

# Sparse Multi-Task Regression and Feature Selection to Identify Brain Imaging Predictors for Memory Performance

Hua Wang<sup>1</sup>, Feiping Nie<sup>1</sup>, Heng Huang<sup>1</sup>, Shannon Risacher<sup>2</sup>,  
Chris Ding<sup>1</sup>, Andrew J Saykin<sup>2</sup>, Li Shen<sup>2</sup>, ADNI\*

<sup>1</sup> Computer Science and Engineering, University of Texas at Arlington, TX

huawangcs@gmail.com, feipingnie@gmail.com, heng@uta.edu, chqding@uta.edu

<sup>2</sup> Radiology and Imaging Sciences, Indiana University School of Medicine, IN

srisache@iupui.edu, asaykin@iupui.edu, shenli@iupui.edu

## Abstract

*Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive impairment of memory and other cognitive functions, which makes regression analysis a suitable model to study whether neuroimaging measures can help predict memory performance and track the progression of AD. Existing memory performance prediction methods via regression, however, do not take into account either the interconnected structures within imaging data or those among memory scores, which inevitably restricts their predictive capabilities. To bridge this gap, we propose a novel Sparse Multi-task Regression and feature selection (SMART) method to jointly analyze all the imaging and clinical data under a single regression framework and with shared underlying sparse representations. Two convex regularizations are combined and used in the model to enable sparsity as well as facilitate multi-task learning. The effectiveness of the proposed method is demonstrated by both clearly improved prediction performances in all empirical test cases and a compact set of selected RAVLT-relevant MRI predictors that accord with prior studies.*

## 1. Introduction

Through employing pattern classification methods, neuroimaging has demonstrating its effectiveness in predicting Alzheimer's disease (AD) status based on individual magnetic resonance imaging (MRI) and/or positron emission tomography (PET) scans [5, 11, 18]. Because AD is a neurodegenerative disorder characterized by progressive impairment of memory and other cognitive functions, it is important to understand how structural and functional changes

in brain can influence the performance of neuropsychological tests. As a result, regression models have been used to study whether neuroimaging measures can help predict clinical scores and track AD progression [19, 22]. For example, in [22], stepwise regression was performed in a pairwise fashion to relate each MRI and FDG-PET measures of the eight candidate regions to each of the four Rey's Auditory Verbal Learning Test (RAVLT) memory scores. This approach was univariate and thereby overlooked the interrelated structures within both imaging data and clinical data. In [19], using relevance vector regression, the voxel-based morphometry (VBM) features extracted from the entire brain were jointly analyzed to predict each selected clinical score, while the investigations of different clinical scores are independent from each other.

In this paper, we embrace, rather than ignore, the complexity of the mapping between interconnected imaging measures and interrelated clinical scores; and propose a novel Sparse Multi-task Regression and feature selection (SMART) method to jointly analyze all the imaging and clinical data within a single regression model and common subspace. Our research focuses on investigating the relationships between MRI measures and RAVLT memory scores using the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort [23]. Instead of including all possible imaging measures to predict memory performance, the proposed SMART method is designed to select the most prominent imaging features that are able to predict memory performance with improved prediction accuracy. Different from LASSO [20] and other related methods that perform feature selection separately for each individual memory score, the proposed sparse multi-task learning model treats each memory score as a cognition task and selects imaging features that can jointly influence multiple scores/tasks. We propose to use the combined  $\ell_{2,1}$ -norm and  $\ell_1$ -norm regularizations to select features with high correlations to

\*Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (U01 AG024904,adni.loni.ucla.edu).

a subset of memory scores. To demonstrate the effectiveness of the proposed SMART method, we apply it to identify relevant MRI markers that can predict multiple RAVLT memory scores. Our empirical results yield not only clearly improved prediction rates in all the test cases, but also a compact set of RAVLT-relevant MRI predictors that are in accordance with prior studies.

