



Longitudinal Analysis for Disease Progression via Simultaneous Multi-Relational Temporal-Fused Learning

Baiying Lei^{1,2,3,4}, Feng Jiang^{1,2,3}, Siping Chen^{1,2,3}, Dong Ni^{1,2,3*} and Tianfu Wang^{1,2,3*}

¹ School of Biomedical Engineering, Shenzhen University, Shenzhen, China, ² National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen University, Shenzhen, China, ³ Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen University, Shenzhen, China, ⁴ Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, China

It is highly desirable to predict the progression of Alzheimer's disease (AD) of patients [e.g., to predict conversion of mild cognitive impairment (MCI) to AD], especially longitudinal prediction of AD is important for its early diagnosis. Currently, most existing methods predict different clinical scores using different models, or separately predict multiple scores at different future time points. Such approaches prevent coordinated learning of multiple predictions that can be used to jointly predict clinical scores at multiple future time points. In this paper, we propose a joint learning method for predicting clinical scores of patients using multiple longitudinal prediction models for various future time points. Three important relationships among training samples, features, and clinical scores are explored. The relationship among different longitudinal prediction models is captured using a common feature set among the multiple prediction models at different time points. Our experimental results based on the Alzheimer's disease neuroimaging initiative (ADNI) database shows that our method achieves considerable improvement over competing methods in predicting multiple clinical scores.

Keywords: Alzheimer's disease (AD), longitudinal analysis, feature selection, joint learning, prediction

OPEN ACCESS

Edited by:

Aurel Popa-Wagner,
University of Rostock, Germany

Reviewed by:

Jingwen Niu,
Temple University, USA
Yan Jin,
University of Texas MD Anderson
Cancer Center, USA

*Correspondence:

Dong Ni
nidong@szu.edu.cn
Tianfu Wang
tfwang@szu.edu.cn

Received: 28 June 2016

Accepted: 11 January 2017

Published: 03 March 2017

Citation:

Lei B, Jiang F, Chen S, Ni D and Wang T (2017) Longitudinal Analysis for Disease Progression via Simultaneous Multi-Relational Temporal-Fused Learning. *Front. Aging Neurosci.* 9:6. doi: 10.3389/fnagi.2017.00006

INTRODUCTION

Alzheimer's disease (AD) imposes heavy social-economic burdens on society (Fan et al., 2008; Alzheimer's Association, 2014; Shi et al., 2015), and patients experience tremendous cognitive decline throughout progression of the AD disease. Tremendous effort have been devoted to improve the understanding and monitoring of AD progression (Brookmeyer et al., 2007; Hinrichs et al., 2009; Liu M. et al., 2014; Wang et al., 2014; Jie et al., 2015; Lei et al., 2015a; Liu et al., 2015; Lei et al., 2016; Zhuo et al., 2016). Modeling disease progression based on cognitive decline in longitudinal analysis has been widely investigated in the neuroimaging field (Fan et al., 2008; Davatzikos et al., 2010; Stonnington et al., 2010; Wang et al., 2010; Hinrichs et al., 2011). In recent decades, neuroimaging-based longitudinal studies have proven to be an important research direction in characterizing the neurodegenerative process of AD, where data at multiple time points are often used (Teipel et al., 2007; Vemuri et al., 2009; Jack et al., 2010; Cuingnet et al., 2011). It has been reported that researchers are able to study the cognitive decline due to the neurodegenerative property of AD

with traditional structural magnetic resonance imaging (MRI) (Davatzikos et al., 2001, 2008; Dickerson et al., 2001; Gaser et al., 2001; Leow et al., 2006; Jack et al., 2008; Vemuri et al., 2008; Frisoni et al., 2010; Stonnington et al., 2010; Wang et al., 2016a), diffusion-weighted MRI (Jin et al., 2015, 2017; Daianu et al., 2016; Wang et al., 2016b; Wu et al., 2016), and functional MRI (Yang et al., 2016). In addition, cognitive decline in the neurodegenerative cognitive measures, e.g., the Alzheimer's disease assessment scale cognitive subscale (ADAS-Cog) and the mini mental state examination (MMSE), can be used to partially reveal AD progression (Davatzikos et al., 2001, 2008; Dickerson et al., 2001; Gaser et al., 2001; Leow et al., 2006; Jack et al., 2008; Vemuri et al., 2008; Frisoni et al., 2010). However, accurate prediction of AD progression still remains a challenging task due to the complicated characteristic of AD progression.

The first challenge in longitudinal studies for AD diagnosis is dimensionality of the data, which is usually much higher than the available number of samples. To address this issue, researchers have developed various feature selection models for different clinical scores (e.g., ADAS-Cog and MMSE) to identify disease-related biomarkers among multiple time points (Yuan and Lin, 2006; Zhang et al., 2012; Zhou et al., 2013). Among these methods, Lasso and its variants are the most popular techniques for feature selection (Tibshirani, 1996; Guyon et al., 2002; Guyon and Elisseeff, 2003; Yuan and Lin, 2006; Wang et al., 2010). For instance, (Wang et al., 2010), linear regression models are applied in high-dimensional pattern recognition problems not only to estimate the stage of AD, but also to construct a stable model. Adaptive regional feature extraction is applied to this model for the prediction of regression variables.

The second challenge is understanding the underlying relationship between features, subjects, and clinical scores at different time points. This relationship is seldom considered even though many longitudinal studies show promising predictive power in AD progression prediction. This relationship could provide inherent high-level information that is useful for studying AD. Therefore, modeling and utilizing this relationship could enhance the learning performance in predicting AD progression.

To tackle the above-mentioned challenges, several methods such as the group Lasso (Yuan and Lin, 2006), the temporally constrained group Lasso (TGL, Caroli and Frisoni, 2010; Jack et al., 2010), and the convex fused sparse group Lasso (cFSGL, Zhou et al., 2013) incorporated cognitive progression information into linear regression models to predict disease progression. Evidently, multi-task learning methods using intrinsic information achieve better performance than single-task learning methods (Liu et al., 2012, 2015; Zhu et al., 2014a; Jie et al., 2015), and this learning framework has shown great success in predicting ADAS-Cog and MMSE due to its good generalization capabilities (Zhang et al., 2011; Zhu et al., 2014a). For example, Zhang and Shen (Zhang et al., 2011) proposed a joint regression and classification scheme to understand the mechanism of AD. This method was further improved by Zhu et al. using a feature and variable graph matching method to jointly identify AD status and predict clinical scores (Zhu et al., 2014a). The inter-modality constraints described by Liu et al.

were also included in a multi-task learning framework for AD diagnosis (Liu F. et al., 2014). Despite promising performance achieved by these methods, most existing methods fail to take advantage of the cognitive progression from multiple time points among features, subjects, and clinical scores. This mission is undesirable in longitudinal analysis and follow-up studies.

To incorporate cognitive progression relationships, longitudinal analysis has been widely explored to model cognitive progression and to exploit the associated imaging markers and cognitive changes across all time points (Misra et al., 2009; Davatzikos et al., 2010; Stonnington et al., 2010; Hinrichs et al., 2011; Zhang et al., 2012; Zhang and Shen, 2012; Zhou et al., 2013). For instance, Zhou et al. (2013) integrated temporal smoothness into their method using multi-task learning techniques to identify biomarkers for disease progression. Remarkable performance is achieved based on temporal-relational constraints and later-time constraints, in which each task is treated separately using a single baseline feature for predicting future-time-point score. Huang et al. (2015, 2016) proposed an improved random forest framework and took advantage of the longitudinal information at multiple time points to further improve the accuracy of AD score prediction. Using complementary information, such as MRI data and clinical scores, is desirable as it might uncover important imaging biomarkers. Zhang et al. found that longitudinal analysis is effective for mild cognitive interference (MCI) prediction (Zhang and Shen, 2012) and proposed the utilization of MRI features at multiple time points with temporal smoothness regularization (Zhang et al., 2012). This method outperformed competing methods, because the disease pattern is better revealed by comprehensive cognitive progression information as compared to methods using only the baseline features. In addition, the AD progression prediction problem in Wang et al. (2012) was addressed by a high-order multi-task learning method that exploits the temporal correlations in imaging and cognitive data with a structured sparsity inducing term. Promising predictive power is achieved via multiple time point features, and the clinical scores are learned independently. This arrangement is undesirable as correlation of the clinical score is ignored.

Most existing methods ignore the relationship among different features, subjects, and clinical scores. Furthermore, the relational information is seldom studied even though there is a strong correlation between the clinical scores and MRI data (Gaser et al., 2001; Leow et al., 2006). Moreover, most of the previous studies only focused on one or two types of relationships without considering the cognitive decline at different time points. The integration of this type of information would better identify the spatial patterns of brain atrophy because the associated feature patterns and the specific patterns of the neighboring time points are highly correlated. It has been also demonstrated in literature (Zhu et al., 2014a; Jie et al., 2015) that multiple relationships between feature-feature and subject-subject boost diagnosis performance.

