

# Discriminative Group Sparse Representation for Mild Cognitive Impairment Classification

Heung-Il Suk, Chong-Yaw Wee, and Dinggang Shen

Department of Radiology and Biomedical Research Imaging Center (BRIC),  
University of North Carolina, Chapel Hill, USA

**Abstract.** Witnessed by recent studies, functional connectivity is a useful tool in extracting brain network features and finding biomarkers for brain disease diagnosis. It still remains, however, challenging for the estimation of a functional connectivity from fMRI due to the high dimensional nature. In order to tackle this problem, we utilize a group sparse representation along with a structural equation model. Unlike the conventional group sparse representation, we devise a novel supervised discriminative group sparse representation by penalizing a large within-class variance and a small between-class variance of features. Thanks to the devised penalization term, we can learn connectivity coefficients that are similar within the same class and distinct between classes, thus helping enhance the diagnostic accuracy. In our experiments on the resting-state fMRI data of 37 subjects (12 mild cognitive impairment patients; 25 healthy normal controls) with a cross-validation technique, we demonstrated the validity and effectiveness of the proposed method, showing the best diagnostic accuracy of 89.19% and the sensitivity of 0.9167.

## 1 Introduction

Although it's still unclear why some people with Mild Cognitive Impairment (MCI) progress to Alzheimer's Disease (AD) and some do not, MCI is considered as an early stage of dementia and it's estimated that approximately 10% to 15% of individuals with MCI progress to AD in one year [6]. While there is no medical treatment to stop or reverse it, recent dementia specific pharmacological advances can slow its progression. Therefore, it has been of great importance for early detection of MCI and a proper treatment.

A lot of studies have witnessed that the functional connectivity, defined as the temporal correlations between spatially distinct brain regions [1], can be a useful tool in finding biomarkers for brain disease diagnosis. Although a large part of the literature has considered the correlation approach to model the functional connectivity, it is hard to interpret the resulting connectivity due to its pairwise computation and full connectedness, while which can be addressed by simple thresholding.

Based on the assumption of the small-world network characteristics in human brain functions, many groups have focused their research on the sparse connectivity [7, 15]. A sparse connectivity can be constructed via the least absolute shrinkage and selection operator (lasso), 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 a challenging topic, since biomedical research is usually conducted at a group level to extract the population features, especially for disease diagnosis. Efficient group analysis requires appropriate handling of expected inter-subject variability without destroying inter-group differences. To this end, Wee *et al.* proposed a constrained sparse functional connectivity network [15] via a group sparse representation [16].

Interestingly, while discrimination is the main goal of the computer-aided brain disease diagnosis, the optimization of the sparse representation is based on regression with a criterion that does not explicitly include a discrimination task. To our best knowledge, there has been no work on brain disease diagnosis and/or medical image analysis with the application of the supervised sparse modeling that explicitly combines the regression and discriminative methods in a unified framework. In this work, we present a novel method of classifying MCI and Normal Control (NC) with sparse modeling in a supervised and discriminative manner. Specifically, we combine a group analysis with a class-discriminative feature extraction by extending the group lasso [16] with the introduction of a label-informed regularization term, which penalizes a large within-class variance and a small between-class variance of connectivity coefficients.

## 2 Materials and Methods

### 2.1 Materials

We use resting state fMRI (rs-fMRI) images acquired from 37 subjects (12 MCI, 25 NC). For each subject, 150 rs-fMRI volumes were acquired per scan. 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 T1-weighted anatomical MRI images were also acquired from the same machine.

We discarded the first 10 fMRI image volumes of each subject for magnetization equilibrium. The remaining 140 fMRI images were preprocessed by applying the typical procedures of slice timing, motion correction, and spatial normalization using SPM8<sup>1</sup>. 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 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 those regions along with 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 [13].

---

<sup>1</sup> Available at <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>

In the spatial normalization, the fMRI images of each subject were coregistered to their respective T1-weighted structure images. The fMRI brain space was then parcellated into 116 ROIs based on the Automated Anatomical Labeling (AAL) template [12]. A mean time series of each ROI was computed from the intensity of all voxels in the ROI. Therefore, we had a set of time series  $\mathbf{X} \in \{X^{(n)} \in \mathbb{R}^{V \times R}\}_{n=1}^N$ , where  $N$  is the number of subjects,  $R$  and  $V$  denote, respectively, the number of ROIs (=116) and the number of volumes (=140).

