Constrained Sparse Functional Connectivity Networks for MCI Classification

Chong-Yaw Wee[1], Pew-Thian Yap[1], Daoqiang Zhang[1],
Lihong Wang[2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, NC, USA
[2] Brain Imaging and Analysis Center,
Duke University Medical Center, Durham, NC, USA

Abstract. Mild cognitive impairment (MCI) is difficult to diagnose due to its subtlety. Recent emergence of advanced network analysis techniques utilizing resting-state functional Magnetic Resonance Imaging (rs-fMRI) has made the understanding of neurological disorders more comprehensively at a whole-brain connectivity level. However, inferring effective brain connectivity from fMRI data is a challenging task, particularly when the ultimate goal is to obtain good control-patient classification performance. Incorporating sparsity into connectivity modeling can potentially produce results that are biologically more meaningful since most biologically networks are formed by a relatively few number of connections. However, this constraint, when applied at an individual level, will degrade classification performance due to inter-subject variability. To address this problem, we consider a constrained sparse linear regression model associated with the least absolute shrinkage and selection operator (LASSO). Specifically, we introduced sparsity into brain connectivity via l_1-norm penalization, and ensured consistent non-zero connections across subjects via l_2-norm penalization. Our results demonstrate that the constrained sparse network gives better classification performance than the conventional correlation-based network, indicating its greater sensitivity to early stage brain pathologies.

1 Introduction

Mild cognitive impairment (MCI) is an intermediate stage of brain cognitive decline between normal aging and dementia. MCI is associated with increased risk of developing Alzheimer's disease (AD), especially when memory loss is the predominant symptom. Some individuals with MCI remain stable or return to normal over time, but more than half progress to dementia within 5 years [6]. According to a latest, long-term study of nearly 4000 participants, cognitive impairment has a significant impact on life expectancy similar to chronic conditions such as diabetes or chronic heart failure [14]. Early detection is important for possible delay of the progression of mild MCI to moderate and severe stages. However, diagnosis of MCI is difficult due to its mild symptoms of cognitive impairment, causing most computer-aided diagnosis to achieve lower than desired performance.

Constructing functional brain connectivity from neuroimaging data holds great promise for identifying image-based markers that are important for distinguishing between

MCI and normal aging. A large body of work on functional connectivity modeling has been proposed based on correlation analysis [18, 19]. However, correlation only captures pairwise information and is unable to provide an adequate and complete account of the interaction between many brain regions. Inferring effective brain connectivity from fMRI data for biologically more meaningful interpretation and better classification performance is a challenging task. Many spurious connections arise due to the low frequency (< 0.1Hz) spontaneous fluctuation of blood oxygen level dependent (BOLD) signals and physiological noise such as cardiac and respiratory cycles. Recent work [16] has shown that certain sparsity constraints can be imposed to elucidate robust connections from a set of noisy connections. The sparsity constraint correlates with the fact that, neurologically, a brain region predominantly interacts only with a small number of other regions. Sparse brain connectivity for fMRI data can be constructed through penalizing the linear regression model with l_1-norm as in the least absolute shrinkage and selection operator (LASSO) [10].

However, sparse modeling is unable to deal with inter-subject variability since l_1 penalization at an individual level will result in different network structures across subjects [20], i.e., the non-zero connections are different for each subject. This will inevitably make the comparison between subjects difficult and thus degrade generalization performance of trained classifiers. To address this issue, we propose to employ a constrained sparse linear regression model that minimizes the effect of inter-subject variability in network representation. By this approach, the connection topology is kept identical among subjects, while at the same time allowing individual connection parameters to vary between subjects. This will allow better and more direct comparison among subjects for patient identification. To the best of our knowledge, the current study is the first attempt to construct functional brain network using constrained sparse linear regression model for the purpose of MCI classification. We seek to validate whether this new network modeling strategy can be used to improve classification performance. We will also identify brain regions that contribute most to the classification performance. This paper sheds new light on the effectiveness of applying constrained sparse functional network for diagnosis of progressive neurodegenerative disorders.

2   Materials and Methods

Resting-state fMRI (rs-fMRI) scans of 12 MCI patients and 25 healthy controls were acquired using a 3 Tesla (Signa EXCITE, GE) scanner with the following parameters: TR/TE = 2000/32 ms, flip angle = 77°, imaging matrix = 64 × 64, FOV = 256 × 256 mm^2, 34 slices, 150 volumes, and voxel thickness = 4 mm. During scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen to prevent them from falling into sleep and the saccade-related activation due to eyes-closed. Informed consent was obtained from all subjects, and the experimental protocols were approved by the institutional ethics board. Confirmation of diagnosis for all subjects was made via expert consensus panels. Demographic and clinical information of the participants is provided in Table 1.

