Temporally-Constrained Group Sparse Learning for Longitudinal Data Analysis in Alzheimer's Disease

Biao Jie,

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Department of Computer Science and Technology, Anhui Normal University, Wuhu, 241000, China

Mingxia Liu,

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Jun Liu,

Imaging and Computer Vision Department, Siemens Corporate Research, Princeton, NJ 08540, USA

Daoqiang Zhang,

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Dinggang Shen, and

Biomedical Research Imaging Center and Department of Radiology, the University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, Republic of Korea

the Alzheimer's Disease Neuroimaging Initiative[1]

Abstract

Sparse learning has been widely investigated for analysis of brain images to assist the diagnosis of Alzheimer's disease (AD) and its prodromal stage, i.e., mild cognitive impairment (MCI). However, most existing sparse learning-based studies only adopt cross-sectional analysis methods, where the sparse model is learned using data from a single time-point. Actually, multiple time-points of data are often available in brain imaging applications, which can be used in some longitudinal analysis methods to better uncover the disease progression patterns. Accordingly, in

Correspondence to: Daoqiang Zhang; Dinggang Shen.

this paper we propose a novel temporally-constrained group sparse learning method aiming for longitudinal analysis with multiple time-points of data. Specifically, we learn a sparse linear regression model by using the imaging data from multiple time-points, where a *group regularization term* is first employed to group the weights for the same brain region across different time-points together. Furthermore, to reflect the smooth changes between data derived from adjacent time-points, we incorporate *two smoothness regularization terms* into the objective function, i.e., *one fused smoothness term* which requires that the differences between two successive weight vectors from adjacent time-points should be small, and another *output smoothness term* which requires the differences between outputs of two successive models from adjacent time-points should also be small. We develop an efficient optimization algorithm to solve the proposed objective function. Experimental results on ADNI database demonstrate that, compared with conventional sparse learning-based methods, our proposed method can achieve improved regression performance and also help in discovering disease-related biomarkers.

Index Terms

Sparse learning; longitudinal data analysis; temporal smoothness; group sparsity; Alzheimer's Disease (AD); Mild Cognitive Impairment (MCI)

I. Introduction

Alzheimer's disease (AD) is the most common form of dementia, which leads to progressive loss of memory and cognition function [1]. As a prodromal stage of AD, mild cognitive impairment (MCI) tends to progress to probable AD at a rate of approximately 10% to 15% per year. Thus, early and accurate diagnosis of AD/MCI is of vital importance for early treatment and possible delay of disease. At present, many pattern classification and regression methods have been proposed for AD or MCI diagnosis and prognosis by using biomarkers from different modalities, e.g., structural brain atrophies measured by magnetic resonance imaging (MRI) [2–5], metabolic brain alterations measured by fluorodeoxyglucose positron emission tomography (FDG-PET) [6, 7], and pathological amyloid depositions measured through cerebrospinal fluid (CSF) [3, 8–10], etc.

Recently, sparse learning techniques have attracted increasing attention due to their excellent performances in a series of neuroimaging applications on different modalities. For example, in a recent study [11], a voxel-based sparse classifier based on a L_1-norm regularized linear regression model, also known as the least absolute shrinkage and selection operator (LASSO) [12], was applied for classification of AD and MCI using MRI data, showing better performance than support vector machine (SVM) which is one of the state-of-the-art methods in brain imaging classification. In the literature, several other advanced sparse learning models (i.e., LASSO variants) have also been developed in neuroimaging applications. For example, researchers in [13] proposed to use elastic net [14] to identify both neuroimaging and proteomic biomarkers for AD and MCI based on MRI and proteomic data, and researchers in [15] proposed a generalized sparse regularization term with domain-specific knowledge for functional MRI (fMRI) based brain decoding. Recently, group LASSO [16] with a $L_{2,1}$-norm regularization term was used for jointly learning multiple

tasks including both classification tasks (e.g., AD/MCI vs. Normal Controls (NC)) and regression tasks (e.g., estimation of clinical cognitive scores) with MRI data in [17] and multimodal data (i.e., MRI, FDG-PET, and CSF) in [18], respectively. It is worth noting that both above-mentioned methods assume that multiple regression/classification variables are inherently related and determined by the same underlying AD pathology (i.e., the diseased brain regions). With such assumption, both regression tasks and classification tasks can be solved jointly.

Most existing sparse learning-based studies focus on using cross-sectional analysis methods, where only the data from a single time-point is used for model construction. However, multiple time-points of data are often available in some brain imaging applications, which can be used in longitudinal analysis to uncover the disease progression patterns. According to the number of time-points in the input and output of learning models, we can categorize the existing sparse models into four types: **1)** Single-time-point Input and Single-time-point Output (SISO), **2)** Single-time-point Input and Multi-time-points Output (SIMO), **3)** Multi-time-points Input and Single-time-point Output (MISO), and **4)** Multi-time-points Input and Multi-time-points Output (MIMO). In Fig. 1, we give an illustration for all these four learning problems, with more details given in next section.

In this paper, we address the above problems (i.e., SIMO, MISO and MIMO) by using sparse learning-based methods, where longitudinal data in *either* output *or* input (or both) can be employed. For that purpose, we develop a novel temporally-constrained group LASSO method, namely tgLASSO, where both the group regularizer and the temporal smoothness regularizer are incorporated into the objective function. Specifically, as in group LASSO (gLASSO), we first learn a sparse linear regression model by using data from each time-point, and further utilize a group regularizer to group the weights corresponding to the same brain region across different time-points together. In addition, to reflect the smooth changes between data from adjacent time-points, we also develop two smoothness regularizers: 1) a *fused smoothness term* (originated from fused LASSO [19, 20]), which requires the differences between two successive weight vectors from adjacent time-points to be small; 2) an *output smoothness term*, which requires that the differences between outputs of two successive models from adjacent time-points to be small. Furthermore, we develop an efficient optimization algorithm for solving the proposed problem. It is worth noting that, in order to capture temporal changing patterns of biomarkers in disease progression [21, 22], some researchers recently have explored to model disease progression via fused LASSO method [23]. However, different from their methods, our method incorporates a new smoothness regularizer (i.e., *output smoothness term*) into the objective function to capture the smoothness of outputs of two successive prediction models from adjacent time-points, which is one of our major contributions and was not investigated before [15, 23].

To validate the efficacy of our proposed method, we first perform a set of experiments (corresponding to the above MIMO, MISO and SIMO learning problems) on estimating clinical scores from MRI data on 445 subjects (including 91 AD, 202 MCI and 152 NC) from the Alzheime's Disease Neuroimaging Initiative (ADNI) database. Here, each subject has MRI data and the corresponding clinical scores, including Mini Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale - Cognitive Subscale

(ADAS-Cog), at 4 different time-points (i.e., baseline, 6-month, 12-month and 24-month). Then, we perform experiments on predicting MCI conversion from baseline MRI data using the biomarkers discovered in the first set of experiments. Our hypothesis is that, using longitudinal data, the proposed temporally-constrained group sparse learning method would perform better in discovering AD-related biomarkers and thus would achieve better performances in subsequent regression and classification tasks than the conventional methods.

II. Method

The data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.adni-info.org). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether the serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California–San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

A. Subjects

In the current study, we use all 445 ADNI subjects (including 91 AD, 202 MCI, and 152 NC) with all corresponding MRI data as well as two cognitive scores (MMSE and ADAS-Cog) at 4 different time-points (i.e., baseline, 6-month, 12-month and 24-month). In particular, for the MCI cohort, it contains 104 MCI converters (MCI-C) and 98 MCI non-converters (MCI-NC). In Table 1, we list the demographic characteristics of all studied subjects.

B. MRI Data Acquisition

In our previous works, we have described in detail on acquiring MRI data from ADNI [18, 24]. In short, structural MR scans were acquired from 1.5T scanners. Raw Digital Imaging and Communications in Medicine (DICOM) MRI scans were downloaded from the public ADNI site (adni.loni.usc.edu) [22], reviewed for quality, and automatically corrected for spatial distortion caused by gradient nonlinearity and B1 field inhomogeneity.

C. Image Analysis

In our experiments, we follow our previous works [18, 24] to perform image pre-processing for all MR images. Specifically, anterior commissure (AC) - posterior commissure (PC) correction is first performed on all images using MIPAV software (http://mipav.cit.nih.gov/index.php), followed by the N3 algorithm [25] which is used to correct the intensity inhomogeneity. Then, we perform skull-stripping on structural MR images, using a learning based method proposed in [26] that includes both brain surface extractor (BSE) [27] and brain extraction tool (BET) [28]. Next, the skull stripping results were further manually reviewed to ensure clean skull and dura removal. After the removal of cerebellum, the FSL package [29] is used to segment structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Afterwards, a fully automatic 4-dimensional atlas warping method called 4D HAMMER [30] is used to register all different time-point images of each subject to a template with 93 manually-labeled Regions of Interest (ROI) [31]. After registration, we can label all images based on the 93 labeled ROIs in the template. For each of the 93 ROIs in the labeled MR image, we compute the total GM volume of that region and use it as feature. In this study, we only use GM for feature extraction, because GM is the most affected by AD and also widely-used in the literature [11,18, 24, [32, 33]. Note that if there is no any GM in a specific region, the feature value for this region will be 0.

D. Temporally-constrained Group Sparse Learning (tgLASSO)

1) Four Different Learning Problems—Since AD (and its prodromal form, MCI) is a progressive neurodegenerative disease, we can obtain a series of temporal changes reflected in MRI data and clinical scores (e.g., MMSE and ADAS-Cog for AD) from studied subjects. In this work, we focus on estimating clinical scores by using MRI data. According to the number of time-points in both MRI data (input) and clinical scores (output), there are four different learning problems as shown in Fig. 1.

