

## **HHS Public Access**

Author manuscript *Proc IEEE Int Symp Biomed Imaging*. Author manuscript; available in PMC 2018 July 16.

Published in final edited form as: *Proc IEEE Int Symp Biomed Imaging*. 2018 April ; 2018: 1406–1410. doi:10.1109/ISBI.2018.8363835.

### MULTI-TASK SPARSE SCREENING FOR PREDICTING FUTURE CLINICAL SCORES USING LONGITUDINAL CORTICAL THICKNESS MEASURES

Jie Zhang<sup>1</sup>, Yanshuai Tu<sup>1</sup>, Qingyang Li<sup>1</sup>, Richard J. Caselli<sup>2</sup>, Paul M. Thompson<sup>3</sup>, Jieping Ye<sup>4</sup>, and Yalin Wang<sup>1</sup>

<sup>1</sup>School of Computing, Informatics, and Decision Systems Engineering, Arizona State Univ., Tempe, AZ

<sup>2</sup>Department of Neurology, Mayo Clinic, Scottsdale, AZ

<sup>3</sup>Imaging Genetics Center, Institute for Neuroimaging and Informatics, Univ. of Southern California, Los Angeles, CA

<sup>4</sup>Department of Electrical Engineering and Computer Science, Univ. of Michigan, Ann Arbor, MI

#### Abstract

Cortical thickness estimation performed *in-vivo* via magnetic resonance imaging (MRI) is an effective measure of brain atrophy in preclinical individuals at high risk for Alzheimer's disease (AD). However, the high dimensionality of individual cortical thickness data coupled with small population samples make it challenging to perform cortical thickness feature selection for AD diagnosis and prognosis. Thus far, there are very few methods that can accurately predict future clinical scores using longitudinal cortical thickness measures. In this paper, we propose an unsupervised dictionary learning algorithm, termed Multi-task Sparse Screening (MSS) that produces improved results over previous methods within this problem domain. Specifically, we formulate and solve a multi-task problem using extracted top-p significant features from the Alzheimer's Disease Neuroimaging Initiative (ADNI) longitudinal data. Empirical studies on publicly available longitudinal data from ADNI dataset (N= 2797) demonstrate improved correlation coefficients and root mean square errors, when compared to other algorithms.

#### **Index Terms**

Alzheimer's Disease; Cortical Thickness; Group Lasso; Multi-task; Dictionary Learning

#### **1. INTRODUCTION**

Cortical thickness estimation in magnetic resonance imaging (MRI) is an important research that has been applied to detect localized neuroanatomical differences. It has been carefully studied in Alzheimer's Disease (AD) research as a potential imaging biomarker to evaluate AD risk, track AD progression, especially in preclinical individuals at high risk for AD (e.g. Mild Cognitive Impairment (MCI)), and facilitate early interventions [5]. A number of cortical thickness estimation methods have been developed [4, 11, 8, 17, 19] and they were well adopted in neuroimaging research. Despite advances in cortical thickness estimation

used to track symptomatic patients, there is still a lack of accurate and reliable brain imaging systems that can predict future clinical decline by analyzing longitudinal cortical thickness changes.

In neuroimaging research, there is also constant interest in building prediction models by analyzing global brain structural changes. A growing number of longitudinal preclinical or clinical AD imaging cohorts are under development. However, the main challenge in developing brain imaging systems to assist AD diagnosis and prognosis arises from the fact that the neuroimaging data dimensionality is intrinsically high while only a small number of samples are typically available. In this regard, feature selection is necessary to overcome this so-called "large p, small n" problem (p = 129, 600 in this paper). Feature selection reduces the feature dimension by selecting a subset of original variables [9], which are considered high quality training data. However, most current feature selection methods either demand careful parameter tuning or generate features that cannot be easily interpreted or even visualized or focus on classification problem [16]. Lasso [18] is one of the most popular feature selection methods due to its computational feasibility and amenability to theoretical analysis, but the Lasso estimator is not variable selection consistent if the irrepresentable condition fails [21], which means the correct sparse subset of the relevant variables cannot be identified asymptotically with large probability. In addition, over-fitting becomes one of the biggest concerns for building prediction models in brain imaging research. Recent developments from dictionary learning [14, 12, 13, 26] can offer valuable insights for above challenges. However, most existing works on dictionary learning focus on the prediction of a target outcome at a single time point [24, 22] or some region-of-interest [23, 25]. To address above challenges, a joint analysis of imaging features from multiple time points is expected to improve AD prediction performance.

