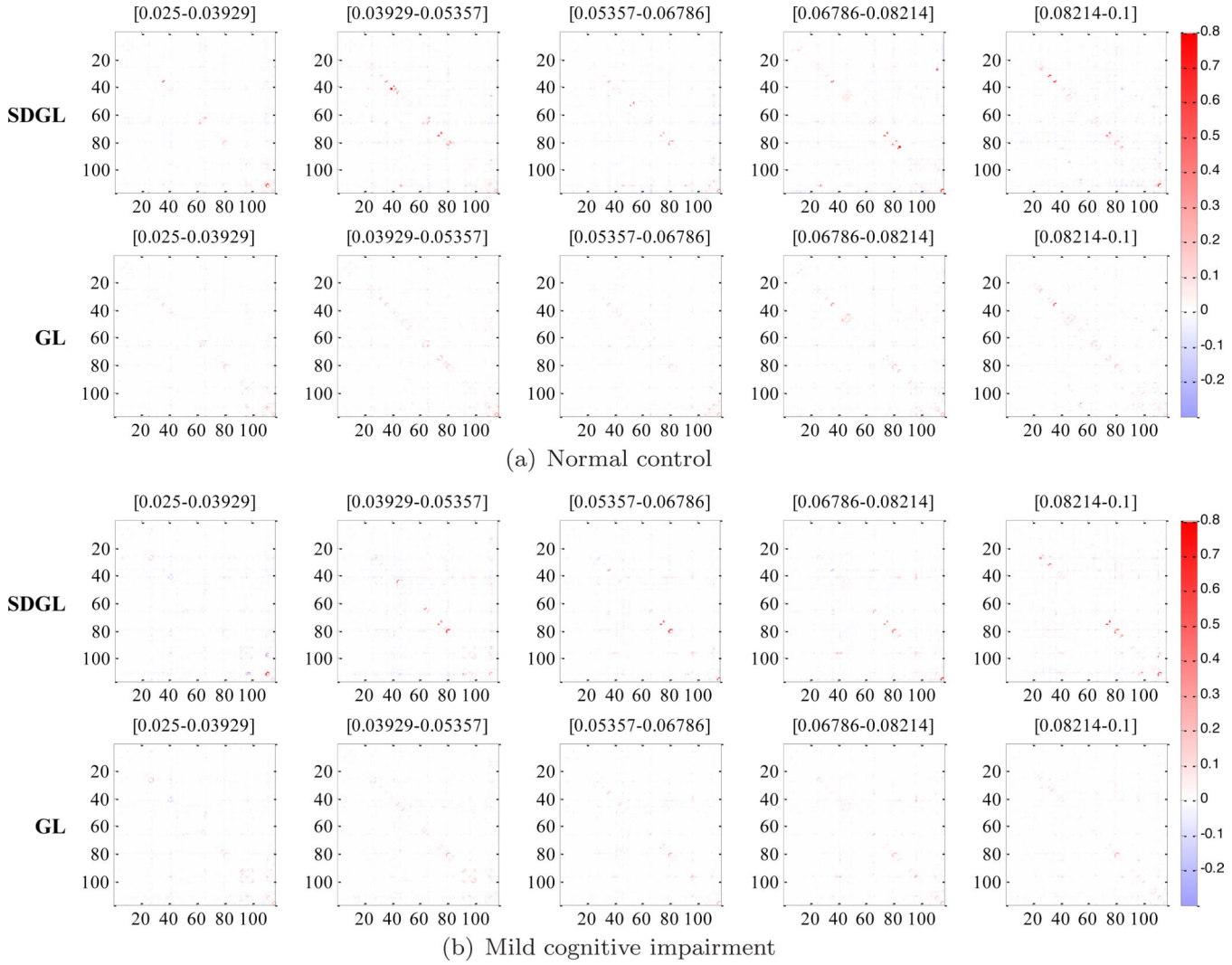



**Fig. 7.**

Comparison of the functional connectivities of a normal control subject and a mild cognitive impairment patient estimated by the proposed Supervised Discriminative Group Lasso (SDGL) and the conventional Group Lasso (GL). For a clear view, refer to the electronic version.