Specifically, as shown in Fig. 1(a), in the first learning problem (i.e., SISO), we want to estimate the clinical scores at a certain time-point, e.g., time-point 1 (baseline), by using imaging data from single time-point (e.g., baseline). Because both input and output are derived from a particular single time-point, the SISO problem contains no longitudinal information, and thus can be easily solved by the existing sparse learning methods (e.g., LASSO [12]). In the second learning problem, i.e., SIMO shown in Fig. 1(b), the clinical scores at each time-point (ranging from 1 to T) can be estimated by using imaging data from single time-point (e.g., baseline). Similarly, in the third learning problem, i.e., MISO shown in Fig. 1(c), we aim to estimate clinical scores at time-point T, by using imaging data from all time-points (from 1 to T). Finally, in the fourth learning problem, i.e., MIMO shown in Fig. 1(d), we want to estimate clinical scores at each time-point j ($j = 1, …, T$), by using imaging data from its corresponding time-point j. It is worth noting that MIMO will degenerate to be SIMO if we set the input (imaging data) $\mathbf{x}_j = \mathbf{x}_1$ ($j = 1, …, T$). Similarly, MIMO will degenerate to be MISO if we set the output (clinical score) $z_j = z_1$ (for $j = 1, …, T$). In the following, we will develop a new temporally-constrained group sparse learning (tgLASSO) method for solving the MIMO (as well as MISO and SIMO) problems.

2) Objective Function—Assume that we have N training subjects, and each subject has imaging data derived from T different time-points, represented as $\{\mathbf{x}_{i1}, ..., \mathbf{x}_{ij}, ..., \mathbf{x}_{iT}\}$ where $\mathbf{x}_{ij} \in \Re^{1 \times D}$ is a D-dimensional row vector. Denote $\mathbf{X}_j = [\mathbf{x}_{1j}; ...; \mathbf{x}_{ij}; ...; \mathbf{x}_{Nj}]$ ($\in \Re^{N \times D}$) and $\mathbf{y}_j$ ($\in \Re^N$) as the training data matrix (input) and the corresponding clinical scores at the j-th time-point, respectively. We use the linear model to estimate the clinical score from the imaging data $\mathbf{x}$ at the j-th time-point as $h_j(\mathbf{x}) = \mathbf{x}\mathbf{w}_j$, where the feature weight vector $\mathbf{w}_j \in \Re^D$. Let $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_j, ..., \mathbf{w}_T]$ ($\in \Re^{D \times T}$) denote the weight vector matrix for all T learning tasks, with each column vector corresponding to one specific task. The objective function of our temporally-constrained group LASSO (tgLASSO) can be defined as follows:

$$\min_{\mathbf{W}} J(\mathbf{W}) = \frac{1}{2}\sum_{j=1}^{T} \|\mathbf{y}_j - \mathbf{X}_j\mathbf{w}_j\|_2^2 + R_g(\mathbf{W}) + R_s(\mathbf{W}) \tag{1}$$

where $R_g(\mathbf{W})$ and $R_s(\mathbf{W})$ are the *group regularization term* and the *smoothness regularization term*, respectively. Specifically, the *group regularization term* is defined as below:

$$R_g(\mathbf{W}) = \lambda_1 \|\mathbf{W}\|_{2,1} = \lambda_1 \sum_{d=1}^{D} \|\mathbf{w}^d\|_2 \tag{2}$$

Here, $\mathbf{w}^d$ is the d-th row vector of $\mathbf{W}$. It is worth noting that the use of L_2-norm on row vectors encourages the weights corresponding to the d-th feature across multiple time-points to be grouped together, and the further use of L_1-norm tends to jointly select features based on the strength of T time-points. The regularization parameter λ_1 controls the group sparsity of the linear models.

In addition, the *smoothness regularization term* is defined as follows:

$$R_s(\mathbf{W}) = \lambda_2 \sum_{j=1}^{T-1} \|\mathbf{w}_j - \mathbf{w}_{j+1}\|_1 + \lambda_3 \sum_{j=1}^{T-1} \|\mathbf{X}_j\mathbf{w}_j - \mathbf{X}_{j+1}\mathbf{w}_{j+1}\|_2^2 \tag{3}$$

where the first term in Eq. (3) is called the *fused smoothness term* which originates from fused LASSO [19, 20], and it constrains the differences between two successive weight vectors from adjacent time-points to be small. Due to the use of L_1-norm in the fused smoothness term that encourages the sparsity on difference of weight vectors, there will be a lot of zero components in the weight difference vectors. In other words, a lot of components from adjacent weight vectors will be identical because of using the fused smoothness regularization. In our study, we will select those features with non-zero weights for subsequent regression or classification tasks. The second term in Eq. (3) is called the *output smoothness term* that requires the differences between outputs of two successive models from adjacent time-points to be small as well. The regularization parameters λ_2 and λ_3 balance the relative contributions of the two terms and also control the smoothness of the linear models. It is easy to know that when both λ_2 and λ_3 are zero, our method will

degenerate to group LASSO [16]. In the next section, we will develop an efficient optimization algorithm to solve the objective function defined in Eq. (1).

3) Efficient Iterative Optimization Algorithm—It is worth noting that the above defined objective function is the first time to simultaneously include both the group and the (fused plus output) smoothness regularizations, which has not been studied before. In the Appendix, we have developed an efficient algorithm to solve the objective function. Here, the key idea is to separate the objective function into the smooth term and the non-smooth term and then use the iterative projected gradient descent approach [34], which combines the gradient descent and proximal mapping to update the iterations for final solution. For more details, please refer to the Appendix.

E. Validation

In our experiments, each of the 445 subjects has the corresponding MRI data and clinical scores (including MMSE and ADAS-Cog) at 4 different time-points, i.e., baseline, 6-month (M06), 12-month (M12), and 24-month (M24). To evaluate the efficacy of our proposed tgLASSO method, we compare our method with existing popular sparse learning methods, including LASSO and group LASSO (gLASSO). In addition, we perform two sets of experiments on longitudinal data from ADNI database, i.e., estimating clinical scores and predicting MCI conversion.

In the first set of experiments, we estimate the clinical scores (i.e., MMSE and ADAS-Cog) from MRI data in three different problem settings, i.e., MIMO, MISO and SIMO, which involve the use of different types of longitudinal information as shown in Fig. 1. To evaluate the regression performance of our proposed method, we use a 10-fold cross-validation strategy by computing the Pearson's correlation coefficient between the predicted and the actual clinical scores, and also computing the root mean square error (RMSE) between the predicted and the actual clinical scores. Specifically, the whole set of samples are first partitioned into 10 subsets (each subset with a roughly equal size). Then, the samples within one subset are selected as the testing data, and samples in the other 9 subsets are combined as the training data. This process is repeated for 10 times independently. In the experiment, we compute both the average value of the Pearson's correlation coefficients and the average value of the RMSEs in all 10-fold cross-validation as the final results.

In the second set of experiments, we predict the MCI conversion from baseline MRI data using the biomarkers discovered by tgLASSO under the MIMO problem setting. Specifically, we first perform feature selection by using our proposed tgLASSO method on longitudinal training data (with MRI data and corresponding clinical scores of MMSE and ADAS-Cog at 4 time-points, i.e., baseline, M06, M12 and M24), in order to select the most discriminative brain regions. Then, a support vector machine (SVM) classifier is constructed based on the baseline training data (with MRI data and corresponding class labels at baseline time-point) with the selected brain regions for the prediction of MCI. Similar to the first set of experiments, we also adopt a 10-fold cross-validation strategy to evaluate the classification performance by three statistical measures, including the classification accuracy (i.e., the proportion of MCI subjects correctly classified), the sensitivity (i.e., the proportion

of MCI converters correctly classified), and the specificity (i.e., the proportion of MCI non-converters correctly classified). Besides, we also calculate the area under receiver operating characteristic (ROC) curve (AUC) as performance measure.

In our experiments, for each extracted feature value, we perform the following feature normalization, i.e., subtracting the mean and then dividing the standard deviation (of all training subjects). For all respective methods, another round of cross-validation on the training data is used for determining the values for parameters (e.g., λ_1, λ_2 and λ_3). Specifically, we, respectively, vary the values of λ_1, λ_2 and λ_3 within the range of {0.25 0.2 0.15 0.1 0.09 0.08 0.07 0.06 0.05 0.04 0.03 0.02 0.01} and compute the prediction performance via the inner 10-fold cross-validation on the training subjects. The parameter values with the best performance (on the inner 10-fold cross-validation) will be used for prediction of the unknown subjects (i.e., testing subjects in each (outer) cross validation). The linear SVM is implemented using LIBSVM toolbox with the default parameter value (i.e., C=1) [35]. It is worth noting that the cross-validation on training subjects is only used to determine the optimal parameter values.

III. Results

A. Estimating Clinical Scores

In this group of experiments, we first estimate two regression variables (i.e., MMSE and ADAS-Cog) in three learning problems (i.e., MIMO, MISO and SIMO) at four time-points, respectively. Before showing the estimation results, we first plot the average longitudinal changes of clinical scores from baseline to M24 in different kinds of subjects (i.e., AD, MCI-C, MCI-NC and NC) in Fig. 2. Fig. 3 and Fig. 4, respectively, show the comparison of correlation coefficients and RMSEs achieved by LASSO, gLASSO and tgLASSO in estimating the clinical scores of MMSE and ADAS-Cog under different longitudinal analysis settings (i.e., MIMO, MISO and SIMO). In addition, for better comparison, Fig. 3 and Fig. 4 also give the estimation results of the proposed tgLASSO method with $\lambda_3 = 0$.

It can be seen from Fig. 2 that, as disease progresses, the cognitive performance of the AD and MCI-C subjects decline gradually as reflected by the decreased MMSE and increased ADAS-Cog scores, while the cognitive performance of the MCI-NC and NC subjects declines much slower than those of the AD and MCI-C subjects.