In this paper, we propose a novel feature selection model integrated with an unsupervised framework that we termed as Multi-task Sparse Screening (MSS) algorithm. Our framework has two components. Stage one selects the top-*p* significant features by screening group Lasso based feature ranking [20], which is theoretically rigorous and more robust against data perturbation and less sensitive to the input parameters. The second stage is formulated as a multi-task dictionary learning problem which uses shared and individual dictionaries to encode both consistent and varying imaging features along longitudinal time points. We evaluate the proposed framework on brain images from Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) (N= 2797) by analyzing longitudinal cortical thickness features to predict future cognitive scores, including the Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale cognitive subscale (ADAS-cog). Our experimental results outperform seven other methods and demonstrate the superiority of the proposed algorithm.

Our main contributions are threefold. First, we select top-*p* stability features by robust group Lasso screening from longitudinal cortical thickness features. Second, we consider the variance of subjects from different time points by an unsupervised dictionary learning method (i.e. MSS). To the best of our knowledge, it is the first learning model that unifies both sparse learning feature screening and multi-task learning with dictionary learning research for brain imaging analysis. Thirdly, in our empirical experiments on a relatively

#### 2. MULTI-TASK SPARSE SCREENING

Let  $X \in \mathbb{R}^{d \times n}$  represents the cortical thickness features and  $y \in \mathbb{R}^n$  as a vector of *n* observations (responses) which are responses of the clinical scores (e.g., MMSE and ADAScog) in regression problem. Because we use three time points (baseline, 6-month, 12-month) to predict the clinical scores at 24-month, we divide total *d* features into *L* groups, with  $d_I$  as

the number of features in group *l*. We use a data matrix,  $[X]_l \in \mathbb{R}^{d_l \times n}$ , to represent the

higher correlation coefficients when compared to seven other algorithms.

predictors corresponding to the *I*th group, with corresponding coefficient vector  $[a]_I$ . We assume that *y* and *X* have been normalized so that all variables have a mean of zero. With the group information available, the group Lasso problem [21] takes the form of solving the convex optimization problem as follows:

$$\inf_{\alpha \in \mathbb{R}^d} \frac{1}{2} \| y - \sum_{l=1}^{L} [\alpha]_l [X]_l \|_2^2 + \lambda \sum_{l=1}^{L} \sqrt{d_l} \| [\alpha]_l \|_2, \quad (1)$$

where  $d = \sum_{l=1}^{L} d_{l}$ . The regularization term accounts for groups of different sizes by the  $\sqrt{d_{l}}$  terms. Eq. 1 behaves like the lasso [18] at the group level and it does reduce to lasso when the group sizes are all one. The optimal solution vector a of the group Lasso problem above is sparse at the group level but could be made sparse within a group by adding an L1 norm penalty term. we consider the non-zero elements as the important features because of the sparsity of the solution vector. Therefore we propose a group lasso screening process in this study to select the top p relevant features to the original ADNI dataset.

#### Algorithm 1

Multi-task Sparse Screening (MSS)

**Input:** Cortical thickness features from different time points  $\{X_1, ..., X_T\}$  and corresponding response  $\{y_1, ..., y_T\}$  where  $d \times n_t$  each  $X_t \in \mathbb{R}$  and  $y_t \in \mathbb{R}^t$ 

Output: Dictionaries and sparse codes

1: for t = 1 to T do

- 2: Set all the elements of  $\Gamma$  to be zero.
- 3: **for** j = 1 to *J* **do**
- 4: Optimize the Group Lasso with  $\lambda_j$ ,  $X_b$ ,  $y_b$  get the optimal solution  $a^*(\lambda_j)$ .
- 5: For each group *I* in the feature space *L*,  $\Gamma_I = \Gamma_I + 1$  if  $[a]_l^*(\lambda_i)$  is not zero.
- 6: end for
- 7: Rank  $\Gamma$  in descending order, select the top *p* features from  $X_t$  to construct  $\widetilde{X}_t$ .

