Published in final edited form as:
Proc IEEE Int Symp Biomed Imaging. 2014 May ; 2014: 665–668. doi:10.1109/ISBI.2014.6867958.

JOINT IDENTIFICATION OF IMAGING AND PROTEOMICS BIOMARKERS OF ALZHEIMER'S DISEASE USING NETWORK-GUIDED SPARSE LEARNING

Jingwen Yan[1], **Heng Huang**[2], **Sungeun Kim**[1], **Jason Moore**[3], **Andrew Saykin**[1], and **Li Shen**[1,†] **for the ADNI**[‡]

[1]Radiology and Imaging Sciences, BioHealth Informatics, Indiana University, IN, USA

[2]Computer Science and Engineering, University of Texas at Arlington, TX, USA

[3]Genetics, Community and Family Medicine, School of Medicine at Dartmouth College, NH, USA

Abstract

Identification of biomarkers for early detection of Alzheimer's disease (AD) is an important research topic. Prior work has shown that multimodal imaging and biomarker data could provide complementary information for prediction of cognitive or AD status. However, the relationship among multiple data modalities are often ignored or oversimplified in prior studies. To address this issue, we propose a network-guided sparse learning model to embrace the complementary information and inter-relationships between modalities. We apply this model to predict cognitive outcome from imaging and proteomic data, and show that the proposed model not only outperforms traditional ones, but also yields stable multimodal biomarkers across cross-validation trials.

Index Terms

Sparse learning; regression; neuroimaging; proteomic biomarker; cognitive outcome

1. INTRODUCTION

The study of Alzheimer's disease (AD) is experiencing an important shift from disease categories to AD-relevant outcomes, facilitating the identification of biomarkers for early detection. One particular example is the extension from disease status to cognitive outcomes. Although beta-amyloid plaques and neurofibrillary tangles are two major hallmarks of AD [1], various cognitive tests remain the most common way to help with clinical diagnosis. Exploring the relationship between multimodal imaging and biomarker measurements and cognitive outcomes is an important research topic.

[†]Correspondence to Li Shen (shenli@iu.edu).

[‡]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.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.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Earlier studies have been performed on single modality data, such as magnetic resonance imaging (MRI), positron emission tomography (PET), or cerebrospinal fluid (CSF) biomarkers. Recent studies have explored multimodal data and reported improved performance (e.g., MRI, FDG-PET and CSF biomarkers were jointly studied in [2]). Since multiple modalities are not completely isolated, there exist inter-correlations among those. Existing multimodal methods typically employ a simple strategy to bundle these data together, and thus ignore or oversimplify their relationships(e.g. [3]).

In this paper, we propose a new network-guided sparse learning model embracing both the complementary information and inter-relationships between modalities. The proposed model is applied to evaluate the predictive power of MRI and CSF proteomic measurements towards cognitive outcomes. The empirical results demonstrate significant improvements over the state-of-the-arts competing models, and also yield stable multimodal biomarkers across cross-validation trials.

2. METHODS

We write matrices as boldface uppercase letters and vectors as boldface lowercase letters. Given a matrix $\mathbf{M} = (m_{ij})$, its i-th row and j-th column are denoted as $\mathbf{m}^i$ and $\mathbf{m}_j$ respectively. The Frobenius norm and $\ell_{2,1}$-norm (also called as $\ell_{1,2}$-norm) of a matrix are defined as $\|\mathbf{M}\|_F = \sqrt{\sum_i \|\mathbf{m}^i\|_2^2}$ and $\|\mathbf{M}\|_{2,1} = \sum_i \sqrt{\sum_j m_{ij}^2} = \sum_i \|\mathbf{m}^i\|_2$, respectively.

We focus on multi-task learning paradigm, where MRI and CSF measures are used to predict one or more cognitive outcomes. Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \subseteq \Re^d$ be MRI and CSF measures and $\{\mathbf{y}_1, \cdots, \mathbf{y}_n\} \subseteq \Re^c$ cognitive outcomes, where n is the number of samples, d is the number of predictors (feature dimensionality) and c is the number of response variables (tasks). Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n]$.