Intuitively, information from clinical scores and subjects at multiple time-points can play an important role in identifying temporal patterns in longitudinal analysis. In this paper, a

new multi-task joint feature learning method is developed to exploit the intrinsic relation of the data to boost performance of disease prediction. Manifold learning and discriminative learning theories have achieved remarkable performance by incorporating (Stanciu et al., 2014; Zhu et al., 2014b; Jie et al., 2015; Lei et al., 2015b; Zhou et al., 2015). Therefore, we explore temporal smoothness and multi-relation graphs among different patterns and cognitive measures to uncover human brain variations for better diagnosis of AD progression. Specifically, we define a novel objective function to impose multi-relation information. A group sparsity regularizer is used to jointly select a small number of specific features across different time points. We also incorporate multi-relation smoothness regularization to capture the relationship among features, subjects, and clinical scores. After the selection of longitudinal feature, the final selected brain regions are employed for clinical score prediction using multi-kernel support vector regression (SVR, Chang and Lin, 2011). To the best of our knowledge, there is no existing sparse model that incorporates multi-relation smoothness in its objective function to estimate the clinical scores (e.g., ADAS-Cog and MMSE). In addition, our method focuses on multiple relationships, which has obvious advantages over existing methods that only exploit feature or sample relationships. It is worth noting that we need to observe the behavioral changes in patients' condition over time in order to model disease progression. However, it is difficult to extract huge amount of information from data that is collected from multiple time points. Therefore, we utilize the aforementioned relational constraints to build a robust regression model by selecting the best and most relevant features to predict patient's clinical behavior at multiple future time points.

In this paper, we propose a joint learning procedure for multiple longitudinal predictions of AD progression by exploiting their inherent relationships. In particular, we propose three novel regularization terms (each modeling a set of crucial relationships at different time points), and incorporate these regularizers in a multi-task sparse feature selection model. We also introduce a specifically designed loss function to jointly predict the patients' clinical scores at multiple future time points, thus condensing the common information shared by data from different time points and permitting the selection of the most meaningful features for multiple prediction tasks. We evaluate our method using the Alzheimer's disease neuroimaging initiative (ADNI) (<http://adni.loni.usc.edu/>) database (Alzheimer's Association, 2014), and our method achieves promising results in estimating multiple clinical scores at multiple future time points using only baseline data. For the ADNI baseline, a total of 445 subjects—91 with AD, 202 with mild cognitive impairment (MCI), and 152 cognitively normal controls (NCs) are investigated in our study to predict the ADAS-Cog/MMSE scores for the next 2 years because the subjects already have completed MRI and clinical score data. In this work, we focus on only using suitable data instead of all the data from the ADNI database to study of disease progression.

Our extensive experimental results show that the proposed joint learning framework obtains state-of-the-art performance for future ADAS-Cog/MMSE score prediction. We observe that

the hippocampal formation, amygdala, temporal pattern, and uncus demonstrate the most definitive patterns in predicting clinical scores at all-time points.

MATERIALS AND METHODS

Materials

Our investigation is based on data obtained from the ADNI database (Alzheimer's Association, 2014), which was created and updated since 2004. This 6-year study is funded \$60 million from the public and private sectors, which include the National Institute on Aging, the National Institute of Biomedical Imaging, and Bioengineering, and the Food and Drug Administration. The principle goal of ADNI is to verify that serial MRI and positron emission tomography (PET) images, along with other biological markers, clinical, and neuropsychological assessment can be used to measure the progression of MCI and early AD. The 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 to lessen the time and cost of clinical trials. ADNI is the collective effort of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. 800 adults aged 55–90 were recruited to participate in this research, which includes approximately 200 cognitively normal older individuals followed for 3 years, 400 people with MCI followed for 3 years, and 200 people with early AD followed for 2 years. For up-to-date information, please refer to <http://www.adni-info.org>.

Subjects and Pre-processing

The general eligibility criteria of ADNI are briefly described in the following. Subjects between 55 and 90 years of age who have a study partner to provide an independent evaluation of functioning were selected. Patients taking certain psychoactive medications were excluded. There are three general inclusion/exclusion criteria: (1) The range of MMSE scores of healthy subjects (non-depressed, non-MCI, and non-demented) is 24–30; (2) The range of MMSE scores of MCI subjects is also between 24 and 30. The subjects having objective memory loss were measured with education adjusted scores using the Wechsler memory scale logical memory II, a CDR of 0.5, an absence of significant levels of impairment education adjusted scores fall between 20 and 26, and satisfy the National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD. The study subjects gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by each participating site

TABLE 1 | Statistical information of clinical scores.

Clinical scores	Mean	Median	Min	Max
MMSE	12.02	10	0	59
ADAS-Cog	26.27	28	5	30

Institutional Review Board (IRB). **Table 1** gives the detailed ADAS-Cog and MMSE information of the subjects used in our study.

The pre-processing of feature extracted from the ROI regions have been widely applied in the literature (Liu et al., 2012, 2015; Zhu et al., 2014a; Jie et al., 2015). For our method, pre-processing was first applied to the T1-weighted MRI brain images of each subject, and then skull stripping was performed to clean the skull. The cerebellum was removed by warping a labeled Jacob atlas to the skull-stripped image (Wang et al., 2014). Segmentation by the FAST method (Zhang et al., 2001) was then applied to segment the brain images into three tissues, which include white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). After segmentation, the brain image was nonlinearly registered with a HAMMER tool (Shen and Davatzikos, 2002). The features used in this study include the volume intensity extracted from the region of interest (ROI) of different brain regions (Zhang et al., 2011). Specifically, the brain image of each subject was partitioned into 93 ROIs by atlas warping, and the volume of GM tissue of each ROI was extracted as a feature. Similar to the study in Wee et al. (2012), the obtained features were normalized to facilitate disease diagnosis and prognosis.

Notation and Problem Statement

For this work, capital bold letters denote matrices, small bold letters denote vectors, and non-bold letters denote regular variables. Let $X \in \mathbb{R}^{S \times F}$ denote the data of S different subjects, where each subject is represented by an F -dimensional feature vector from the baseline MRI image. Let X denote the data generated from baseline time point. We denote $x_{u, \cdot}$ and $x_{\cdot, v}$ as the u -th row vector and the v -th column vector of X , respectively. Let $Y = \{Y^{(t)} \in \mathbb{R}^{S \times C}, t = 1, \dots, T\}$ denote C types of clinical cognitive scores (e.g., ADAS-Cog and MMSE) for S subjects

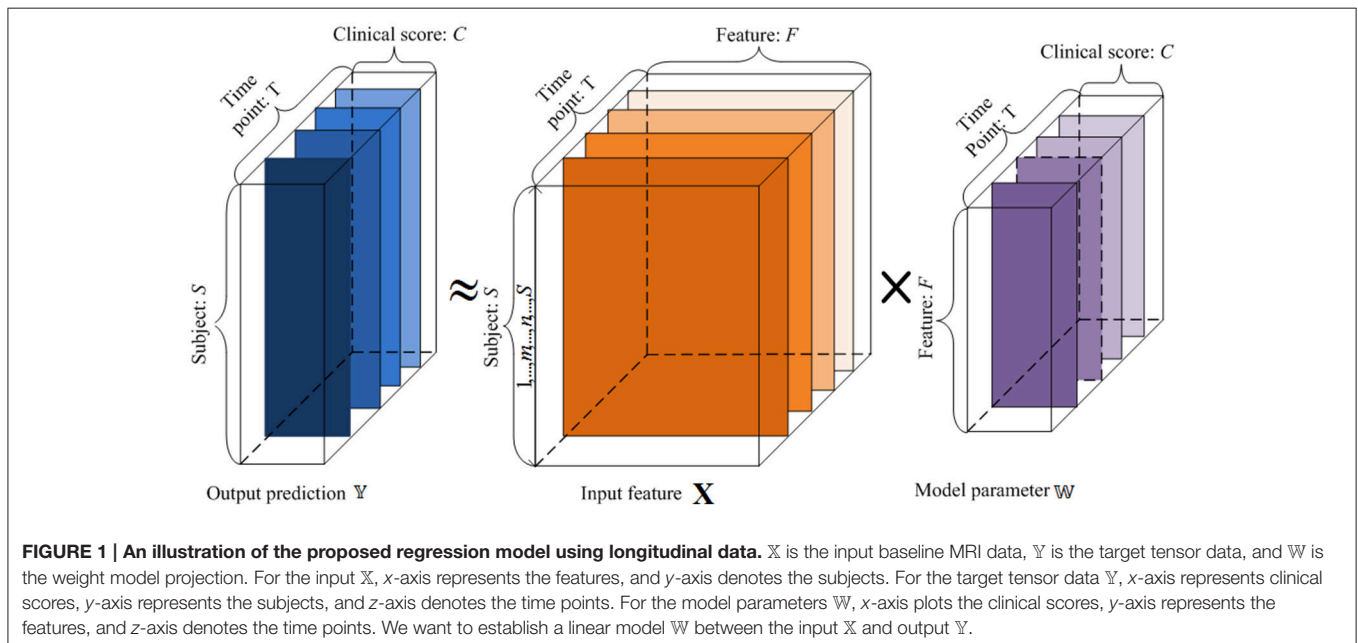
at T time points, where $Y^{(t)} \in \mathbb{R}^{S \times C}$ is the corresponding clinical scores at the t -th time point for S subjects. Let $\mathbb{W} = \{W^{(t)} \in \mathbb{R}^{F \times C}, t = 1, \dots, T\}$ as denote the set of weight matrices that map the original features to clinical scores, where $W^{(t)}$ represents the weight matrix for the t -th time point.