8: end for 9: for k = 1 to x do 10: for t = 1 to T do 11: Get feature matrix  $\widetilde{X}_t$ . 12: Update  $\widehat{D}_t^k: \widehat{D}_t^k = \Phi$ . 13: Update  $z_t^{k+1}$  and index set  $I_t^{k+1}$  by CCD. 14: Update the  $\widehat{D}_t$  and  $\overline{D}_t$  by one step SGD. 15: Normalize  $\widehat{D}_t^{k+1}$  and  $\overline{D}_t^{k+1}$  based on the index set  $I_t^{k+1}$ 16: Update the shared dictionary  $\Phi: \Phi = \widehat{D}_t^{k+1}$ 17: end for 18: end for 18: end for

We optimize the group Lasso screening in a sequence of parameter values  $\lambda_0 > \lambda_1 > ... > \lambda_J$ . For each parameter value  $\lambda_j$  where  $j \in (0, J)$ , we obtain the optimal solution vector  $a^*(\lambda_j)$ . Then we record the index of non-zero elements in the solution vector. We use an index factor  $\Gamma$  to record the frequency of non-zero elements in each  $a^*(\lambda_j)$  where  $\Gamma \in \mathbb{R}^L$ . For each group *I* in the feature space *L*, we increase the  $\Gamma_I$  by  $\Gamma_I = \Gamma_I + 1$  when the elements of  $[a]_l^*(\lambda_j)$  are not equal to zero in the *I* group. Finally, after obtaining the index frequency  $\Gamma$ , we rank  $\Gamma$  by descend and select the top *p* frequent ones to construct the feature matrix  $\tilde{X}$  from *X*.

After the screening procedure, to use features from different time points, we defined our multi-sparse learning model. Given the reconstructed feature matrices from *T* different time points:  $\{\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_T\}$  our objective is to learn a set of sparse codes  $\{Z_1, Z_2, ..., Z_T\}$  for each time point where  $\tilde{X}_t \in \mathbb{R}^{p \times n_t}$ ,  $Z_t \in \mathbb{R}^{l_t \times n_t}$  and  $t \in \{1, ..., T\}$   $n_t$  is the number of subjects for  $\tilde{X}_t$  and  $l_t$  is the dimension of each sparse code in  $Z_t$ . When employing sparse coding [15] to learn the sparse codes  $Z_t$  by  $\tilde{X}_t$  individually, we obtain a set of dictionaries  $\{D_1, ..., D_T\}$  but there are no correlations between the learned dictionaries. One solution is to construct the feature matrices  $\{\tilde{X}_1, ..., \tilde{X}_T\}$  into one feature matrix  $\tilde{X}$  to obtain the dictionary D. However, only one dictionary D is not enough to show the variation among subjects from different time points. To address this challenge, we integrate the idea of multi-task learning into sparse coding and propose a novel multi-task algorithm termed Multi-task Sparse Screening (MSS), to learn features from different time points.

For the feature matrix  $\widetilde{X}_t$  of a particular time point, MSS learns a dictionary  $D_t$  and sparse codes  $Z_t$ .  $D_t$  is composed of two parts:  $D_t = [\widehat{D}, \overline{D}_t]$  where  $\widehat{D} \in \mathbb{R}^{p \times \hat{l}}$ , and  $\hat{l} + \overline{l}_t = l_t$ .  $\widehat{D}$  is

common among all the learnt dictionaries while  $\overline{D}_t$  is different from each other and only learnt from  $\widetilde{X}_t$ . Thus, objective function can be reformulated as follows:

$$\min_{D_1,...,D_T,Z_1,...,Z_T} \sum_{t=1}^T \frac{1}{2} \| \widetilde{X}_t - [\widehat{D},\overline{D}_t] Z_t \|_F^2 + \lambda \sum_{t=1}^T \| Z_t \|_1 : D_t \in \Psi_t, \quad (2)$$

where  $\hat{D}_1 = \dots = \hat{D}_T$ ,  $\Psi_t = \left\{ D_t \in \mathbb{R}^{p \times l_t} : \forall_j \in 1, \dots, l_t, \|[D_t]_j\|_2 \le 1 \right\}$  and  $[D_t]_j$  is the *j*th column of  $D_t$ ,  $\|\cdot\|_F$  is Fibonacci norm and  $\lambda$  is the positive regularization parameter.

Fig. 1 illustrates the MSS process example with subjects of ADNI from three different time points: baseline, 6- and 12-months. We used different colors to denote different important features selected for different time slots. We initialize  $\hat{D}$  by randomly selecting  $\hat{l}$  samples from feature matrices across different time points  $\{\tilde{X}_1, ..., \tilde{X}_T\}$ , to construct it. For the individual part of each dictionary, we randomly select  $\bar{l}$  subjects from the corresponding matrix  $\tilde{X}_t$  to construct  $\bar{D}_t$  After initializing dictionary  $D_t$  for each time point, we set all the sparse code  $Z_t$  to be zero at the beginning. The key steps of MSS are summarized in Algorithm 1.