## 2. Sparse Multi-Task Regression and Feature Selection

Recently sparse regularizations have been applied to classification based feature selection studies. LASSO [20] was shown to efficiently select useful features for a single task. However, in our work, we expect to estimate predictive models for several related memory performance scores together, not an individual one. The multi-task feature learning [2, 15] used the  $\ell_{2,1}$ -norm regularization to couple feature selection across tasks using a strict assumption - all tasks share a common underlying representation. However, in many cases, the common pattern is shared by many tasks, but not all.

To address this issue, we propose a new Sparse Multi-Task Regression and feaTure selection (SMART) model to include both  $\ell_{2,1}$ -norm and  $\ell_1$ -norm regularizations for selecting imaging features, *i.e.*, morphometric variables, and predicting memory performance. The combined convex norms help us pick up the features with high correlations to a subset of tasks. The new objective leads to a more difficult optimization problem. To address this problem, we derive a new efficient algorithm with proved global convergence. In this paper, given a matrix  $M$ , we denoted its  $i$ -th row and  $j$ -th column as  $m^i$  and  $m_j$ , respectively.

### 2.1. Joint Sparse Regularizations Using Mixed Non-Smooth Norms

To identify the predictable correlations between memory performance scores and morphometric variables, the linear (least square) regression method is a standard way in medical image analysis research. Given the morphometric variables of  $n$  training samples  $\{x_i \in \mathbb{R}^d\}_{i=1}^n$  and the associated memory scores  $\{y_i \in \mathbb{R}^c\}_{i=1}^n$ , traditional least square regression solves the following optimization problem to obtain the projection matrix  $W \in \mathbb{R}^{d \times c}$  (the bias  $b$  is absorbed into  $W$  when the constant value 1 is added as an additional dimension for each data):

$$\min_W \sum_{i=1}^n \|W^T x_i - y_i\|_2^2 = \|X^T W - Y\|_F^2, \quad (1)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix,  $X = [x_1, \dots, x_n] \in \mathbb{R}^{d \times n}$  and  $Y = [y_1, \dots, y_n]^T \in \mathbb{R}^{n \times c}$ .

In the regular linear regression, the weight matrix  $W$  is not sparse. All morphometric variables are involved to the

memory scores prediction. However, some of them are irrelevant to memory performance prediction. Therefore, it is desirable to select the important morphometric variables for more accurate scores prediction. To this end, instead of imposing the squared  $\ell_2$ -norm regularization as in traditional ridge regression, we impose the  $\ell_{2,1}$ -norm regularization. Because the  $\ell_{2,1}$ -norm regularization penalizes each row of  $W$  as a whole and enforce sparsity among the rows, it is able to select the most prominent morphometric variables [14]. Specifically, we solve the following convex optimization problem:

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

where  $\|\cdot\|_{2,1}$  denotes the  $\ell_{2,1}$ -norm of a matrix.

We further consider some important morphometric variables are only correlated to a subset of tasks. The  $\ell_{2,1}$ -norm cannot handle them properly. Thus, we add an  $\ell_1$ -norm regularizer to impose the sparsity among all elements in  $W$  and propose our new Sparse Multi-tAsk Regression and feaTure selection (SMART) model as<sup>1</sup>:

$$\min_W \|X^T W - Y\|_F^2 + \gamma_1 \|W\|_1 + \gamma_2 \|W\|_{2,1}. \quad (3)$$

Although our objective function is convex, it is difficult to be solved, because the both regularization terms are non-smooth. Here, we propose an efficient algorithm to solve our objective function in Eq. (3).

Taking the derivative with respect to  $w_i$  ( $1 \leq i \leq c$ ), and setting it to zero, we have

$$X X^T w_i - X y_i + \gamma_1 D_i w_i + \gamma_2 \tilde{D} w_i = 0, \quad (4)$$

where  $D_i$  ( $1 \leq i \leq c$ ) is a diagonal matrix with the  $k$ -th diagonal element as  $\frac{1}{2\|w_{ki}\|}$ ,  $\tilde{D}$  is a diagonal matrix with the  $k$ -th diagonal element as  $\frac{1}{2\|w^k\|_2}$ . Thus,

$$w_i = (X X^T + \gamma_1 D_i + \gamma_2 \tilde{D})^{-1} X y_i. \quad (5)$$

Note that  $D_i$  and  $\tilde{D}$  depend on  $W$  and thus is also unknown variables. We propose an iterative algorithm to solve this problem, which is as listed in Algorithm 1.

