Structural Connectivity Guided Sparse Effective Connectivity for MCI Identification

Yang Li[1], **Jingyu Liu**[1], **Meilin Luo**[1], **Ke Li**[2], **Pew-Thian Yap**[3], **Minjeong Kim**[3], **Chong-Yaw Wee**[4], and **Dinggang Shen**[3]

[1]School of Automation Sciences and Electrical Engineering, Beihang University, Beijing, China

[2]School of Aeronautic Science and Engineering, Beihang University, Beijing, China

[3]Department of Radiology and BRIC, UNC Chapel Hill, Chapel Hill, NC, USA

[4]Department of Biomedical Engineering, National University of Singapore, Singapore, Singapore

Abstract

Recent advances in network modelling techniques have enabled the study of neurological disorders at a whole-brain level based on functional connectivity inferred from resting-state magnetic resonance imaging (rs-fMRI) scan possible. However, constructing a directed effective connectivity, which provides a more comprehensive characterization of functional interactions among the brain regions, is still a challenging task particularly when the ultimate goal is to identify disease associated brain functional interaction anomalies. In this paper, we propose a novel method for inferring effective connectivity from multimodal neuroimaging data for brain disease classification. Specifically, we apply a newly devised weighted sparse regression model on rs-fMRI data to determine the network structure of effective connectivity with the guidance from diffusion tensor imaging (DTI) data. We further employ a regression algorithm to estimate the effective connectivity strengths based on the previously identified network structure. We finally utilize a bagging classifier to evaluate the performance of the proposed sparse effective connectivity network through identifying mild cognitive impairment from healthy aging.

1 Introduction

Mild cognitive impairment (MCI), as the intermediate state of cognitive function between normal aging and dementia, has attracted a great deal of attention recently due to its high progression rate to dementia. More than 50% of individuals with MCI progress to dementia within 5 years [5]. MCI is thus an appropriate target for early diagnosis and intervention of Alzheimer's disease (AD). However, its mild symptoms cause most existing computer-aided diagnosis frameworks perform relatively inferior with low sensitivity [7].

Recently, neuroimaging-based techniques have been shown to be a powerful tool for predicting the progression of MCI to AD [12]. For example, functional connectivity inferred from resting-state magnetic resonance image (rs-fMRI) data can reflect temporal

Correspondence to: Meilin Luo.

interactions between distinct region-of-interest (ROI) in the brain [13]. However, functional connectivity conveys only the pairwise correlation [4], ignoring the directed causal influence between ROIs [10]. Such analysis is susceptible to noise due to the low frequency (< 0.1 Hz) spontaneous fluctuation of blood oxygen level dependent (BOLD) signals, thus may not accurately reveal the brain states at rest. This limits the capability of correlation-based functional connectivity to provide an adequate and complete account of the interactions among multiple brain regions. As an alternative, effective connectivity, has been employed to reflect the causal interactions between a pair of ROIs [4].

Most of the biological networks are intrinsically sparse [3]. Sparse modeling methods such as the least absolute shrinkage and selection operator (Lasso) have been applied to construct sparse brain connectivity networks [2]. However, Lasso which is normally applied on a single modality of neuroimaging data (e.g., rs-fMRI), inevitably ignores the important complementary information from other modalities. Additionally, sparse models through penalizing the linear regression are mainly focused on the resulted squared loss, ignoring the relations between the signals of different time points, which are of great importance for revealing the characteristics of dynamic model. To resolve those issues, we proposed, in this paper, an Ultra-Weighted-Lasso approach to construct a more accurate sparse effective connectivity network and used it for MCI classification. Specifically, the proposed approach was a modified version of Lasso to incorporate structural connectivity information derived from diffusion tensor imaging (DTI) data for deriving the brain network structure. The regression loss of the proposed modified Lasso model involves a term that conveys the relation between different time points, *i.e.*, the derivative. In this way, a more accurate effective connectivity structure can be obtained. The connectivity strength of each edge in the identified network can then be inferred using an Ultra-Orthogonal-Forward regression (UOFR) model which also takes into consideration the derivatives of the signal. We seek to evaluate the capability of this new effective connectivity network for improving the MCI classification performance.

