Structural Brain Network Constrained Neuroimaging Marker Identification for Predicting Cognitive Functions

De Wang[1], Feiping Nie[1], Heng Huang[1,⋆], Jingwen Yan[2], Shannon L. Risacher[2], Andrew J. Saykin[2], Li Shen[2,⋆], and for the Alzheimer's Disease Neuroimaging Initiative[⋆⋆]

[1] Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA
{wangdelp,feipingnie}@gmail.com,heng@uta.edu
[2] Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
jingyan@umail.iu.edu,{srisache,asaykin,shenli}@iu.edu

Abstract. Neuroimaging markers have been widely used to predict the cognitive functions relevant to the progression of Alzheimer's disease (AD). Most previous studies identify the imaging markers without considering the brain structural correlations between neuroimaging measures. However, many neuroimaging markers interrelate and work together to reveal the cognitive functions, such that these relevant markers should be selected together as the phenotypic markers. To solve this problem, in this paper, we propose a novel network constrained feature selection (NCFS) model to identify the neuroimaging markers guided by the structural brain network, which is constructed by the sparse representation method such that the interrelations between neuroimaging features are encoded into probabilities. Our new methods are evaluated by the MRI and AV45-PET data from ADNI-GO and ADNI-2 (Alzheimer's Disease Neuroimaging Initiative). In all cognitive function prediction tasks, our new NCFS method outperforms other state-of-the-art regression approaches. Meanwhile, we show that the new method can select the correlated imaging markers, which are ignored by the competing approaches.

Keywords: Neuroimaging Marker Identification, Brain Network Based Feature Selection, Correlated Marker Selection, Imaging Genetics.

⋆ Corresponding author.

⋆⋆ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

J.C. Gee et al. (Eds.): IPMI 2013, LNCS 7917, pp. 536–547, 2013.

1 Introduction

Alzheimer's disease (AD) is the most common dementia that has acquired tremendous research attention due to its wide range of effects and mysterious underlying mechanisms. AD is a neurodegenerative disorder characterized by progressive impairment of cognitive functions, and thus it is crucial to understand how structural and functional changes in brain can influence cognitve performance. As a powerful tool for capturing neurodegenerative process in AD progression, neuroimaging data have been widely studied for classification of disease status [9,3] and more recently for prediction of disease-relevant cognitive scores [13,10].

Regression analyses were commonly used to predict cognitive scores from imaging measures. In [13], stepwise regression was performed in a pairwise way to associate the Magnetic Resonance Imaging (MRI) and FDG-PET (Positron Emission Tomography) measures to memory scores. Because the univariate model was used in this study, their results neglected the interrelated structures within both imaging and clinical data. As a special case of a sparse linear model, the relevance vector regression method was applied to relate the Voxel-Based Morphometry (VBM) features to selected clinical scores [10]. This approach was connected to the combination of Bayesian model and least absolute shrinkage and selection operator (LASSO) model [12]. Another study [14] used the structured sparsity-inducing norm, $\ell_{2,1}$-norm, to select neuroimaging markers that are important to most prediction tasks. Later work [15,16] presented joint sparse multitask models to identify the imaging markers related to both cognitive scores and outcomes. In most recent work, we introduced the high-order low-rank sparse learning models to select the longitudinal neuroimaging markers associated to genetic basis [17] or cognitive scores [18].

Most existing studies selected the neuroimaging features totally based on their influences on the prediction results, *i.e.* in a model-driven way. However, from the functional brain circuitry point of view, many neuroimaging markers interrelate with each other and work together to reveal the brain cognitive functions. Thus, it is desired to explore and utilize such interrelation structures and select these important and structurally correlated features together.

In this paper, we propose a new brain network constrained feature selection (NCFS) method to naturally integrate the structural brain network into cognitive function prediction model. As a result, our selected imaging markers not only have prediction power, but also indicate the feature interrelations in structural brain network. We apply the new method to analyze the MRI, AV45-PET and cognitive data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The neuroimaging markers are identified to predict the cognitive scores from five sets of different neuropsychological tests.