The $\ell_{2,1}$ norm [4] is a multi-task version of traditional lasso. While lasso only focuses on the feature level sparsity, The $\ell_{2,1}$ norm is proposed to couple multiple tasks together in addition to the original sparsity property:

$$\min_{\mathbf{W}} \|\mathbf{W}^T\mathbf{X} - \mathbf{Y}\|_F^2 + \gamma\|\mathbf{W}\|_{2,1}. \quad (1)$$

Yet in this model the rows of $\mathbf{W}$ are equally treated, which ignores the structures among predictors. Group-Sparse Multitask Regression and Feature Selection (G-SMuRFS) method [5] was proposed to exploit the structures within and between the predictors and response variables. It assumes 1) a partition scheme exists among predictors, and 2) predictors within one partition should have similar weights. G-SMuRFS can be thought of as a multi-task version of group lasso.

In practice, the relationship among predictors may not be as simple as a straightforward partition used by G-SMuRFS. Many studies have shown that the human brain can be modeled as a complex network. To model these complicated structures among predictors, we propose a new *Network-Guided* $\ell_{2,1}$ Sparse Learning (NG-L21) model as follows.

Let **R** be the predictor correlation matrix with $\mathbf{R_{ij}}$ indicating correlation between predictors *i* and *j*. Given a threshold *t*, we can construct a network by connecting predictors with high positive and negative correlations (i.e., $|\mathbf{R_{ij}}| \geq t$). We hypothesize that positively (negatively) correlated predictors share positively (negatively) similar weights across tasks. Therefore, we try to minimize

$$\sum_{1\leq i<j\leq d \& |\mathbf{R_{ij}}|\geq t} \|\mathbf{W}_i - sign(\mathbf{R_{ij}})\mathbf{W}_j\|_F^2. \quad (2)$$

Since the selection of a proper threshold may be trivial, we also propose a threshold-free or weighted method as follows:

$$\sum_{1\leq i<j\leq d} \mathbf{R_{ij}}\|\mathbf{W}_i - sign(\mathbf{R_{ij}})\mathbf{W}_j\|_F^2. \quad (3)$$

Here *sign*($\mathbf{R_{ij}}$) deals with positive and negative correlations while $\mathbf{R_{ij}}$ itself is applied to reduce the constraint impact on the less correlated pairs. Thus, the whole regularization term can be reformatted as $\|\mathbf{NW}\|_F^2$, in which **N** is a neighboring matrix transformed from the symmetric correlation matrix **R**. In **N**, each row corresponds to one element in the correlation matrix **R** (Fig. 1). For each pair of predictors *i*, *j*, we create a row in **N** with *i*-th entry as $-\mathbf{R_{ij}}$, *j*-th entry as $-\mathbf{R_{ij}}\mathbf{sign}(\mathbf{R_{ij}})$ and all the other entries as zeros. The intuition is that the weight difference between two correlated predictors should be minimized, which is reflected by the new regularization term of $\|\mathbf{NW}\|_F^2$. Also by including the weight, the more correlated a feature pair is, the more constraint the pair is imposed by. We call this model *NG-L21*. Although this model is similar to graph-guided fusion [6] and group weighted fusion [7], neither of the prior models explored the correlation of predictors and multi-task responses together. The final objective function is formulated as:

$$\min_{\mathbf{W}} \|\mathbf{W}^T\mathbf{X} - \boldsymbol{Y}\|_F^2 + \gamma_1\|\mathbf{NW}\|_F^2 + \gamma_2\|\mathbf{W}\|_{2,1}, \quad (4)$$

where **N** is a sparse matrix where each row indicates a neighborhood relationship (i.e., edge) in the predictor network.

Well known to be non-derivable, ℓ_1-norm can be easily solved by approximate lasso where an extremely small value is added to enable the smoothness. Eq. (4) can then be solved by simply taking the derivative w.r.t **W** and setting it to 0:

$$\mathbf{X}\mathbf{X}^T\mathbf{W} - \mathbf{X}\mathbf{Y}^T + \gamma_1\mathbf{D_1}\mathbf{W} + \gamma_2\mathbf{D_2}\mathbf{W} = 0, \quad (5)$$

where $\mathbf{D_1} = \mathbf{N^T N}$, a matrix with each row integrating all neighboring relations. For the *i*-th row, it is the sum of all the rows in **N** whose *i*-th element is nonzero. $\mathbf{D_2}$ is a diagonal matrix with the *i*-th diagonal element as $\frac{1}{2\|\mathbf{w}^i\|_2}$. We have

$$\mathbf{W}=(\mathbf{X}\mathbf{X}^T+\gamma_1\mathbf{D_1}+\gamma_2\mathbf{D_2})^{-1}\mathbf{X}\mathbf{Y}^T, \quad (6)$$

where $\mathbf{W}$ can be efficiently obtained by solving $(\mathbf{X}\mathbf{X}^T + \gamma_1\mathbf{D}_1 + \gamma_2\mathbf{D}_2)\mathbf{W} = \mathbf{X}\mathbf{Y}^T$. Following [5], an efficient iterative algorithm based on Eq. (6) can be developed as follows, and can be shown to converge to the global optimum.