2 Materials and Methodology

A dataset with 27 participants (10 MCI and 17 healthy controls) was used in this study. There are no significant differences in terms of age, gender, years of education, and Mini Mental State Examination (MMSE) score between MCI and healthy subjects. Both rs-fMRI and DTI scans were acquired using a 3 Tesla (Signa EXCITE, GE) scanner with following parameters: rs-fMRI: TR/TE = 2000/32 ms, flip angle = 77°, imaging matrix = 64×64, FOV = 256×256 mm^2, voxel thickness = 4 mm, 34 slices, and 150 volumes; DTI: b=0 and 1000 s/mm^2, flip angle = 90°, TR/TE = 17000/78 ms, imaging matrix = 128×128, FOV = 256×256 mm^2, voxel thickness = 2 mm, and 72 continuous slices. The same scanner was used to acquire the T1-weighted anatomical MRI images using the following parameters: TE = 2.976 ms, TR = 7.460 ms, flip angle = 12°. The imaging matrix = 256×224 with a rectangular FOV of (256×256 mm^2), voxel thickness = 1 mm, and 216 continuous slices. Subjects were instructed to keep eyes open and stare at a fixation cross in the middle of the screen to prevent them falling into sleep and avoid the saccade-related activation due to eyes-closed during scanning.

For DTI data, all images were first parcellated into 90 regions based on the automated anatomical labeling (AAL) template [9] using a deformable DTI registration algorithm. Whole-brain streamline fiber tractography was then performed on each image with minimal seed point fractional anisotropy (FA) of 0.45, stopping FA of 0.25, minimal fiber length of 20 mm, and maximal fiber length of 400 mm. The number of fibers passing through each pair of regions was counted and two regions were considered as anatomically connected if fibers passing through their respective masks were present. For rs-fMRI data, standard preprocessing procedure was carried out using the statistical parametric mapping (SPM8) software, including the removal of the first 10 volumes, slice timing correction, head-motion artifact correction, regression of nuisance signals (ventricle, white matter, and 6 head-motion parameters), signal de-trending, and band-pass filtering (0.0250.1 Hz). All fMRI images were coregistered to their own T1-weighted image before parcellated into 90 regions based on AAL template.

2.1 Effective Connectivity Inference via Ultra-Weighted-Lasso

Suppose we have M ROIs, the mean time series of i-th ROI for the n-th subject, $\mathbf{y}_i$, can be regarded as a response vector that can be estimated as a linear combination of the mean time series of other ROIs as

$$\mathbf{y}_i = \mathbf{A}_i \alpha_i + \mathbf{e}_i, \quad (1)$$

where $\mathbf{e}_i$ is the noise, $\mathbf{y}_i = [y_i(1), y_i(2), \cdots, y_i(N)]^T$ is the time series with N being the number of time points, $\mathbf{A}_i = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \cdots, \mathbf{y}_M]$ is the data matrix of the i-th ROI, and $a_i = [a_1, \cdots, a_{i-1}, a_{i+1}, \cdots, a_M]^T$ is the weight vector. The solution of the l_1-norm regularized optimization problem results in a sparse weight vector, a_i, which can be computed as follows:

$$\alpha_i = \arg\min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{A}_i \alpha_i\|_2^2 + \lambda \|\alpha_i\|_1, \quad (2)$$

where $\lambda > 0$ is the regularization parameter, controlling the sparsity of the model. Note that the first term in Eq. (2) uses only the least squares loss of the regression result, ignoring the relations between different time points in signal $\mathbf{y}$, *i.e.*, derivative. To resolve the issue, the Eq. (2) can be modified as

$$\alpha_i = \arg\min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{A}_i \alpha_i\|_2^2 + \sum_{l=1}^{K} \left\| D^l \mathbf{y}_i - \sum_{j=1, j\neq i}^{M} \alpha_j D^l \mathbf{y}_j \right\|_2^2 + \lambda \|\alpha_i\|_1, \quad (3)$$

where $D^l \mathbf{y}$ is a measurement of the l-order weak derivative of the signal $\mathbf{y}$ and satisfies

$$\int_{[0,T]} \mathbf{y}_i D^l \varphi(t) \mathrm{dt} = (-1)^l \int_{[0,T]} \varphi(t) D^l \mathbf{y}_i \mathrm{dt}, \qquad (4)$$

where the test function $\varphi(t) \in C_0^\infty(0, T)$ is smooth and exhibits the properties of compact support within [0, T]. Thus, for simplicity, $D^l\mathbf{y}$ in Eq. (4) can be replaced by

$$R_{\mathbf{y}_i}^l(\tau) = \int_{[\tau,T+1]} \mathbf{y}_i D^l \varphi(t) \mathrm{dt}, \quad \tau \in [0, N - T] \qquad (5)$$

where τ is the time shift of the test function. Hence, we can obtain the model as

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{A}_i \alpha_i\|_2^2 + \sum_{l=1}^K \left\| R_{\mathbf{y}_i}^l(\tau) - \sum_{j=1,j\neq i}^M \alpha_j R_{\mathbf{y}_j}^l(\tau) \right\|_2^2 + \lambda \|\alpha_i\|_1. \qquad (6)$$

Please refer to [6] for more information about the detail discussion of both weak derivatives and test functions.