2 Structural Brain Network Construction Using A New Sparse Representation Model

To understand and characterize the underlying architectures of complex brain networks, previous neuroscience studies utilized the Pearson correlation

coefficients of cortical thickness measurements from MRI to create the network of anatomical connections [2,6]. Thus, given the neuroimaging measures, we can create the structural brain network using their correlation coefficients. In this constructed brain network, each vertex represents one neuroimaging feature and each edge between two vertices encodes their correlation. However, this simple correlation analysis may not effectively construct the brain anatomical structure. To solve this problem, we propose a new sparse representation model to construct an effective structural brain network among neuroimaging measures.

Denote the neuroimaging features are $F = [f_1, f_2, \cdots, f_d] \in \Re^{n \times d}$, where n is the number of subject and d is the number of features. Our goal is to construct a connectivity matrix A, in which $A_{i,j}$ encodes the correlation between features f_i and f_j. A popular method to compute A is using the Gaussian kernel function. The major disadvantage of this method is that the hyper-parameter σ in the Gaussian kernel function is very sensitive and is difficult to tune in practice.

Recently, as an approximate formulation for Gaussian graphical modeling [7], the sparse coding method was applied to compute the similarity matrix A. The similarity vector α_i includes the similarities between the i-th feature and the rest features, and is calculated by the sparse representation as follows:

$$\min_{\alpha_i} ||F^{-i}\alpha_i - f_i||_2^2 + \lambda||\alpha_i||_1, \tag{1}$$

where $F^{-i} = [f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_d] \in \Re^{n \times (d-1)}$. However, the above model is not shift-invariant. We propose to impose two new constraints: $\alpha_i^T 1 = 1$ and $\alpha_i \geq 0$, such that $||(F^{-i} + t1^T)\alpha_i - (f_i + t)||_2^2 = ||F^{-i}\alpha_i - x_i||_2^2$. More important, based on these two new constraints, the learned similarities in α_i can be interpreted as probabilities. Interestingly, the new constraints make the second term be constant. So the sparse representation model is to solve:

$$\min_{\alpha_i \geq 0, \alpha_i^T 1 = 1} ||F^{-i}\alpha_i - f_i||_2^2. \tag{2}$$

Because the parameter λ is canceled, our new model doesn't require any parameter tuning and is suitable for practical applications. Our new objective is a non-smooth convex problem. We use the accelerated projected gradient method to solve this problem.

3 Structural Brain Network Constrained Neuroimaging Feature Learning

To identify biologically meaningful markers, we integrate the structural brain network into cognitive function prediction model. The structural brain network can constrain and guide the machine learning model to identify important and correlated biomarkers, which potentially play the key roles in memory and cognition circuitry. To predict cognitive scores, we use regression methods, but it is challenging to incorporate the network connectivity matrix A into a regression

model. To solve this problem, we introduce a novel structural brain network constrained feature selection (NCFS) method with new optimization algorithm.

To select the correlated features, in the regression model, if the i-th and j-th features have high similarity in A, their coefficient vectors w^i and w^j ($w^i \in \Re^{c\times 1}$ is the transpose of the i-th row of parameter matrix W) should be both large or both small, *i.e.*, similar to each other. Because the similarity vector α_i encodes the similarity between the i-th feature and other features, the coefficient vector w^i and the rest coefficient vectors $W^{-i} = [w^1, ..., w^{i-1}, w^{i+1}, ..., w^d] \in \Re^{c\times(d-1)}$ should also have similar sparse representation with α_i. Thus, we should minimize:

$$\sum_{i=1}^{d} ||W^{-i}\alpha_i - w^i||^2 = Tr(W^T(I-A)(I-A)^T W), \tag{3}$$

where $A = [\tilde{\alpha}_1, ..., \tilde{\alpha}_d] \in \Re^{d\times d}$, in which $\tilde{\alpha}_i \in \Re^{d\times 1}$ is an augmented $\alpha_i \in \Re^{(d-1)\times 1}$ with the i-th element as 0.

Denoting $L = (I-A)^T(I-A)$, our new NCFS objective is to solve:

$$\min_{W,b} ||X^T W + 1b^T - Y||_F^2 + \gamma_1 Tr(W^T L W) + \gamma_2 ||W||_{2,1}. \tag{4}$$

In our new objective, the first regularization term imposes the structural brain network constraint, and the second regularization term is the structured sparse mixed norm to select features which are important to all prediction tasks. Thus, the features identified by our NCFS model are important to all cognitive functions prediction and also correlated in the structural brain network.