### 2.2. Algorithm Analysis

**Theorem 1** *Algorithm 1 decreases the objective value in each iteration.*

**Proof:** According to Step 2 in the algorithm, we have

$$\begin{aligned} W^{(t+1)} &= \min_W \text{Tr}(X^T W - Y)^T (X^T W - Y) \\ &\quad + \gamma_1 \sum_{i=1}^c w_i^T D_i^{(t)} w_i + \gamma_2 \text{Tr} W^T \tilde{D}^{(t)} W, \end{aligned} \quad (6)$$

<sup>1</sup>This paper was first submitted to a conference in 2010 and we notice a similar objective in [12] when we prepare the camera-ready draft of this paper. This research is an independent work.

---

**Algorithm 1:** Algorithm

---

**Input:**  $X, Y$ Initialize  $W^1 \in \mathbb{R}^{d \times c}$ ,  $t = 1$ ;**while not converge do**

1. Calculate the diagonal matrices

 $D_i^{(t)}$  ( $1 \leq i \leq c$ ) and  $\tilde{D}^{(t)}$ , where the  $k$ -th diagonal element of  $D_i^{(t)}$  is  $\frac{1}{2|w_{ki}^{(t)}|}$ , the  $k$ -th diagonal element of  $\tilde{D}^{(t)}$  is  $\frac{1}{2\|(w^{(t)})^k\|_2}$ ;2. For each  $i$  ( $1 \leq i \leq c$ ), $w_i^{(t+1)} = (X X^T + \gamma_1 D_i^{(t)} + \gamma_2 \tilde{D}^{(t)})^{-1} X y_i$ ;3.  $t = t + 1$ ;**Output:**  $W^{(t)} \in \mathbb{R}^{d \times c}$ .

---

therefore we have

$$\begin{aligned} & Tr(X^T W^{(t+1)} - Y)^T (X^T W^{(t+1)} - Y) \\ & + \gamma_1 \sum_{i=1}^c (w_i^{(t+1)})^T D_i^{(t)} w_i^{(t+1)} + \gamma_2 Tr(W^{(t+1)})^T \tilde{D}^{(t)} W^{(t+1)} \\ & \leq Tr(X^T W^{(t)} - Y)^T (X^T W^{(t)} - Y) \\ & + \gamma_1 \sum_{i=1}^c (w_i^{(t)})^T D_i^{(t)} w_i^{(t)} + \gamma_2 Tr(W^{(t)})^T \tilde{D}^{(t)} W^{(t)} \\ \Rightarrow & Tr(X^T W^{(t+1)} - Y)^T (X^T W^{(t+1)} - Y) \\ & + \gamma_1 \sum_{i=1}^d \sum_{j=1}^c \left( \frac{(w_{ij}^{(t+1)})^2}{2\|w_{ij}^{(t)}\|} - \|w_{ij}^{(t+1)}\| + \|w_{ij}^{(t)}\| \right) \\ & + \gamma_2 \sum_{k=1}^d \left( \frac{\|(w^{(t+1)})^k\|_2^2}{2\|(w^{(t)})^k\|_2} - \|(w^{(t+1)})^k\|_2 + \|(w^{(t)})^k\|_2 \right) \\ & \leq Tr(X^T W^{(t)} - Y)^T (X^T W^{(t)} - Y) \\ & + \gamma_1 \sum_{i=1}^d \sum_{j=1}^c \left( \|w_{ij}^{(t)}\| + \frac{(w_{ij}^{(t)})^2}{2\|w_{ij}^{(t+1)}\|} - \|w_{ij}^{(t)}\| \right) \\ & + \gamma_2 \sum_{k=1}^d \left( \|(w^{(t)})^k\|_2 + \frac{\|(w^{(t)})^k\|_2^2}{2\|(w^{(t+1)})^k\|_2} - \|(w^{(t)})^k\|_2 \right) \\ \Rightarrow & Tr(X^T W^{(t+1)} - Y)^T (X^T W^{(t+1)} - Y) \\ & + \gamma_1 \sum_{i=1}^d \sum_{j=1}^c \|w_{ij}^{(t+1)}\| + \gamma_2 \sum_{k=1}^d \|(w^{(t+1)})^k\|_2 \\ & \leq Tr(X^T W^{(t)} - Y)^T (X^T W^{(t)} - Y) \\ & + \gamma_1 \sum_{i=1}^d \sum_{j=1}^c \|w_{ij}^{(t)}\| + \gamma_2 \sum_{k=1}^d \|(w^{(t)})^k\|_2 \end{aligned}$$