Table 1. Demographic and clinical information of the participants

Group	MCI	Control	p-value
No. of subjects (Male/Female)	6/6	9/16	-
Age (mean $\pm$ SD)	75.0 ± 8.0	72.9 ± 7.9	0.3598^a
Years of education (mean $\pm$ SD)	18.0 ± 4.1	15.8 ± 2.4	0.0491^a
MMSE (mean $\pm$ SD)	28.5 ± 1.5^b	29.3 ± 1.1	0.1201^a

[a] The p value was obtained by two-sample two-tailed t-test.
[b] One of the patients does not have a MMSE score.

Post-processing of the fMRI images including slice timing correction and head-motion correction were performed using the Statistical Parametric Mapping (SPM8[1]) software package. The images were then masked with their respective gray matter (GM) masks, created by segmenting the GM regions from their T1-weighted images to eliminate the physiological noise caused by cardiac and respiratory cycles in white matter and cerebrospinal fluid [17]. Then, we parcellated the brain space into 116 ROIs by warping the automatic anatomical labeling (AAL) template to the GM-masked fMRI images. The mean time series of each ROI was computed for each subject. Temporal band-pass filtering of frequency interval ($0.025 \leq f \leq 0.100$Hz) was then performed to minimize the effects of low-frequency drift and high-frequency noise. This frequency interval was further decomposed into 5 equal-length spectral, enabling a more frequency specific analysis of the regional mean time series [19].

2.1 Constrained Sparse Functional Brain Connectivity

Suppose we have N training subjects and M ROIs, the mean time series of p-th ROI for n-th subject, $\mathbf{y}_p^n$, can be regarded as a response vector that is modeled by a linear combination of time courses of other ROIs as

$$\mathbf{y}_p^n = \mathbf{A}_p^n \boldsymbol{\alpha}_p^n + e_p^n, \tag{1}$$

where e_p^n is the noise, $\mathbf{y}_p^n = [y_p^n(1); y_p^n(2); \ldots; y_p^n(T)]$ with T being the number of time points in the time series, $\mathbf{A}_p^n = [\mathbf{y}_1^n, \ldots, \mathbf{y}_{p-1}^n, \mathbf{y}_{p+1}^n, \ldots, \mathbf{y}_M^n]$ is data matrix of the p-th ROI, and $\boldsymbol{\alpha}_p^n = [\alpha_1^n; \ldots; \alpha_{p-1}^n; \alpha_{p+1}^n; \ldots; \alpha_M^n]$ is the weight vector. The sparse brain connectivity modeling of the n-th subject and p-th ROI can be considered as a standard l_1-norm regularized optimization problem with the following objective function

$$f(\boldsymbol{\alpha}_p^n) = \frac{1}{2} \left\| \mathbf{y}_p^n - \mathbf{A}_p^n \boldsymbol{\alpha}_p^n \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_p^n \right\|_1, \tag{2}$$

where $\lambda > 0$ is the regularization parameter controlling the "sparsity" of the model, with a higher value corresponding to a sparser model, i.e., more elements in $\boldsymbol{\alpha}^n$ are zero. Note the l_1-norm penalization is imposed individually on different $\boldsymbol{\alpha}_p^n$ vectors.

For multiple subjects, the objective function in Eq. (2) can be modified as

$$f(\boldsymbol{\alpha}_p) = \sum_{n=1}^{N} \left(\frac{1}{2} \left\| \mathbf{y}_p^n - \mathbf{A}_p^n \boldsymbol{\alpha}_p^n \right\|_2^2 \right) + \lambda \left\| \boldsymbol{\alpha}_p \right\|_{2,1} \tag{3}$$

[1] http://www.fil.ion.ucl.ac.uk.spm

where $\|\boldsymbol{\alpha}_p\|_{2,1}$ is the summation of l_2-norms of non-overlapping $\|\boldsymbol{\alpha}_p^n\|_1$. Specifically, the l_2-norm penalization is imposed on the same elements across different matrices $\boldsymbol{\alpha}_p$ which forces the weights corresponding to certain feature across different subjects to be grouped together. This constraint promotes group-based sparsity by keeping the topology to be identical among subjects, while at the same time allowing variation between subjects. This reduces inter-subject variability and allows for better and more consistent inter-subject comparison for patient identification. The nonzero coefficients in $\boldsymbol{\alpha}$ matrix are treated as an indicator of functional brain connectivity. We use the SLEP toolbox [11] to solve Eq. (3).

