

Functioning

Cortical surface biomarkers for predicting cognitive outcomes using group $l_{2,1}$ norm

Jingwen Yan^{a,b}, Taiyong Li^c, Hua Wang^d, Heng Huang^e, Jing Wan^{a,f}, Kwangsik Nho^a, Sungeun Kim^a, Shannon L. Risacher^a, Andrew J. Saykin^a, Li Shen^{a,b,f,*},
for the Alzheimer's Disease Neuroimaging Initiative¹

^a Department of Radiology and Imaging Sciences, School of Medicine, Indiana University, Indianapolis, IN, USA

^b Department of Biohealth, School of Informatics and Computing, Indiana University, Indianapolis, IN, USA

^c School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China

^d Department of Electrical Engineering and Computer Science, Colorado School of Mines, Golden, CO, USA

^e Department of Computer Science and Engineering, University of Texas at Arlington, TX, USA

^f Department of Computer and Information Science, Purdue University, Indianapolis, IN, USA

ARTICLE INFO

Article history:

Received 7 May 2013

Received in revised form 28 June 2014

Accepted 2 July 2014

Available online 29 August 2014

Keywords:

Alzheimer's disease neuroimaging initiative (ADNI)

Magnetic resonance imaging (MRI)

Cortical thickness

Sparse learning

Cognitive function prediction

ABSTRACT

Regression models have been widely studied to investigate the prediction power of neuroimaging measures as biomarkers for inferring cognitive outcomes in the Alzheimer's disease study. Most of these models ignore the interrelated structures either within neuroimaging measures or between cognitive outcomes, and thus may have limited power to yield optimal solutions. To address this issue, we propose to use a new sparse multitask learning model called Group-Sparse Multi-task Regression and Feature Selection (G-SMuRFS) and demonstrate its effectiveness by examining the predictive power of detailed cortical thickness measures toward 3 types of cognitive scores in a large cohort. G-SMuRFS proposes a group-level $l_{2,1}$ -norm strategy to group relevant features together in an anatomically meaningful manner and use this prior knowledge to guide the learning process. This approach also takes into account the correlation among cognitive outcomes for building a more appropriate predictive model. Compared with traditional methods, G-SMuRFS not only demonstrates a superior performance but also identifies a small set of surface markers that are biologically meaningful.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

With the growing prevalence of Alzheimer's disease (AD) worldwide, it is of great importance to identify valid biomarkers which can help with early detection and monitoring of therapeutic responses. Despite the 2 well-known hallmarks of AD, beta-amyloid plaques and neurofibrillary tangles, various cognitive tests remain the most common clinical routine for diagnosis. Compared with the binary disease status, AD-relevant cognitive outcomes may provide

additional valuable information for studying the underlying disease mechanisms.

The release of the large-scale imaging and biomarker data of the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort has provided great opportunities for people to develop advanced computational methods for better understanding of the underlying neurodegenerative mechanism in relation to cognitive decline in AD. For example, regression analysis has become a widely used approach for the exploration of the relationship between imaging measures and cognitive outcomes. Using the ADNI data, many regression models have been used to investigate the relationships between multimodal imaging measures and cognitive scores (Wagner et al., 2005; Wan et al., 2012, 2014; Wang et al., 2011). Most of the existing models have used summary statistics (e.g., average intensity) of each region of interest (ROI) as input features. Although voxel-based image measures or vertex-based surface measures could provide more detailed morphometric information than ROI summary statistics, direct application of conventional regression models to these measures may be inadequate to yield biologically meaningful results. For example, standard linear or

* Corresponding author at: Department of Radiology and Imaging Sciences, IU Center for Neuroimaging, Indiana University School of Medicine, 355 W 16th Street, Suite 4100, Indianapolis, IN 46202, USA. Tel.: +1 317 963 7504; fax: +1 317 963 7547.

E-mail address: shenli@iu.edu (L. Shen).

¹ 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 the 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.

ridge regression model typically produces nonsparse results that are not ideal for biomarker discovery. Conventional sparse models such as Lasso (Tibshirani, 1996) are likely to yield scattered patterns hard to interpret, because of the lacking of proper handling of the spatial correlation and prior anatomic knowledge in these models. To address this issue, we propose to use a new sparse multitask learning model called Group-Sparse Multi-task Regression and Feature Selection (G-SMuRFS) (Wang et al., 2012) for identifying effective surface biomarkers that can predict cognitive outcomes. We demonstrate its effectiveness by examining the predictive power of detailed cortical thickness measures toward 3 types of cognitive scores (Alzheimer's Disease Assessment Scale [ADAS], Mini-Mental State Examination [MMSE], and Rey Auditory Verbal Learning Test [RAVLT]) in the ADNI cohort.

Enormous efforts have been made to evaluate the power of sparse learning methods in the neuroimaging field, such as identifying structural (Avants et al., 2010; Batmanghelich et al., 2012; Sabuncu and Van Leemput, 2012; Wan et al., 2012, 2014; Wang et al., 2011) or functional (Grosenick et al., 2013; Jenatton et al., 2012; Michel et al., 2011; Varoquaux et al., 2012) imaging biomarkers associated with other imaging modality (Avants et al., 2010), cognitive scores (Jenatton et al., 2012; Varoquaux et al., 2012; Wan et al., 2012, 2014; Wang et al., 2011), behavior (Grosenick et al., 2013; Michel et al., 2011), as well as diagnostic conditions (Batmanghelich et al., 2012; Sabuncu and Van Leemput, 2012). However, using detailed cortical surface measures to predict cognitive outcomes is an under-explored area. In this study, we attempt to explore a novel application of G-SMuRFS to the identification of detailed surface-based cortical biomarkers that are relevant to cognitive outcomes. G-SMuRFS proposes a group-level $l_{2,1}$ -norm strategy to achieve 3 goals: (1) group relevant surface features together in an anatomically meaningful manner (i.e., ROI information is incorporated) and use this prior knowledge to guide the learning process (i.e., spatial correlation within each ROI is addressed); (2) take into account the correlation among cognitive outcomes for building a more appropriate predictive model (i.e., multiple correlated cognitive scores are predicted together); and (3) optimize the selection of cognition-relevant surface biomarkers while maintaining high prediction accuracy. The high dimensionality of the vertex-based cortical surface data (e.g., 327,684 vertices in our study) introduces major computational challenges. To address this issue, we introduce a down-sampling technique to merge neighboring vertices into small surface patches to reduce the computational cost while preserving detailed surface information. Our overarching goal is to examine and validate the predictive power of these detailed cortical thickness measures toward cognitive outcomes while considering the group structures defined by anatomically meaningful ROIs. The results may provide important information about potential surrogate biomarkers for early detection and/or therapeutic trials in AD.

2. Methods

2.1. Neuroimaging and cognition data

All the data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu) (Weiner et al., 2010). One goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography, 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, we refer interested readers to www.adni-info.org/.