The last step holds, because [14] for any vector  $w$  and  $w_0$ , we have  $\|w\|_2 - \frac{\|w\|_2^2}{2\|w_0\|_2} \leq \|w_0\|_2 - \frac{\|w_0\|_2^2}{2\|w_0\|_2}$ . Thus, the algorithm decreases the objective value in each iteration.  $\square$

At the convergence,  $W^{(t)}$ ,  $D_i^{(t)}$  ( $1 \leq i \leq c$ ) and  $\tilde{D}^{(t)}$  will satisfy the Eq. (5). As the problem (3) is a convex problem, satisfying the Eq. (5) indicates that  $W$  is a global optimum

solution to the problem (3). Therefore, Algorithm 1 will converge to the global optimum of the problem (3). Because we have closed form solution in each iteration, our algorithm converges very fast.

### 3. Imaging and Memory Data

Both MRI and memory data used in this study were obtained from the ADNI database<sup>2</sup>. ADNI is a landmark investigation sponsored by the NIH and industrial partners designed to collect longitudinal neuroimaging, biological and clinical information from 822 participants that will track the neural correlates of memory loss from an early stage. Further information can be found in [13] and at [www.adni-info.org](http://www.adni-info.org). Following a previous imaging genetics study [17], 708 out of 733 non-Hispanic Caucasian participants with no missing MRI morphometric and RAVLT information were included in this study. The 708 participants are categorized by three baseline diagnostic groups: healthy control (HC,  $n = 199$ ), mild cognitive impairment (MCI,  $n = 346$ ) (thought to be a preclinical stage of AD), and AD ( $n = 163$ ).

Two widely employed automated MRI analysis techniques were used to process and extract imaging measures across the entire brain from all baseline scans of ADNI participants as previously described [17]. First, voxel-based morphometry (VBM) [22] was performed to define global gray matter (GM) density maps and extract local GM density values for 86 target regions. Second, automated parcellation via FreeSurfer V4 [9] was conducted to define 56 volumetric and cortical thickness values and to extract total intracranial volume (ICV). The full descriptions about these measures are available in [17]. All these measures were adjusted for the baseline age, gender, education, handedness, and baseline ICV using the regression weights derived from the healthy control participants.

The cognitive measures we use to test the proposed SMART method are the baseline RAVLT memory scores from all ADNI participants [1]. The standard RAVLT format starts with a list of 15 unrelated words (List A) repeated over five different trials and participants are asked to repeat. Then the examiner presents a second list of 15 words (List B), and the participant is asked to remember as many words as possible from List A. Trial 6, termed as 5 minute recall, requests the participant again to recall as many words as possible from List A, without reading it again. Trial 7, termed as 30 minute recall, is administrated in the same way as Trial 6, but after a 30 minute delay. Finally, a recognition test with 30 words read aloud, requesting the participant to indicate whether or not each word is on List A. The RAVLT has proven useful in evaluating verbal learning and memory. The five RAVLT scores are summarized in Table 1.

---

<sup>2</sup><http://www.loni.ucla.edu/ADNI>

Table 1. Descriptions of RAVLT cognitive measures.

Task 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 total number of words recalled
RECOG	30 minute delay recognition score

## 4. Experimental Results and Discussions

In this section, we evaluate the proposed SMART method by applying it to the ADNI cohort, where a wide range of MRI morphometric features are examined and selected to predict memory performance measured by five RAVLT scores shown in Table 1. The goal is to select a compact set of RAVLT-relevant MRI features while maintaining high predictive power.