Our goal is to create a linear regression model to reveal the longitudinal associations between the original features and the cognitive trajectories through time, and predict the clinical scores at multiple future time points from the baseline data ($t = 1$). This is illustrated in **Figure 1**. Each subject's features are assigned as a row in the matrix X . By learning the weight vectors in each $W^{(t)}$, we can reconstruct the corresponding clinical scores in each $Y^{(t)}$, as explained below.

To simplify the problem and design the objective function, we unfolded the projected weight matrix by concatenating all $W^{(t)}$ s as $\hat{W} = [W^{(1)}, W^{(2)}, \dots, W^{(t)}, \dots, W^{(T)}] \in \mathbb{R}^{F \times CT}$. Hence, the model parameters can be incorporated in the objective function to extract the common features across different time points. This unfolded simplified weighting matrix is shown in **Figure 2**.

Methodology

A key advantage of longitudinal studies is the ability to observe the patients' changes through time, and to effectively utilize the shared common information in different time points to select the best set of features for monitoring the progression of MCI patients and thereafter predicts their future status. There are three different aspects in which this common information could be leveraged, i.e., the relationships among features, subjects, and clinical scores. Intuitively, the pairwise similarities among features, subjects, and clinical scores should be preserved in the predictions via the regression model. In this section, we introduce a method to incorporate such relation information into



a multi-task learning framework. Specifically, we define a linear regression model for each time point using the baseline data as a single task, and then formulate the global regression model in a multi-task learning framework with a $l_{2,1}$ sparsity constraint, where the above three relational aspects are incorporated as regularization terms. **Figure 3** shows the flowchart of the proposed method.

To clearly illustrate the relationship among features, subjects, and clinical scores, we adopted a graph matching technique in our proposed method, where the feature-feature, subject-subject and clinical score-clinical score relationships are represented in terms of graph (shown in **Figure 4**). In **Figure 4A**, a node represents one feature and an edge represents the

relationship between the connected nodes. **Figures 4B,C** show the relationship among subjects and among clinical scores, respectively. In these graphs, the length of an edge between features (subjects or clinical scores) represents the similarity among features (subjects or clinical scores), where the similarity increases with the length of the edge. These graphs are built based on the information of training data, which are then used as regularization terms for subsequent multi-task learning procedure.

Previous studies (e.g., Zhou et al., 2013) revealed the associations among imaging features and cognitive scores at each time point separately, under the assumption that each task at each time point is independent. However, this assumption does not always hold because clinical scores possess temporal correlation. In fact, harnessing the temporal correlation could potentially help predict the clinical cognitive scores. This motivates us to use a joint learning regression model across all time points, which would help identify the most relevant imaging markers for the prediction of cognitive scores. Specifically, we seek to learn the weight coefficient matrices to uncover the clinical scores progression, through which the information from each learning task and the common structures among multiple time points can be jointly discovered.

To select the most relevant and discriminant features at each time point, a correlation-induced sparsity model with a least-squares loss function is proposed. The loss function controls the prediction error, while the sparsity assumption leads to the least number of contributing features. As a result, for each time point, we would select the features that are most correlated with the actual clinical scores. A general form of the proposed objective function is defined as below:

$$\min_{\{W^{(t)}, t=1, \dots, T\}} \sum_{t=1}^T \|Y^{(t)} - XW^{(t)}\|_F^2 + \lambda_1 \sum_{t=1}^T \Phi(W^{(t)}) + \lambda_2 \|\hat{W}\|_{2,1}, \quad (1)$$

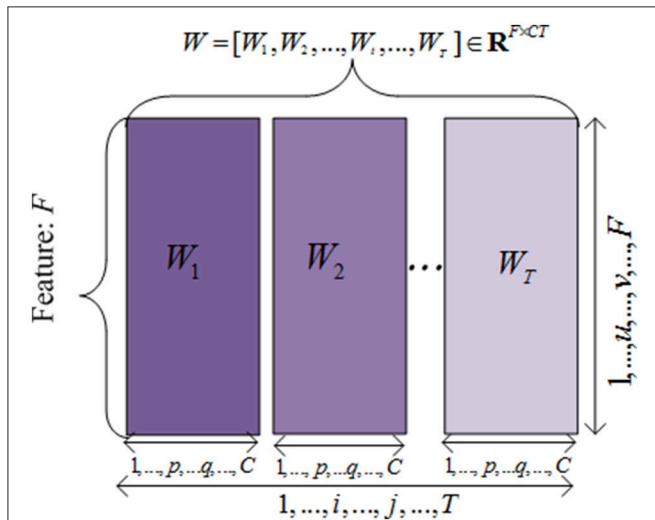


FIGURE 2 | The illustration of the unfolded weight matrices. Each row is corresponding to a specific feature vector among different time points, and each column is corresponding to a clinical score vector at one time point.

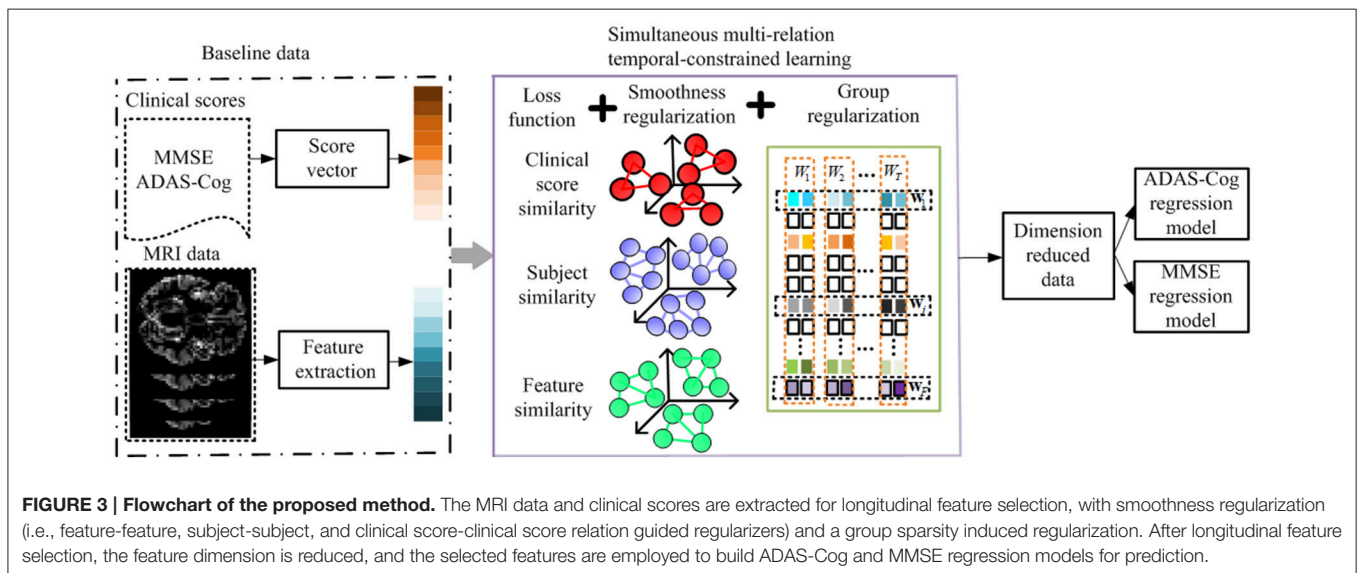
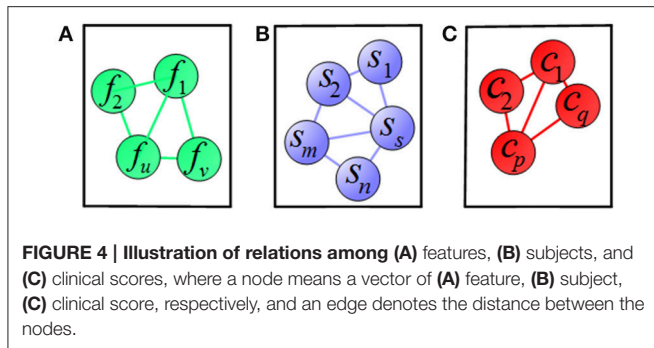