We downloaded the baseline 1.5 T MRI scans, demographic information, and baseline diagnosis for all the ADNI-phase 1 (ADNI-1)

participants. We also downloaded 3 types of baseline cognitive scores: ADAS, MMSE, and RAVLT. For each participant, FreeSurfer V4, an automatic brain segmentation and cortical parcellation tool, was applied to automatically label cortical and subcortical tissue classes (Dale et al., 1999; Fischl et al., 1999) and to extract surface-based thickness measures. We focused our study on examining the thickness measures from surface locations labeled with any of the 34 FreeSurfer cortical ROIs (shown in Table 1) in both hemispheres. The measures from surface locations labeled with “unknown” were excluded in this study.

Following a previous imaging genetics study (Shen et al., 2010), in this work, we concentrated our analyses on the Caucasian subjects determined by population stratification analysis using the ADNI genetics data (Saykin et al., 2010). A total of 718 of 745 Caucasian participants with no missing MRI morphometric and the cognitive outcome information were included in the study. The 718 participants were categorized by 3 baseline diagnostic groups: healthy control (HC, $n = 197$), MCI ($n = 349$), and AD ($n = 172$). Demographics information of these subjects can be found in Table 2. All the imaging and cognitive outcome measurements were adjusted for age, gender, education, and handedness, although intracranial volume was applied as an extra covariate for imaging measurements.

2.2. Group-sparse multitask regression and feature selection

Throughout this section, 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 $l_{2,1}$ -norm (also called as $l_{2,1}$ -

Table 1

Thickness measures at surface locations from the following 34 pairs of bilateral FreeSurfer cortical regions (68 ROIs in total) were analyzed in this study

ID	ROI name
1	Banks of the superior temporal sulcus
2	Caudal anterior cingulate
3	Caudal middle frontal gyri
4	Corpus callosum
5	Cuneus
6	Entorhinal cortex
7	Fusiform gyri
8	Inferior parietal gyri
9	Inferior temporal gyri
10	Isthmus cingulate
11	Lateral occipital gyri
12	Lateral orbitofrontal gyri
13	Lingual gyri
14	Medial orbitofrontal
15	Middle temporal gyri
16	Parahippocampal gyri
17	Paracentral lobule
18	Pars opercularis
19	Pars orbitalis
20	Pars triangularis
21	Pericalcarine gyri
22	Postcentral gyri
23	Posterior cingulate
24	Precentral gyri
25	Precuneus
26	Rostral anterior cingulate
27	Rostral middle frontal gyri
28	Superior frontal gyri
29	Superior parietal gyri
30	Superior temporal gyri
31	Supramarginal gyri
32	Frontal pole
33	Temporal pole
34	Transverse temporal pole

Key: ROIs, regions of interest.

Table 2
Participant characteristics

Category	HC	MCI	AD
Number of subjects	197	349	172
Gender (M/F)	107/90	224/125	94/78
Handedness (R/L)	183/14	316/33	160/12
Baseline age (y, mean ± SD)	76.2 ± 5.0	75 ± 7.3	75.6 ± 7.5
Education (y, mean ± SD)	16.2 ± 2.7	15.7 ± 3	14.9 ± 3.1

Key: AD, Alzheimer's disease; F, female; HC, healthy control; L, left; M, male; MCI, mild cognitive impairment; R, right; SD, standard deviation.

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.

This study is centered on the multitask learning paradigm, where multimodal imaging measures are used to predict one or more cognitive outcomes. Let $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subseteq \mathbb{R}^d$ be imaging measures and $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\} \subseteq \mathbb{R}^c$ 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 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$.

To investigate the correlation between imaging measures and cognitive outcomes, linear and ridge regression models are 2 standard methods. Whereas linear regression (least square) may yield unstable results for correlated predictors, ridge regression have one more regularization term, the Frobenius norm of trained weights, to successfully solve the problem and ascertain the numerical stability simultaneously (Equation (1)).

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

where, the entry w_{ij} of weight matrix \mathbf{W} measures the relative importance of the i -th predictor in predicting the j -th response, and $\gamma > 0$ is a tradeoff parameter.

Regression weights brought by Frobenius norm are typically nonsparse, which makes the results hard to interpret and unsuitable for biomarker discovery. To produce sparse solutions, the following traditional Lasso model (Tibshirani, 1996) can be used:

$$\min_{\mathbf{W}} \|\mathbf{W}^T \mathbf{X} - \mathbf{Y}\|_F^2 + \gamma \|\mathbf{W}\|_1$$

However, this multitask Lasso model is equivalent to applying Lasso to each outcome variable independently and ignores the correlation among the outcome variables. As a result, although the outcome variables are correlated, the features selected by the previously mentioned Lasso model could be relevant to some outcomes (i.e., regression weights $\neq 0$) but not to others (i.e., regression weights = 0).

$l_{2,1}$ -norm (Equation (2)), motivated by ridge and Lasso, is proposed as follows.

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

where $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^d \|\mathbf{w}^i\|_2$ (see also Fig. 1). $l_{2,1}$ -norm is one of the advanced techniques that addresses both the outcome correlation and sparsity issues, by enforcing an l_2 -norm across the tasks and an l_1 -norm across the features. Although the l_2 -norm ascertains the similarity pattern across the tasks, the l_1 -norm ascertains the sparsity across the features.

In all the previously mentioned methods, imaging features were all treated separately, where the underlying brain structures were not taken into account. In many cases, different brain structures may be responsible for different brain functions. Therefore, it would be

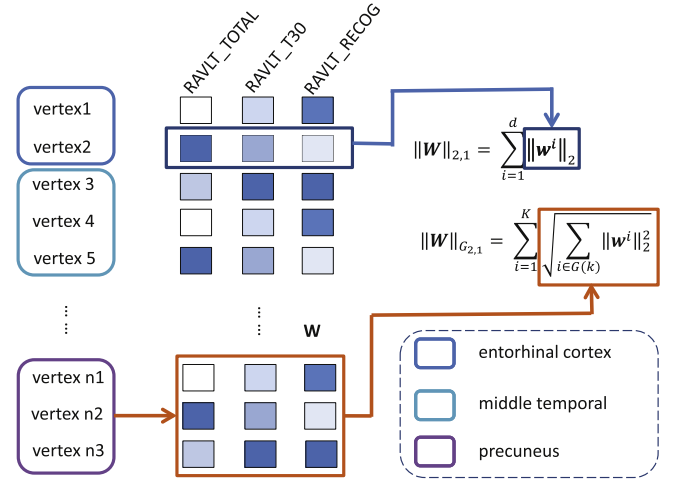


Fig. 1. Illustration of the G-SMuRFS method. Two regularization terms, group $l_{2,1}$ -norm ($\|\mathbf{W}\|_{2,1}$) and $l_{2,1}$ -norm ($\|\mathbf{W}\|_{2,1}$), are integrated to group surface vertices by ROIs and to jointly select prominent vertices across all cognitive scores. Abbreviations: G-SMuRFS, group-sparse multitask regression and feature selection; ROIs, regions of interest.

much more meaningful to include the structural information in the regression procedure. G-SMuRFS (Wang et al., 2012), a newly proposed regression model derived from the $l_{2,1}$ -norm, takes into account the group information in the regression procedure and has yielded promising results in a previous imaging genetics study (Wang et al., 2012). In this work, we apply this algorithm to group all the vertices within each ROI together and incorporate the anatomic boundary information into the regression procedure. As illustrated in Fig. 1, this method can be applied to address this issue by grouping cortical vertices using ROI boundary information (i.e., group $l_{2,1}$ -norm, or $G_{2,1}$ -norm), where cortical measures from the same ROI tend to be selected together as joint predictors and yield an anatomically meaningful biomarker discovery result. On the other hand, the $l_{2,1}$ -norm can help select imaging features that can predict all or most of the cognitive outcomes. As a result, the learned regression model and the selected cortical biomarkers should be more biologically meaningful and more informative.