As can be seen from both Fig. 3 and Fig. 4, our proposed tgLASSO method consistently outperforms other methods in estimating clinical scores. Specifically, tgLASSO achieves the average (i.e., across four time-points) correlation coefficients of 0.613, 0.657 and 0.594 for estimating MMSE scores in the MIMO, MISO and SIMO learning problems, respectively, while the best average correlation coefficients of the competing methods are 0.607, 0.647 and 0.589, respectively. Similarly, for estimating ADAS-Cog scores, tgLASSO achieves the average correlation coefficients of 0.639, 0.676 and 0.623 in the three learning problems, while the best average correlation coefficients of the competing methods are 0.635, 0.665 and 0.622, respectively. Also, tgLASSO achieves the average (i.e., across four time-points) RMSEs of 2.988, 2.845 and 3.022 for estimating MMSE scores in the MIMO, MISO and SIMO learning problems, respectively, while the best average RMSEs of the competing

methods are 3.011, 2.867 and 3.032, respectively. Similarly, for estimating ADAS-Cog scores, tgLASSO achieves the average RMSEs of 6.080, 5.853 and 6.181 in the three learning problems, while the best average RMSEs of the competing methods are 6.107, 5.925 and 6.179, respectively. Moreover, we perform, respectively, the paired t-test between correlation coefficients of the proposed tgLASSO method and the correlation coefficients of the competing methods (i.e., gLASSO and LASSO), and between RMSEs of the proposed tgLASSO method and RMSEs of the competing methods (i.e., gLASSO and LASSO). The results in both tests show that the proposed tgLASSO method is significantly better than gLASSO and LASSO methods in three learning problems (i.e., with all p-values less than 0.05). Also, we perform the paired t-test over squared residuals between the proposed tgLASSO method and each competing method, and show results in Table 2. From Table 2, we can see that most of p-values in three learning problems are also less than 0.05. These results validate the efficacy of our proposed method in jointly estimating the clinical scores based on longitudinal analysis. Besides, both Fig. 3 and Fig. 4 also indicate that estimating later time-point scores often achieves better performance than estimating earlier time-point scores. This may be because the relationship between imaging features and clinical scores becomes much stronger with progress of disease or brain aging, e.g., atrophy in the brain is more obvious in advanced disease and thus the related features are more distinctive and correlated to the clinical scores. In addition, from Fig. 3 and Fig. 4, we can further observe that the prediction results of tgLASSO with $\lambda_3 = 0$ are worse than tgLASSO, while better than any other competing methods. These further show the advantage of using two smoothness items (i.e., fused smoothness and output smoothness). Besides, the prediction results of MISO learning model are usually superior to those of MIMO learning model in the first three time-points, which indicates that the clinical scores at last time-point (i.e., M24) may help induce more important features (i.e., brain atrophy regions) for prediction. It is worth noting that MIMO and MISO are the two different types of learning model, and should be used for different longitudinal analysis settings, respectively. Specifically, MIMO is a *multi-time-points* input and *multi-time-points* output learning model, which can be used to estimate clinical scores at multiple time-points by using imaging data from the same time-point. On the other hand, MISO is a *multi-time-points* input and *single-time-point* output learning model, which can be used to predict clinical score at the last time-point by using imaging data from all previous time-points.

Fig. 5 shows feature weight maps of three different methods in a certain cross-validation case when estimating MMSE scores for MIMO learning problem. In addition, both Fig. S1 and Fig. S2 in Supplementary Material also show the corresponding feature weight maps of different methods for the MISO and SIMO learning problems, respectively. Here, it is worth noting that both gLASSO and tgLASSO jointly learn weight vectors for the four time-points, while LASSO learns each weight vector independently for each time-point.

As can be seen from Fig. 5 and Figs. S1–S2 in Supplementary Material, due to the use of group regularization, gLASSO and tgLASSO obtain more grouped weights across different time-points than LASSO. Furthermore, due to the use of smoothness regularization, tgLASSO achieves more smooth weights across different time-points than other two methods. These properties are helpful to discover those intrinsic biomarkers relevant to brain diseases. For example, as shown in Fig. 5, both left and right hippocampal regions, the well-

known AD-relevant biomarkers, are detected by tgLASSO, while only the left hippocampal region is detected by other two methods.

B. Predicting MCI Conversion

In this set of experiments, we predict the future conversion of MCI patients based on baseline data, using the biomarkers discovered in the first set of experiments corresponding to MIMO learning problem. Here, for both joint learning methods (i.e., gLASSO and tgLASSO), we first learn the corresponding gLASSO and tgLASSO models using longitudinal training MRI data (with longitudinal MMSE/ADAS-Cog scores) at four time-points to select the important brain regions (with respect to MMSE/ADAS-Cog scores), and then train SVM classifiers on the baseline training MRI data with above-selected brain regions, respectively. On the other hand, since LASSO cannot deal with longitudinal data, we learn a LASSO model using only the baseline training MRI data (with baseline MMSE/ADAS-Cog scores) to select the important brain regions, and then train a SVM classifier on the baseline training MRI data with above-selected brain regions. Table 3 gives the results of different methods in predicting the MCI conversion.

As can be seen from Table 3, our proposed tgLASSO method consistently outperforms the other two methods in all performance measures. Specifically, our proposed method achieves a classification accuracy of 75.7%, a sensitivity of 72.9%, and a specificity of 82.0% when learning the tgLASSO model with guidance from MMSE clinical score, while achieves a classification accuracy of 74.7%, a sensitivity of 73.9%, and a specificity of 76.1% when learning the tgLASSO model with guidance from ADAS-Cog clinical score. These results are consistently better than other methods on each performance measure. In addition, Table 3 also indicates that, by using longitudinal data, the gLASSO method can obtain better performance than the LASSO method, but it is still inferior to our proposed method (tgLASSO).

C. The Most Important Brain Regions

In this subsection, we investigate the top selected brain regions by our proposed tgLASSO method in the MIMO learning problem. Since the selected brain regions are different in each 10-fold cross-validation, we chose the brain regions with top occurrence frequency in all cross-validation as the most important brain regions, when learning models using the clinical scores of MMSE and ADAS-Cog, respectively. Table 4 lists the 16 most important brain regions detected by the proposed tgLASSO method. Also, in Table 4, we give the average of each selected ROI's weights across all folds and time-points, as well as the corresponding standard deviation. The result shows that the most important regions obtained by our method include *hippocampal, amygdala, temporal pole, uncus* and *middle temporal regions*, which are consistent with previous studies. In addition, from Table 4, we can see the obtained standard deviations are very small, indicating that the weight maps of each selected ROI across different time-points are very smooth. This furthermore shows the advantage of using our proposed smoothness regularizations. For visual inspection, in Fig. 6, we also highlight those selected brain regions listed in Table 4.

IV. Discussion

In this paper we have proposed a novel temporally-constrained group sparse learning method for longitudinal analysis with multiple time-points of data. Our proposed method has been validated on 445 subjects (including 91 AD, 202 MCI and 152 NC) with cognitive scores at 4 different time-points (i.e., baseline, 6-month, 12-month and 24-month) through two sets of experiments, i.e., **1)** estimating MMSE and ADAS-Cog scores at each time point in three learning problems (including SIMO, MISO and MIMO), and **2)** predicting future conversion of MCI subjects using baseline data. The experimental results show that our proposed method can *not only* significantly improve regression performance *but also* help in discovering disease-related biomarkers useful for MCI conversion prediction, compared with the conventional sparse learning methods.

A. Significance of Results

Recently, sparse learning methods have been widely used for diagnosis of AD/MCI. However, multiple time-points of data, which are often available and may potentially further improve performance, are not fully utilized in existing methods. Our study demonstrated that, by embedding the longitudinal information of data, our proposed method can achieve better performance in estimating the clinical scores as well as predicting the MCI conversion. It is worth noting that, some recent works, e.g., methods in [36] and [23], also adopted the sparse feature learning method for analyzing longitudinal data. Different from both above-mentioned methods, we propose to use both the group and the (fused + output) smoothness regularizations in sparse learning to better reflect the longitudinal change patterns of the brain with the progression of disease. The experimental results also show the advantage of our proposed method compared with existing sparse learning methods.

The brain regions selected by our proposed method are known to be related to the AD by many studies using group comparison methods, which include *hippocampal* [37–41], *amygdala* [38], *temporal pole* [42], *uncus* [43] and *middle temporal regions* [38, 39]. For example, it has been reported that there exists a strong correlation between hippocampal volume and dementia severity [44].

B. Predicting Clinical Scores and MCI Conversion

A lot of works have studied the relationship between cognitive scores and imaging markers with neuroimaging data [45–47]. A variety of high-dimensional regression methods have been used for estimating or predicting clinical scores for AD/MCI subjects, based on the neuroimaging data. For example, [48] used a principal component analysis (PCA) based model to predict the 12-month change in MMSE score based on the baseline MRI data of 49 MCI subjects. In [49], researchers used a joint Bayesian classifier by sharing the same hyper-parameters for model parameters to estimate the MMSE and ADAS-Cog scores from the ADNI baseline MRI data of 264 subjects. Recently, in [50, 51], researchers used the sparse learning methods to predict scores of MMSE and ADAS-cog based on MRI data from ADNI dataset. Table 5 summarizes the results of these methods. As can be seen from Table 5, our proposed method achieves comparable results in estimating clinical scores of MMSE and ADAS-Cog, compared with those recently published results in AD/MCI studies.