Taking the derivative w.r.t. b and setting to zero, we have $b = \frac{1}{n}Y^T 1 - \frac{1}{n}W^T X 1$. Thus, our objective becomes:

$$\min_{W} ||CX^T W - CY||_F^2 + \gamma_1 Tr(W^T L W) + \gamma_2 ||W||_{2,1}, \tag{5}$$

where $C = I - \frac{1}{n}11^T$ is the centering matrix. This new objective is a non-smooth convex problem with two regularization terms. We will derive an efficient algorithm to optimize the new objective.

3.1 New Optimization Algorithm to Solve Problem (5)

By taking the derivative of Eq. (5) w.r.t. W and setting to zero, we have:

$$XCX^T W - XCY + (\gamma_1 L + \gamma_2 P)W = 0, \tag{6}$$

where P is a diagonal matrix, the i-th diagonal element is $\frac{1}{2||w^i||_2}$. Eq. (6) leads to the solution W as:

$$W = (XCX^T + \gamma_1 L + \gamma_2 P)^{-1} XCY. \tag{7}$$

Note that P is not a constant and depends on the unknown W. We propose an iterative algorithm to solve W based on Eq. (7). As initialization, we guess a

Initialize $W \in \Re^{d \times c}$
repeat
1. Calculate diagonal matrix P, the i-th diagonal element is $\frac{1}{2||w^i||_2}$.
2. Update W by $W = (XCX^T + \gamma_1 L + \gamma_2 P)^{-1} XCY$

until *Converges*

Algorithm 1. Optimization algorithm to solve problem (5)

solution W, then we calculate P with current W and update W with Eq. (7). The detailed algorithm is summarized in Algorithm 1.

If Algorithm 1 converges, then the converged solution W satisfies Eq. (6). Since problem (5) is convex, the converged solution W is the global optimal solution to problem (5). In next subsection, we will prove that Algorithm 1 indeed converges.

3.2 Convergence Analysis on Our New Algorithm

For the Algorithm 1, we have the following theorem:

Theorem 1. *In each iteration, Algorithm 1 will decrease the objective value of problem (5) till the algorithm converges.*

Proof: In the Step 2 of Algorithm 1, we denote the updated solution W as $\tilde{W}$. According to Step 2, we have

$$\tilde{W} = \arg\min_W ||CX^T W - CY||_F^2 + Tr(W^T(\gamma_1 L + \gamma_2 P)W). \tag{8}$$

Since $\tilde{W}$ is the optimal solution in Eq. (8), we have

$$||CX^T\tilde{W}-CY||_F^2+Tr(\tilde{W}^T(\gamma_1 L+\gamma_2 P)\tilde{W}) \leq ||CX^T W-CY||_F^2+Tr(W^T(\gamma_1 L+\gamma_2 P)W)$$

According to the definition of P in Step 1 of Algorithm 1, we have

$$\begin{aligned} &||CX^T\tilde{W} - CY||_F^2 + \gamma_1 Tr(\tilde{W}^T L\tilde{W}) + \gamma_2 \sum_{i=1}^{d} \frac{||\tilde{w}^i||_2^2}{2||w^i||_2} \\ &\leq ||CX^T W - CY||_F^2 + \gamma_1 Tr(W^T LW) + \gamma_2 \sum_{i=1}^{d} \frac{||w^i||_2^2}{2||w^i||_2} \end{aligned} \tag{9}$$

Based on a inequality $||\tilde{w}||_2 - \frac{||\tilde{w}||_2^2}{2||w||_2} \leq ||w||_2 - \frac{||w||_2^2}{2||w||_2}$ for any vectors w and $\tilde{w}$, we have

$$\gamma_2 \sum_{i=1}^{d} ||\tilde{w}^i||_2 - \gamma_2 \sum_{i=1}^{d} \frac{||\tilde{w}^i||_2^2}{2||w^i||_2} \leq \gamma_2 \sum_{i=1}^{d} ||w^i||_2 - \gamma_2 \sum_{i=1}^{d} \frac{||w^i||_2^2}{2||w^i||_2} \tag{10}$$