Mainly motivated by sparse learning, such as Lasso (Tibshirani, 1996) and group Lasso (Yuan and Lin, 2006), the new regularization term was applied in G-SMuRFS to consider both the group sparsity through the $G_{2,1}$ -norm and the individual biomarker sparsity for joint learning via an $l_{2,1}$ -norm regularization (Puniani et al., 2010). In the objective function Equation (3), whereas the second term couples all the regression coefficients of a group of features across all the c tasks together, the third term penalizes all c regression coefficient of each individual feature as whole to select features across multiple learning tasks.

$$\min_{\mathbf{W}} \|\mathbf{W}^T \mathbf{X} - \mathbf{Y}\|_F^2 + \gamma_1 \|\mathbf{W}\|_{G_{2,1}} + \gamma_2 \|\mathbf{W}\|_{2,1} \quad (3)$$

where $\|\mathbf{W}\|_{G_{2,1}} = \sum_{i=1}^k \sqrt{\sum_{j \in G(k)} \|\mathbf{w}^j\|_2^2}$ is the $G_{2,1}$ -norm, and

$\|\mathbf{W}\|_{2,1} = \sum_{i=1}^d \|\mathbf{w}^i\|_2$ is the $l_{2,1}$ -norm (see also Fig. 1). The solution of the objective function (Equation (3)) can be obtained through an iterative optimization procedure. By setting the derivative with respect to \mathbf{W} to zero, \mathbf{W} can be solved as in (Equation (4)).

$$\mathbf{W} = (\mathbf{X}\mathbf{X}^T + \gamma_1 \mathbf{D}_1 + \gamma_2 \mathbf{D}_2)^{-1} \mathbf{X}\mathbf{Y}^T \quad (4)$$

where \mathbf{D}_1 is a block diagonal matrix with the k -th diagonal block as $\frac{1}{2\|\mathbf{W}^k\|_F} \mathbf{I}_k$, \mathbf{I}_k is an identity matrix with size of m_k , m_k is the total

Table 3
Cross-validation performance comparison of linear regression, ridge regression, PLS, $l_{2,1}$ -norm, and G-SMuRFS; performance is measured by the root mean squared error (RMSE) and correlation coefficient (CC) between the actual and predicted scores of the test subjects. The average (avg) and standard deviation (std) of performance measures across 5 cross-validation trials are shown as “avg \pm std” for each experiment

	ADAS	MMSE	RAVLT		
			TOTAL	T30	RECOG
Performance comparison using RMSE					
G-SMuRFS	0.7663 \pm 0.0375	0.8325 \pm 0.0399	0.8490 \pm 0.0654	0.8776 \pm 0.0701	0.9167 \pm 0.0471
$l_{2,1}$ -norm	0.7631 \pm 0.0346	0.8322 \pm 0.0411	0.8464 \pm 0.0697	0.8807 \pm 0.0657	0.9215 \pm 0.0440
PLS	0.8844 \pm 0.0425	0.8999 \pm 0.0453	0.9246 \pm 0.0533	0.9515 \pm 0.0578	0.9706 \pm 0.0512
Ridge	0.7859 \pm 0.0327	0.8579 \pm 0.0332	0.8453 \pm 0.0697	0.8977 \pm 0.0553	0.9315 \pm 0.0430
Linear	1.0524 \pm 0.0559	1.1342 \pm 0.0739	1.1726 \pm 0.0956	1.3042 \pm 0.0803	1.2775 \pm 0.1168
Performance comparison using CC					
G-SMuRFS	0.6438 \pm 0.0258	0.5552 \pm 0.0078	0.5277 \pm 0.0539	0.4753 \pm 0.0591	0.3985 \pm 0.0533
$l_{2,1}$ -norm	0.6468 \pm 0.0175	0.5548 \pm 0.0058	0.5301 \pm 0.0544	0.4692 \pm 0.0522	0.3889 \pm 0.0588
Partial least square	0.4608 \pm 0.0551	0.4339 \pm 0.0406	0.3782 \pm 0.0591	0.3037 \pm 0.0565	0.2390 \pm 0.0382
Ridge	0.6167 \pm 0.0171	0.5127 \pm 0.0272	0.5296 \pm 0.0540	0.4364 \pm 0.0416	0.3632 \pm 0.0681
Linear	0.4902 \pm 0.0544	0.3779 \pm 0.0246	0.3685 \pm 0.1005	0.2958 \pm 0.1359	0.2533 \pm 0.0726

Key: G-SMuRFS, group-sparse multitask regression and feature selection; PLS, partial least square; RAVLT, Rey Auditory Verbal Learning Test; RECOG, RAVLT recognition score; T30, RAVLT 30 minutes delay score; TOTAL, RAVLT total score.

feature numbers included in group k , D_2 is a diagonal matrix with the i -th diagonal element as $\frac{1}{2\|W^i\|_2}$. Detailed optimization procedure and algorithm can be found in Wang et al. (2012).

Generally, the advantage of this model is 3-fold, namely: (1) it addresses the highly correlated nature of the cortical vertices within each surface ROI; (2) takes into account the correlation of multiple scores of the same cognitive function test; and (3) achieves both the global biomarker sparsity as well as the ROI group sparsity.

3. Experimental results and discussion

3.1. Experimental setting

In this study, we examined all the cortical thickness measures across 34 pairs of bilateral cortical surface ROIs (68 ROIs in total) (Table 1) regarding their power for predicting the ADAS, MMSE, and RAVLT cognitive scores. Our cortical surface data, generated by FreeSurfer, contains 327,684 vertices per surface. For the efficiency purpose, we completed a preprocessing step to downsample 327,684 vertex-based thickness measures to 3133 surface-patch-based measures using the following approach. First, we randomly selected cortical surface data from 50 HC participants. Second, for each ROI (say, with m vertices), we performed the k -mean clustering using this preselected HC subset to partition the ROI into roughly $m/100$ surface patches, where each patch was formed by a set of neighboring vertices with similar thickness. As a result, 3133 patches were defined on the cortical surface. Third, excluding 320 patches from the region labeled as “unknown,” we got 2813 patches from the ROIs shown in Table 1. Finally, we applied this patch scheme to the entire data set. The cortical thickness measures of all vertices within one patch were averaged to represent the patch-level thickness measure. These 2813 patch-level measures were used as predictors in our regression analysis.