2.2   Feature Extraction and Feature Selection

Weighted local clustering coefficient, a measure that quantifies the cliquishness of the nodes, is extracted from all connectivity maps as

$$C_p = \frac{2 \times \sum_{q:q \neq p \in \zeta} e_{p,q}}{k_p(k_p - 1)}, \tag{4}$$

where k_p is the number of ROIs that are connected to the p-th ROI, ζ is the subnetwork comprising nodes directly connected to the p-th ROI, and $e_{p,q}$ is the parameter value between the p-th and q-th ROIs. A total of 116 features are obtained from each map, producing a pool of 580 features for each subject.

After feature extraction, we utilized a hybrid method to select the most relevant features for classification. Two filter-based approaches are initially used to reduce the number of features, followed by a wrapper-based approach to further select a subset of features that is favorable to MCI classification. Specifically, in the first filter-based approach, only those features with p-values smaller than the predefined threshold, measured via between-group t-test, will be retained for subsequent feature selection. Then, the minimum redundancy and maximum relevance (mRMR) algorithm [4] was employed to further exclude redundant features. Finally, the support vector machine (SVM) recursive feature elimination [9], a wrapper-based method, was used to select a subset of most discriminative features for MCI classification.

2.3   Classification

SVM with linear kernel was employed to evaluate the discriminative power of the selected features derived from constrained sparse networks. The optimal SVM models as well as an unbiased estimation of the generalization performance of the complete framework were obtained via a nested cross-validation scheme. For N total number of subjects involved in the study, one was left out for testing, and the remaining $N - 1$ were used for training. From these $N - 1$ samples, $N - 1$ different training subsets were formed by each time leaving one more sample out, giving us $N - 2$ subjects in each training subset. For each training subset, feature extraction and feature selection were performed. The performance of each combination of SVM parameters along with the selected features was evaluated using the second left out subject. The combination that gives the best performance was used to construct the optimal SVM model for future

classification. This procedure was repeated $N - 1$ times, once for each training subset. When the completely unseen (totally left out during the entire training and parameter optimization process) test sample was to be classified, all $N - 1$ classifiers were used, and their outcomes were combined using an averaging operator to provide the final classification decision. This process was repeated N times, each time leaving out a different subject, finally leading to an overall cross-validation classification accuracy. In this study, the optimization of λ parameter in Eq. (3) was performed via grid search.

3   Experimental Results

Constrained sparse connectivity maps of one healthy control and one MCI patient are shown in Figure 1. Spatial connection topology of our sparse networks and the fully-connected correlation-based networks are shown in Figure 2. It can be clearly observed that the generated connectivity networks are significantly sparser than the fully connected correlation-based networks. There are a number of findings that are biologically interesting. First, the bilateral temporal lobes show a relatively smaller amount of intra-lobe connections than other lobes, which has been extensively reported in the literature [16, 18]. Second, there are significantly more inter-lobe connections between parietal and occipital lobes than any other lobe pairs [16], possibly due to compensatory effect. These selected non-zero elements in the constrained sparse matrices reflect the connections that are crucial for discriminating MCI patients from healthy controls.

  MCI classification performance of the proposed constrained sparse network was compared with correlation-based connectivity using single and multi-spectral characterization. In the single spectrum case, feature extraction was directly performed on the band-pass filtered BOLD signal without further frequency sub-band decomposition. Sub-band decomposition in the multi-spectral case enables more detailed characterization of subtle changes in BOLD signal, and hence better discriminative power [19]. During evaluation, SVM classifier with the same linear kernel but different hyperparameters was used in a leave-one-out fashion due to the limited number of available samples. MCI classification performance for constrained sparse and correlation-based
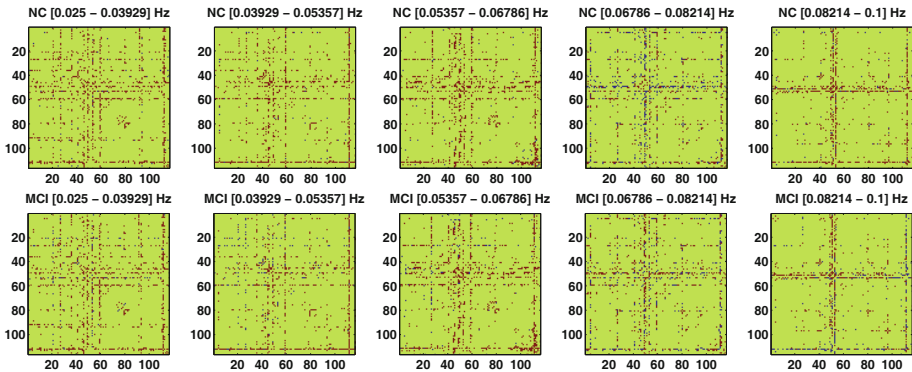