Additionally, it is notable that the Lasso model may produce a suboptimal estimation by inflicting the same penalization on every regressor. To resolve this issue, we can, on one hand, rely on the time series for the penalization adjustment of different regressors using adaptive-Lasso [14]. On the other hand, since the structural connectivity is highly correlated with functional connectivity [4], with a stronger structural connectivity indicating a higher opportunity of the functional connectivity, we opt, in this paper, to utilize the structural connectivity derived from DTI to modify the penalization weight for the rs-fMRI time series as

$$\alpha_i = \arg \min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \mathbf{A}_i \alpha_i\|_2^2 + \sum_{l=1}^K \left\| R_{\mathbf{y}_i}^l(\tau) - \sum_{j=1,j\neq i}^M \alpha_j R_{\mathbf{y}_j}^l(\tau) \right\|_2^2 + \lambda \sum_{j=1,j\neq i}^M w_i(j)|\alpha_i(j)|, \qquad (7)$$

where $w_i(j)$ is the penalization weight for the j-th regressor. Let $S_{i,j}$ be the structural connectivity between the i-th ROI and the j-th ROI, $w_i(j)$ can then be defined as

$$w_i(j) = \exp\{1 - n_i \cdot f_{i,j}\}, \qquad (8)$$

where n_i is the total number of structural connectivity for the i-th ROI, and $f_{i,j} = S_{i,j} / \sum_{j=1,j\neq i}^M S_{i,j}$ is the proportion of structural connectivity between i-th ROI and j-th ROI to the sum of connectivity of i-th ROI. Equation (8) indicates that for a pair of brain regions that with a strong structural connectivity, there is high probability that an effective connectivity exists between them and a small penalization should be imposed. The

regression model, which is referred as Ultra-Weighted-Lasso, incorporates the weak derivatives and penalization to derive a more accurate sparse effective connectivity structure based on the resulted a.

2.2 Effective Network Construction via UOFR

The non-zero element $a_i(j)$ in a_i indicates the existence of effective influence the j-th ROI exerting on the i-th ROI. Herein, an algorithm of Ultra Orthogonal Forward Regression (UOFR) which takes into consideration the derivatives, is used to calculate the effective connectivity strength. For the time series $\mathbf{y}_i$ and $\{\mathbf{y}_j | a_i(j) > 0; j \in [1, 2, \cdots, M]\}$, the series are firstly extended as

$\tilde{\mathbf{y}} = [y(1), y(2), \cdots, y(N), R_{\mathbf{y}}^1(1), R_{\mathbf{y}}^1(2), \cdots, R_{\mathbf{y}}^1(N - T), \cdots, R_{\mathbf{y}}^K(N - T)]^T$, which is referred as ultra time series. Then, a forward orthogonalization is applied to the ultra time series $\{\tilde{\mathbf{y}}_j | a_i(j) > 0; j \in [1, 2, \cdots, M]\}$ to calculate the in error reduction ratio (ERR) as [6]

$$\mathrm{ERR}_i(j) = \frac{<\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_j>^2}{<\tilde{\mathbf{y}}_i, \tilde{\mathbf{y}}_i><\tilde{\mathbf{y}}_j, \tilde{\mathbf{y}}_j>} \quad (9)$$

where $< \cdot >$ is the inner product. For each round of orthogonalization, the maximum ERR_i, i.e., $ERR_i(j)$, which reflects the effect the j-th ROI exerting on the i-th ROI, is referred as the effective connectivity strength of $g_{i,j}$, which reflects the effect the j-th ROI exerting to the i-th ROI. Note that $g_{i,j}$ is nonnegative and varies from 0 to 1.