The response variables in the multivariate multiple regression analysis included the following 5 cognitive scores: ADAS-cog total score (ADAS), MMSE score (MMSE), RAVLT total score (TOTAL), RAVLT 30 minutes delay score (T30), and RAVLT recognition score (RECOG). To provide an unbiased estimate of the prediction performance of each method tested in the experiments, we used a 5-fold cross-validation, where each fold contained a similar portion of AD, MCI, and HC participants. We calculated the following 2 metrics to compare the prediction performance across different methods: (1) root mean square error (RMSE) between the actual and predicted scores of all the test subjects; and (2) Pearson correlation

coefficient (CC) between the actual and predicted scores of all the test subjects.

In our experiments, we compared G-SMuRFS with 4 competing multivariate regression methods: (1) $l_{2,1}$ -norm; (2) partial least square (PLS); (3) ridge; and (4) linear regression. Parameters for these models were optimally tuned using a nested cross-validation strategy on the training data, with search grid in the range of 5×10^{-3} to 5×10^3 . For these regression analyses, the input data included 2813 surface patch-level thickness measures as predictors and cognitive scores as response variables. We also performed univariate surface-based analysis using SurfStat (Chung et al., 2010) to cross check whether univariate and multivariate methods could yield similar patterns.

3.2. Results and discussion

Prediction performance, measured by RMSE and CC, of the cortical thickness measurement under 5 different regression models is shown in Table 3, where the average (avg) and standard deviation (std) of performance measures across 5 cross-validation trials are shown as “avg \pm std” for each experiment. The prediction performances using those features selected by G-SMuRFS and $l_{2,1}$ -norm are higher (i.e., lower RMSE and higher CC) than those of linear, ridge, and PLS regression models. In particular, G-SMuRFS demonstrates clear performance improvement over PLS and linear regression on predicting all 5 scores and over ridge regression on predicting MMSE and RAVLT-RECOG. Prediction performances of G-SMuRFS and $l_{2,1}$ -norm are similar. Fig. 2 shows scatter plots of actual and predicted (by G-SMuRFS) cognitive scores.

Fig. 3 shows the histogram of regression weights associated with all the cortical measures for each method, in an example cross-validation trial, and the cortical maps of these regression weights are shown in Fig. 4A–E. Note that all the PLS weights are very small, and thus, a different color scale is used. From these plots, we observe the following: (1) PLS, ridge, and linear regression models yielded nonsparse results where most surface measures shared relatively similar impact on the prediction performance; and (2) G-SMuRFS and $l_{2,1}$ -norm presented a much better sparsity across all the cortical measures, where a small portion of the cortical surface was identified to be relevant to the outcome.

Besides the sparsity at the cortical patch level, we also examined the group sparsity of all 5 models at the ROI level. In Fig. 5, ROI level sparsity is demonstrated through the histogram of “high impact” (i.e., top 50) cortical markers against each of the 34 pairs of bilateral ROIs for: (1) G-SMuRFS; (2) $l_{2,1}$ -norm; (3) PLS; (4) ridge regression;

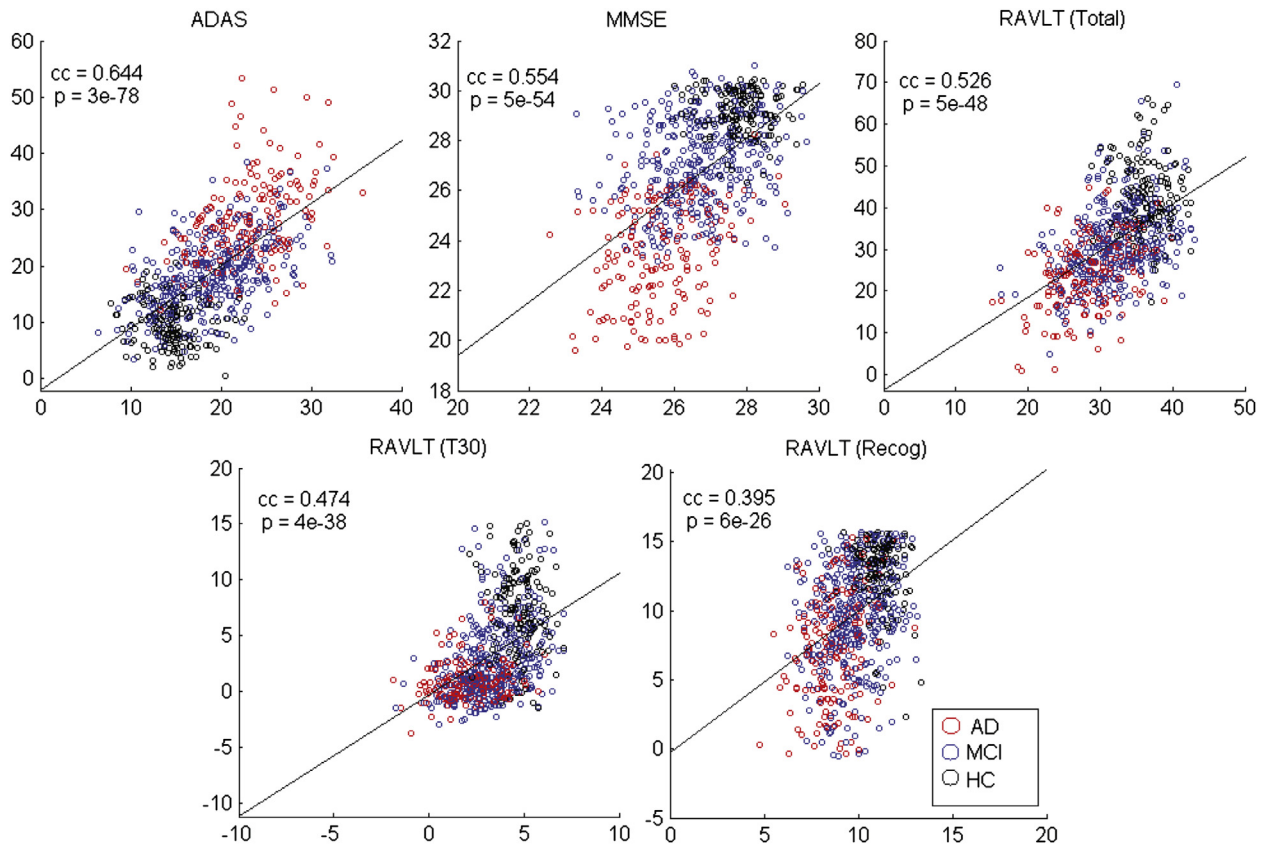


Fig. 2. Scatter plots of actual (on y-axis) and predicted (by G-SMuRFS, on x-axis) cognitive scores. Note that the actual cognitive scores are pre-adjusted and thus may have negative values. The testing samples across 5 cross-validation trials were pulled together to calculate the correlation coefficients (CC) and the p-values. Thus, the CCs shown here are slightly different from the CCs shown in Table 3 that were calculated separately for each cross-validation trial. Abbreviations: ADAS, Alzheimer’s Disease Assessment Scale; MMSE, Mini-Mental State Examination; RAVLT, Rey Auditory Verbal Learning Test.