Following research in the literature, we utilize the low frequency fluctuation features in rs-fMRI with a frequency interval of  $0.025 \leq f \leq 0.100$  Hz on  $\mathbf{X}$ . Based on Wee *et al.*'s work [15], we further decomposed this frequency interval into five equally spaced non-overlapping frequency bands (0.025-0.039 Hz, 0.039-0.054 Hz, 0.054-0.068 Hz, 0.068-0.082 Hz, 0.082-0.100 Hz).

## 2.2 Methods

In this section, we describe a novel method of jointly learning common functional brain networks across subjects via group sparse representation and class-discriminative connectivity coefficients with a label-informed regularization term. We exploit a Structural Equation Model (SEM) [5], assuming that the brain activity of a ROI can be represented by a linear combination of the activity of the other ROIs. Given a set of time series of  $R$  ROIs for  $N$  subjects,  $\{X^{(n)} = [\mathbf{y}_1^{(n)}, \dots, \mathbf{y}_r^{(n)}, \dots, \mathbf{y}_R^{(n)}]\}_{n=1}^N$ , where  $\mathbf{y}_r^{(n)} = [y_r^{(n)}(1), y_r^{(n)}(2), \dots, y_r^{(n)}(V)]^T$  is a  $V$ -length time series of  $r$ -th ROI for  $n$ -th subject, let us consider the SEM for  $r$ -th ROI formulated as follows:

$$L(\mathbf{W}_r) = \frac{1}{2} \sum_n \left\| \mathbf{y}_r^{(n)} - \mathbf{A}_r^{(n)} \mathbf{w}_r^{(n)} \right\|_2^2 \quad (1)$$

where  $\mathbf{A}_r^{(n)} = [\mathbf{y}_1^{(n)} \dots \mathbf{y}_{r-1}^{(n)}, \mathbf{y}_{r+1}^{(n)} \dots \mathbf{y}_R^{(n)}] \in \mathbb{R}^{V \times (R-1)}$  is a data matrix composed of time series of all ROIs except for  $r$ -th ROI,  $\mathbf{w}_r^{(n)} = [w_r^{(n)}(1), \dots, w_r^{(n)}(r-1), w_r^{(n)}(r+1), \dots, w_r^{(n)}(R)]^T \in \mathbb{R}^{(R-1) \times 1}$  is a regression coefficient vector, and  $\mathbf{W}_r = [\mathbf{w}_r^{(1)} \dots \mathbf{w}_r^{(n)} \dots \mathbf{w}_r^{(N)}] \in \mathbb{R}^{(R-1) \times N}$  is a coefficient matrix of  $r$ -th ROI over  $N$  subjects. Note that an element of the coefficient vector  $\mathbf{w}_r^{(n)}$  represents the respective ROI's relationship to  $r$ -th ROI for  $n$ -th subject. Therefore, we can consider the coefficients as the connective strengths between ROIs. Hereafter, we use the regression coefficients and the connectivity coefficients interchangeably.

Let  $\mathbf{W}_r[g] = [w_r^{(1)}(g), \dots, w_r^{(n)}(g), \dots, w_r^{(N)}(g)]$  denote the  $g$ -th row of the coefficient matrix  $\mathbf{W}_r$ . In order to incorporate the class-label information, we utilize the metric of within-class-variance (WCV)  $f_W(\mathbf{W}_r[g])$  and between-class-variance (BCV)  $f_B(\mathbf{W}_r[g])$  defined as follows:

$$f_W(\mathbf{W}_r[g]) = \frac{1}{|\mathbb{N}^+|} \sum_{n \in \mathbb{N}^+} \left( w_r^{(n)}(g) - \hat{w}_r^+[g] \right)^2 + \frac{1}{|\mathbb{N}^-|} \sum_{n \in \mathbb{N}^-} \left( w_r^{(n)}(g) - \hat{w}_r^-[g] \right)^2 \quad (2)$$