2.3 Feature Extraction

Weighted-clustering coefficient (GC_i) and betweenness centrality (GB_i) are extracted from the constructed sparse effective connectivity network for the MCI classification and are defined as

$$\mathrm{GC}_i = \frac{\sum_{j,k \neq i} \left(g_{i,j}^{1/3} + g_{j,i}^{1/3}\right) \left(g_{i,k}^{1/3} + g_{k,i}^{1/3}\right) \left(g_{j,k}^{1/3} + g_{k,j}^{1/3}\right)}{2(|\Delta_i|(|\Delta_i| - 1) - 2|\Delta_i^{\leftrightarrow}|)}, \quad (10)$$

where $g_{i,j}$ denotes the effective connectivity from the i-th ROI to the j-th ROI, $|\Delta_i|$ denotes the number of ROIs that are adjacent to the i-th ROI, and $|\Delta_i^{\leftrightarrow}|$ is the number of bilateral edges between the i-th ROI and its neighbors, respectively.

$$\mathrm{GB}_i = \sum_{j \neq i \neq k} \frac{\sigma_{j,k}(i)}{\sigma_{j,k}}, \quad (11)$$

where $\sigma_{j,k}$ denotes the total number of shortest paths from the j-th ROI to the k-th ROI, and $\sigma_{j,k}(i)$ is the number of those paths that pass through the i-th ROI.

2.4 Classification and Evaluation

Decision tree (DT) with a bagging strategy [1] and leave-one-out cross-validation (LOOCV) is used as classifier for MCI classification. Specifically, for N_s total number of subjects involved, one is first left out for testing, and the remaining $N_s - 1$ are used for feature selection and DT construction, where $N_s - 1$ decision trees (DTs) are constructed using different $N_s - 2$ training subjects and validated using the second left out validation subject. The predicated label of the first left out testing subject is obtained based on the majority voting of the constructed $N_s - 1$ DTs. The process is repeated for N_s times, each time left out different subject as the testing subject. Finally, the overall classification performance is obtained by comparing the predicted labels of all subjects with the ground truth.

3 Experimental Results

The constructed effective connectivity maps of one healthy control (NC) and one MCI patient are shown in Fig. 1(a)–(b), respectively. The constructed effective networks are sparse and asymmetric, indicating that the interaction within a pair of ROIs, *i.e.*, $g_{i,j}$ and $g_{j,i}$ are not restricted to be equal. The sums of ERR for effective connectivity of each ROI for corresponding MCI and NC subjects are shown in Fig. 1(c)–(d). Almost all ROIs show a sum larger than 0.95, indicating that the rs-fMRI time-series of an ROI can be well represented by the time series of other selected. It also assure the convergence of the network.

In this work, the proposed sparse effective connectivity network based framework is compared with other related works using the same dataset, including the framework that uses single modality data, either DTI or rs-fMRI data, separately [12], and the framework that directly fuses features from multiple modalities at the feature level (Direct) [12]. We further compared the performance of our proposed framework to the results of the Weighted-Lasso with OFR procedure (Weighted-Lasso-OFR) which omits the derivative information, and the results of Ultra-Lasso with UOFR procedure (Ultra-Lasso-UOFR) which without penalization weights from DTI-based structural connectivity. The LOOCV classification results of all compared frameworks are summarized in Table 1. The proposed framework, which combines the information from multimodal neuroimaging data and the Ultra-Weighted-Lasso and the UOFR approaches by yielding a cross-validation accuracy of 96.3%, which is at least 7.4% improvement compared to the second best performed framework (Direct). Furthermore, it also outperforms all other frameworks in terms of the rest of the computed statistical measurements. Particularly, an area of 0.994 under the receiver operating characteristic curve (AUC) demonstrates an excellent generalization power of our proposed Sparse Effective Connectivity for MCI Identification 305 framework. Our method achieved a gain of at least 11% in accuracy while requiring only extra 15% more of computation time if compared to the single-modal Ultra-Lasso-UOFR and multimodal Weighted-Lasso-OFR approaches, implying its efficacy in accurate inference of effective connectivity.

The most discriminant regions selected by our framework are mainly located in the frontal lobe (e.g., superior frontal gyrus [13]), the temporal lobe (e.g., temporal pole [8]), the occipital lobe (e.g., middle occipital gyrus and superior occipital gyrus [8]) and other