### 4.1. Improved Memory Performance Prediction

In our experiments, we examine three different sets of morphometric variables  $\{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^d$  for each participant, where  $d = 86$  for VBM morphometric variables,  $d = 56$  for FreeSurfer morphometric variables, and  $d = 144$  for the combined set of VBM and FreeSurfer variables. Evaluating the memory performance prediction on the three baseline diagnosis groups (HC, MCI, AD) and the group with all participants (HC + MCI + AD) using the three types of morphometric variables, we end up with a total of twelve test cases as in Table 2, where, e.g., “FreeSurfer\_HC” denotes the test case conducted on the participants of MCI group using FreeSurfer morphometric variables, and “VBM+FreeSurfer\_all” denotes the test case conducted on all the participants using the combined morphometric variables by VBM and FreeSurfer.

We compare SMART against multivariate regression (MRV) in memory performance prediction. For each test case, we randomly pick 80% participants and use their morphometric variables and memory scores as training data, and perform the prediction for the remaining participants. The prediction performances assessed by root mean square error (RMSE), a widely used measurement for statistical regression analysis, are reported in Table 2.

A first observation on the results in Table 2 shows that the proposed SMART method consistently outperforms the conventional multivariate regression method in all the test cases for all the cognitive tasks. The FreeSurfer measures, VBM measures, and combined measures have similar predictive powers.

A more careful analysis shows that, using our method, it is easier to predict memory performance for AD than HC, while MCI shows an intermediate pattern. This partially agrees with the findings in [22], which claims that MR morphometry is not related to memory in HC, but positively

Table 2. Prediction performance measured by RMSE.

Test cases		TOTAL	TOT6	TOTB	T30	RECOG
FreeSurfer_HC	MVR	8.762	4.362	3.281	4.305	4.021
	SMART	6.645	2.940	2.235	2.806	3.621
FreeSurfer_MCI	MVR	6.998	2.765	2.399	2.480	3.427
	SMART	5.600	1.990	1.953	1.709	3.181
FreeSurfer_AD	MVR	5.897	1.768	2.058	1.382	3.390
	SMART	5.042	1.452	1.716	1.050	2.830
FreeSurfer_all	MVR	5.926	2.238	2.036	2.090	3.342
	SMART	5.736	2.139	1.961	1.966	3.196
VBM_HC	MVR	8.651	3.772	2.885	3.496	4.776
	SMART	6.705	2.844	2.139	2.656	3.584
VBM_MCI	MVR	11.495	4.256	4.621	4.032	5.598
	SMART	5.584	1.832	1.931	1.669	3.017
VBM_AD	MVR	7.223	2.162	2.622	1.479	4.163
	SMART	5.120	1.518	1.826	0.904	2.781
VBM_all	MVR	6.090	2.290	2.140	2.141	3.396
	SMART	5.718	2.103	1.993	1.921	3.182
VBM+FreeSurfer_HC	MVR	12.265	5.416	4.349	5.089	6.703
	SMART	6.664	2.829	2.230	2.683	3.577
VBM+FreeSurfer_MCI	MVR	68.222	26.146	23.489	30.033	34.306
	SMART	5.533	1.901	1.869	1.606	3.114
VBM+FreeSurfer_AD	MVR	14.552	4.307	5.141	4.297	8.430
	SMART	4.805	1.218	1.731	0.858	2.865
VBM+FreeSurfer_all	MVR	6.505	2.596	2.258	2.540	3.582
	SMART	5.809	2.208	2.000	2.051	3.214

related to memory functions in MCI and AD. Using multivariate regression, the above trend holds only for FreeSurfer measures. For VBM and combined cases, it is far more difficult to predict memory performance in MCI than HC and AD (11.495 vs. 8.651 and 7.233 for VBM, and 68.22 vs. 12.265 and 14.552 for VBM + FreeSurfer).

Finally, we can see that the most predictable outcome is T30 for AD group with RMSE of 1.050 for FreeSurfer, 0.904 for VBM, and 0.858 for the combined measures. Considering TOTAL is the sum of the 5 scores, the performance for AD group is decent with RMSE of 5.042 for FreeSurfer, 5.120 for VBM, and 4.805 for the combined measures. The least predictable outcome is RECOG, whose RMSEs are generally greater than 2.7.