$$f_B(\mathbf{W}_r[g]) = (\hat{w}_r^+[g] - \hat{w}_r^-[g])^2 \quad (3)$$

where  $\mathbb{N}^+$  and  $\mathbb{N}^-$  denote, respectively, the set of subjects belonging to the class ‘+’ and ‘-’,  $|\mathbb{N}^+|$  and  $|\mathbb{N}^-|$  denote, respectively, the cardinality of the sets  $\mathbb{N}^+$  and  $\mathbb{N}^-$ ,  $\hat{w}_r^+[g] = \frac{1}{|\mathbb{N}^+|} \sum_{n \in \mathbb{N}^+} w_r^{(n)}(g)$ , and  $\hat{w}_r^-[g] = \frac{1}{|\mathbb{N}^-|} \sum_{n \in \mathbb{N}^-} w_r^{(n)}(g)$ .

The idea of exploiting WCV and BCV to extract class-discriminative features is similar to the Linear Discriminant Analysis (LDA) [2]. Unlike LDA, in this paper, we take the difference of WCV and BCV for computational efficiency. Incorporating the functions of  $f_W(\mathbf{W}_r[g])$  and  $f_B(\mathbf{W}_r[g])$  into the conventional group sparse representation, we devise a new objective function formulated as follows:

$$J(\mathbf{W}_r) = \min_{\mathbf{W}_r} L(\mathbf{W}_r) + \lambda_1 \|\mathbf{W}_r\|_{2,1} + \lambda_2 \left( \sum_g \|f_W(\mathbf{W}_r[g])\|_2 - \sum_g \|f_B(\mathbf{W}_r[g])\|_2 \right). \quad (4)$$

In this objective function, we penalize the high WCV and the low BCV. With the introduction of the newly devised penalty terms, the connectivity for the subjects within a class are imposed to be similar to each other, while those between classes to be distinct. We call this novel label-informed sparse model as ‘*Supervised Discriminative Group Lasso*’ (SDGL).

With appropriate algebraic operations, we can simplify the variance related terms in Eq. (4) as follows

$$\sum_g \|f_W(\mathbf{W}_r[g])\|_2 = \|\mathbf{W}_r D_1\|_{2,1}^2 \quad (5)$$

$$\sum_g \|f_B(\mathbf{W}_r[g])\|_2 = \|\mathbf{W}_r D_2\|_{2,1}^2 \quad (6)$$

where  $D_1 \in \mathbb{R}^{N \times N}$  and  $D_2 \in \mathbb{R}^{N \times N}$  denote, respectively, definitive matrices to compute WCV and BCV of the connectivity coefficients in  $\mathbf{W}_r$ . Specifically,  $D_1$  is a composite matrix that computes the sum of the differences between the connectivity coefficients and their mean in each class, and  $D_2$  is a matrix that computes the difference between the mean of the connectivity coefficients of two different classes. In our experiments, we used a SLEP toolbox<sup>2</sup> [4] to optimize the objective function.

### 2.3 Functional Connectivity and Feature Selection

In this work, we benefit from the brain functional information of the test samples 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. However, since we do not have the label information for the test samples, the composite matrices of  $D_1$  and

<sup>2</sup> Available at ‘<http://www.public.asu.edu/~jye02/Software/SLEP/index.htm>’

$D_2$  cannot be defined, and thus the optimization problem in Eq. (4) cannot be solved in its current form. To this end, we define composite matrices  $\hat{D}_1$  and  $\hat{D}_2$  by concatenating zero-vectors to  $D_1$  and  $D_2$  in Eq. (5) and Eq. (6) as follows:

$$\hat{D}_i = \begin{bmatrix} D_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(K+L) \times (K+L)} \quad (7)$$

where  $i \in \{1, 2\}$ , and  $K$  and  $L$  denote, respectively, the number of training and test samples. By setting the row and column vectors corresponding to the test samples zero<sup>3</sup>, and solving the optimization problem of Eq. (4) 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 during the optimization, we use the label information of only the training samples, and optimize Eq. (4) for each ROI and then concatenate the optimized coefficient vectors across ROIs to construct a connectivity matrix for each subject.