Input: $\mathbf{X}$, $\mathbf{Y}$
Initialize $\mathbf{W}^1 \in \mathbb{R}^{d\times c}, t = 1$;
while *not converge* **do**
1. Calculate the diagonal matrices $\mathbf{D}_2^{(t)}$, where the i-th diagonal element of $\mathbf{D}_2^{(t)}$ is $\frac{1}{2\|\mathbf{w}_i^t\|_2}$;
2. $\mathbf{W}^{(t+1)} = (\mathbf{X}\mathbf{X}^T + \gamma_1\mathbf{D}_1 + \gamma_2\mathbf{D}_2^{(t)})^{-1}\mathbf{X}\mathbf{Y}^T$;
3. $t = t + 1$;
end
Output: $\mathbf{W}^{(t)} \in \mathbb{R}^{d\times c}$.

3. RESULTS

3.1. Data and Experimental Setting

The MRI, CSF proteomic, and cognitive data were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. 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 [8]. For up-to-date information, see www.adni-info.org.

This study (N=204) included 66 AD, 57 MCI and 81 healthy control (HC) participants (Table 1). For each baseline MRI scan, FreeSurfer (FS) V4 was employed to extract 73 cortical thickness measures and 26 volume measures. 82 CSF proteomic analytes, evaluated by Rules Based Medicine, Inc. (RBM) proteomic panel [9] and surviving quality control process, were also included in this work. The 99 imaging measures and 82 proteomic analytes were used to predict a set of cognitive scores [8]: Rey Auditory Verbal Learning Test (RAVLT, 5 scores shown in Table 2 as joint outcomes). Using the regression weights from HC participants, all the MRI, CSF, and cognitive measures were pre-adjusted for the baseline age, gender, education, and handedness, with intracranial volume as an additional covariate for MRI only.

3.2. Experimental Results

We denote the weighted network model as NG-L21$_{\mathbf{w}}$, and the thresholded one as NG-L21$_{\mathbf{t}}$. For comparing performances between these two models and competing methods (i.e., Linear, Ridge, elastic net and L21), regression analysis was conducted jointly on all five RAVLT scores. Based on the assumption that FS and CSF measures could provide complementary information, we performed 18 experiments based on six different methods and three datasets (FS, CSF, FS+CSF). In each experiment, Pearson's correlation coefficients (CCs) between the actual and predicted cognitive scores were computed to measure the prediction performances. Using 10-fold cross-validation, parameters were estimated and average CCs over 10 trials were reported.

In our experiments, CSF proteomic analytes were found to have limited prediction power by itself (typically CC<0.4). But combining CSF and FS yielded improved results than using FS alone (Table 3), indicating possible complementary information provided by the two modalities. Both NG-L21 models outperformed the other methods in most cases. Ridge obtained comparable and sometimes better performances than NG-L21; but Ridge's root mean square error (not shown due to space limit) tended to be higher than NG-L21. Fig. 2(a) and Fig. 2(b) show the regression weights in heatmaps and in brain space respectively. Ridge produced non-sparse patterns, which made the results less interpretable. Both NG-L21 and L21 identified a small number of imaging markers, including AmygVol, EntCtx, and HippVol, which were known to be related to RAVLT scores.

NG-L21$_{\mathbf{w}}$ achieved similar or slightly better performance than NG-L21$_{\mathbf{t}}$. This indicates that the NG-L21 performance is mainly determined by the correlations of high values and small weights (those were included in NG-L21$_{\mathbf{w}}$ but excluded in NG-L21$_{\mathbf{t}}$) have just modest effect on improving the performance. In addition, we also compared NG-L21 with G-SMuRFS using symmetric information as grouping strategy. Generally they achieved similar performance, but tuned parameters of $\ell_{2,1}$-norm in GSMuRFS shrunk to almost 0, and led to non-sparse results.