Fig. 1. Constrained sparse connectivity maps with $\lambda = 0.15$. (Red = positive connection, blue = negative connection, green = no connection)
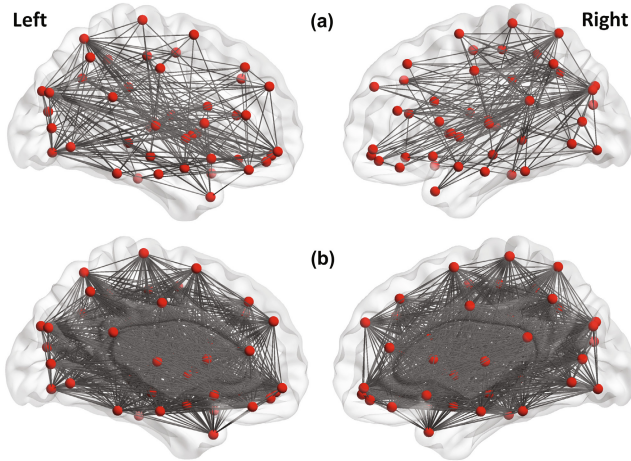
Fig. 2. Constrained sparse networks with $\lambda = 0.15$ (a) and the fully-connected correlation-based networks (b), after excluding the cerebellum

networks are summarized in Table 2. The proposed constrained sparse network with multi-spectral characterization yields the best classification performance with an accuracy of 86.49%, which is an increment of at least 10% from that of the correlation-based approach. A cross-validation estimation of the generalization performance shows an area of 0.8333 under the receiver operating characteristic curve (AUC), indicating good diagnostic power. Note that $\lambda = 0.15$ was found to give the optimal performance.

The most discriminant regions that were selected in the classification process include regions located in frontal lobes (e.g. orbitofrontal cortex [8], frontal gyri [1] and rectus gyrus [5]), temporal lobes (e.g. temporal gyri [2, 5, 15] and temporal pole [13]), and other regions such as cingulate gyri [7], amygdala [3], angular gyrus [15], and occipital gyri [12], which is in line with the findings that AD, strongly related to episodic memory impairment, causes atrophies in temporal and frontal lobes at the beginning stages of the disease. The selected regions are shown graphically in Figure 3.

Table 2. Classification performance for constrained-sparse and correlation-based networks using single and multi-spectral characterization. (ACC = Accuracy; SEN = SENsitivity; SPE = SPEcificity)

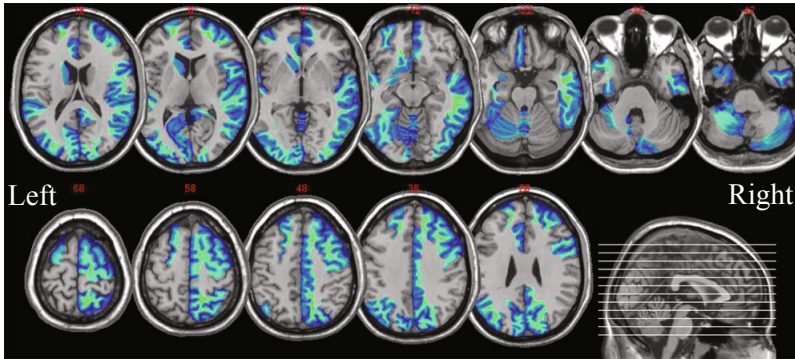| Approach	ACC (%)	AUC	SEN	SPE
Correlation + Single Spectrum	67.57	0.6633	0.0833	0.9600
Sparse + Single Spectrum	72.97	0.6200	0.2500	0.9600
Correlation + Multi-Spectral	75.68	0.7070	0.4167	0.9200
Sparse + Multi-Spectral	86.49	0.8333	0.6667	0.9600