FIGURE 3 | Flowchart of the proposed method. The MRI data and clinical scores are extracted for longitudinal feature selection, with smoothness regularization (i.e., feature-feature, subject-subject, and clinical score-clinical score relation guided regularizers) and a group sparsity induced regularization. After longitudinal feature selection, the feature dimension is reduced, and the selected features are employed to build ADAS-Cog and MMSE regression models for prediction.



where $\| \cdot \|_F^2$ is the Frobenius norm of a matrix, \mathbf{W} is the coefficient weight matrix, λ_1 and λ_2 are the regularization parameters. The last term in Equation (1) is a group regularizer to uncover the correlation among different features and jointly select features for multiple tasks, which is defined as $\| \widehat{\mathbf{W}} \|_{2,1} = \sum_{i=1}^F \| \widehat{\mathbf{w}}_{i,:} \|_2$, where $\widehat{\mathbf{w}}_{i,:}$ is the i -th row vector of $\widehat{\mathbf{W}}$, and $\| \cdot \|_F^2$ is $l_{2,1}$ -norm. It is worth noting that $l_{2,1}$ -norm computes the sum of the l_2 -norm of each row of $\widehat{\mathbf{W}}$, which enforces many rows to be zero, and hence it is suitable for feature selection. Features corresponding to those non-zero rows in $\widehat{\mathbf{W}}$ are regarded as the most predictive features in subsequent learning models. The second term $\Phi(\mathbf{W}^{(t)})$ is our proposed regularizer, which is comprised of multiple parts corresponding to three relationships among features, among subjects, and among clinical scores. To define the proposed regularization term, $\Phi(\mathbf{W}^{(t)})$, the “feature-feature,” the “subject-subject,” and the “clinical score-clinical score” relationships at each different time point are incorporated. We use the idea of Laplacian matrices and graphs to obtain the similarity in the local structures (Belkin and Niyogi, 2003). In the rest of this section, we explain it in detail and discuss all its characteristics.

The “feature-feature” relation information is imposed as the relationship between columns of the input matrix \mathbf{X} , and is reflected in the relation between corresponding rows in the coefficients weight matrix $\mathbf{W}^{(t)}$. Hence, the widely used graph Laplacian (Hinrichs et al., 2009; Wang et al., 2012; Zhu et al., 2013) is leveraged. To measure the similarity between the u -th and the v -th features of \mathbf{X} . in the original feature space, we use the heat kernel defined as below:

$$f_{uv} = \exp(-\| \mathbf{x}_{:,u} - \mathbf{x}_{:,v} \|_2^2), \tag{2}$$

where $\mathbf{x}_{:,u}$ is the u -th column of the input data \mathbf{X} . Based on the similarity, we develop the first feature-feature relation based regularization term as

$$R_f(\mathbf{W}^{(t)}) = \sum_{u,v=1}^F f_{uv} \| \mathbf{w}_{u,:}^{(t)} - \mathbf{w}_{v,:}^{(t)} \|_2^2, \tag{3}$$

where $\mathbf{w}_{u,:}^{(t)}$ is the u -th row of $\mathbf{W}^{(t)}$ at time point t . Thus, the highly correlated features produce large weights in the above sparsity regularization.

The second regularization is based on the “subject-subject” relation graph. We know that the output clinical scores of similar

subjects should be similar. Therefore, similar to the previous term, we use a heat kernel to exploit the “subject-subject” similarities and define the similarity between the m -th and the n -th subject as

$$\phi_{mn} = \exp(-\| \mathbf{x}_{m,:} - \mathbf{x}_{n,:} \|_2^2), \tag{4}$$

where $\mathbf{x}_{m,:}$ is the m -th row of input \mathbf{X} . Here, “subject-subject” relation regularizer is defined as

$$R_s(\mathbf{W}^{(t)}) = \sum_{m,n=1}^S \phi_{mn} \| \mathbf{x}_{m,:} \mathbf{W}^{(t)} - \mathbf{x}_{n,:} \mathbf{W}^{(t)} \|_2^2. \tag{5}$$

The last regularization is based on the “clinical score-clinical score” relation. For each subject’s feature vector $\mathbf{x}_{p,:}$ in our regression framework, different sets of weight coefficients are used to regress the output clinical scores $\mathbf{y}_{:,p}^{(t)}$. In other words, the elements in each column of $\mathbf{W}^{(t)}$ are related to the elements in each column of $\mathbf{Y}^{(t)}$ through the feature vectors. As a result, if two clinical scores are correlated, the corresponding weight columns in matrix $\mathbf{W}^{(t)}$ should be correlated too. Similarly, we use a heat kernel to exploit the “clinical score-clinical score” relation. The similarity between the p -th clinical score and the q -th clinical score is defined using a heat kernel as

$$\psi_{pqt} = \exp(-\| \mathbf{y}_{:,p}^{(t)} - \mathbf{y}_{:,q}^{(t)} \|_2^2), \tag{6}$$

where $\mathbf{y}_{:,p}^{(t)}$ is the p -th column vector of $\mathbf{Y}^{(t)}$. To this end, we define the clinical score relational regularization term as

$$R_c(\mathbf{W}^{(t)}) = \sum_{p,q=1}^C \psi_{pqt} \| \mathbf{w}_{:,p}^{(t)} - \mathbf{w}_{:,q}^{(t)} \|_2^2. \tag{7}$$

Therefore, our proposed joint learning model using the relation information, as discussed above, is

$$\min_{\{\mathbf{W}^{(t)}, t=1, \dots, T\}} \sum_{t=1}^T \| \mathbf{Y}^{(t)} - \mathbf{X} \mathbf{W}^{(t)} \|_F^2 + \lambda_1 \sum_{t=1}^T (R_f(\mathbf{W}^{(t)}) + R_s(\mathbf{W}^{(t)}) + R_c(\mathbf{W}^{(t)})) + \lambda_2 \| \widehat{\mathbf{W}} \|_{2,1}. \tag{8}$$

Using Equation (8), we incorporate three types of relationship data at multiple time points into a unified objective function. This method is referred as the simultaneous multi-relation temporally constrained learning (SMTL). To the best of our knowledge, this is the first work to simultaneously incorporate multi-relation information such as “feature-feature,” “subject-subject,” and “clinical score-clinical score” in fused regularizations, which is difficult to solve in current sparse models. In addition, no previous studies jointly apply the multi-relation information across multiple time points as additional regularizers.

Motivated by Zhu et al. (2013), the optimization problem in Equation (8) could be solved in an alternative way (i.e., finding the optimal solution for one variable while the others are fixed). The optimization steps are discussed in the following section. After selecting the most meaningful features, we use a SVR model to predict the clinical scores of patients at multiple future time points. Considering how AD changes over time, our algorithm

benefits from a joint multi-task learning framework, in which multiple relationships are introduced as regularization terms to take advantage of the local structure similarity in the data. Because of the sparsity property of the $l_{2,1}$ -norm regularization on weight vectors, the optimal weights contain some zero or close to zero row vectors. Structured sparsity is then imposed through penalizing all the regression coefficients corresponding to each single feature, at multiple time points. Thus, the most distinctive and predictive features will have similar large weights across all time points.

Optimization Algorithms

Although our objective function is convex, it is difficult to solve because regularization terms are based on non-smooth sparsity-inducing norms in the objective function (Zhu et al., 2013, 2014a). $l_{2,1}$ norm minimization is more challenging to solve than the l_1 -norm minimization problem. As most existing optimization algorithms are too computation costly to solve our problem (Wee et al., 2012; Zhang and Shen, 2012), an efficient iterative algorithm is developed in this work.

In the similarity measurement, a Laplacian graph at each time point is built based on a diagonal matrix and formulated as: $D_f = f_{uv}, D_s = \phi_{mn}, D_c^{(t)} = \psi_{pqt}$. Let S_f, S_s and $S_c^{(t)}$ denote the summation of the diagonal entry of D_f, D_s , and $D_c^{(t)}$, respectively. The graph Laplacian L_f for the feature space at each time point is: $L_f = D_f - S_f$. Similarly, we have the Laplacian graph for the subject and clinical score, $L_s = D_s - S_s$ and $L_c^{(t)} = D_c^{(t)} - S_c^{(t)}$, respectively. As any regularization term $R_s(W^{(t)})$ can be reformulated as $R_f(W^{(t)}) = Tr((W^{(t)})^T L_f W^{(t)}), R_s(W^{(t)}) = Tr((W^{(t)})^T L_s W^{(t)}), R_c(W^{(t)}) = Tr((W^{(t)})^T L_c^{(t)} W^{(t)})$. Assuming L_D is Laplacian graph built based on a diagonal matrix of \widehat{W} , the objective is first reformulated as:

$$J(W^{(t)}) = \min_{\{W^{(t)}, t=1, \dots, T\}} \sum_{t=1}^T \|Y^{(t)} - XW^{(t)}\|_F^2 + \lambda_1 \sum_{t=1}^T Tr((W^{(t)})^T L_f W^{(t)} + (XW^{(t)})^T L_s XW^{(t)} + W^{(t)T} L_c^{(t)} (W^{(t)})^T) + \lambda_2 Tr(\widehat{W}^T L_D \widehat{W}), \tag{9}$$

where L_D is Laplacian graph built based on a diagonal matrix of \widehat{W} . The optimal solution of $W^{(t)}$ is obtained by taking the derivative of the objective function with respect to $W^{(t)}$.