In Algorithm 1,  $\Phi$  represents the shared part of each dictionary. Then, we learn the sparse code  $z_t^{k+1}$  from  $Z_t$  by several steps of Cyclic Coordinate Descent (CCD) [2] where  $z_t^{k+1} \in \mathbb{R}^{l_t}$ . Then we use learnt sparse codes  $z_t^{k+1}$  to update the dictionary  $\hat{D}_t^{k+1}$  and  $\bar{D}_t^{k+1}$  by one step Stochastic Gradient Descent (SGD)[27]. Since  $z_t^{k+1}$  is very sparse, we use the index set  $I_t^{k+1}$  to record the location of non-zero entries in  $z_t^{k+1}$  to accelerate the update of sparse codes and dictionaries.  $\Phi$  is updated at the end of *k*th iteration to ensure  $\hat{D}_t^{k+1}$  is the same among all the dictionaries.

#### 3. EXPERIMENT

We used all available brain longitudinal imaging data from ADNI. The subject demography table is shown in Table 1 including AD, MCI and Cognitively Unimpaired (CU). FreeSurfer [7] is adopted to compute the cortical thickness by deforming the white matter surface to pial surface and then measuring the deformation distance as the cortical thickness. FreeSurfer also produces a spherical parameterizations for each pial surface, which is used by weighted spherical harmonics representation (WSHR) [3] to register pial surfaces across subjects. WSHR fixes the Gibbs phenomenon (ringing effects) associated with the traditional Fourier descriptors and spherical harmonic representation by weighting the series expansion with exponential weights [3]. The exponential weights make the representation converges faster and reduces the amount of wiggling. We used WSHR to help create consistent features across subjects.

#### 3.1. Experiment setting

We set the columns of  $D_t$  to be 1000; 500 for the shared part and 500 for the individual part. We test whether MSS provides a statistically powerful clinical score prediction solution on three groups of experiments, including (1) The whole ADNI dataset N=2797 (Whole), (2) AD and MCI, (3) MCI and CU. We did four times 5-fold cross validation to estimate the prediction power and report the averaged results. For each set of experiments, we selected 5000 features along a sequence of  $\lambda$  values equally spaced on the linear scale of from 0.1 to 1 and selected  $\kappa = 10$  epochs. We used output of MSS as regressors and 2 clinical scores (MMSE and ADAS-cog) as responses in the Lasso regression. For testing, we used baseline dictionary to predict 24-months clinical scores since this dictionary also has the common dictionary already including baseline, 6-month and 12-month information. We evaluated the overall regression performance measures by using Correlation Coefficient (CC) and root Mean Square Error (rMSE)as employed in [28]. The definitions of two measures are as

follows:  $rMSE(y, \hat{y}) = \sqrt{\frac{\|y - \hat{y}\|_2^2}{n}}$ , and  $Corr(y, \hat{y}) = \frac{cov(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}}$  where y is the ground truth of target

at a single time point and  $\hat{y}$  is the corresponding prediction by a prediction model.  $cov(y, \hat{y})$  is the covariance and  $\sigma_y$  and  $\sigma_{\hat{y}}$  are the standard deviations. We release our source code of MSS on Github (https://github.com/zj00377/multi-task-sparse-screening).

To validate the effectiveness and accuracy of the proposed method, we consider three different conditions: 1) To show the validity of the feature screening strategy, we performed MSS without prior feature screening and we used MS-N to indicate that no feature screening was involved in. 2) Another concern is whether our MSS has better power to predict clinical decline or whether it has better performance than single task sparse coding (SSC). To this end, we compare the MSS with online dictionary learning [14] referred as SSC for single-task based sparse coding. 3) We also want to know whether our group Lasso screening produces powerful features. We used SSC-N to refer single-task sparse coding without features screening. Besides, we compared with four methods: two single-task regression methods including ridge regression (Ridge) [10] and Lasso and two multi-task methods including Multi-Task Learning (MTL) [6] and  $L_{2,1}$  norm regularization with least square loss [1].

#### 3.2. Experimental Results

Table 2 shows regression performance on three different experiments, we marked the first and second rank of all sets of the experiments. Firstly, we can see the feature selection on the high-dimensional features before learning sparse codes is important because the MS-N and SSC-N have poorer performance compared to MSS and SSC. Secondly, for sparse coding and the dictionary learning method, our MSS and MS-N have better performance than SSC and SSC-N, which shows our multi-task dictionary learning is superior to single-task dictionary learning methods. Meanwhile, comparing MS-N to two single-task regression methods, Lasso and Ridge, our proposed method also performs better in term of correlation and rMSE. Thirdly, our proposed method has better performance than two multi-task regression methods MTL and L21, which may be because we selected the top-*p* features and used the common and individual features of different time points for different tasks.