Summing over Eq. (9) and Eq. (10) on both sides, we arrive at

$$\begin{aligned} &||CX^T\tilde{W} - CY||_F^2 + \gamma_1 Tr(\tilde{W}^T L\tilde{W}) + \gamma_2 ||\tilde{W}||_{2,1} \\ &\leq ||CX^T W - CY||_F^2 + \gamma_1 Tr(W^T LW) + \gamma_2 ||W||_{2,1}\,. \end{aligned} \tag{11}$$

Table 1. Participant characteristics

Category	HC	MCI	AD
Number	105	237	18
Gender(M/F)	53/52	137/100	11/7
Handness(R/L)	98/7	208/29	16/2
Baseline age (mean±std)	74.2±5.7	71.1±7.5	76.2±11.0
Education (mean±std)	16.4±2.6	16.2±2.6	15.3±2.7

Thus, our algorithm will decrease the objective value of problem (5) in each iteration till it converges. □

Because the objective value of problem (5) has a lower bound 0, the algorithm will converge. As mentioned before, the algorithm will converge to the global optimal solution of problem (5).

4 Experimental Results and Discussions

4.1 Neuroimaging and Cognition Data Descriptions

We have applied and evaluated the proposed new models on the neuroimaging and cognitive data downloaded from the ADNI database (`adni.loni.ucla.edu`). One goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see `www.adni-info.org`.

We downloaded corrected 3T structural MRI scans [4], pre-processed AV-45 PET scans [5], and cognitive data from the ADNI website. Our analysis focused on the baseline MRI, AV45-PET and cognitive data at the ADNI-GO/2 phase. All the participants with a baseline diagnosis were involved in the study, including 105 health control (HC), 237 MCI and 18 AD participants (Table 1).

VBM in SPM8 [1] was applied to preprocess structural MRI scans, as previously described [8]. Briefly, scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) maps, normalized to MNI space, and smoothed with an 8mm FWHM kernel. Besides, all scans were also processed through automated segmentation and parcellation using Freesurfer version 5.1.

AV-45 PET scans were pre-processed using techniques identical to the previous techniques for processing ADNI PiB PET scans [5]. Standardized uptake value ratio (SUVR) AV-45 PET images were created by intensity normalizing to a mean cerebellar GM region of interest (ROI). Downloaded scans were co-registered to the structural MRI scan from the corresponding visit and normalized to MNI space using SPM8, as described in [11]. Mean AV-45 measures were calculated for ten ROIs, including frontal lobe, parietal lobe, temporal lobe,Limbic lobe, occipital lobe, anterior cingulate, poster cingulate, precuneus, and cerebellum. In addition, a summarized measure were calculated based on the ROIs that best differentiate AD and HC in previous experiment.

Table 2. Comparisons of cognitive score prediction performance of five methods using the average RMSEs (mean+std)