By taking the derivative of the objective function in Equation (8) with respect to $W^{(t)}$ and set to 0, we obtain:

$$X^T X - X(Y^{(t)})^T + \lambda_1 (L_f W^{(t)} + X^T L_s XW^{(t)} + W^{(t)T} L_c^{(t)}) + \lambda_2 L_D W^{(t)} = 0. \tag{10}$$

We can rewrite Equation (10) as

$$(X^T X + \lambda_1 L_f + \lambda_1 X^T L_s X + \lambda_2 L_D)W^{(t)} + W^{(t)} (\lambda_1 L_c^{(t)}) = X(Y^{(t)})^T. \tag{11}$$

This equation is regarded as Sylvester equation and solvable in the closed form Zhu et al. (2014a) using

$$AW^{(t)} + W^{(t)}B = Q, \tag{12}$$

where $A = X^T X + \lambda_1 L_f + \lambda_1 X^T L_s X + \lambda_2 L_D, B = \lambda_1 L_c^{(t)}$ and Q is $X(Y^{(t)})^T, W^{(t)}, (1 \leq t \leq T)$ can be obtained by solving the Sylvester equations (Zhu et al., 2014a) when the time t changes from 1 to T .

Since $L_f, L_s, L_c^{(t)}$, and L_D are obtained from $W^{(t)}$ and are dependent of $W^{(t)}$, an iterative optimization is proposed to efficiently obtain the global solutions of $W^{(t)}, (1 \leq t \leq T)$ (Zhu et al., 2013). The solution of $W^{(t)}$ for $1 \leq t \leq T$ is summarized in **Algorithm 1**. The iterative optimization method updates $W^{(t)}$ until the objective function converged.

EXPERIMENTAL RESULTS

A 10-fold cross validation strategy is employed to avoid any bias introduced in the data and experiments. A set of 445 subjects comprising of 91 NC, 202 patients with MCI, and 152 AD patients are included in our study. We also used subjects from the ANDI database with T1-weighted MRI data from a 1.5 T scanner. The entire set of subjects is equally partitioned into 10 subsets, and the subjects of one subset are selected as the testing samples and the subjects in the remaining nine subsets are used for training regression models. In our experiments, the regression model is implemented using the LIBSVM toolbox with default parameters, and a linear kernel is adopted after normalizing each feature vector into a unit norm. In the pre-processing step, the features are z-normalized by removing their mean and dividing the result by its standard deviation. The two cognitive measurements, ADAS-Cog and MMSE, are computed

Algorithm 1 | An iterative algorithm to solve the optimization problem in Equation (8)

Input: Baseline MRI training data of S subjects and F dimensional feature: $X \in \mathbf{R}^{S \times F}$
 T time points clinical scores of S subjects and C dimensional clinical score vector: $Y = \{Y^{(t)} \in \mathbf{R}^{S \times C}, t = 1, \dots, T\}$ Parameters: regularization parameters and iteration times

Output: Weight projection matrix: $W = \{W^{(t)} \in \mathbf{R}^{F \times C}, t = 1, \dots, T\}$

Set iteration $r = 0$ and initialize $W^{(t)} \in \mathbf{R}^{F \times C}$ according to the linear model for each time point;

Initialization: $\widehat{W}^{(0)} = [W^{(1)}, W^{(2)}, \dots, W^{(t)}, \dots, W^{(T)}]$

Repeat

for $t = 1$ to T

 Calculate $L_f, L_s, L_c^{(t)}$ and L_D , according to the above definitions;

 Update $\widehat{W}_f^{(t)}$ by solving the Sylvester problem in equation (13);

End for

$\widehat{W}_{r+1}^{(t)} = [W^{(1)}, W^{(2)}, \dots, W^{(t)}, \dots, W^{(T)}];$

$r = r + 1;$

until ($r = 50$ or $\widehat{W} - W_2 < 10^{-6}$)

Return $W^{(t)}, (1 \leq t \leq T)$

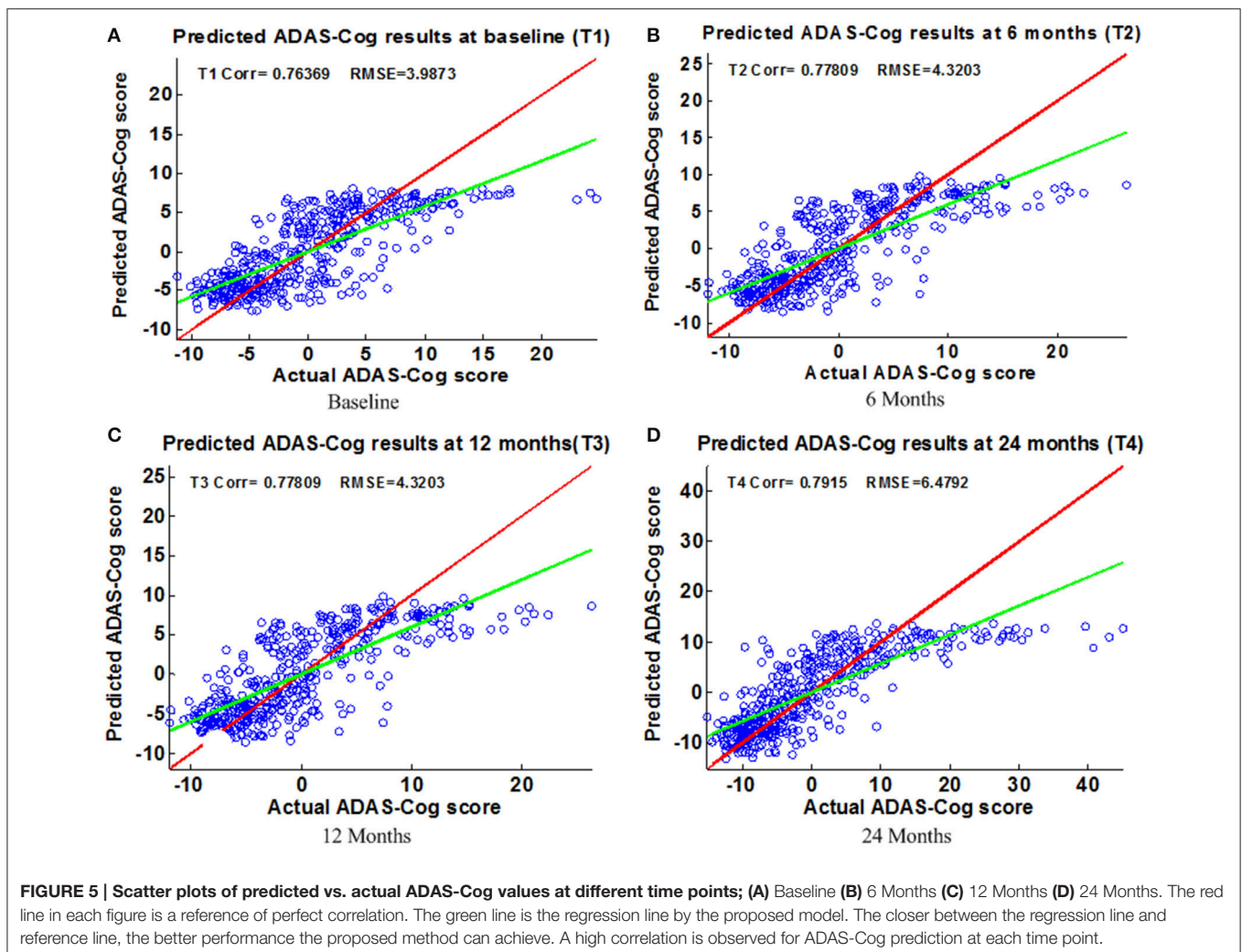
from the MRI data collected at four different time points. The experimental setup from Zhou et al. (2013) is adopted. The regularization parameters in the feature-selection model of Equation (8) (i.e., λ_1 and λ_2) are determined by performing another round of cross-validation on the training data.