Moreover, our proposed MSS method consistently outperformed the competing methods for the cases of different pairs of clinical labels. In the ADNI study, CU individuals and stable MCI patients (i.e. nondemented subjects) are less likely to have significant changes on the cognitive scores. We apply our models to the subgroup that consists of AD and MCI, MCI and CU patients, respectively, and the results show that our MSS still achieved the best results among all methods tested. For MCI vs. CU group, our method showed the best CC of 0.80 on MMSE and 0.77 on ADAS-cog and best rMSE of 4.27 for MMSE and 6.52 on ADAS-cog. In AD vs. MCI group, our method achieved the best CC of 0.76 on MMSE and 0.81 on ADAS-cog and best rMSE of 3.13 on MMSE and 6.51 on ADAS-cog. We show the scatter plots for the predicted values versus the actual values for MMSE and ADAS-Cog on the whole dataset in Fig. 2. We see the predicted values and actual clinical scores have a high correlation. Our experimental results validated the effectiveness of our MSS algorithm.

#### 4. CONCLUSION AND FUTURE WORK

We propose a novel algorithm Multi-task Sparse Screening. Our extensive experimental results demonstrate that the proposed framework is more effective than seven other standard methods. In future work, we will investigate our algorithm on visualizing our selected cortical thickness features.

#### Acknowledgments

The research was supported in part by NIH (R21AG049216, RF1AG051710, R01EB025032 and U54EB020403) and NSF (DMS-1413417 and IIS-1421165).

#### References

- 1. Argyriou A, Evgeniou T, Pontil M. Convex multi-task feature learning. Machine Learning. 2008; 73(3):243–272.
- Canutescu AA, Dunbrack RL. Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein science. 2003; 12(5):963–972. [PubMed: 12717019]
- Chung MK, Dalton KM, Shen L, Evans AC, Davidson RJ. Weighted fourier series representation and its application to quantifying the amount of gray matter. Medical Imaging, IEEE Transactions on. 2007; 26(4):566–581.
- Clarkson MJ, Cardoso MJ, Ridgway GR, Modat M, Leung KK, Rohrer JD, Fox NC, Ourselin S. A comparison of voxel and surface based cortical thickness estimation methods. Neuroimage. 2011; 57(3):856–865. [PubMed: 21640841]
- Cuingnet R, Gerardin E, Tessieras J, Auzias G, Lehéricy S, Habert MO, Chupin M, Benali H, Colliot O. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the adni database. neuroimage. 2011; 56(2):766–781. [PubMed: 20542124]
- Evgeniou T, , Pontil M. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining ACM; 2004Regularized multi-task learning; 109117
- 7. Fischl B. Freesurfer. Neuroimage. 2012; 62(2):774-781. [PubMed: 22248573]
- Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA. Sep; 2000 97(20):11050–11055. [PubMed: 10984517]
- 9. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. Mar.2003 3:1157–1182.
- 10. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970; 12(1):55–67.