Fig. 2(c) shows regression coefficients of CSF proteomic markers for all 5 RAVLT scores. Since all the scores measure memory performance, the coefficients across these scores are similar for the top proteomic markers. Some proteins known to be related to AD (e.g., APOE, ApoC-III, CD40, FRTN) are detected for all these scores. For example, FRTN (ferettin) is the main iron-storage protein capable of containing thousands of iron atoms. Recently it has been reported that ferritin from AD patients has a content of aluminum higher than that of controls [10]. And the irregular iron accumulative and disrupted iron metabolism have also been previously identified to be related with brain disorders. Detailed analysis of these identified proteomic markers warrants further investigation.

4. CONCLUSIONS

We proposed a new network-guided sparse learning framework, NG-L21, aiming to flexibly incorporate and model structure among predictors. Unlike traditional methods, this model could provide advantages in several folds: 1) explicitly incorporating the relationships among predictors in a more general way, 2) using data-driven patterns without any predefined parameters, 3) effectively identifying biomarkers influencing multiple responses, and 4) selection of correlated markers together rather than picking only one of them to improve the stability. With the application to the ADNI multimodal data (predicting memory scores from MRI and CSF proteomic measures), NG-L21 demonstrated improved prediction performance over the state-of-the-art competing methods, and identified stable and meaningful multimodal biomarkers. Combining MRI and CSF proteomic data yielded enhanced prediction performance than each single modality.

Acknowledgments

This research was supported by NIH R01 LM011360, U01 AG024904, RC2 AG036535, R01 AG19771, P30 AG10133, and NSF IIS-1117335 at IU, by NSF CCF-0830780, CCF-0917274, DMS-0915228, and IIS-1117965 at UTA, and by NIH R01 LM011360, R01 LM009012, and R01 LM010098 at Dartmouth.

References

1. Weiner MW, et al. The Alzheimers Disease Neuroimaging Initiative: A review of papers published since its inception. Alzheimer's dimentia. 2012:1–68.
2. Walhovd KB, et al. Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiology of Aging. 2010
3. Yan J, et al. Multimodal neuroimaging predictors for cognitive performance using structured sparse learning. MBIA'12, LNCS. 2012; 7509:1–17.
4. Obozinski, G., et al. Technical Report, Technical report. Statistics Department; UC Berkeley: 2006. Multi-task feature selection.
5. Wang H, et al. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics. 2012; 28(2):229–37. [PubMed: 22155867]
6. Kim S, et al. A multivariate regression approach to association analysis of a quantitative trait network. Bioinformatics. 2009; 25:i204–i212. [PubMed: 19477989]
7. Gaudes, CC., et al. Structured sparse deconvolution for paradigm free mapping of functional MRI data. 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI); 2012. p. 322-325.
8. Aisen PS, et al. Clinical core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. Alzheimers Dement. 2010; 6(3):239–46. [PubMed: 20451872]
9. Rebecca C, et al. Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimers disease diagnosis and prognosis. Plos one. 2011; 6(4):e18850. [PubMed: 21526197]
10. Solea PD, et al. Possible relationship between Al/ferritin complex and Alzheimer's disease. Clinical Biochemistry. 2013; 46(1–2):89–93. [PubMed: 23103708]

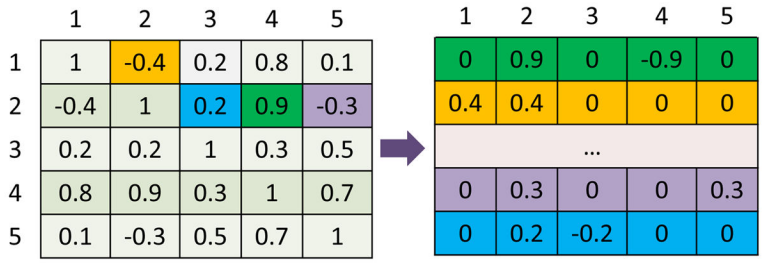

Fig. 1.

Each row in network matrix **N** (Right) corresponds to an element in correlation matrix **R** (Left).