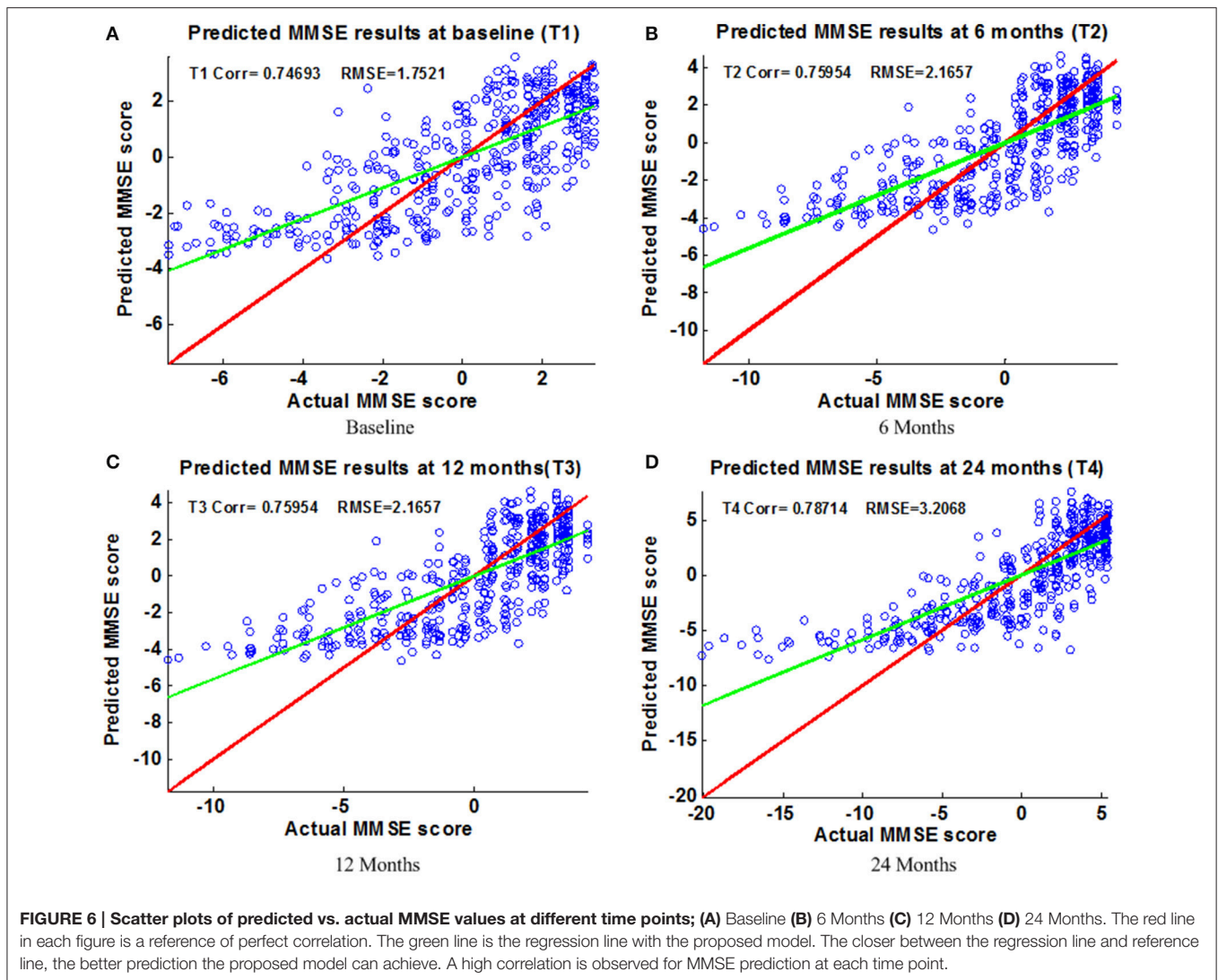
Figure 5 summarizes the objective function values at different iterations. The objective function values monotonically decrease as the number of iterations increases, which is consistent with our convergence analysis. The objective value quickly converged after a few iterations, which demonstrates the effectiveness of the proposed optimization method and the efficacy of our feature selection algorithms.

Extensive experiments are performed to evaluate the effectiveness of the proposed feature selection method. Two regression tasks are constructed to predict the changes in ADAS-Cog and MMSE for the baseline and a two-year follow-up study. The widely used Pearson's correlation coefficient (Corr) and root-mean-square error (RMSE) metrics are used to measure performance (Duchesne et al., 2009; Ito et al., 2010; Stonnington et al., 2010; Zhou et al., 2013).

ADAS-cog and MMSE Prediction Results

Figure 6 shows the scatter plots of the estimated ADAS-Cog scores vs. the actual ADAS-Cog scores obtained with the proposed method at baseline, 6, 12, and 24 months. **Figure 7** shows the scatter plots of the predicted MMSE vs. the actual MMSE scores obtained with the proposed method at baseline, 6, 12, and 24 months. The linear model adopted in the proposed method is illustrated with a red line, and the perfect regression method is shown by a green line for comparison. Although predicting future clinical scores is quite challenging, it is evident that our proposed method achieves remarkable results in terms of Corr and RMSE results. Similar to the previous studies (Zhang and Shen, 2012; Zhang et al., 2012; Zhou et al., 2013), we observed that predicting early and changes of MCI up to 1 year is more difficult than later time points as less distinct information is available to separate early MCI. The low correlation values in the early time periods are mainly due to the failure to detect progression from MCI to early AD, and it is even harder to uncover essential changes of brain regions in early MCI. Our experimental results show that the proposed multiple time point





joint learning with multiple relationships information is better at predicting ADAS-Cog and MMSE scores than single-task learning and the separate learning methods. We also address the problem of predicting the future cognitive decline of MCI subjects.

Algorithm Comparison

A comparison of our proposed method with three feature selection methods, namely, Lasso, temporal group Lasso (TGL, Zhang et al., 2012), and convex fused temporally constrained group Lasso (cFSGL, Zhou et al., 2013) is performed. **Figure 8** shows the performance of these methods on predicting ADAS-Cog/MMSE scores at baseline (T1), 6 months (T2), 12 months (T3), and 24 months (T4). We also present the detailed algorithm comparison results in terms of Corr and RMSE in **Table 2**. Our experimental results demonstrate that the proposed method performs better than the separate learning method in predicting ADAS-Cog and MMSE scores. We observed that the prediction of the changes of early and first year clinical score

is significantly harder than the later time points since there is less distinct information available for the earlier prediction, which was also confirmed by previous studies (Zhang et al., 2012; Zhou et al., 2013). The main reason for low correlation in the early time frame is that there is not sufficient time for MCI to progress to early AD, thus it is more challenging to uncover the essential changes of brain regions in early MCI. Our proposed method achieves stable and promising results for both ADAS-Cog and MMSE prediction, and outperforms several state-of-the-art methods (Zhang et al., 2012; Zhou et al., 2013). From the experimental results, the promising prediction results clearly suggest the effectiveness of the proposed method.

Next, we compare different sub-groups containing all subjects including AD, MCI, NC (ALL), and only MCI patients (MCI). **Figure 9** presents the detailed comparison results of the different sub-groups for ADAS-Cog and MMSE prediction. Since our method is jointly learned, better results are observed from the ALL group. This group provides the opportunity

for the proposed algorithm to discover multi-relationship information using multi-task learning. **Table 3** shows the detailed comparison of the state-of-the-art methods. The proposed

method with multi-relationship information obtains remarkable results, especially for the prediction of later time points for both ADAS-Cog and MMSE, which are consistent with the findings in the previous publications as well (Zhang et al., 2012; Zhou et al., 2013).

Top Selected Brain Regions

The hippocampus plays an important role in identifying brain conditions through AD modeling and measuring the cognitive outcomes such as ADAS-Cog and MMSE. By extracting features from the associated brain regions, the multi-task mechanism accounts for the hippocampus region's sensitivity in cognitive score prediction via MRI baseline data. MCI is characterized by the temporal lobe neocortical regions during dementia decline. Therefore, it is of great significance to find biomarkers for the diagnosis of AD. In this section, we investigate multiple regression variables and their relationships with brain variables such as ADAS-Cog and MMSE. Based on the cross validation, the top predictive regions were selected in terms of the frequency of feature appearance. **Figure 10** shows the top 30 most predictive brain regions with the highest weights for ADAS-Cog and MMSE predictions of our proposed and cFSGL methods, where darker colors denote larger weights and vice versa. The most predictive