regions such as cingulum gyri [11,13], hippocampus [8], and thalamus [8], in line with those reported in the AD/MCI literature. The selected most discriminant regions are graphically displayed in Fig. 2.
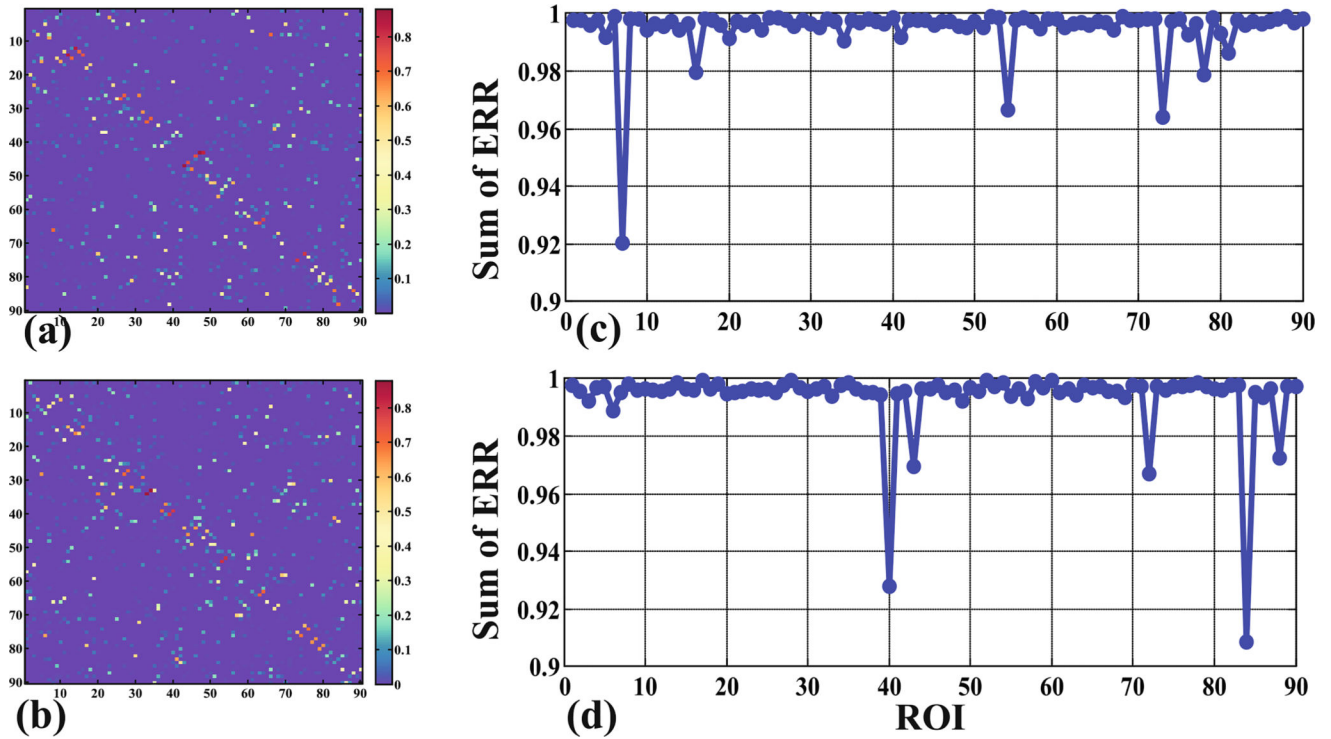
4 Discussions and Conclusion

In this study, we propose an Ultra-Weighted-Lasso-UOFR based effective connectivity network inference method using multimodal DTI and rs-fMRI data, and explore its diagnostic power for distinguishing MCI patients from healthy subjects. In this framework, multimodal integration is achieved via an Ultra Weighted-Lasso method where weighted penalties derived from the DTI data are incorporated into a sparse regression procedure for identifying the topology of effective connectivity network. This method provides a more accurate detection of the high dimensional effective interaction architecture among brain regions via the returned non-zero coefficients. Additionally, effective connectivity strength is estimated using an orthogonal forward regression procedure based on the identified network structure. Experimental results on MCI classification demonstrate the superiority of the constructed sparse effective connectivity network over other competing methods. In conclusion, the proposed approach sheds a light on integrating information from multimodal neuroimaging data to infer sparse effective connectivity for brain disease diagnosis.
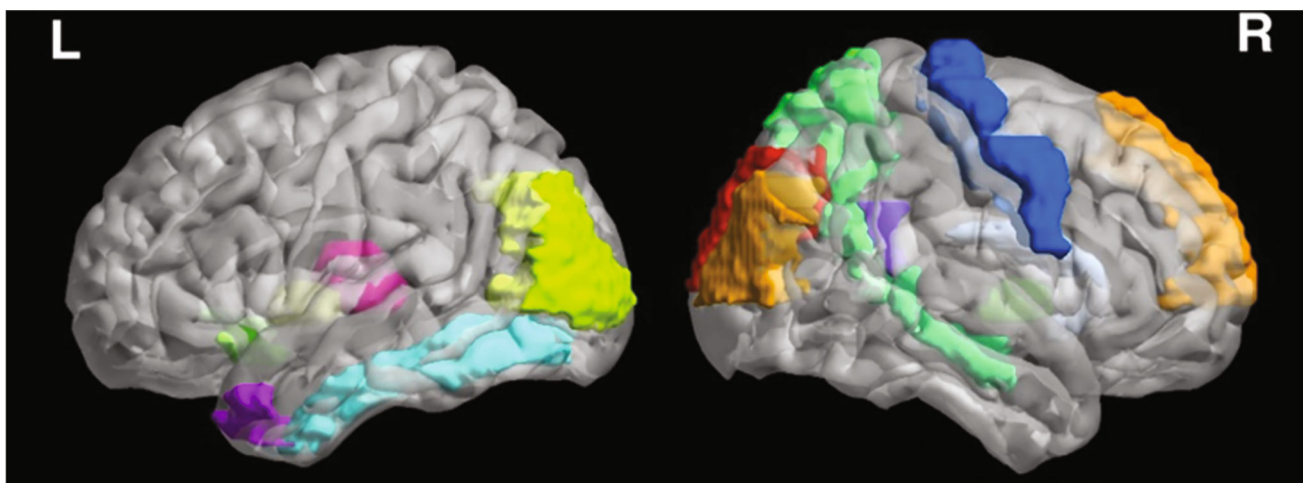
References