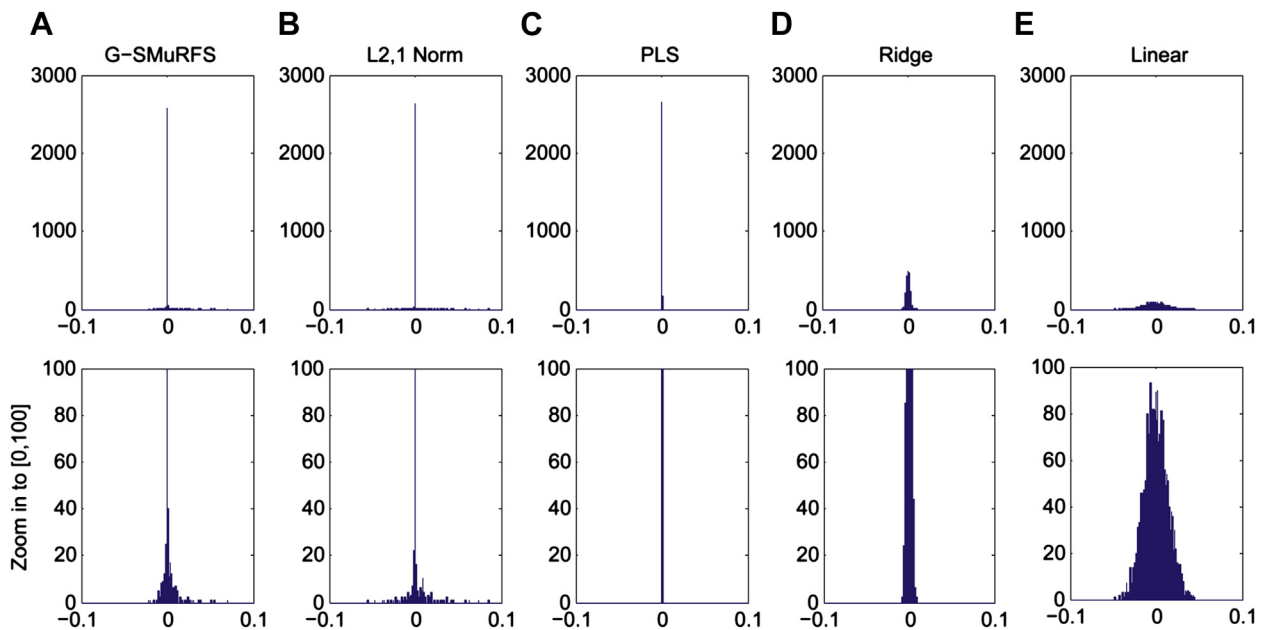


Fig. 3. Histogram of regression weights of all cortical measures for predicting the RAVLT total score in an example cross-validation trial. Shown from left to right are the results of (A) G-SMuRFS, (B) L2,1-norm, (C) PLS, (D) ridge regression, and (E) linear regression. The top row shows the complete histograms, and the bottom row shows the zoom in view of the partial histograms for $y \in [0, 100]$. Abbreviations: G-SMuRFS, group-sparse multitask regression and feature selection; RAVLT, Rey Auditory Verbal Learning Test.

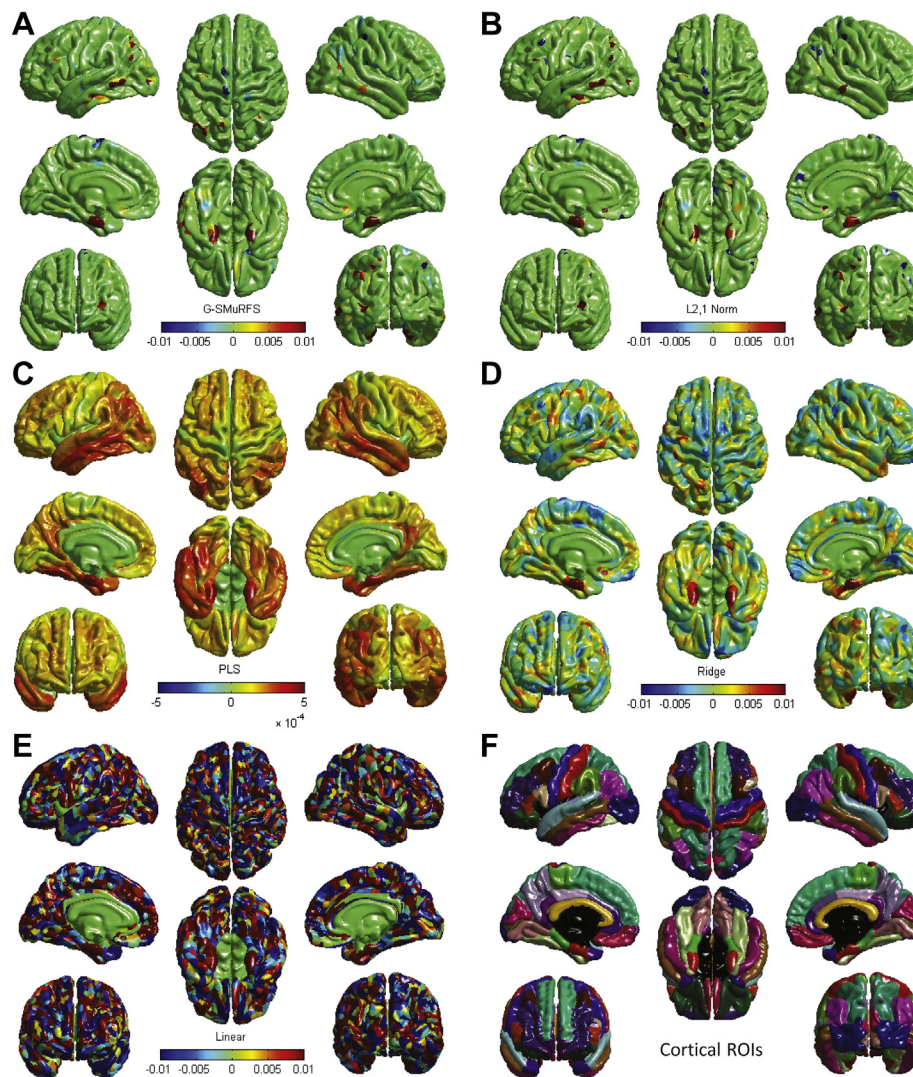


Fig. 4. Cortical map of regression weights for predicting the RAVLT total score in an example cross-validation trial using 5 different models: (A) G-SMuRFS, (B) $l_{2,1}$ -norm, (C) PLS, (D) ridge regression, and (E) linear regression. The red color indicates regions where the thickness is positively correlated with the RAVLT total score, and the blue color indicates regions where the thickness is negatively correlated with the score. Shown in (F) are 34 pairs of color-coded bilateral cortical ROIs. Abbreviations: G-SMuRFS, group-sparse multitask regression and feature selection; PLS, partial least square; RAVLT, Rey Auditory Verbal Learning Test; ROIs, regions of interest. (For interpretation of the references to color in this Figure, the reader is referred to the web version of this article.)