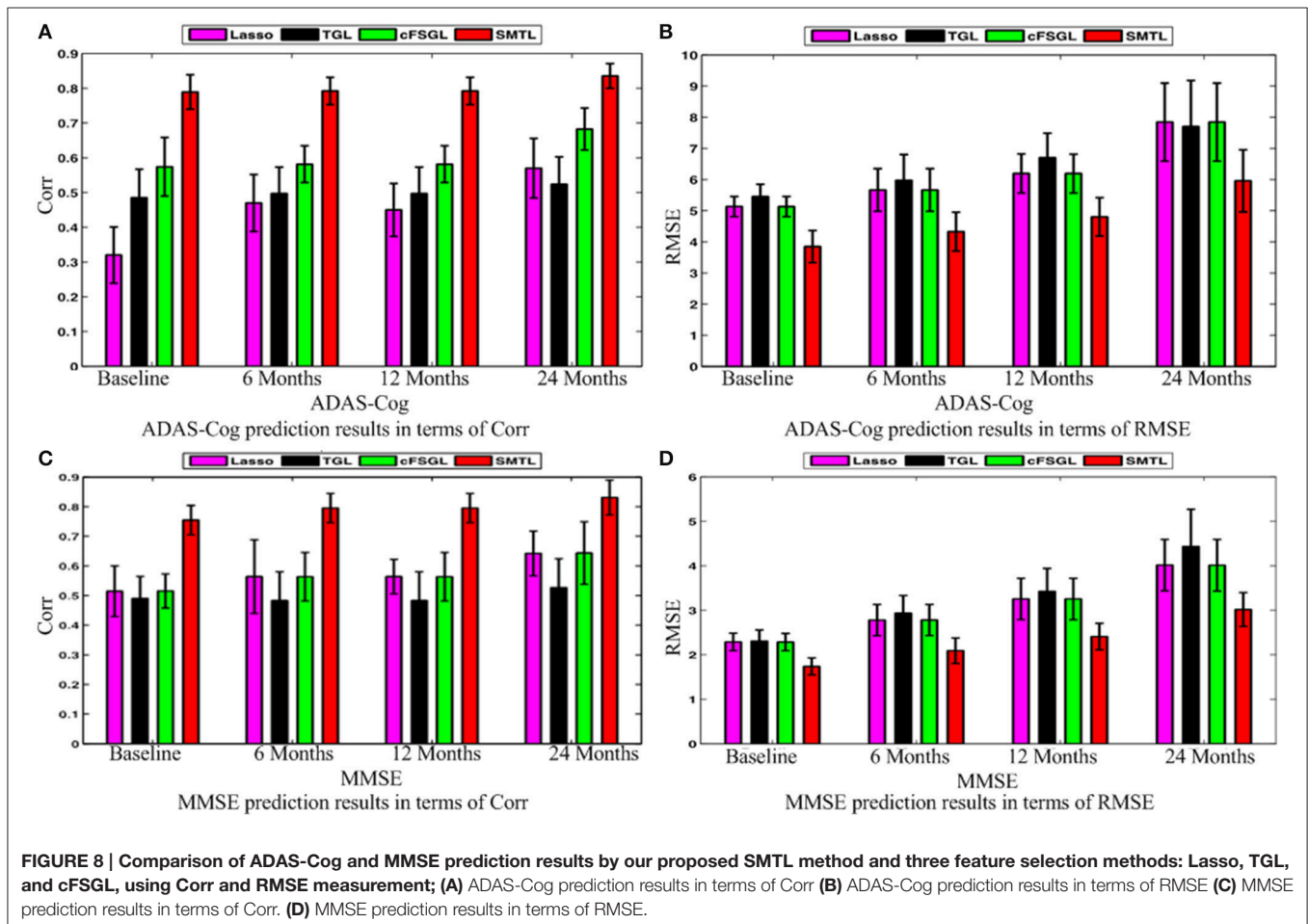
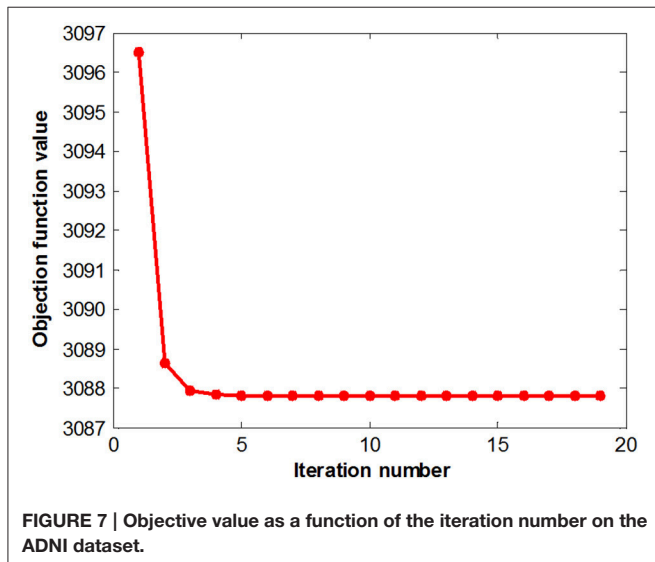


Table 2 | Comparison of ADAS-Cog and MMSE prediction results by our proposed SMTL method and three feature selection methods: Lasso, TGL, and cFSGL, using Corr and RMSE measurement.

	ADAS-Cog				MMSE			
	Lasso	TGL	cFSGL	SMTL	Lasso	TGL	cFSGL	SMTL
Baseline Corr	0.32 ± 0.08	0.49 ± 0.08	0.58 ± 0.08	0.77 ± 0.05	0.51 ± 0.09	0.56 ± 0.12	0.64 ± 0.08	0.75 ± 0.08
M06 Corr	0.48 ± 0.07	0.50 ± 0.07	0.59 ± 0.05	0.78 ± 0.04	0.49 ± 0.07	0.48 ± 0.10	0.63 ± 0.10	0.79 ± 0.10
M12 Corr	0.46 ± 0.12	0.52 ± 0.09	0.62 ± 0.05	0.79 ± 0.04	0.52 ± 0.06	0.57 ± 0.08	0.64 ± 0.12	0.79 ± 0.12
M24 Corr	0.57 ± 0.08	0.53 ± 0.08	0.69 ± 0.05	0.84 ± 0.04	0.61 ± 0.05	0.58 ± 0.05	0.65 ± 0.05	0.83 ± 0.06
Baseline RMSE	5.19 ± 0.33	5.53 ± 0.32	5.20 ± 0.33	3.81 ± 0.45	2.25 ± 0.20	2.30 ± 0.24	2.24 ± 0.21	1.75 ± 0.20
M06 RMSE	5.48 ± 0.62	5.90 ± 0.65	5.48 ± 0.63	4.36 ± 0.46	2.75 ± 0.31	2.85 ± 0.33	2.76 ± 0.32	2.31 ± 0.29
M12 RMSE	6.25 ± 0.58	6.74 ± 0.63	6.24 ± 0.59	4.91 ± 0.56	3.42 ± 0.49	3.59 ± 0.52	3.42 ± 0.50	2.48 ± 0.40
M24 RMSE	7.95 ± 1.20	7.81 ± 1.39	7.96 ± 1.23	6.00 ± 1.00	4.05 ± 0.58	4.48 ± 0.82	4.04 ± 0.58	3.00 ± 0.38