### 4.2. Feature Selection Capabilities

The main advantage of the proposed SMART method lies in its capability to simultaneously perform regression analysis and feature selection. Besides reducing the computational complexity of the learning model as in other applications, feature selection is of significant importance in the study of neuroimaging, because it has a potential to identify the relevant imaging predictors and explain the effects of morphometric changes in relation to memory performance.

The heat map of the regression coefficients of each FreeSurfer measure w.r.t. each cognitive task ( $W$  in Eq. (3))

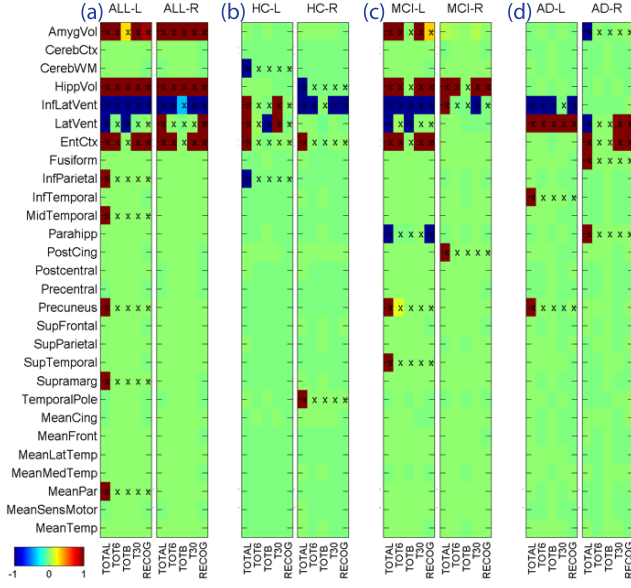


Figure 1. Heat map of selected features for prediction using FreeSurfer measures in (a) the total sample, (b) HC, (c) MCI, and (d) AD. In each of (a-d), regression weights (*i.e.*, coefficients) for left and right measures are visualized as two separate panels, where columns in each panel correspond to different memory scores. Since our method selects features with absolute values  $\geq 1$ , the range of the color map is limited to  $[-1,1]$  for a more effective visualization. All selected features are marked with “x”.

learned by SMART is shown in Fig. 1. The bigger the magnitude of an coefficient is, the more important the feature is in predicting the corresponding memory score. For example, “HippVol” (hippocampal volume) plays the most important role in memory performance prediction when testing on all participants, while “LatVent” (volume of lateral ventricle) is the most effective predictor when the test is conducted on AD group. The selected features by our method are marked with “x”. The heat map of regression coefficients of VBM measures are shown in Fig. 2. Fig. 3 visualizes the cortical map of selected features for prediction of TOTAL score using FreeSurfer measures in the total sample (left) and the AD sample (right).

Fig. 1 shows that “HippVol” is consistently selected in all the groups except AD, implicating that it is an important indicator for cognitive decline and has a potential for early detection of AD. This perfectly accords with many evidences in existing literatures [3, 4, 7, 8, 16, 21]. In addition, “EntCtx” (thickness of entorhinal cortex), “Parahipp” (thickness of parahippocampal gyrus), “Precuneus” (thickness of precuneus) and “InfParietal” (thickness of inferior parietal gyrus) are also selected in different test conditions. These areas are important components of the brain’s episodic memory network [22], which has been proved to be normally engaged during episodic recall and heavily impact the memory performance [6, 10, 22]. Similar observations that our selections match literature evidences can also