The connectivity matrix  $\mathbf{Q}^{(n)}$  that represents inter-regional correlations in neuronal variability for  $n$ -th subject can then be estimated from the trained sparse regression coefficients over  $R$  ROIs, *i.e.*,  $\mathbf{Q}^{(n)} = [\mathbf{q}_1^{(n)}, \dots, \mathbf{q}_r^{(n)}, \dots, \mathbf{q}_R^{(n)}]$ , where  $\mathbf{q}_r^{(n)} = [w_r^{(n)}(1), \dots, w_r^{(n)}(r-1), 0, w_r^{(n)}(r+1), \dots, w_r^{(n)}(R)]^T$ . In order to obtain a symmetric functional connectivity representation, we take the average of the connectivity matrix and its transposed one,  $\mathbf{C} = (\mathbf{Q} + \mathbf{Q}^T)/2$ . 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 finally represented by a  $z$ -map. In this work, we utilize the weighted local clustering coefficients computed from  $\mathbf{Z}$  as features.

Given training samples from  $N$  subjects<sup>4</sup>, 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. We select features by applying three methods sequentially:  $t$ -test, minimum redundancy and maximum relevance [8], and recursive feature elimination with a linear Support Vector Machine (SVM) [9], and find an optimal parameter for SVM with a grid search algorithm.

### 3 Experimental Results and Discussions

The most direct comparison between two methods can be the accuracy, which counts the number of correctly classified samples in a test set. Table 1 presents that the proposed method outperforms the conventional group lasso in both single- and multi-spectrum approaches, showing the diagnostic accuracies of 86.49% and 89.19% in single- and multi-spectrum, respectively. Here, we should

<sup>3</sup> In Eq. (7), it's assumed that the last  $L$  samples are for the test.

<sup>4</sup> In our case, we have one sample from each subject.

**Table 1.** A summary of the performances of the competing methods

Methods		Accuracy (%)	AUC	Sensitivity	Specificity
Single-spectrum	Group lasso ( $\lambda=0.15$ )	75.68	0.67	0.3333	0.96
	Proposed SDGL ( $\lambda_1=0.15, \lambda_2=0.15$ )	<b>86.49</b>	0.81	0.5833	<b>1.0</b>
Multi-spectrum	Group lasso ( $\lambda=0.15$ )	78.38	0.8	0.5	0.88
	Proposed SDGL ( $\lambda_1=0.05, \lambda_2=0.5$ )	<b>89.19</b>	<b>0.9567</b>	<b>0.9167</b>	0.88

note that the accuracy of the group lasso in multi-spectrum is lower than the one reported in [15]. The main reason for that comes from the difference in pre-processing. In this work, we regressed out the nuisance signals from the regions of CSF, WM, GM along with the six head-motion profiles, which were not performed in Wee *et al.*'s work. From a signal processing point of view, the regression step allows us to acquire more noise-free signals to be analyzed. Henceforth, we believe that the results from our experiment are more faithful.

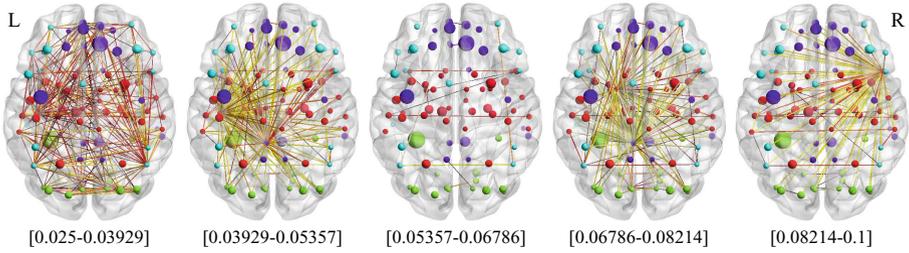
Regarding the sensitivity and specificity, the higher the sensitivity, the lower the chance of mis-diagnosing MCI patients, and the higher the specificity, the lower the chance of mis-diagnosing normal to MCI. Although the specificity of the proposed method is similar or slightly better than the other methods, the proposed SDGL in multi-spectrum overwhelms the competing methods, reporting a sensitivity of 0.9167. Clinically, it's much more beneficial to have a high sensitivity, *i.e.*, correct identification of MCI patients, which can result in taking proper treatments and, to the end, slowing the risk of progressing to AD.

One of the most effective methods of evaluating the performance of diagnostic tests in brain disease as well as other medical areas is the Area Under the receiver operating characteristic Curve (AUC), a combined measure of sensitivity and specificity. 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.