and (5) linear regression, respectively. Each vertical bar here indicates an ROI in the brain. Although the top 50 biomarkers identified by G-SMuRFS are associated with a small number of ROIs, the same number of high impact biomarkers identified through $l_{2,1}$ -norm, ridge regression, and linear regression are scattered across a large portion of cortical surface regions, making the result hard to interpret. G-SMuRFS yielded sparse patterns at the ROI level that have the potential for identifying relevant biomarkers. Although PLS also yielded sparse patterns, the predictive power of its top 50 markers (RMSE = 1.045, CC = 0.377) is lower than that of G-SMuRFS's (RMSE = 0.938, CC = 0.46).

Fig. 6 shows example of G-SMuRFS regression weights that were averaged over the 5 cross-validation trials and were then mapped back onto the cortical surface. Our multitask regression experiment was performed to identify thickness measures for jointly predicting ADAS, MMSE, RAVLT TOTAL, RAVLT RECOG, and RAVLT T30 scores. The weight maps for ADAS (Fig. 6A), MMSE (Fig. 6B), TOTAL (Fig. 6C), RECOG (Fig. 6D), and T30 (not shown) are very similar to one another except that the ADAS pattern is in the opposition direction.

Thickness measures from left and right entorhinal cortex, left middle temporal gyri, left inferior parietal gyri, right medial orbitofrontal gyri, and right precunes are positively correlated to the MMSE and RAVLT scores, and negatively correlated to ADAS. The measures from left fusiform are correlated to ADAS, MMSE, TOTAL, and T30, and the measures from right middle temporal gyri are correlated to ADAS, MMSE, and RECOG. These patterns identified by our multivariate G-SMuRFS regression analysis match well with the weight map patterns computed by the univariate SurfStat analysis shown in Fig. 7.

The ROIs identified in this work are either related to AD or in accordance with findings in similar prior studies. For example, entorhinal cortex (part of medial temporal cortex) and precuneus are among the cortical signature of AD studied in (Bakkour et al., 2009; Dickerson et al., 2009). Wan et al. (2012), (2014) and Wang et al. (2011) performed similar regression studies for predicting cognitive outcomes using MRI measures. However, they examined only summary statistics (volume, thickness, or gray matter density) of both cortical and subcortical ROIs instead of detailed cortical thickness measures. The mean thickness of entorhinal cortex was found to be

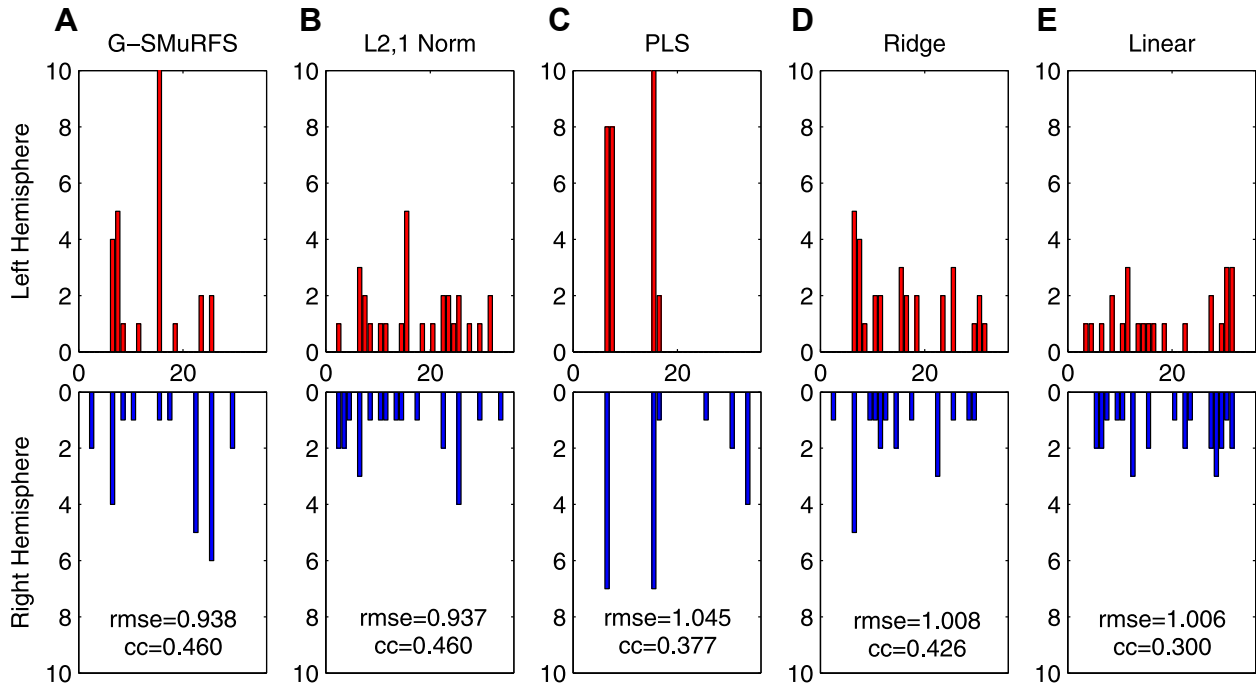


Fig. 5. Number of “high impact” (i.e., top 50) cortical markers for predicting the RAVLT total score, in an example cross-validation trial, is plotted against the corresponding ROI (34 ROIs in total). The x-axis shows the ROI IDs (see Table 1 for the corresponding ROI names). The y-axis shows the number of top markers in the left hemisphere ROI (top row) or the right hemisphere ROI (bottom row). Shown from left to right are the results of (A) G-SMuRFS, (B) $l_{2,1}$ -norm, (C) PLS, (D) ridge regression, and (E) linear regression. The cross-validation performance using these top 50 markers measured by root mean square error (RMSE) and correlation coefficient (CC) between the actual and predicted RAVLT total scores of all the test subjects is shown in each panel. Abbreviations: G-SMuRFS, group-sparse multitask regression and feature selection; PLS, partial least square; RAVLT, Rey Auditory Verbal Learning Test; ROIs, regions of interest.