In addition, MCI is a prodromal stage of AD, with high likelihood of conversion to AD. There was a strong association between the structural pattern of atrophy identified in AD and the pattern of atrophy found in MCI converters. It has been proposed in [52] that there is a long preclinical phase of AD with no symptoms of cognitive dysfunction but with an ongoing AD pathology, and recent study [53] has suggested that the structural changes detected by MRI may be evident even ten years before clinical diagnosis of AD. Therefore, a lot of recent studies in early diagnosis of AD have been focused on predicting the conversion of MCI to AD, i.e., identifying the MCI converters (MCI-C) from MCI non-converters (MCI-NC) [18, 54–60]. For example, in a recent work [54], an accuracy between 67.4% and 74.7% was reported on 21 MCI-C and 98 MCI-NC subjects using MRI data. More recently, in [60], an accuracy of 0.68 was reported on 97 MCI-C and 93 MCI-NC subjects based on MRI data in the ADNI dataset. In contrast, our method achieves the accuracy between 74.7% and 75.7% on 104 MCI-C and 98 MCI-NC subjects from ADNI, which are comparable to the best results reported in those recent studies.

C. Effect of Parameters

In the objective function of our proposed tgLASSO method, there are two regularization terms, i.e., the *group regularization* term and *smoothness regularization* term, where the second one consists of two parts including the *fused smoothness* term and the *output smoothness* term. The regularization parameters λ_1, λ_2 and λ_3 balance the relative contribution of these regularization terms. Here, the larger λ_1 value means few features preserved for estimating the clinical scores due to the imposed 'group sparsity' constraint via the $L_{2,1}$-norm. The parameters λ_2 λ_3 and control the contributions of two smooth regularization items.

To investigate the effect of two smoothness regularization terms on the performance of our proposed method, we first fix the value of (i.e., setting λ_1 to 0.1 and 0.07 for estimating regression scores of MMSE and ADAS-Cog, respectively), and test the values of λ_2 and λ_3 from a set of [0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.15 0.20 0.25]. Fig. 7 gives the regression performance of MMSE and ADAS-Cog scores using different values of λ_2 and λ_3 in our proposed method for the MIMO learning problem. It is worth noting that, for each plot, the bottom row and the right column denote the results when using only the output smoothness regularization ($\lambda_2 = 0$) or only the fused smoothness regularization ($\lambda_3 = 0$), respectively. As we can see from Fig. 7, the larger values (i.e., better estimation performance) mainly focus on the inner intervals of the square, which indicates the effectiveness of combining two smoothness regularization terms for predicting clinical scores. This also implies that each term is indispensable for achieving good performance.

Furthermore, we test the performance of our proposed method with different values of λ_1. Specifically, we vary the value of λ_1 from the range of [0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.10 0.15 0.20 0.25], and compute the prediction results of our proposed tgLASSO with the optimal λ_2 and λ_{3s} obtained by using the inner cross-validation on training data. Fig. 8 graphically shows the obtained results for the MIMO learning problem. For comparison, we also give the prediction performance of gLASSO method, where only the group regularization term is included (i.e., setting both λ_2 and λ_3 to zero). It is worth noting

that, for each plot, the leftmost points denote the results with no feature selection (i.e., using all features for estimating clinical scores). As can be seen from Fig. 8, for estimating two kinds of clinical scores (i.e., MMSE and ADAS-Cog), our proposed tgLASSO method consistently achieves better performance than gLASSO method for all λ_1 values. Specifically, at each time point, our method yields relatively high correlation coefficients for all λ_1 values (except for zero), showing its robustness to regularization parameter λ_1 and also the advantage of including the smoothness regularization terms.

D. Limitations

The current study is limited by the following factors. First, our proposed method performs prediction based on longitudinal data and thus requires each subject having the corresponding data, i.e., MRI data and corresponding clinical scores, at each time point, which limits the size of subjects that can be used for study. For example, there are more than 400 MCI subjects in the ADNI dataset, while there are only 202 MCI subjects with MRI data and corresponding MMSE and ADAS-Cog scores at multiple time-points (including baseline, 6-month, 12-month and 24-month). Second, there also exist other modalities of data, e.g., PET and CSF. However, since the number of subjects with all modality data (including MRI, PET and CSF) is too small for reasonable learning, the current study does not consider using multi-modality data. In the future work, we will study how to utilize subjects with incomplete multi-modality data (i.e., missing of certain modality data) for further performance improvement. Third, selecting important and stabilized features (i.e., brain regions) and determining the optimal regularization parameters are the two important problems for sparse-based methods. However, some brain regions reported in the literature, such as the precuneus and (posterior) cingulate, the entorhinal, perirhinal and parahippocampal regions, and the lateral ventricles, were not found by our proposed method. In future work, we will explore some techniques, such as performing more cross-validations on the training subjects to select most frequently occurring features as stabilized features, and also using Bayesian models (instead of the grid-searching approach) to determine the optimal parameter values, to address the above problems. Finally, during image pre-processing, the brain region parcellation is also a very important step for the subsequent feature extraction and prediction. Also, previous studies have demonstrated that other methods, such as voxel-based methods, still obtained comparable results to region-based methods [55]. But this paper does not analyze the impact of different brain parcellation atlases on regression performance.

V. Conclusions

In this paper, we propose a new sparse learning method called tgLASSO for longitudinal data analysis with multiple time-points of data, which is different from most existing sparse learning methods that focus on cross-sectional data analysis such as using only the data from single time-point. Our methodological contributions include two parts, i.e., **1)** we propose to simultaneously use group and (fused + output) smoothness regularizations in sparse learning models; and **2)** we develop an efficient iterative optimization algorithm for solving the new objective function. Experimental results on estimating clinical scores from imaging data at

multiple time-points illustrate the advantages of our method over existing sparse methods in both regression performance and ability in discovering disease-related imaging biomarkers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Ron B, et al. Forecasting the global burden of Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association. 2007; 3(3):186–191.

2. de Leon MJ, et al. Longitudinal CSF isoprostane and MRI atrophy in the progression to AD. J Neurol. 2007 Dec; 254(12):1666–1675. [PubMed: 17994313]

3. Fjell AM, et al. CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. J Neurosci. 2010 Feb 10; 30(6):2088–2101. [PubMed: 20147537]

4. Du AT, et al. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain. 2007 Apr; 130(Pt 4):1159–1166. [PubMed: 17353226]

5. McEvoy LK, et al. Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology. 2009 Apr; 251(1):195–205. [PubMed: 19201945]

6. De Santi S, et al. Hippocampal formation glucose metabolism and volume losses in MCI and AD. Neurobiology of Aging. 2001 Jul-Aug;22(4):529–539. [PubMed: 11445252]

7. Morris JC, et al. Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease. Archives of Neurology. 2001 Mar 1; 58(3):397–405. 2001. [PubMed: 11255443]

8. Shaw LM, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009 Apr; 65(4):403–413. [PubMed: 19296504]

9. Mattsson N, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. Jama. 2009 Jul 22; 302(4):385–393. [PubMed: 19622817]

10. Bouwman FH, et al. Longitudinal changes of CSF biomarkers in memory clinic patients. Neurology. 2007 Sep 4; 69(10):1006–1011. [PubMed: 17785669]

11. Liu M, et al. Ensemble sparse classification of Alzheimer's disease. Neuroimage. 2012 Apr 2; 60(2):1106–1116. [PubMed: 22270352]

12. Tibshirani R. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B-Methodological. 1996; 58(1):267–288.

13. Shen L, et al. Identifying Neuroimaging and Proteomic Biomarkers for MCI and AD via the Elastic Net. Lect Notes Comput Sci. 2011; 7012:27–34.

14. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2005; 67:301–320.

15. Ng B, Abugharbieh R. Generalized sparse regularization with application to fMRI brain decoding. Inf Process Med Imaging. 2011:612–623. [PubMed: 21761690]

16. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2006; 68:49–67.

17. Wang H, et al. Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. Med Image Comput Comput Assist Interv. 2011:115–123. [PubMed: 22003691]

18. Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage. 2012 Jan 16; 59(2):895–907. [PubMed: 21992749]

19. Liu, J., et al. An efficient algorithm for a class of fused lasso problems; Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining; 2010. p. 323-332.

20. Tibshirani R, et al. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society Series B-Statistical Methodology. 2005; 67:91–108.

21. Caroli A, Frisoni GB. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. Neurobiol Aging. 2010 Aug; 31(8):1263–1274. [PubMed: 20538373]

22. Jack CR Jr, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008 Apr; 27(4):685–691. [PubMed: 18302232]

23. Zhou J, et al. Modeling disease progression via multi-task learning. Neuroimage. 2013 Sep. 78:233–248. [PubMed: 23583359]

24. Zhang D, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011 Apr 1; 55(3):856–867. [PubMed: 21236349]

25. Sled JG, et al. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging. 1998 Feb; 17(1):87–97. [PubMed: 9617910]

26. Shi F, et al. LABEL: Pediatric brain extraction using learning-based meta-algorithm. Neuroimage. 2012 Sep; 62(3):1975–1986. [PubMed: 22634859]

27. Shattuck DW, et al. Magnetic resonance image tissue classification using a partial volume model. Neuroimage. 2001 May; 13(5):856–876. [PubMed: 11304082]

28. Smith SM. Fast robust automated brain extraction. Human Brain Mapping. 2002 Nov; 17(3):143–155. 2002. [PubMed: 12391568]

29. Zhang Y, et al. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Transactions on Medical Imaging. 2001; 20(1): 45–57. [PubMed: 11293691]

30. Shen D, et al. 4D HAMMER Image Registration Method for Longitudinal Study of Brain Changes. Proceedings of the Human Brain Mapping. 2003:1–8.

31. Kabani N, et al. A 3D atlas of the human brain. Neuroimage. 1998; 7(4):S717.

32. Liu M, et al. View-centralized multi-atlas classification for Alzheimer's disease diagnosis. Hum Brain Mapp. 2015 May; 36(5):1847–1865. [PubMed: 25624081]

33. Jie B, et al. Manifold regularized multitask feature learning for multimodality disease classification. Hum Brain Mapp. 2015 Feb; 36(2):489–507. [PubMed: 25277605]

34. Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM J. Img. Sci. 2009; 2(1):183–202.

35. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. 2001 Available: http://www.csie.ntu.edu.tw/~cjlin/.

36. Zhang D, Shen D. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS One. 2012; 7(3):e33182. [PubMed: 22457741]

37. Chetelat G, et al. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. Neuroreport. 2002 Oct 28; 13(15):1939–1943. 2002. [PubMed: 12395096]

38. Misra C, et al. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage. 2009 Feb 15; 44(4): 1415–1422. [PubMed: 19027862]

39. Fox NC, Schott JM. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet. 2004 Jan 31; 363(9406):392–394. [PubMed: 15074306]

40. Chincarini A, et al. Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease. Neuroimage. 2011 Sep 15; 58(2):469–480. [PubMed: 21718788]

41. La Joie R, et al. Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. Neuroimage Clin. 2013; 3:155–162. [PubMed: 24179859]

42. Convit A, et al. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiology of Aging. 2000; 21(1):19–26. 2000/0. [PubMed: 10794844]

43. Devanand DP, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. Neurology. 2007 Mar 13; 68(11):828–836. [PubMed: 17353470]

44. Csernansky JG, et al. Correlations between antemortem hippocampal volume and postmortem neuropathology in AD subjects. Alzheimer Disease & Associated Disorders. 2004 Oct-Dec;18(4): 190–195. [PubMed: 15592129]

45. Thompson PM, et al. Mapping hippocampal and ventricular change in Alzheimer disease. Neuroimage. 2004 Aug; 22(4):1754–1766. [PubMed: 15275931]

46. Stonnington CM, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage. Jul 15; 51(4):1405–1413.

47. Frisoni GB, et al. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology. 2010 Feb; 6(2):67–77. [PubMed: 20139996]

48. Duchesne S, et al. Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. Neuroimage. 2009 Oct 1; 47(4):1363–1370. [PubMed: 19371783]

49. Fan, Y., et al. Joint Estimation of Multiple Clinical Variables of Neurological Diseases from Imaging Patterns; 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2010. p. 852-855.

50. Wan J, et al. Identifying the neuroanatomical basis of cognitive impairment in Alzheimer's disease by correlation- and nonlinearity-aware sparse Bayesian learning. IEEE Trans Med Imaging. 2014 Jul; 33(7):1475–1487. [PubMed: 24710828]

51. Yan J, et al. Cortical surface biomarkers for predicting cognitive outcomes using group l2,1 norm. Neurobiol Aging. 2015 Jan; 36(Suppl 1):S185–S193. [PubMed: 25444599]

52. Jack CR Jr, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010 Jan; 9(1):119–128. [PubMed: 20083042]

53. Tondelli M, et al. Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. Neurobiology of Aging. 2012 Apr; 33(4):e25–e36.

54. Aguilar C, et al. Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. Psychiatry Res. 2013 May 30; 212(2):89–98. [PubMed: 23541334]

55. Cuingnet R, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. Neuroimage. 2011 May 15; 56(2): 766–781. [PubMed: 20542124]

56. Duchesne S, Mouiha A. Morphological Factor Estimation via High-Dimensional Reduction: Prediction of MCI Conversion to Probable AD. Int J Alzheimers Dis. 2011; 2011:1–8.

57. Davatzikos C, et al. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging. 2011 Dec; 32(12):2322 e19–2322 e27.

58. Lehmann M, et al. Visual ratings of atrophy in MCI: prediction of conversion and relationship with CSF biomarkers. Neurobiol Aging. 2013 Jan; 34(1):73–82. [PubMed: 22516280]

59. Cho Y, et al. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. Neuroimage. 2012 Feb 1; 59(3): 2217–2230. [PubMed: 22008371]

60. Liu X, et al. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. Neuroimage. 2013 Dec.83:148–157. [PubMed: 23792982]
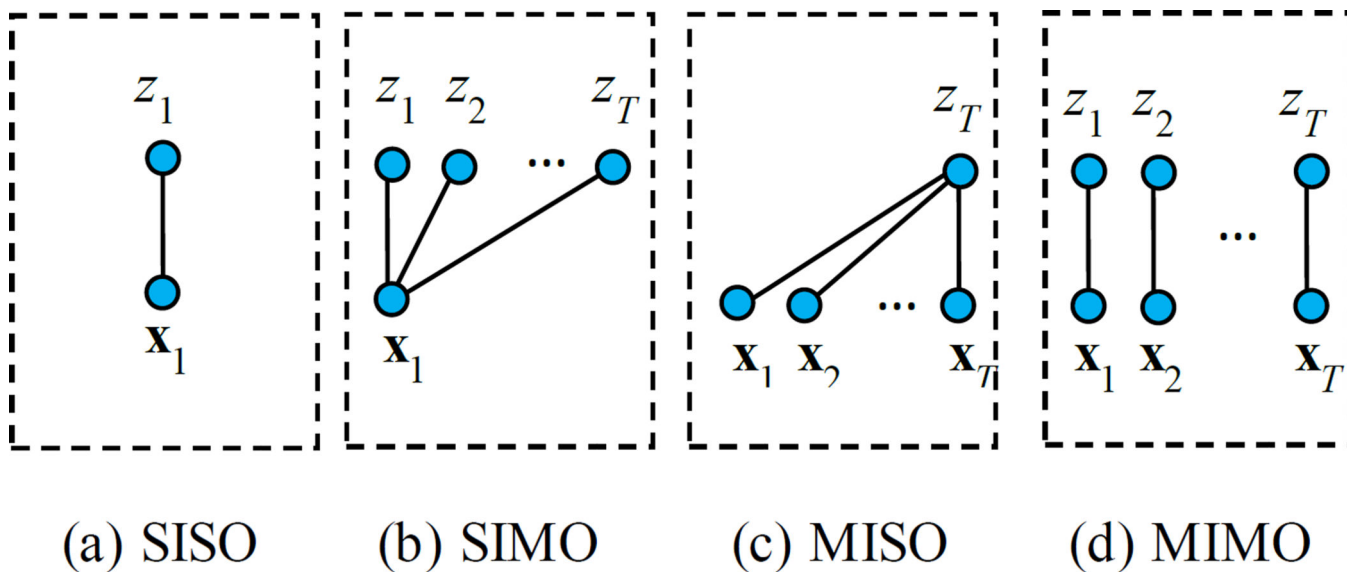
Fig. 1.
Illustration of four different learning problems: (a) Single-time-point Input and Single-time-point Output (SISO), (b) Single-time-point Input and Multi-time-points Output (SIMO), (c) Multi-time-points Input and Single-time-point Output (MISO), and (d) Multi-time-points Input and Multi-time-points Output (MIMO). Here, each edge represents a model, and the nodes x_j and z_j denote the imaging data (input) and clinical score (output) at the j-th time-point, respectively.
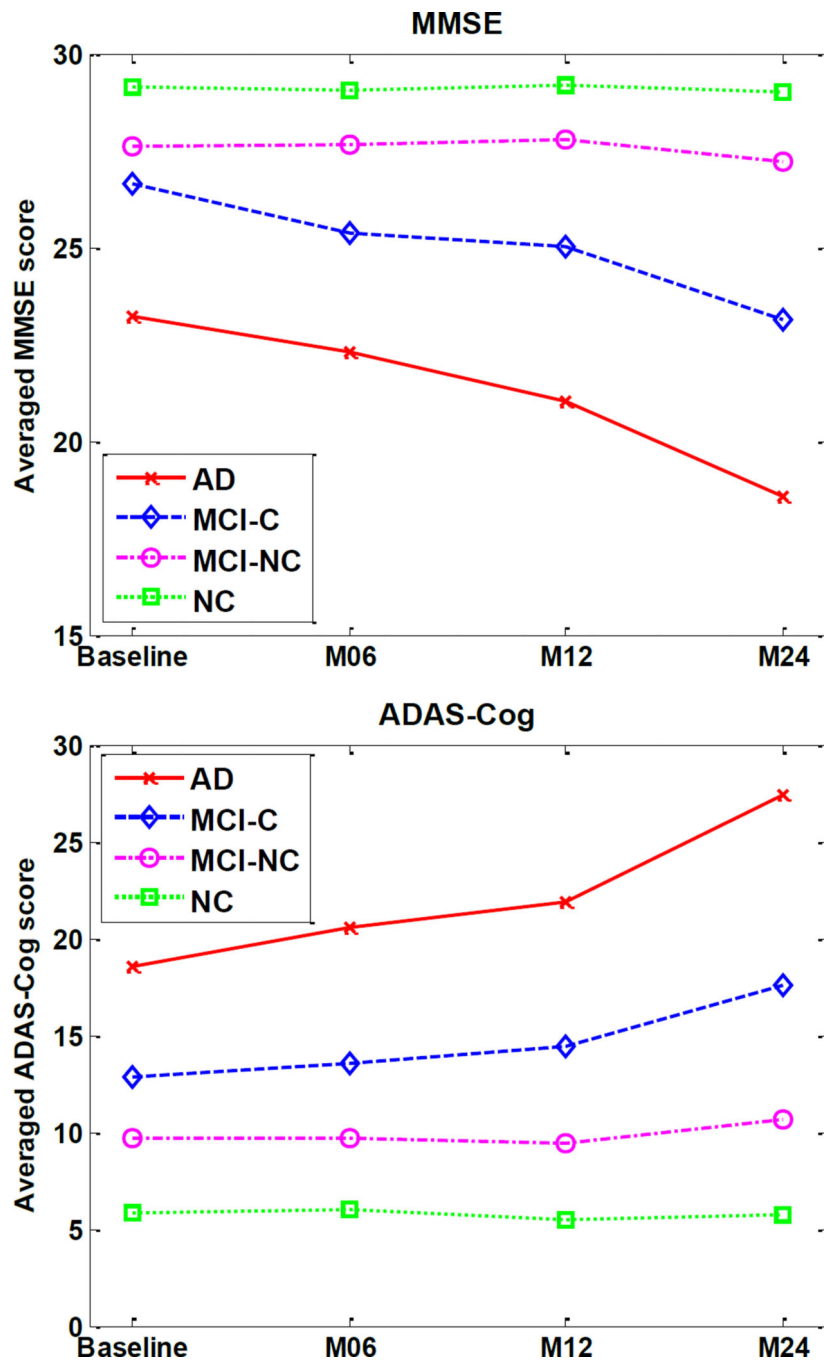
Fig. 2.