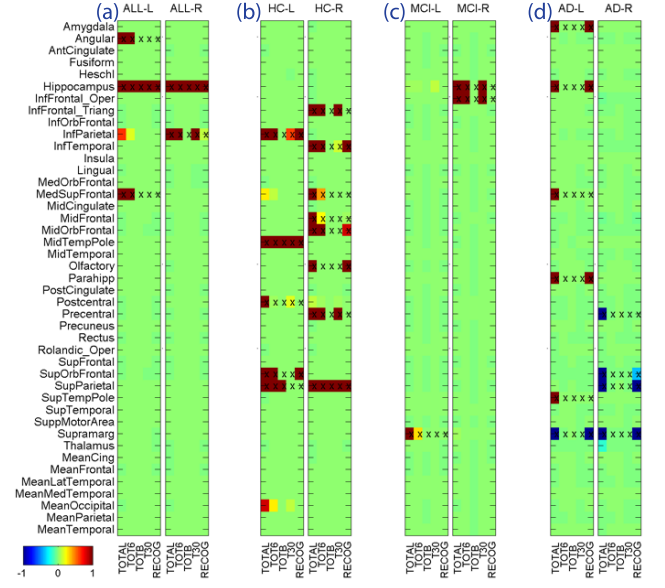


Figure 2. Heat map of selected features for prediction using VBM measures in (a) the total sample, (b) HC, (c) MCI, and (d) AD. In each of (a-d), regression weights (*i.e.*, coefficients) for left and right measures are visualized as two separate panels, where columns in each panel correspond to different memory scores. Since our method selects features with absolute values  $\geq 1$ , the range of the color map is limited to  $[-1,1]$  for a more effective visualization. All selected features are marked with “x”.

be found in both Fig. 1 and Fig. 2, which concretely confirm the effectiveness of the proposed method from neurobiological perspective.

Moreover, besides selecting common prominent features across all cognitive tasks through imposing  $\ell_{2,1}$  regularization as in Eq. (3), we also enforce sparsity on  $W$  through  $\ell_1$  regularization, such that the relative importance of the selected features are properly weighted. For example, as in Fig. 2, the “Hippocampus” (GM density) is only selected in MCI and AD groups, but not selected by HC group. This observation, again, is extensively supported in literature. It has been shown that, in normal aging, memory, including listing learning measures with clinically applied retention intervals ( $< 1h$ ), appears weakly related to medial temporal lobe (MTL) [16], whereas memory has consistently been related to MTL volumes in MCI and AD [16]. This provides one more evidence showing the ability of SMART for properly identifying relevant features.

## 5. Conclusions

In this paper, we proposed a new SMART model to perform both regression analysis for memory performance prediction and morphometric variables selection in an MCI/AD study. Different from related existing methods that ignore the interrelated structures within imaging data or those within clinical data, SMART analyzes all the imaging and

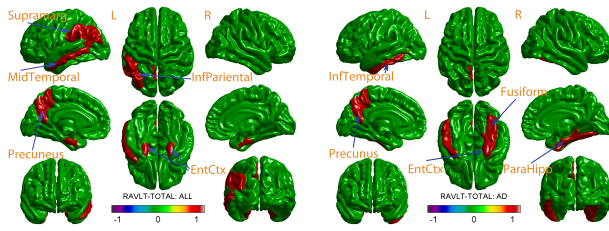


Figure 3. Cortical map of selected features for prediction using FreeSurfer measures in the total sample (left) and the AD sample (right). Each map only visualizes the regression weights for RAVLT-TOTAL score for individual cortical thickness measures (i.e., volume measures and mean thickness measures of larger regions are not included). Since our method selects features with absolute values  $\geq 1$ , the range of the color map is limited to  $[-1, 1]$  for a more effective visualization.

clinical data within a single regression framework and common subspace, such that the predictive performance can be improved by these correlations. Our experiments using the MRI and RAVLT data of the ADNI cohort yielded promising results: (1) the prediction performance of SMART was consistently better than conventional multi-variate regression, (2) a compact set of imaging predictors were identified in each test case and were in accordance with prior findings, and (3) these selected imaging features could predict multiple memory scores at the same time and had a potential to play an important role in determine cognitive functions and characterizing AD progression. These promising results were consistent with our theoretical foundation and prior studies, which demonstrated the effectiveness of the proposed method.

## Acknowledgement.

HW and HH were supported by NSF-IIS 1117965, NSF-CNS 0923494, NSF-IIS 1041637, NSF-CNS 1035913. SR, AS and LS were supported in part by NSF-IIS 1117335, NIA 1RC 2AG036535, CTSI-IUSM/CTR (RR025761), NIA P30 AG10133, and NIA R01 AG19771.