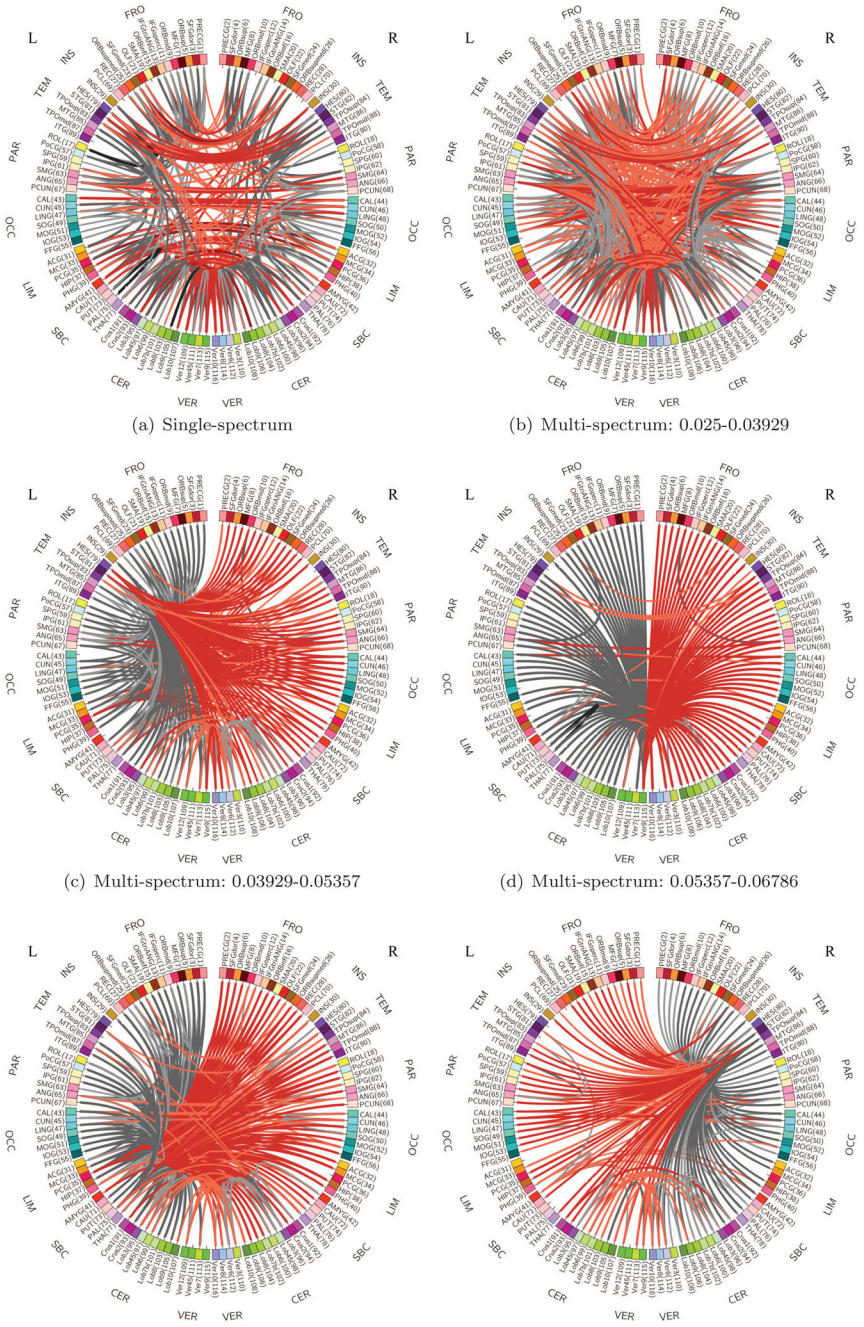



**Fig. 8.**

Connectogram of the functional connectivities. (FRO: FROntal, INS: INSula, TEM: TEMporal, PAR: PARietal, OCC: OCCipital, LIM: LIMbic, SBC: SuBCortical, CER: CERebellum, VER: VERmis)