Average longitudinal changes of clinical scores in different kinds of subjects: MMSE (up) and ADAS-Cog (down).
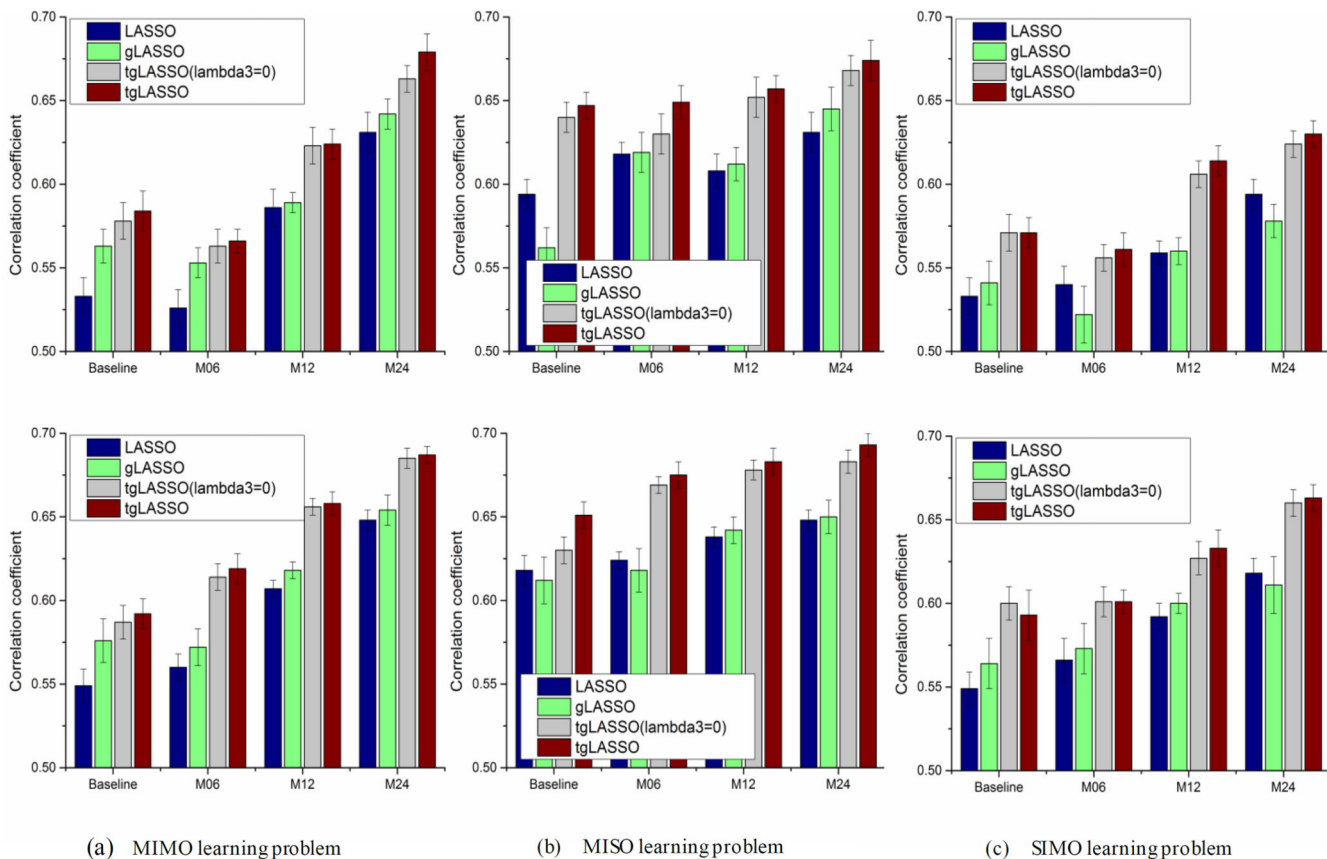
Fig. 3.