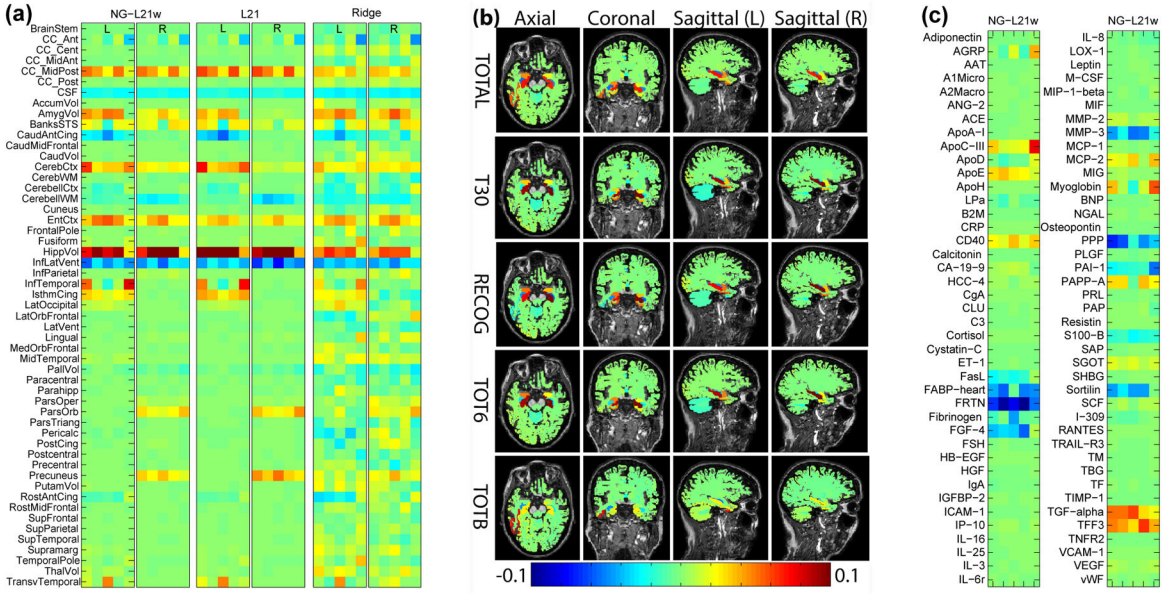

Fig. 2.

Heat maps of regression weights (average over 10-fold cross-validation) for predicting RAVLT scores using FS + CSF measures. (a) FS weights from NG-L21$_{\mathbf{w}}$, L21, and Ridge respectively. Results from left (L) and right (R) sides are shown in a pair of panels. For each panel, 5 columns show results for TOTAL, T30, RECOG, TOT6, and TOTB respectively. The measures shown in the first seven row are unilateral, and the remaining ones are bilateral. (b) FS weights from NG-L21$_{\mathbf{w}}$ mapped onto brain. (c) CSF weights from NG-L21$_{\mathbf{w}}$: 5 columns correspond to TOTAL, T30, RECOG, TOT6, and TOTB respectively.

Table 1

Participant characteristics (all from ADNI-1).

Category	AD	MCI	HC
Number	66	57	81
Gender(M/F)	37/29	36/21	42/39
Handness(R/L)	63/3	55/2	77/4
Age(mean± std)	75.14± 7.66	74.59± 7.42	76.24± 5.35
Education	15.04± 2.97	15.79± 2.88	15.86± 2.85

Table 2

RAVLT scores.

Score ID	Description
TOTAL	Total score of the first 5 learning trials
TOT6	Trial 6 total number of words recalled
TOTB	List B total number of words recalled
T30	30 minute delay number of words recalled
RECOG	30 minute delay recognition score

Table 3

Average correlation coefficient between predicted and actual scores over 10 cross-validation trials: FS results (top panel) and FS+CSF results (bottom panel) are shown.

	TOTAL	T30	RECOG	TOT6	TOTB
NG-L21$_w$	0.6084	0.5395	0.55	0.5337	0.4888
NG-L21$_t$	0.5879	0.5173	0.5497	0.5052	0.4775
L21	0.574	0.5007	0.5227	0.4915	0.4771
ElasticNet	0.6032	0.4971	0.5462	0.5061	0.4953
Ridge	0.5996	0.5365	0.5316	0.5335	0.4828
Linear	0.4015	0.3505	0.3778	0.3141	0.2041
NG-L21$_w$	0.6314	0.523	0.5885	0.5872	0.4991
NG-L21$_t$	0.6312	0.5223	0.5908	0.5883	0.4954
L21	0.6207	0.5178	0.575	0.5744	0.4858
ElasticNet	0.6293	0.5154	0.5691	0.5397	0.5002
Ridge	0.6428	0.5445	0.5595	0.5935	0.4936
Linear	0.1935	0.0826	0.0944	0.1632	0.0374