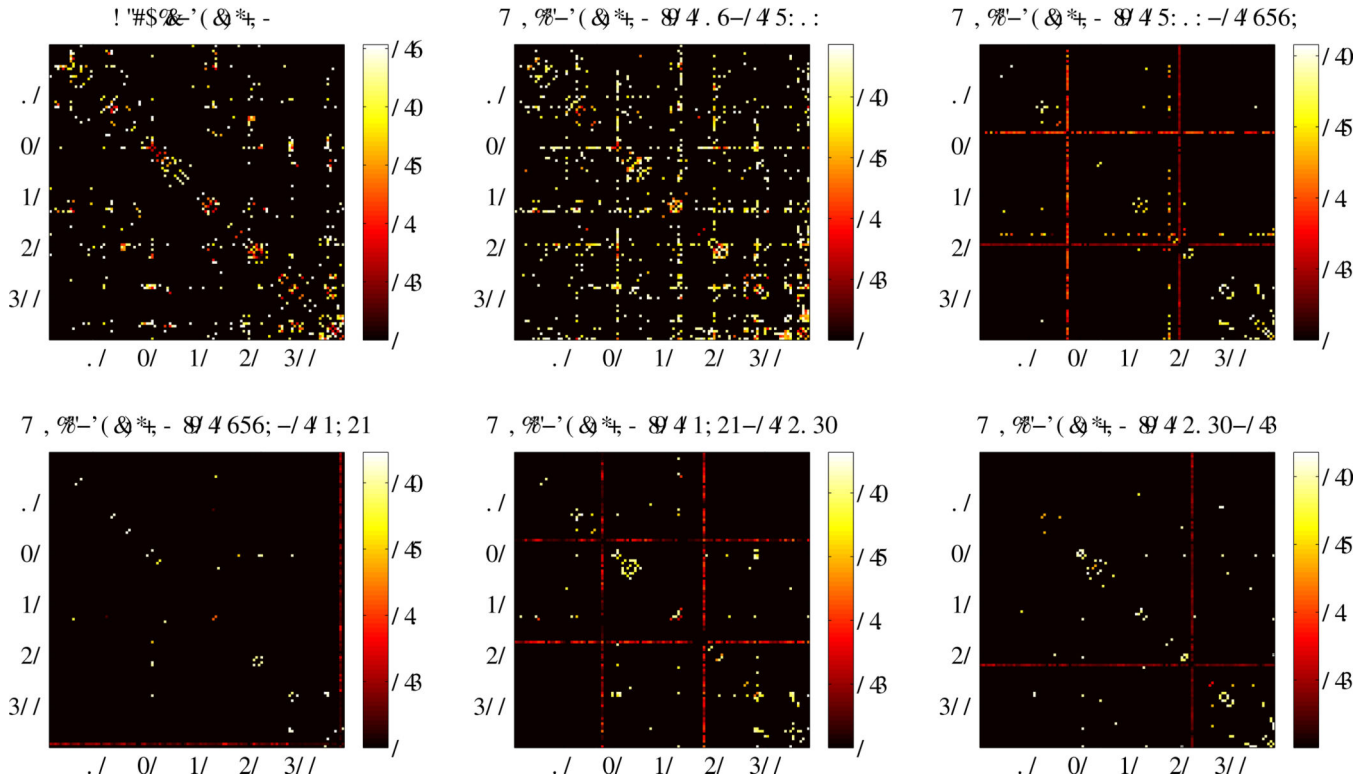


**Fig. 9.**

Connectivity variation maps. Each element in a map denotes the standard deviation of the functional connectivity between the corresponding ROIs estimated by the proposed method over cross-validation trials.

**Table 1**

Demographic and clinical information of the participants. (The *p*-value was obtained by a paired two-sample *t*-test.)

| **Group** | **MCI** | **NC** | ***p*-value** |
|---|---|---|---|
| Number of subjects (Male/Female) | 12 (6/6) | 25 (9/16) | - |
| Age (Mean±SD) | 75.0±8.0 | 72.9±7.9 | 0.3598 |
| Years of education (Mean±SD) | 18.0±4.1 | 15.8±2.4 | 0.0491 |
| MMSE (Mean±SD) | 28.5±1.5 | 29.3 ±1.1 | 0.1201 |

**Table 2**

A summary of the performances of the competing methods. The boldface denotes the best performance in each metric. (HMP: Head-Motion Profiles)

| Methods | | | ACC (%) | AUC | SEN | SPEC | YI | F-Score | BAC |
|---|---|---|---|---|---|---|---|---|---|
| With HMP in regression | Group ICA | | 72.97 | 0.61 | 0.4167 | 0.88 | 0.2967 | 0.5 | 0.6483 |
| | Single-spectrum | Group lasso ($\lambda$=0.15) | 75.68 | 0.67 | 0.3333 | 0.96 | 0.2993 | 0.4709 | 0.6468 |
| | | Proposed method ($\lambda_1$=0.15, $\lambda_2$=0.15) | 86.49 | 0.81 | 0.5833 | **1.0** | 0.5833 | 0.7368 | 0.7917 |
| | Multi-spectrum | Group lasso ($\lambda$=0.15) | 78.38 | 0.8 | 0.5 | 0.88 | 0.38 | 0.5714 | 0.69 |
| | | Proposed method ($\lambda_1$=0.05, $\lambda_2$=0.5) | **89.19** | **0.9567** | **0.9167** | 0.88 | **0.7967** | **0.8462** | **0.8983** |
| Without HMP in regression | Group ICA | | 70.27 | 0.58 | 0.3333 | 0.88 | 0.2133 | 0.4211 | 0.6067 |
| | Single-spectrum | Group lasso ($\lambda$=0.55) | 72.97 | 0.62 | 0.25 | 0.96 | 0.21 | 0.375 | 0.605 |
| | | Proposed method ($\lambda_1$=0.55, $\lambda_2$=0.3) | 83.78 | 0.8533 | 0.5 | 1.0 | 0.5 | 0.6667 | 0.75 |
| | Multi-spectrum | Group lasso ($\lambda$=0.15) | 86.49 | 0.8333 | 0.6667 | 0.96 | 0.6267 | 0.7619 | 0.8133 |
| | | Proposed method ($\lambda_1$=0.05, $\lambda_2$=0.5) | **89.19** | 0.9233 | 0.6667 | 1.0 | 0.6667 | 0.8 | 0.8333 |