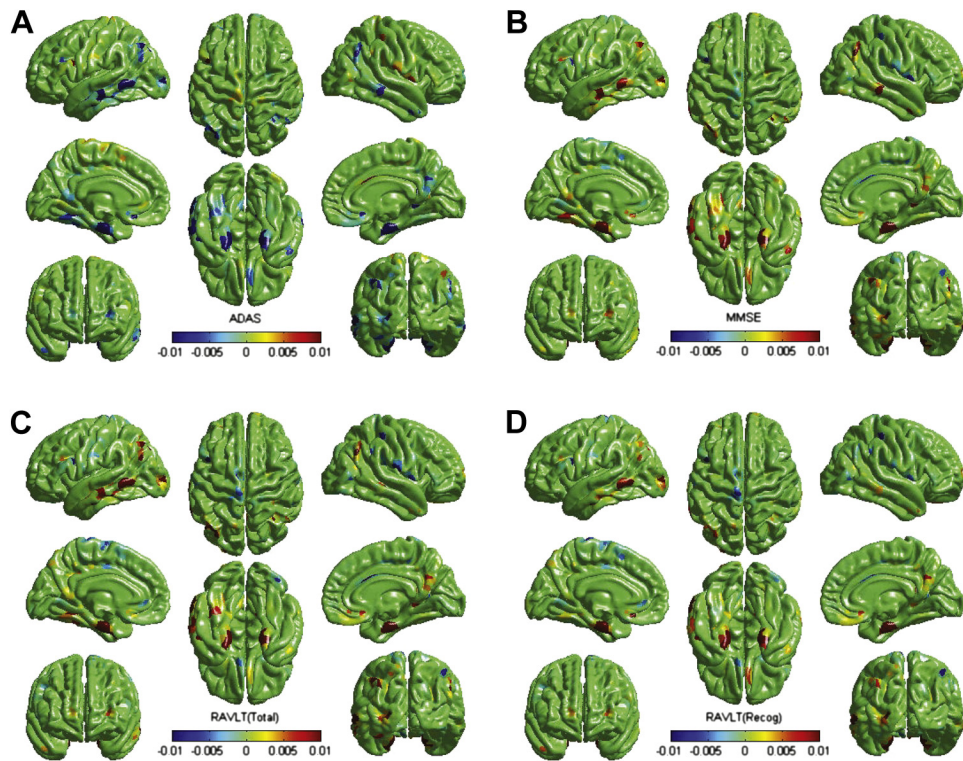


Fig. 6. Example G-SMuRFS regression weights are color coded and mapped onto the cortical surface. The red color indicates regions where the thickness is positively correlated with the corresponding cognitive score ((A) ADAS, (B) MMSE, (C) RAVLT-TOTAL, or (D) RAVLT-RECOG), and the blue color indicates regions where the thickness is negatively correlated with the score. Abbreviations: ADAS, Alzheimer’s Disease Assessment Scale; MMSE, Mini-Mental State Examination; RAVLT, Rey Auditory Verbal Learning Test. (For interpretation of the references to color in this Figure, the reader is referred to the web version of this article.)

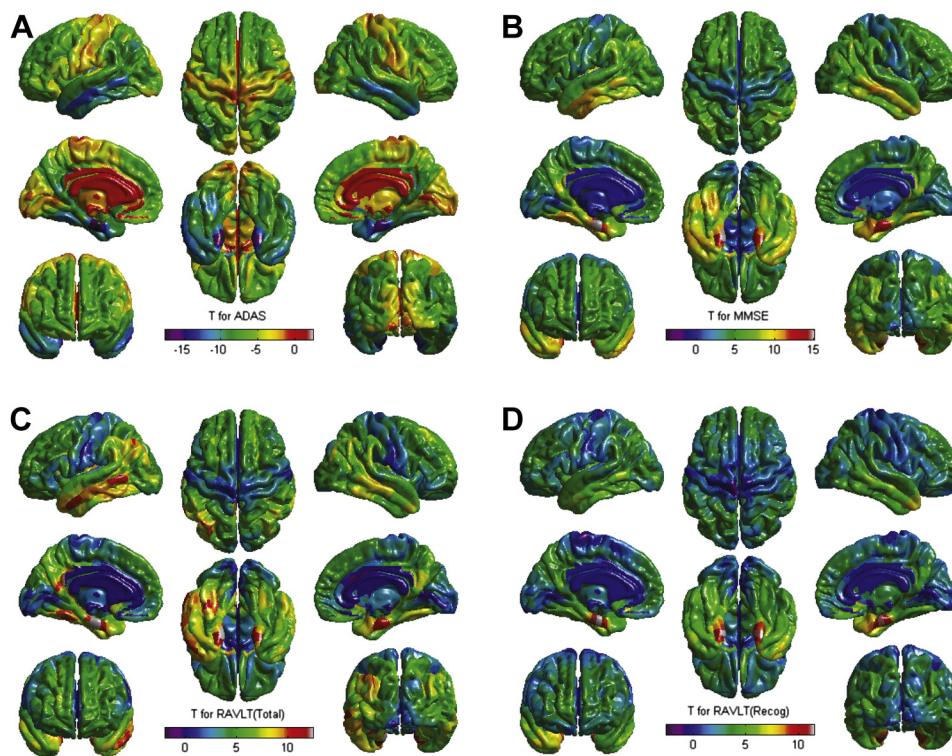


Fig. 7. Example SurfStat t-statistic map: t-statistics for ADAS (A), MMSE (B), RAVLT Total (C) and RAVLT Recog (D) are color coded and mapped onto the cortical surface. Abbreviations: ADAS, Alzheimer's Disease Assessment Scale; MMSE, Mini-Mental State Examination; RAVLT, Rey Auditory Verbal Learning Test.

correlated with ADAS (Wan et al., 2014), MMSE (Wan et al., 2014), and RAVLT (Wan et al., 2014; Wang et al., 2011) scores. The mean thickness of inferior parietal gyri was found to be correlated with ADAS (Wan et al., 2014) and RAVLT (Wang et al., 2011) scores. The mean thickness of middle temporal gyri was found to be correlated with RAVLT scores (Wan et al., 2012). Partly because of the detailed cortical analysis, this work identified some additional ROIs associated with the studied cognitive scores. Replication of these results in independent samples will remain of critical importance for confirmation.

The computational cost of G-SMuRFS was similar to that of the $l_{2,1}$ -norm model but more expensive than linear, ridge, and PLS regressions. We implemented all the regression models using Matlab. For 1 cross-validation trial in our experiments, G-SMuRFS and $l_{2,1}$ -norm took 75–77 seconds, whereas linear regression took 48 seconds, and ridge and PLS took <2 seconds. One interesting future direction is to develop more efficient implementation of G-SMuRFS and make it applicable to the analysis of larger scale data sets.

To sum up, our empirical results are very encouraging and have demonstrated the promise of the G-SMuRFS method in the application of relating cortical morphology to cognitive outcomes: (1) G-SMuRFS regression model outperformed linear, ridge, and PLS regression models and performed similarly to the multitask $l_{2,1}$ -norm model in terms of overall RMSE and CC (Table 3); (2) the biomarkers identified by the G-SMuRFS method were sparser at the patch level than linear regression and ridge regression, and yielded a more stable performance for predicting cognitive scores; and (3) both G-SMuRFS and $l_{2,1}$ methods yielded sparse results at the vertex level (Fig. 3 and Fig. 4); however, the G-SMuRFS model presented a sparser pattern at the ROI level (Fig. 5) than the $l_{2,1}$ -norm model. Taking into account, the spatial information makes the best use of the detailed surface information yet leading to a clustered group level result instead, which is more visible and interpretable.