- Jones SE, Buchbinder BR, Aharon I. Three-dimensional mapping of cortical thickness using Laplace's equation. Hum Brain Mapp. Sep; 2000 11(1):12–32. [PubMed: 10997850]
- 12. Lv J, Jiang X, Li X, Zhu D, Zhang S, Zhao S, Chen H, Zhang T, Hu X, Han J, Ye J, Guo L, Liu T. Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. IEEE Trans Biomed Eng. Apr; 2015 62(4):1120–1131. [PubMed: 25420254]
- Lv J, Lin B, Li Q, Zhang W, Zhao Y, Jiang X, Guo L, Han J, Hu X, Guo C, Ye J, Liu T. Task fMRI data analysis based on supervised stochastic coordinate coding. Med Image Anal. May.2017 38:1– 16. [PubMed: 28242473]
- 14. Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. Proceedings of the 26th annual international conference on machine learning, pages. Acm;2009 :689–696.
- 15. Olshausen BA, Field DJ. Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision research. 1997; 37(23):3311–3325. [PubMed: 9425546]
- Suk HI, Lee SW, Shen D, Initiative ADN, et al. Deep sparse multi-task learning for feature selection in alzheimers disease diagnosis. Brain Structure and Function. 2016; 221(5):2569–2587. [PubMed: 25993900]
- Thompson PM, Hayashi KM, Sowell ER, Gogtay N, Giedd JN, Rapoport JL, de Zubicaray GI, Janke AL, Rose SE, Semple J, Doddrell DM, Wang Y, van Erp TG, Cannon TD, Toga AW. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. Neuroimage. 2004; 23(Suppl 1):2–18.
- Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996:267–288.
- Wang G, Wang Y. Towards a Holistic Cortical Thickness Descriptor: Heat Kernel-Based Grey Matter Morphology Signatures. Neuroimage. Feb.2017 147:360–380. [PubMed: 28033566]
- 20. Wang J, Zhou J, Wonka P, Ye J. Lasso screening rules via dual polytope projection. Advances in Neural Information Processing Systems. 2013:1070–1078.
- Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2006; 68(1):49–67.
- 22. Zhang J, , Fan Y, , Li Q, , Thompson PM, , Ye J, , Wang Y. Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on IEEE; 2017Empowering cortical thickness measures in clinical diagnosis of alzheimer's disease with spherical sparse coding; 446450
- 23. Zhang J, , Shi J, , Stonnington C, , Li Q, , Gutman BA, , Chen K, , Reiman EM, , Caselli R, , Thompson PM, , Ye J. , et al. MICCAI Springer; 2016Hyperbolic space sparse coding with its application on prediction of Alzheimer's disease in mild cognitive impairment; 326334
- 24. Zhang J, Stonnington C, Li Q, Shi J, Bauer RJ III, Gutman BA, Chen K, Reiman EM, Thompson PM, Ye J, Wang Y. Applying sparse coding to surface multivariate tensor-based morphometry to predict future cognitive decline. IEEE International Symposium on BIOMEDICAL IMAGING: From Nano to Macro(ISBI) 2016. 2016
- 25. Zhang Jea. Patch-based sparse coding and multivariate surface morphometry for predicting amnestic mild cognitive impairment and alzheimers disease in cognitively unimpaired individuals. Alzheimer's & Dementia: The Journal of the Alzheimer's Association. 2016; 12(7):947.
- Zhang Q, Li B. Discriminative k-svd for dictionary learning in face recognition. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Jun.2010 :2691– 2698.
- Zhang T. Proceedings of the twenty-first international conference on Machine learning Vol. 116. ACM; 2004Solving large scale linear prediction problems using stochastic gradient descent algorithms.
- Zhou J, Liu J, Narayan VA, Ye J. Modeling disease progression via multi-task learning. Neuroimage. Sep.2013 78:233–248. [PubMed: 23583359]





The pipeline of our proposed Multi-task Sparse Screening (MSS) algorithm.



**Fig. 2.** The scatter plots of whole dataset on 24-month.

#### Table 1

Statistics of studied subjects.

Group	Baseline	6-month	12-month	24-month
AD	199	165	144	109
MCI	407	359	338	254
CU	230	213	201	178
Total	836	737	683	541

# Table 2

Comparison of regression performances with some other methods in terms of Correlation Coefficient (CC) and root Mean Square Error (rMSE) on 24months.

Zhang et al.

		łw	ole			MCI a	nd CU			AD an	d MCI	
Method	M	MSE	AD/	AS-cog	W	MSE	ADA	AS-cog	Μ	MSE	AD/	AS-cog
	cc	rMSE	CC	rMSE	cc	rMSE	CC	rMSE	cc	rMSE	CC	rMSE
MSS	0.82	3.40	0.79	6.39	0.80	4.27	0.77	6.52	0.76	3.13	0.81	6.51
N-SM	0.73	3.81	0.66	6.82	0.59	4.38	0.76	7.11	0.64	3.55	0.66	7.14
SSC	0.65	3.60	0.67	7.02	0.62	4.40	0.65	<u>6.94</u>	0.75	3.31	0.72	6.80
SSC-N	0.52	3.82	0.53	7.10	0.52	4.41	0.57	7.71	0.58	3.58	0.61	7.28
Lasso	0.58	4.22	0.52	7.79	0.59	5.55	0.46	8.14	0.47	4.59	0.56	7.93
Ridge	0.61	4.08	0.63	6.91	0.53	5.44	0.53	8.80	0.58	4.79	0.57	8.80
MTL	0.59	4.09	0.53	7.27	0.63	5.29	0.56	8.86	0.61	4.67	0.54	7.37
L21	0.60	4.14	0.57	7.10	0.61	4.97	0.60	7.62	0.60	4.08	0.62	6.79