Fig. 3. Most discriminant regions that were selected during MCI classification

4   Discussions and Conclusion

We proposed a novel method to infer functional connectivity networks from rs-fMRI data for the purpose of classification. By imposing group-based sparsity, we minimize spurious connections and inter-subject variability. This is accomplished by considering a constrained sparse linear regression model. Specifically, we incorporate sparsity into brain connectivity estimation via l_1-norm penalization, and ensure inter-subject stability of network structure via l_2-norm penalization. This constrained sparse representation generates topologically consistent functional connectivity networks that allow for better comparison between subjects for classification. The experiment results validate that the proposed method yields markedly improved classification performance compared with the correlation-based network.

References

1. Bell-McGinty, S., Lopez, O.L., Meltzer, C.C., Scanlon, J.M., Whyte, E.M., Dekosky, S.T., Becker, J.T.: Differential cortical atrophy in subgroups of mild cognitive impairment. Arch. Neurol. 62(9), 1393–1397 (2005)
2. Convit, A., de Asis, J., de Leon, M.J., Tarshish, C.Y., De Santi, S., Rusinek, H.: Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiol. Aging 21(1), 19–26 (2000)
3. Dai, W., Lopez, O.L., Carmichael, O.T., Becker, J.T., Kuller, L.H., Gach, H.M.: Mild cognitive impairment and Alzheimer disease: Patterns of altered cerebral blood flow at MR imaging. Radiology 250, 856–866 (2009)
4. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 3(2), 185–205 (2005)
5. Fleisher, A.S., Sherzai, A., Taylor, C., Langbaum, J.B., Chen, K., Buxton, R.B.: Resting-state BOLD networks versus task-associated functional mri for distinguishing Alzheimer's disease risk groups. Neuroimage 47(4), 1678–1690 (2009)

6. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J.L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M.C., Whitehouse, P., Winblad, B.: On behalf of the participants of the International Psychogeriatric Association Expert Conference on mild cognitive impairment. Lancet 367, 1262–1270 (2006)
7. Gold, B.T., Jiang, Y., Jicha, G.A., Smith, C.D.: Functional response in ventral temporal cortex differentiates mild cognitive impairment from normal aging. Hum. Brain. Mapp. 31(8), 1249–1259 (2010)
8. Grady, C.L., McIntosh, A.R., Beig, S., Keightley, M.L., Burian, H., Black, S.E.: Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J. Neurosci. 23(3), 986–993 (2003)
9. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning 46(1-3), 389–422 (2004)
10. Lee, H., Lee, D.S., Kang, H., Kim, B.N., Chung, M.K.: Sparse brain network recovery under compressed sensing. IEEE Trans. Med. Imaging 30(5), 1154–1165 (2011)
11. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009), http://www.public.asu.edu/~jye02/Software/SLEP
12. Nobili, F., Mazzei, D., Dessi, B., Morbelli, S., Brugnolo, A., Barbieri, P., Girtler, N., Sambuceti, G., Rodriguez, G., Pagani, M.: Unawareness of memory deficit in amnestic mci: FDG-PET findings. J. Alzheimers Dis. 22(3), 993–1003 (2010)
13. Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S.A., Rodriguez, G., Pagani, M.: Principal component analysis of FDG PET in amnestic MCI. Eur. J. Nucl. Med. Mol. Imaging 35(12), 2191–2202 (2008)
14. Sachs, G.A., Carter, R., Holtz, L.R., Smith, F., Stump, T.E., Tu, W., Callahan, C.M.: Cognitive impairment: An independent predictor of excess mortality: A cohort study. Ann. Intern. Med. 155(5), 300–308 (2011)
15. Smith, C.D., Chebrolu, H., Wekstein, D.R., Schmitt, F.A., Jicha, G.A., Cooper, G., Markesbery, W.R.: Brain structural alterations before mild cognitive impairment. Neurology 68(16), 1268–1273 (2007)
16. Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D.: Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol 4, e1000100 (2008)
17. 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. J. Neurophysiol. 103, 297–321 (2010)
18. 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. Hum. Brain Mapp. 28(10), 967–978 (2007)
19. Wee, C.Y., Yap, P.T., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PLoS ONE 7(5), e37828 (2012)
20. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. Roy. Stat. Soc. B 68(1), 49–67 (2006)