4. Conclusions

We have investigated the power of detailed cortical thickness measurements for predicting ADAS, MMSE, and RAVLT cognitive scores using the data from the ADNI cohort. We have proposed to use a newly developed sparse multitask learning algorithm called G-SMuRFS and have observed the following strengths of this approach that could greatly improve the prediction performance: (1) seamless integration of anatomic knowledge in the learning process by coupling cortical measures from the same ROI together; (2) sparsity at both patch level and ROI level; and (3) multitask learning scheme for addressing correlation among response variables.

Compared with linear, ridge, PLS, or $l_{2,1}$ -norm regression, combining the group $l_{2,1}$ -norm in the regularization term has not only helped select the potential biomarkers in a few ROIs but also improved overall predictive power. Its application to multimodal imaging data would be promising future directions for biomarker discovery and better mechanistic understanding in AD research. Exploration of other imaging modalities as well as the combination of multiple modalities warrants further investigation. Further effort may be made to include more complicated prior structure, like multiple layer groups or networks, to guide the learning procedure. Another possible future topic could be to investigate whether nonlinear models can help improve the prediction rates as well as derive biologically meaningful results.

Disclosure statement

All the authors declare no conflicts of interest.

Acknowledgements

This research was supported by National Science Foundation Division of Information and Intelligent Systems 1117335, National

Institutes of Health R01 LM011360, K99/R00 LM011384, UL1 RR025761, and U01 AG024904, National Institute on Aging RC2 AG036535, R01 AG19771, and P30 AG10133 at IU; and by National Science Foundation Division of Information and Intelligent Systems 1117965, 1302675, 1344152, Division of Biological Infrastructure 1356628 at UTA.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health grant U01 AG024904) and U.S. Department of Defense ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer Drug Discovery Foundation; BioClinica, Inc; Biogen; Bristol-Myers Squibb Foundation; Eisai; Elan; Eli Lilly and Company; F. Hoffmann-La Roche Ltd, and its affiliated company Genentech Foundation; 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; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer; Piramal Imaging; Servier; Synarc Inc; and Takeda Pharmaceutical North America. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (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 the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

- Avants, B.B., Cook, P.A., Ungar, L., Gee, J.C., Grossman, M., 2010. Dementia induces correlated reductions in white matter integrity and cortical thickness: a multivariate neuroimaging study with sparse canonical correlation analysis. *Neuroimage* 50, 1004–1016.
- Bakkour, A., Morris, J.C., Dickerson, B.C., 2009. The cortical signature of prodromal AD: regional thinning predicts mild AD dementia. *Neurology* 72, 1048–1055.
- Batmanghelich, N.K., Taskar, B., Davatzikos, C., 2012. Generative-discriminative basis learning for medical imaging. *IEEE Trans. Med. Imaging* 31, 51–69.
- Chung, M.K., Worsley, K.J., Nacewicz, B.M., Dalton, K.M., Davidson, R.J., 2010. General multivariate linear modeling of surface shapes using SurfStat. *Neuroimage* 53, 491–505.
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L., 2009. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb. Cortex* 19, 497–510.
- Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis - II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
- Grosenick, L., Klingenberg, B., Katovich, K., Knutson, B., Taylor, J.E., 2013. Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage* 72, 304–321.
- Jenatton, R., Gramfort, A., Michel, V., Obozinski, G., Eger, E., Bach, F., Thirion, B., 2012. Multiscale mining of fMRI data with hierarchical structured sparsity. *SIAM J. Imaging Sci.* 5, 835–856.
- Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Thirion, B., 2011. Total variation regularization for fMRI-based prediction of behavior. *IEEE Trans. Med. Imaging* 30, 1328–1340.
- Puniyani, K., Kim, S., Xing, E.P., 2010. Multi-population GWA mapping via multi-task regularized regression. *Bioinformatics* 26, i208–i216.
- Sabuncu, M.R., Van Leemput, K., Alzheimer's Disease Neuroimaging Initiative, 2012. The relevance voxel machine (RVoxM): a self-tuning Bayesian model for informative image-based prediction. *IEEE Trans. Med. Imaging* 31, 2290–2306.
- Saykin, A.J., Shen, L., Foroud, T.M., Potkin, S.G., Swaminathan, S., Kim, S., Risacher, S.L., Nho, K., Huentelman, M.J., Craig, D.W., Thompson, P.M., Stein, J.L., Moore, J.H., Farrer, L.A., Green, R.C., Bertram, L., Jack Jr., C.R., Weiner, M.W., Alzheimer's Disease Neuroimaging Initiative, 2010. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers Dement.* 6, 265–273.
- Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.D., Foroud, T., Pankratz, N., Moore, J.H., Sloan, C.D., Huentelman, M.J., Craig, D.W., DeChairo, B.M., Potkin, S.G., Jack Jr., C.R., Weiner, M.W., Saykin, A.J., Alzheimer's Disease Neuroimaging Initiative, 2010. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *Neuroimage* 53, 1051–1063.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288.
- Varoquaux, G., Gramfort, A., Thirion, B., 2012. Small-sample Brain Mapping: Sparse Recovery on Spatially Correlated Designs with Randomization and Clustering. ICM, Edinburgh, Scotland, 2012.
- Wagner, A.D., Shannon, B.J., Kahn, I., Buckner, R.L., 2005. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9, 445–453.
- Wan, J., Zhang, Z., Rao, B., Fang, S., Yan, J., Saykin, A., Shen, L., 2014. Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation- and nonlinearity-aware sparse Bayesian learning. *IEEE Trans. Med. Imaging* 33, 1475–1477.
- Wan, J., Zhang, Z., Yan, J., Li, T., Rao, B., Fang, S., Kim, S., Risacher, S., Saykin, A., Shen, L., ADNI, f.t. 2012. Sparse Bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in Alzheimer's disease. CVPR'12: IEEE Int Conf on Computer Vision and Pattern Recognition, Providence, Rhode Island.
- Wang, H., Nie, F., Huang, H., Kim, S., Nho, K., Risacher, S., Saykin, A., Shen, L., Alzheimer's Disease Neuroimaging Initiative, 2012. Identifying quantitative trait loci via group-sparse multi-task regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* 28, 229–237.
- Wang, H., Nie, F., Huang, H., Risacher, S., Ding, C., Saykin, A., Shen, L., 2011. Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. *IEEE Conference on Computer Vision, Barcelona, Spain*, pp 557–562.
- Weiner, M.W., Aisen, P.S., Jack Jr., C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., Toga, A., Green, R., Walter, S., Soares, H., Snyder, P., Siemers, E., Potter, W., Cole, P.E., Schmidt, M., Alzheimer's Disease Neuroimaging Initiative, 2010. The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement.* 6, 202–211 e7.
- Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B* 68, 49–77.