		FLU	ADAS	TRAILS	RAVLT	MMSE
VBM	NCFS	**5.0772±0.6240**	**6.2277±0.5672**	**41.8300±2.7196**	**5.4153±0.2972**	1.8822±0.0821
	$\ell_{2,1}$	5.0819±0.6452	6.3236±0.6508	43.3156±2.9257	5.4931±0.3308	1.8657±0.112
	Lasso	5.0820±0.6452	6.3239±0.6505	42.5423±2.2569	5.4760±0.3057	1.8657±0.1112
	Ridge	5.1109±0.6410	6.2382±0.5833	42.1515±2.8509	5.4517±0.2611	**1.8437±0.0743**
	MVLR	5.8757±0.6087	7.0166±0.5415	48.6541±3.9337	6.2138±0.3610	2.1477±0.0920
FS	NCFS	**4.7605±0.5210**	**5.8323±0.3874**	**40.3119±3.3601**	**5.2145±0.3722**	**1.7694±0.1230**
	$\ell_{2,1}$	4.7608±0.4870	5.8441±0.4006	40.9824±3.3570	5.2553±0.33593	1.7719±0.1261
	Lasso	4.7611±0.4869	5.8443±0.4007	40.9353±3.6106	5.2530±0.3814	1.7719±0.1262
	Ridge	4.8796±0.5050	5.8887±0.4198	40.7526±3.0069	5.2245±0.3180	1.7830±0.1017
	MVLR	5.6263±0.3744	6.6088±0.2690	49.0891±2.9285	6.0465±0.4743	2.0364±0.1004
AV45	NCFS	**5.0983±0.5416**	**6.5406±0.2506**	**41.0382±3.2921**	**5.6701±0.5030**	**1.8664±0.1170**
	$\ell_{2,1}$	5.1015±0.5557	6.5554±0.2456	41.9643±3.1609	5.7316±0.6920	1.8673±0.1241
	Lasso	5.1015±0.5557	6.5554±0.2456	41.8831±3.2199	5.7454±0.6409	1.8673±0.1241
	Ridge	5.1448±0.5889	6.5491±0.2529	42.0099±3.2689	5.8452±0.6344	1.8776±0.1291
	MVLR	5.1907±0.6372	6.5793±0.2652	42.9643±3.3477	5.9513±0.7656	1.9226±0.1679

Overall, we had 90 VBM measures, 95 Freesurfer measures, and eleven AV45 measures for each subject. Based on these imaging measures, we performed different regression models to predict five types of cognitive scores: (1) Alzheimer's Disease Assessment Scale-Cognitive test (ADAS); (2) Rey Auditory Verbal Learning Test (RAVLT); (3) FLUENCY; (4) Mini-Mental State Examination (MMSE); and (5) Trail Making Test (TRAILS).

4.2 Experimental Setting

We compared our new method (NCFS) with several competing regression models: multivariate linear regression (MVLR), ridge regression, structured sparsity-inducing norm regularized regression [14] (denoted as $\ell_{2,1}$), and Lasso regularized regression. The evaluation metric used was the standard root mean square error (RMSE), which has been widely used in performance evaluation of regression analysis. For each method, five-fold cross-validation was performed to obtain the average RMSE. In each of five trials, an internal five-fold cross-validation was done to tune the parameters using grid search for different methods. The range of each parameter varied from 10^{-5} to 10^{5}. The reported results were the best results of each method with the optimal parameter.

4.3 Cognitive Functions Prediction Performance Comparisons

The cognitive scores prediction results were reported in Table 2. The proposed method NCFS consistently outperformed other methods by predicting five cognitive scores more accurately with lower RMSEs. The key difference between NCFS and other methods was that the NCFS method utilized the correlation between features by incorporating the structural brain network as the regularization term, whereas other methods tended to select individual feature and ignored other less important but correlated features. Thus, NCFS had a

Table 3. Average correlation values of top ranked features in structural brain network (mean+std) constructed by the sparse representation model

		Top 5	Top 10	Top 20		Top 2
VBM	NCFS	**0.0508±0.0472**	**0.0282±0.0179**	**0.0200±0.0070**	AV45	**0.1980±0.1182**
	$\ell_{2,1}$	0.0227±0.0078	0.0119±0.0097	0.0126±0.0037		0.1342±0.1225
	Lasso	0.0240±0.0083	0.0142±0.0074	0.0131±0.0039		0.1342±0.1225
	Ridge	0.0491±0.0298	0.0277±0.0146	0.0161±0.0047		0.1533±0.1453
	MVLR	0.0196±0.0200	0.0156±0.0093	0.0157±0.0032		0.1395±0.1278
FS	NCFS	**0.0625±0.0332**	**0.0334±0.0165**	0.0196±0.0057	-	-
	$\ell_{2,1}$	0.0299±0.0359	0.0237±0.0112	0.0153±0.0060		
	Lasso	0.0269±0.0311	0.0227±0.0113	0.0131±0.0039		
	Ridge	0.0395±0.0331	0.0258±0.0147	**0.0217±0.0080**		
	MVLR	0.0278±0.0316	0.0207±0.0132	0.0162±0.0044		

built-in mechanism to select structural correlated features which acted together to impact human cognitive functions. MVLR had the largest RMSEs than all other methods, because it was prone to overfit the training data. Table 2 also indicated that FreeSurfer measures were more powerful for predicting cognitive performance than VBM and AV45 measures.