## References

- [1] P. S. Aisen, R. C. Petersen, and *et al.* Clinical core of the alzheimer's disease neuroimaging initiative: progress and plans. *Alzheimers Dement*, 6(3):239–46, 2010.
- [2] A. Argyriou, T. Evgeniou, and M. Pontil. Multi-task feature learning. *NIPS*, pages 41–48, 2007.
- [3] M. Ball, V. Hachinski, A. Fox, A. Kirshen, M. Fisman, W. Blume, V. Kral, H. Fox, and H. Merskey. A new definition of Alzheimer's disease: a hippocampal dementia. *Lancet*, 325(8419):14–16, 1985.
- [4] J. Barnes, R. Scallan, and *et al.* Increased hippocampal atrophy rates in AD over 6 months using serial MR imaging. *Neurobiol Aging*, 29(8):1199–1203, 2008.
- [5] N. Batmanghelich, B. Taskar, and C. Davatzikos. A general and unifying framework for feature construction, in image-based pattern classification. *Inf Process Med Imaging*, 21:423–34, 2009.
- [6] R. Buckner and D. Carroll. Self-projection and the brain. *Trends Cogn. Sci.*, 11(2):49–57, 2007.
- [7] A. Convit, M. de Leon, C. Tarshish, S. De Santi, A. Kluger, H. Rusinek, and A. George. Hippocampal volume losses in minimally impaired elderly. *Lancet*, 345(8944):266, 1995.
- [8] S. De Santi, M. de Leon, and *et al.* Hippocampal formation glucose metabolism and volume losses in MCI and AD. *Neurobiol of aging*, 22(4):529–539, 2001.
- [9] B. Fischl, D. H. Salat, and *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–55, 2002.
- [10] D. Hassabis and E. Maguire. Deconstructing episodic memory with construction. *Trends Cogn. Sci.*, 11(7):299–306, 2007.
- [11] C. Hinrichs, V. Singh, et al. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, 48(1):138–49, 2009.
- [12] S. Lee, J. Zhu, and E. Xing. Adaptive Multi-Task Lasso: with Application to eQTL Detection. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *NIPS*, 2010.
- [13] S. G. Mueller, M. W. Weiner, and *et al.* Ways toward an early diagnosis in alzheimer's disease: The alzheimer's disease neuroimaging initiative (adni). *Alzheimers Dement*, 1(1):55–66, 2005.
- [14] F. Nie, H. Huang, X. Cai, and C. Ding. Efficient and Robust Feature Selection via Joint  $\ell_2, \ell_1$ -Norms Minimization. In *NIPS*, 2010.
- [15] G. Obozinski, B. Taskar, and M. Jordan. Multi-task feature selection. *Technical report, Department of Statistics, University of California, Berkeley*, 2006.
- [16] R. Petersen, C. Jack Jr, and *et al.* Memory and MRI-based hippocampal volumes in aging and AD. *Neurology*, 54(3):581, 2000.
- [17] L. Shen, S. Kim, and *et al.* 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(3):1051 – 1063, 2010.
- [18] L. Shen, Y. Qi, and *et al.* Sparse bayesian learning for identifying imaging biomarkers in AD prediction. In *MICCAI*, 2010.
- [19] C. M. Stonnington, C. Chu, and *et al.* Predicting clinical scores from magnetic resonance scans in alzheimer's disease. *Neuroimage*, 51(4):1405–13, 2010.
- [20] R. Tibshirani. Regression shrinkage and selection via the LASSO. *J. Royal. Statist. Soc B.*, 58:267–288, 1996.
- [21] C. Van Petten. Relationship between hippocampal volume and memory ability in healthy individuals across the lifespan: review and meta-analysis. *Neuropsychologia*, 42(10):1394–1413, 2004.
- [22] K. Walhovd, A. Fjell, and *et al.* Multi-modal imaging predicts memory performance in normal aging and cognitive decline. *Neurobiol Aging*, 31(7):1107–1121, 2010.
- [23] M. W. Weiner, P. S. Aisen, and *et al.* The alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimers Dement*, 6(3):202–11 e7, 2010.