Comparison of correlation coefficients of different methods in estimating the MMSE (up) and ADAS-Cog (down) scores in three different learning problems, i.e., (a) MIMO learning problem, (b) MISO learning problem, and (c) SIMO learning problem.
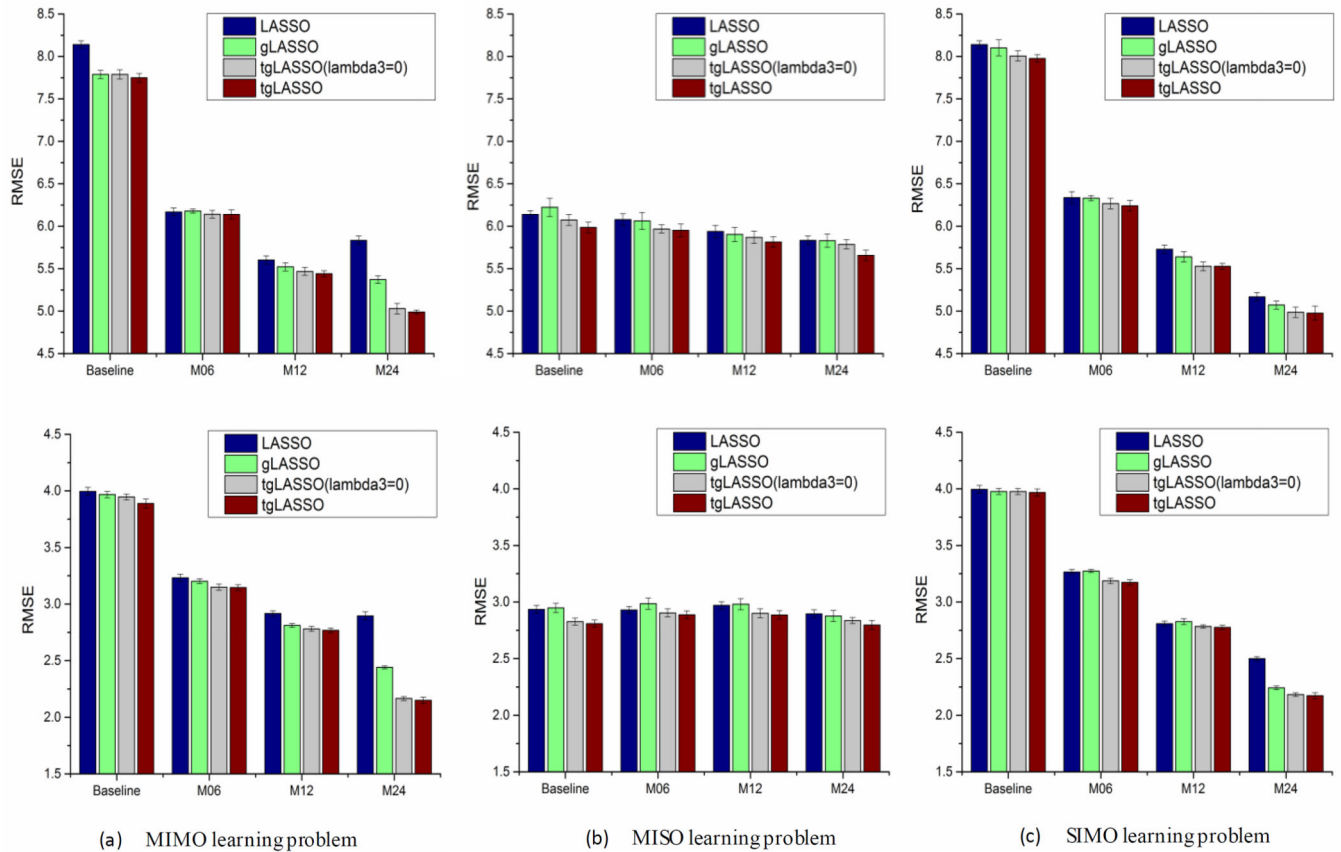
Fig. 4.
Comparison of RMSEs of different methods in estimating the MMSE (up) and ADAS-Cog (down) scores in three different learning problems, i.e., (a) MIMO learning problem, (b) MISO learning problem, and (c) SIMO learning problem.
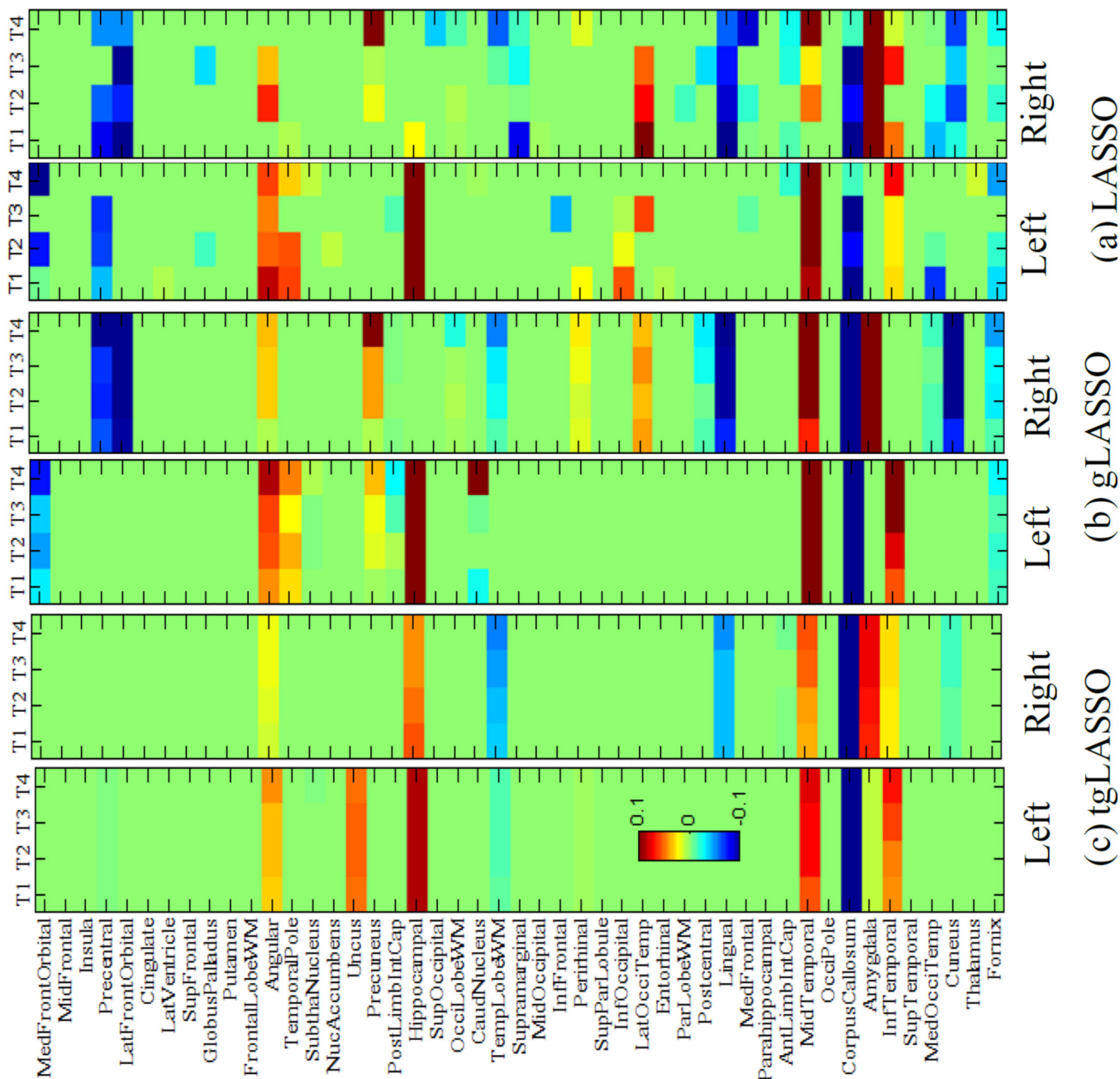
Fig. 5.
Comparison of the feature weight maps of three different methods in the MIMO learning problem: (a) LASSO, (b) gLASSO, and (c) tgLASSO.
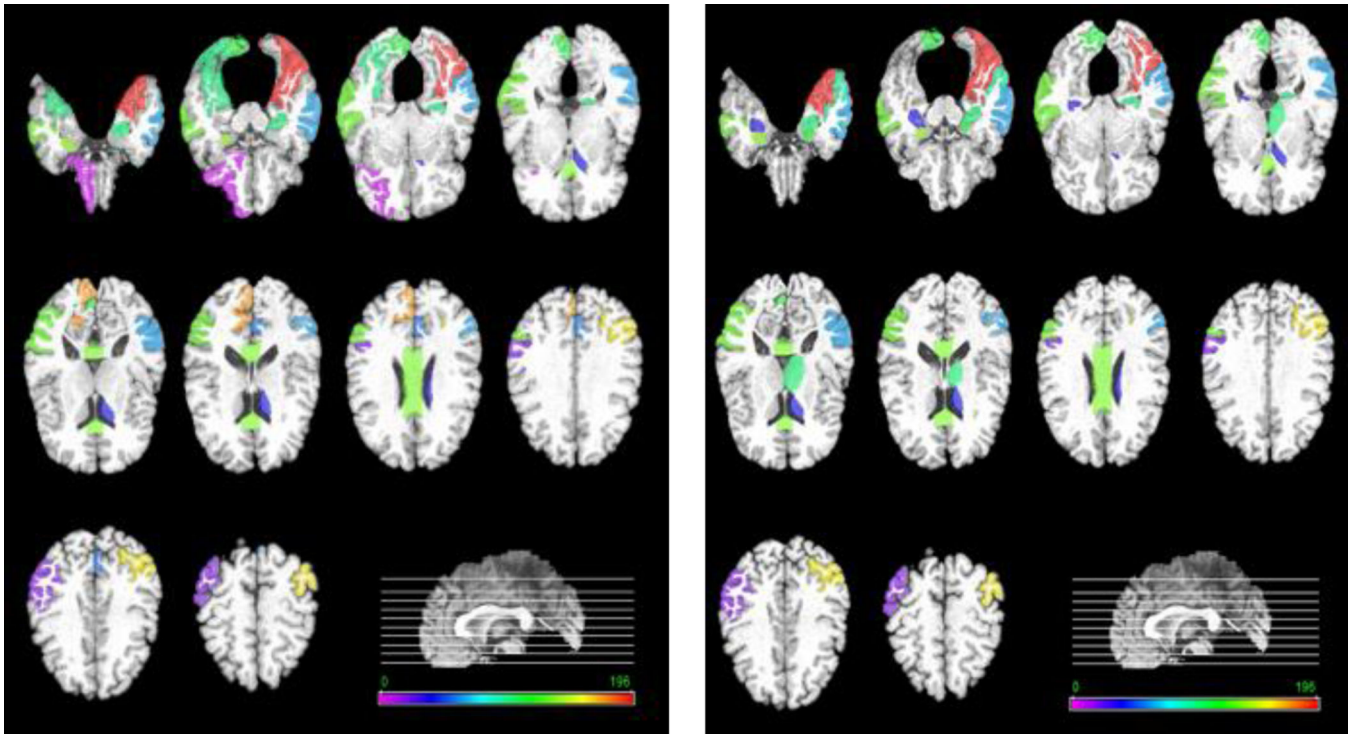
Fig. 6.
The important brain regions detected by the proposed tgLASSO method when estimating the MMSE score (left) and ADAS-Cog score (right).
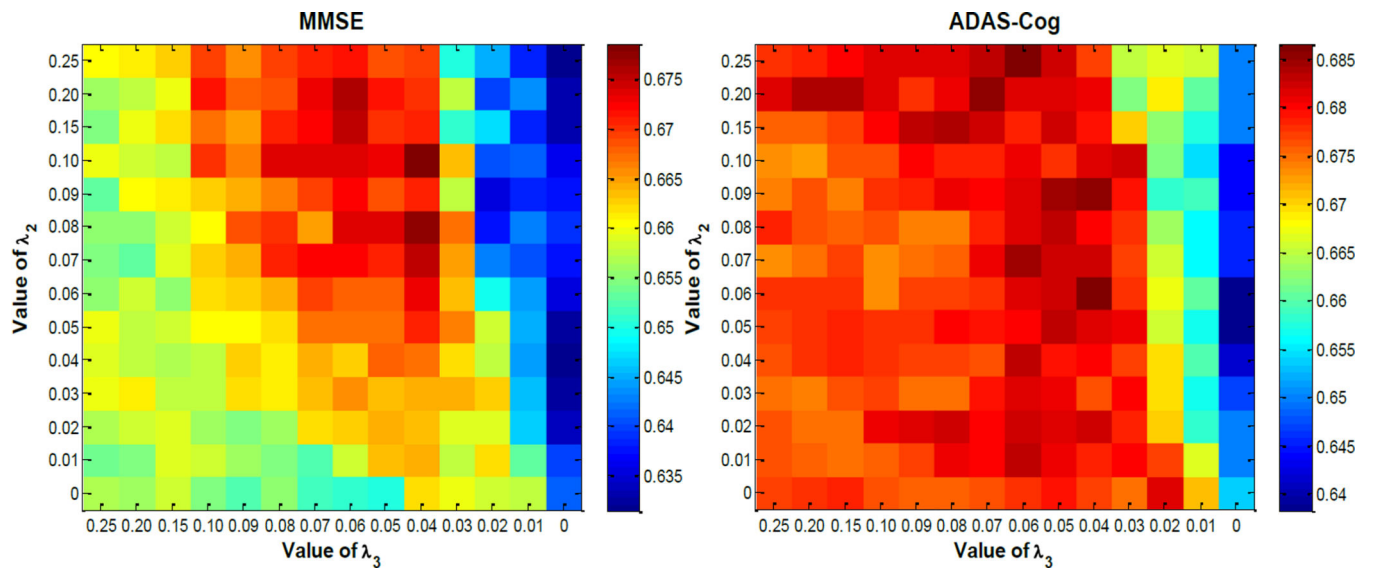
Fig. 7.
Regression performance of MMSE (left) and ADAS-Cog (right) scores under different combinations of λ_2 and λ_3 values.
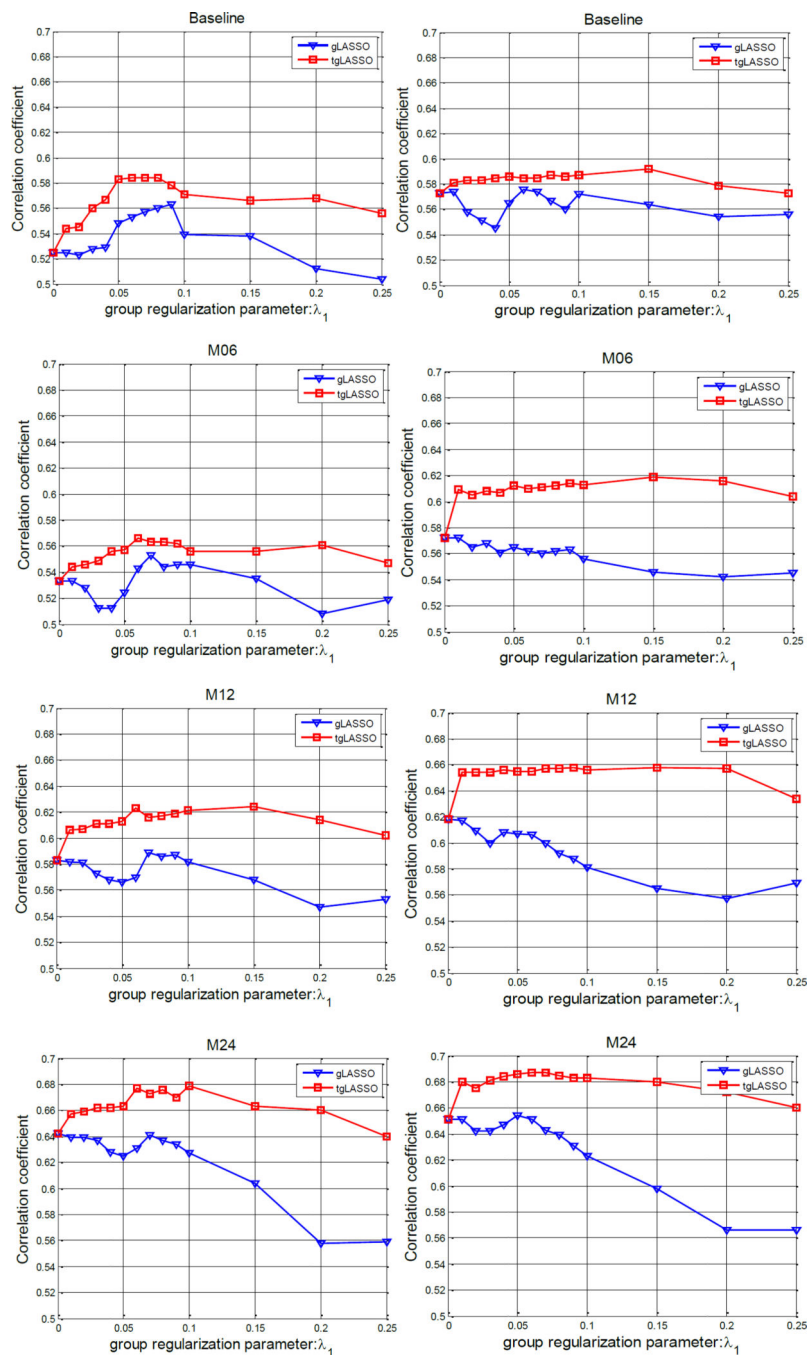
Fig. 8.
Correlation coefficients of proposed method (tgLASSO) and gLASSO method w.r.t. to the select ion of λ_1 value. Left: estimating the MMSE score; Right : estimating the ADAS-Cog score.