We also evaluated the average correlation values of top ranked features in all methods. For each method, we extracted top 5, 10 and 20 features for FreeSurfer and VBM measures. Because AV45 only had 11 features, we selected top 2 features (*i.e.* top 20% of all features) for each method. The correlation of features was computed from the structural brain network obtained by the sparse representation model, which reflected the similarity between features. Table 3 showed the average correlations of top ranked features by different methods. In most case, top ranked features by NCFS method were more correlated than that of other methods. The correlations of top ranked features in $\ell_{2,1}$ and Lasso approaches were relatively low because they tended to select a single feature from a group of correlated features. NCFS selected more (correlated) features in structural brain network, which helped achieve better prediction performance.

4.4 Imaging Markers Analysis and Discussions

Figure 1 visualized the resulting structural brain network of 95 FreeSurfer measures as a connectivity matrix. Indices 1-7 corresponded to seven unilateral measures, which were listed in blue (from BrainStem to CSF) on the x axis of Figures 2. Indices 8-51 (52-95) corresponded to 44 bilateral measures on the left (right) hemisphere, which were listed in black (from AccumVol to TransvTemporal) on the x axis of Figures 2. The white lines separated the unilateral and bilateral measures as well as left and right measures. It was obvious that the left and right hemispheres demonstrated a very similar pattern.

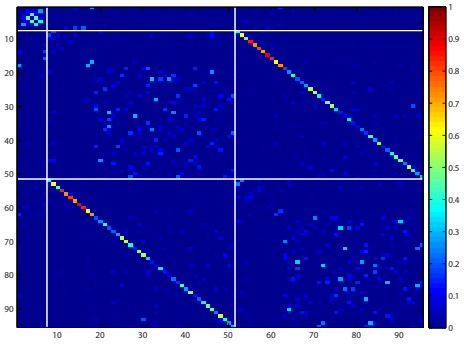

Fig. 1. The resulting structural brain network of 95 freesurfer measures is shown as a connectivity matrix. Indices 1-7 correspond to seven unilateral measures, which are listed in blue (from BrainStem to CSF) on the x axis of Figure 2. Indices 8-51 (52-95) correspond to 44 bilateral measures on the left (right) hemisphere, which are listed in black (from AccumVol to TransvTemporal) on the x axis of Figure 2.

The results of linear regression and ridge regression were not sparse, making it difficult for biomarker identification. Thus, here we only compared the feature selection results between NCFS and $\ell_{2,1}$. Both NCFS and $\ell_{2,1}$ were sparse models that were able to identify a compact set of relevant imaging markers and to explain the underlying brain structural changes related to cognitive status. Shown in Figure 2 were the maps of regression weights for predicting various cognitive scores (from top to bottom: Fluency, ADAS, TRAIL, RAVLT and MMSE) using the FreeSurfer measures. Average regression weights of 5-fold cross-validation trials were plotted for NCFS (the even panels) and $\ell_{2,1}$ (the odd panels). In each panel, the odd (even) rows showed the weights from left (right) hemisphere. Note that the first seven measures (from BrainStem to CSF, colored in blue) were unilateral, and thus their left and right measures were set to be the same. Blue indicated negative correlation, while red indicated positive correlation. The bigger the magnitude of a coefficient was, the more important its imaging measure was in predicting the corresponding cognitive score.

Clearly, $\ell_{2,1}$ yielded more sparse patterns than NCFS. However, for highly correlated features, $\ell_{2,1}$ tended to identify one and ignore the others. This was inadequate for yielding a biologically meaningful interpretation. In contrast, NCFS did seem to work in terms of structuring the identified patterns. For example, the hippocampal volume and amygdala volume were correlated (Fig. 1). They were selected together by NCFS for predicting ADAS and RAVLT scores, while $\ell_{2,1}$ only selected hippocampal volume in this case. In addition, bilateral measures of the same structure were often highly correlated. The NCFS yielded a more symmetric pattern than $\ell_{2,1}$ in many cases.