1. Akhoondzadeh M. Decision tree, bagging and random forest methods detect TEC seismo-ionospheric anomalies around the time of the chile, (m−w = 8.8) earthquake of 27 february 2010. Adv. Space Res. 2016; 57(12):856–867.

2. Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex. 2012; 24:663–676. [PubMed: 23146964]

3. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. Phys. Rep. 2006; 424(4):175–308.

4. Friston KJ. Functional and effective connectivity: a review. Brain Connect. 2011; 1(1):13–36. [PubMed: 22432952]

5. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H. Mild cognitive impairment. The Lancet. 2006; 367(9518):1262–1270.

6. Li Y, Cui WG, Guo YZ, Huang T, Yang XF, Wei HL. Time-varying system identification using an ultra-orthogonal forward regression and multiwavelet basis functions with applications to EEG. IEEE Trans. Neural Netw. Learn. Syst. 2017; PP(99):1–13.

7. Liu F, Wee CY, Chen H, Shen D. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. Neuroimage. 2014; 84:466–475. [PubMed: 24045077]

8. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front. Neurosci. 2015; 9:307. [PubMed: 26388719]

9. 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]

10. Wang K, Liang M, Wang L, Tian L, Zhang X, Li K, Jiang T. Altered functional connectivity in early Alzheimer's disease: a restingstate fMRI study. Hum. Brain Mapp. 2007; 28(10):967–978. [PubMed: 17133390]

11. Wang Z, Nie B, Li D, Zhao Z, Han Y, Song H, Xu J, Shan B, Lu J, Li K. Effect of acupuncture in mild cognitive impairment and Alzheimer disease: a functional MRI study. PLoS One. 2012; 7(8):e42730. [PubMed: 22916152]

12. 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]

13. Xu L, Wu X, Li R, Chen K, Long Z, Zhang J, Guo X, Yao L. Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers. J. Alzheimer's Dis. 2016; 51(4): 1045–1056. [PubMed: 26923024]

14. Zou H. The adaptive lasso and its oracle properties. JASA. 2006; 101(476):1418–1429.

Fig. 1.
Effective connectivity maps based on Ultra-Weighted-Lasso and UOFR for one (a) MCI and one (b) NC subjects. Sums of ERR for effective connectivity of each ROI for the corresponding (c) MCI and (d) NC subjects.

Fig. 2.
Most discriminant regions selected during MCI classification

Table 1

Classification performance of the proposed and competing frameworks.

Methods	ACC (%)	AUC	SEN (%)	SPE (%)	BAC (%)
DTI	88.9	0.935	80.0	94.1	87.1
fMRI	70.4	0.788	70.0	70.6	70.3
Direct	88.9	0.912	90.0	88.2	89.1
Weighted-Lasso-OFR	85.2	0.859	80.0	88.2	84.1
Ultra-Lasso-UOFR	81.5	0.824	60.0	94.1	77.1
Proposed	**96.3**	**0.994**	**100.0**	**94.1**	**97.1**

(ACC = Accuracy; SEN = Sensitivity; SPE = Specificity; BAC = Balanced Accuracy)