Table 1

Demographic characteristics of the studied population from the ADNI database. The values are denoted as mean ± standard deviation.

	AD (n=91)	MCI-C (n=104)	MCI-NC (n=98)	NC (n=152)
Female/Male	38/53	38/66	30/68	76/76
Age	75.4 ± 7.5	75.1 ± 6.8	74.3 ± 7.2	76.1 ± 4.8
Education	15.1 ± 2.9	15.8 ± 3.1	16.2 ± 2.9	16.0 ± 2.9
MMSE (bl)	23.2 ± 2.0	26.7 ± 1.7	27.6 ± 1.7	29.2 ± 0.9
MMSE (M06)	22.3 ± 3.2	25.4 ± 2.7	27.7 ± 2.1	29.1 ± 1.0
MMSE (M12)	21.0 ± 4.3	25.0 ± 2.7	27.8 ± 2.5	29.2 ± 1.1
MMSE (M24)	18.6 ± 6.0	23.1 ± 4.2	27.2 ± 3.2	29.0 ± 1.2
ADAS-Cog (bl)	18.6 ± 5.7	12.9 ± 4.0	9.7 ± 4.2	5.8 ± 2.9
ADAS-Cog (M06)	20.6 ± 6.5	13.6 ± 5.1	9.7 ± 4.1	6.0 ± 3.0
ADAS-Cog (M12)	21.9 ± 8.2	14.4 ± 5.8	9.4 ± 4.9	5.5 ± 2.8
ADAS-Cog (M24)	27.5 ± 11.8	17.6 ± 8.0	10.7 ± 5.7	5.7 ± 3.1

AD = Alzheimer's Disease; MCI = Mild Cognitive Impairment; MCI-C = MCI converter; MCI-NC = MCI non-converter; NC = Normal Controls; MMSE = Mini-Mental State Examination; ADAS-Cog = Alzheimer's Disease Assessment Scale - Cognitive Subscale.

Table 2

The p-values on squared residuals between the proposed tglasso method and competing methods.

Method		MIMO		MISO		SIMO	
		LASSO	gLASSO	LASSO	gLASSO	LASSO	gLASSO
MMSE	Base line	<0.001	0.008	<0.001	0.226	<0.001	0.002
	M06	<0.001	0.490	0.077	0.008	<0.001	0.026
	M12	0.011	0.091	<0.001	0.020	0.010	0.003
	M24	<0.001	0.029	<0.001	0.004	<0.001	0.029
ADAS-Cog	Base line	<0.001	0.094	0.049	0.001	<0.001	0.004
	M06	<0.001	0.024	0.009	0.001	<0.001	0.002
	M12	0.071	0.013	0.243	0.328	0.050	0.020
	M24	<0.001	<0.001	0.013	0.001	<0.001	<0.001

Table 3

Comparison of performance achieved by different methods in predicting the conversion of MCI patients (mean±std).

Method		LASSO	gLASSO	tgLASSO O ($\lambda_3=0$)	tgLASSO O
MMSE	ACC(%)	70.2±0.2	72.2±1.9	73.7±1.1	75.7±1.4
	SEN(%)	69.3±3.1	69.4±2.7	72.8±1.7	72.9±1.9
	SPE(%)	72.7±1.4	79.0±2.6	74.5±1.7	82.0±2.4
ADAS-Cog	ACC(%)	69.7±1.5	71.2±1.1	72.2±1.0	74.7±1.1
	SEN(%)	70.6±2.9	72.0±2.2	71.8±2.2	73.9±1.7
	SPE(%)	70.9±1.6	72.1±0.7	72.4±2.2	76.1±2.9

ACC= ACCuracy, SEN= SENsitivity, SPE= SPEcificity.

Table 4

The most important brain regions detected by the proposed tgLASSO method.

MMSE	AVG OF WEIGHTS ±STD
CORPUS CALLOSUM	−0.155±1.06E-02
LINGUAL GYRUS RIGHT	−0.104±7.48E-05
LATERAL FRONT-ORBITAL GYRUS RIGHT	−0.084±9.69E-06
CUNEUS RIGHT	−0.067±2.24E-04
PERIRHINAL CORTEX LEFT	0.026±1.30E-06
LATERAL OCCIPITOTEMPORAL GYRUS LEFT	0.026±5.84E-06
CAUDATE NUCLEUS LEFT	0.027±1.61E-02
PRECUNEUS LEFT	0.029±4.54E-05
LATERAL OCCIPITOTEMPORAL GYRUS RIGHT	0.040±1.14E-05
TEMPORAL POLE LEFT	0.048±1.36E-04
ANGULAR GYRUS RIGHT	0.049±1.30E-03
ANGULAR GYRUS LEFT	0.057±6.68E-05
MIDDLE TEMPORAL GYRUS LEFT	0.094±3.96E-03
HIPPOCAMPAL FORMATION LEFT	0.111±1.13E-03
MIDDLE TEMPORAL GYRUS RIGHT	0.119±9.82E-03
AMYGDALA RIGHT	0.120±2.30E-04

ADAS-COG	AVG OF WEIGHTS ±STD
INFERIOR TEMPORAL GYRUS LEFT	−0.062±1.44E-03
HIPPOCAMPAL FORMATION RIGHT	−0.058±1.19E-03
CORPUS CALLOSUM	−0.039±4.46E-04
ANGULAR GYRUS LEFT	−0.039±4.04E-04
MIDDLE TEMPORAL GYRUS LEFT	−0.032±4.50E-04
PERIRHINAL CORTEX LEFT	−0.029±2.82E-04
ANGULAR GYRUS RIGHT	−0.024±2.03E-04
AMYGDALA RIGHT	−0.023±5.62E-04
AMYGDALA LEFT	−0.02±6.79E-04
LINGUAL GYRUS RIGHT	−0.019±3.76E-04
PERIRHINAL CORTEX RIGHT	−0.019±2.81E-04
HIPPOCAMPAL FORMATION LEFT	−0.009±7.40E-04
LATERAL OCCIPITOTEMPORAL GYRUS LEFT	−0.009±3.22E-04
CAUDATE NUCLEUS LEFT	0.002±2.83E-04
THALAMUS LEFT	0.024±2.68E-04
MIDDLE TEMPORAL GYRUS RIGHT	0.038±4.03E-03

Table 5

Comparison with correlation coefficient of the state-of-the-art methods.

Method	Subjects	MMSE	ADAS-cog
(Duchesne et al., 2009) [48]	20 MCI-C+ 29 MCI-NC	0.31	-
(Fan et al., 2010) [49]	52 AD+148 MCI+ 64 NC	0.57	0.52
(Wan et al., 2014) [50]	171 AD+222 NC	0.758±0.011	0.767±0.026
(Yan et al., 2015) [51]	172 AD+349 MCI+197 NC	0.5552±0.0078	0.6438±0.0258
Proposed (MIMO)	91 AD+202 MCI+152 NC	0.613±0.010	0.639±0.008