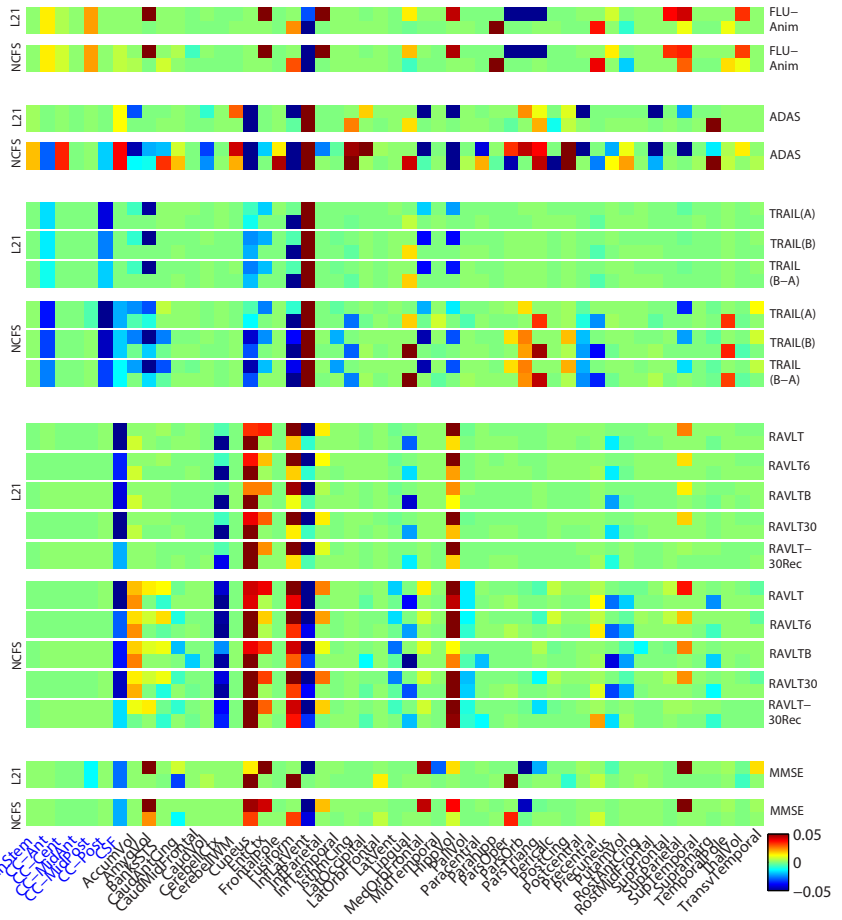

Fig. 2. Heat maps of regression weights for predicting cognitive scores (from top to bottom: Fluency, ADAS, TRAIL, RAVLT and MMSE) from FreeSurfer measures using NCFS (the even panels) and $\ell_{2,1}$ (the odd panels) methods. In each panel, the odd (even) rows show the weights from left (right) hemisphere. Note that the first seven measures (from BrainStem to CSF, colored in blue) are unilateral, and thus their left and right measures are set to be the same. Blue indicates negative correlation, while red indicates positive correlation.

Finally, the identified patterns were in fact expected based on our prior knowledge. For example, RAVLT measured verbal learning memory; and the identified regions included hippocampus, fusiform, entorhinal cortex, and other regions relevant to learning and memory.

5 Conclusions

In this paper, we have proposed two new machine learning models, one for constructing a structural brain network and the other for identifying

cognition-releveant neuroimaging markers using the constructed brain network. This framework discovers imaging biomarkers that are not only important to the cognitive outcomes but also interrelating with each other in the structural brain network. We have applied the new computational models to predicting cognitive outcomes using FreeSurfer, VBM, and AV45 data from the ADNI database. All the empirical results have demonstrated consistently improved performance of our method over the state-of-the-art competing approaches.

Acknowledgement. This research was supported by NSF CCF-0830780, CCF-0917274, DMS-0915228, and IIS-1117965 at UTA; and by NSF IIS-1117335, NIH R01 LM011360, UL1 RR025761, U01 AG024904, RC2 AG036535, R01 AG19771, and P30 AG10133-18S1 at IU.