In order to see which ROIs are discriminative for MCI identification, we define the Most Discriminant ROIs (MDRs) based on the following rules:

$$\text{MDRs} = \{r : F_i(r) > \mu_i + 2\sigma_i, \forall i\}$$

where  $F_i(r)$  is the frequency of the  $r$ -th ROI being selected in the  $i$ -th frequency band,  $\mu_i = 1/R \sum_{r=1}^R F_i(r)$  and  $\sigma = \left[1/R \sum_{r=1}^R (F_i(r) - \mu_i)^2\right]^{1/2}$  denote, respectively, the mean and the standard deviation of the frequencies. The selected MDRs are Left Posterior Cingulate Gyrus [10], Left Postcentral Gyrus [14], Left Putamen [3], Left Lobule IV, V of Cerebellar Hemisphere, Left Lobule VI of Cerebellar Hemisphere, and Lobule VI of Vermis [11].



**Fig. 1.** Functional connectivities in the five decomposed frequency bands

We also illustrated the functional connectivity estimated by the proposed method with a multi-spectrum approach in Fig. 1. From the figure, we can see that the connectivity varies across decomposed frequency bands. Interestingly, the connections are the densest in the frequency band of  $[0.025-0.03929]$ , which means that a huge amount of the functional connectivities 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.

## 4 Conclusion

We propose a novel method of identifying MCI with group sparse representation in a supervised and discriminative manner. Specifically, in order to reflect the class-label information in the model, we utilize a well-known discriminative information of the within-class-variance and the between-class-variance [2] for penalization. We should note that the proposed method jointly learns the coherent brain network structures across subjects regardless of the classes, while imposing similar connective 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 showing 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  $[0.025-0.03929]$ .

While we did not consider the joint graphical lasso [7] due to the limited space, it's another widely used method to estimate a sparse functional connectivity at a group level as the group lasso does. We believe that the proposed regularization term can be also applied to this method, possibly enhancing its performance in brain disease diagnosis.

## References

1. Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S.: Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism* 13, 5–14 (1993)
2. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, 2nd edn. Academic Press Professional (1990)
3. Han, S.D., Arfanakis, K., Fleischman, D.A., Leurgans, S.E., Tuminello, E.R., Edmonds, E.C., Bennett, D.A.: Functional connectivity variations in mild cognitive impairment: associations with cognitive function. *Journal of the International Neuropsychological Society* 18, 39–48 (2012)
4. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009)
5. McIntosh, A.R., Grady, C.L., Ungerleider, L.G., Haxby, J.V., Rapoport, S.I., Horwitzl, B.: Network analysis of cortical visual pathways mapped with PET. *Journal of Neuroscience* 14, 655–666 (1994)
6. Misra, C., Fan, Y., Davatzikos, C.: Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *NeuroImage* 44, 1414–1422 (2009)
7. Ng, B., Varoquaux, G., Poline, J.-B., Thirion, B.: A novel sparse graphical approach for multimodal brain connectivity inference. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 707–714. Springer, Heidelberg (2012)
8. 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* 27(8), 1226–1285 (2005)
9. Rakotomamonjy, A.: Variable selection using SVM based criteria. *Journal of Machine Learning Research* 3, 1357–1370 (2003)
10. Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D.: Network analysis of intrinsic functional brain connectivity in Alzheimer’s disease. *PLoS Computational Biology* 4, e1000100 (2008)
11. Thomann, P.A., Schäfer, C., Seidl, U., Santos, V.D., Essig, M., Schröder, J.: The cerebellum in mild cognitive impairment and Alzheimer’s disease - a structural MRI study. *Journal of Psychiatric Research* 42(14), 1198–1202 (2008)
12. 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* 15(1), 273–289 (2002)
13. Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L.: Intrinsic functional connectivity as a tool for human connectomics: Theory, properties and optimization. *Journal of Neurophysiology* 103, 297–321 (2010)
14. 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* 7(8), e42730 (2012)
15. Wee, C.-Y., Yap, P.-T., Zhang, D., Wang, L., Shen, D.: Constrained sparse functional connectivity networks for MCI classification. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 212–219. Springer, Heidelberg (2012)
16. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B* 68(1), 49–67 (2006)