Data used for this project was funded by the ADNI (U01AG024904). ADNI is funded by the NIA, NIBIB, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd; Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The CIHR is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the FNIH (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at UCSD. ADNI data are disseminated by LONI at UCLA.

References

1. Ashburner, J., Friston, K.: Voxel-based morphometry–the methods. Neuroimage 11(6), 805–821 (2000)
2. Chen, Z.J., He, Y., Rosa-Neto, P., Germann, J., Evans, A.C.: Revealing modular architecture of human brain structural networks by using cortical thickness from mri. Cerebral Cortex 18, 2374–2381 (2008)
3. Hinrichs, C., Singh, V., et al.: Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage 48(1), 138–149 (2009)
4. Jack Jr., C.R., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., Killiany, R.J., Decarli, C.S., Dale, A.M., Carmichael, O.W., Tosun, D., Weiner, M.W.: Update on the magnetic resonance imaging core of the alzheimer's disease neuroimaging initiative. Alzheimers Dement 6(3), 212–220 (2010)
5. Jagust, W.J., Bandy, D., Chen, K., Foster, N.L., Landau, S.M., Mathis, C.A., Price, J.C., Reiman, E.M., Skovronsky, D., Koeppe, R.A.: The alzheimer's disease neuroimaging initiative positron emission tomography core. Alzheimers Dement 6(3), 221–229 (2010)

6. Khundrakpam, B.S., Reid, A., Brauer, J., et al.: Developmental changes in organization of structural brain networks. Cerebral Cortex (2012)
7. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the Lasso. The Annals of Statistics 34(3), 1436–1462 (2006)
8. Risacher, S.L., Saykin, A.J., West, J.D., Shen, L., Firpi, H.A., McDonald, B.C.: Baseline mri predictors of conversion from mci to probable ad in the adni cohort. Curr. Alzheimer Res. 6(4), 347–361 (2009)
9. Shen, L., Qi, Y., Kim, S., Nho, K., Wan, J., Risacher, S.L., Saykin, A.J., ADNI: Sparse bayesian learning for identifying imaging biomarkers in AD prediction. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 611–618. Springer, Heidelberg (2010)
10. Stonnington, C.M., Chu, C., et al.: Predicting clinical scores from magnetic resonance scans in alzheimer's disease. Neuroimage 51(4), 1405–1413 (2010)
11. Swaminathan, S., Shen, L., Risacher, S.L., Yoder, K.K., West, J.D., Kim, S., Nho, K., Foroud, T., Inlow, M., Potkin, S.G., Huentelman, M.J., Craig, D.W., Jagust, W.J., Koeppe, R.A., Mathis, C.A., Jack Jr., C.R., Weiner, M.W., Saykin, A.J.: Amyloid pathway-based candidate gene analysis of [(11)c]pib-pet in the Alzheimer's disease neuroimaging initiative (adni) cohort. Brain Imaging Behav. 6(1), 1–15 (2012)
12. Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. Royal. Statist. Soc. B 58, 267–288 (1996)
13. Walhovd, K., Fjell, A., et al.: Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiol. Aging 31(7), 1107–1121 (2010)
14. Wang, H., Nie, F., Huang, H., Risacher, S.L., Ding, C., Saykin, A.J., Shen, L.: Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In: IEEE Conference on Computer Vision, pp. 557–562 (2011)
15. Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, A.J., Shen, L., ADNI: Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 115–123. Springer, Heidelberg (2011)
16. Wang, H., Nie, F., Huang, H., Risacher, S.L., Saykin, A.J., Shen, L., ADNI: Identifying disease sensitive and quantitative trait relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multi-modal multitask learning. In: 20th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB), vol. 28, pp. i127–i136 (2012)
17. Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Nho, K., Risacher, S.L., Saykin, A.J., Shen, L.: From Phenotype to Genotype: An Association Study of Candidate Phenotypic Markers to Alzheimer's Disease Relevant SNPs. Bioinformatics 28, i619–i625 (2012)
18. Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Risacher, S., Saykin, A., Shen, L.: High-Order Multi-Task Feature Learning to Identify Longitudinal Phenotypic Markers for Alzheimer Disease Progression Prediction. In: Advances in Neural Information Processing Systems, NIPS (2012)