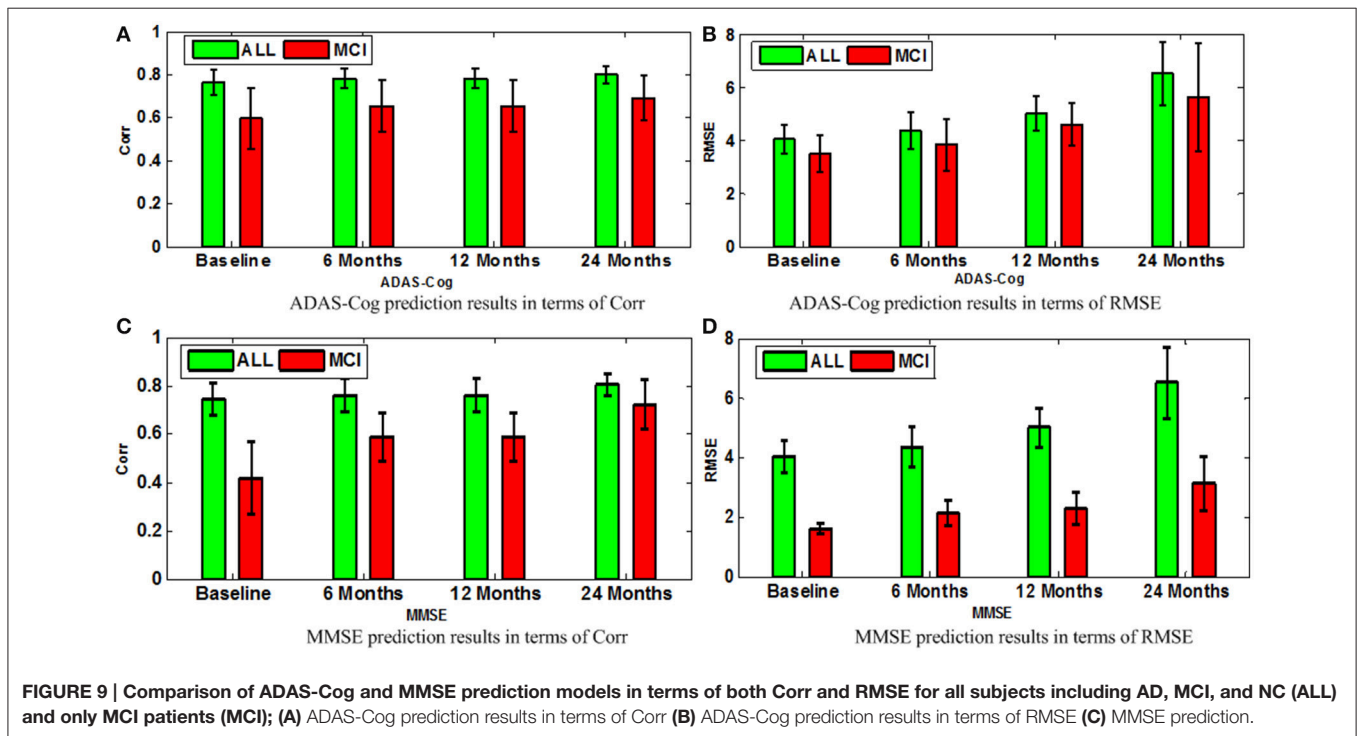


FIGURE 9 | Comparison of ADAS-Cog and MMSE prediction models in terms of both Corr and RMSE for all subjects including AD, MCI, and NC (ALL) and only MCI patients (MCI); (A) ADAS-Cog prediction results in terms of Corr (B) ADAS-Cog prediction results in terms of RMSE (C) MMSE prediction.

brain regions, namely, the hippocampus and amygdala, are commonly selected for regression tasks. Most of the commonly selected top regions such as hippocampal formation, amygdala, and uncus regions proved to be sensitive and provides AD biomarkers in many studies (Zhang et al., 2012; Zhou et al., 2013).

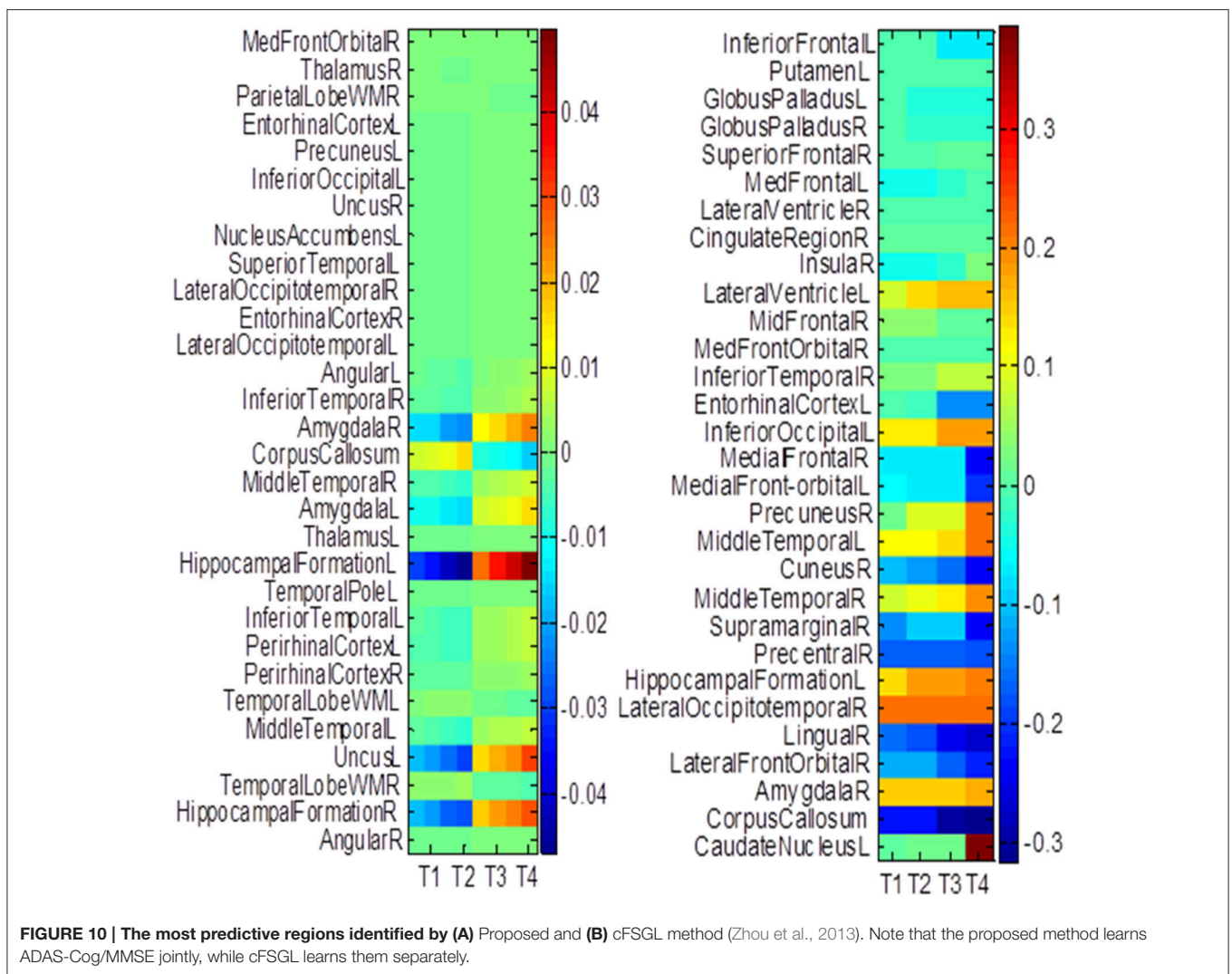
Biomarkers from different time-points are consistently identified, and important MRI patterns are localized, which suggest that the MRI biomarkers are able to predict the ADAS-Cog and MMSE results effectively. The distinctive and important biomarkers selected in our study included the hippocampal formation, amygdala, middle temporal lobe, uncus, and corpus callosum. The hippocampus is the most important region affecting AD as it possesses significant structural lesions (Convit et al., 2000; Laakso et al., 2000; Wolf et al., 2003; Del Sole et al., 2008; Knafo et al., 2009; Derflinger et al., 2011; Poulin

et al., 2011; Ota et al., 2014). Both cFSGL and proposed method selected the important biomarkers for AD diagnosis, while our proposed method identifies more stable and related features for the clinical score measurement. Moreover, the distinctive information selected by the proposed method is more promising for AD modeling than the cFSGL method, due to the joint learning adopted in a multi-task framework, and weights for both ADAS-Cog and MMSE are considered jointly, rather than separately. The common and shared high-level information of ADAS-Cog and MMSE scores can be further explored for disease progression modeling. As a result, the most predictive brain regions such as the hippocampus, amygdala, and temporal patterns are commonly selected in our regression models. The identified brain regions are consistent across multiple time points. The weights for ADAS-Cog and MMSE have similar patterns as well. Moreover, our proposed method confirms that

Table 3 | Algorithm comparisons of the proposed method with the related works for ADAS-Cog and MMSE prediction.

Method	Target	Subject	Feature	Result (Corr)
Duchesne et al., 2009	M12 MMSE	75 NC, 49 MCI, 75 AD	Baseline MRI, age, gender, years of education	MMSE: 0.31 ($p = 0.03$)
Stonnington et al., 2010	Baseline ADAS-Cog and MMSE	Set 1: 73 AD, 91 NC Set 2: (ADNI) 113 AD, 351 MCI, 122 NC	BaselineMRI, CSF	MMSE Set1: 0.7 ($p < 10e-5$) Set 2: 0.48 ($p < 10e-5$) ADAS-Cog Set 2: 0.57 ($p < 10e-5$)
Zhang et al., 2012	Baseline, M06, M12, M24 ADAS-Cog, MMSE	ADNI: 91 AD, 202 MCI, 152 NC	Baseline, M06, M12, M24 MRI	Average MMSE: 0.613 ($p < 10e-5$) Average ADAS-Cog: 0.639 ($p < 10e-5$)
Proposed	Baseline, M06, M12, M24 ADAS-Cog, MMSE	ADNI: 91 AD, 202 MCI, 152 NC	Baseline MRI	Average MMSE: 0.7538 ($p < 10e-5$) Average ADAS-Cog: 0.7875 ($p < 10e-5$)

AD, MCI, and NC denote Alzheimer's disease patients, mild cognitive impairment patients, and normal controls, respectively.



the hippocampus and amygdala are highly important for studies on AD.

In general, higher weights lead to the selection of better features, namely, large weight values correspond to the

effectiveness of features to characterize the brain atrophy through quantitative measurements. The experiments on progression from MCI to AD produce similar results from predicting AD and NC in the training steps. ADAS-Cog and MMSE predictions

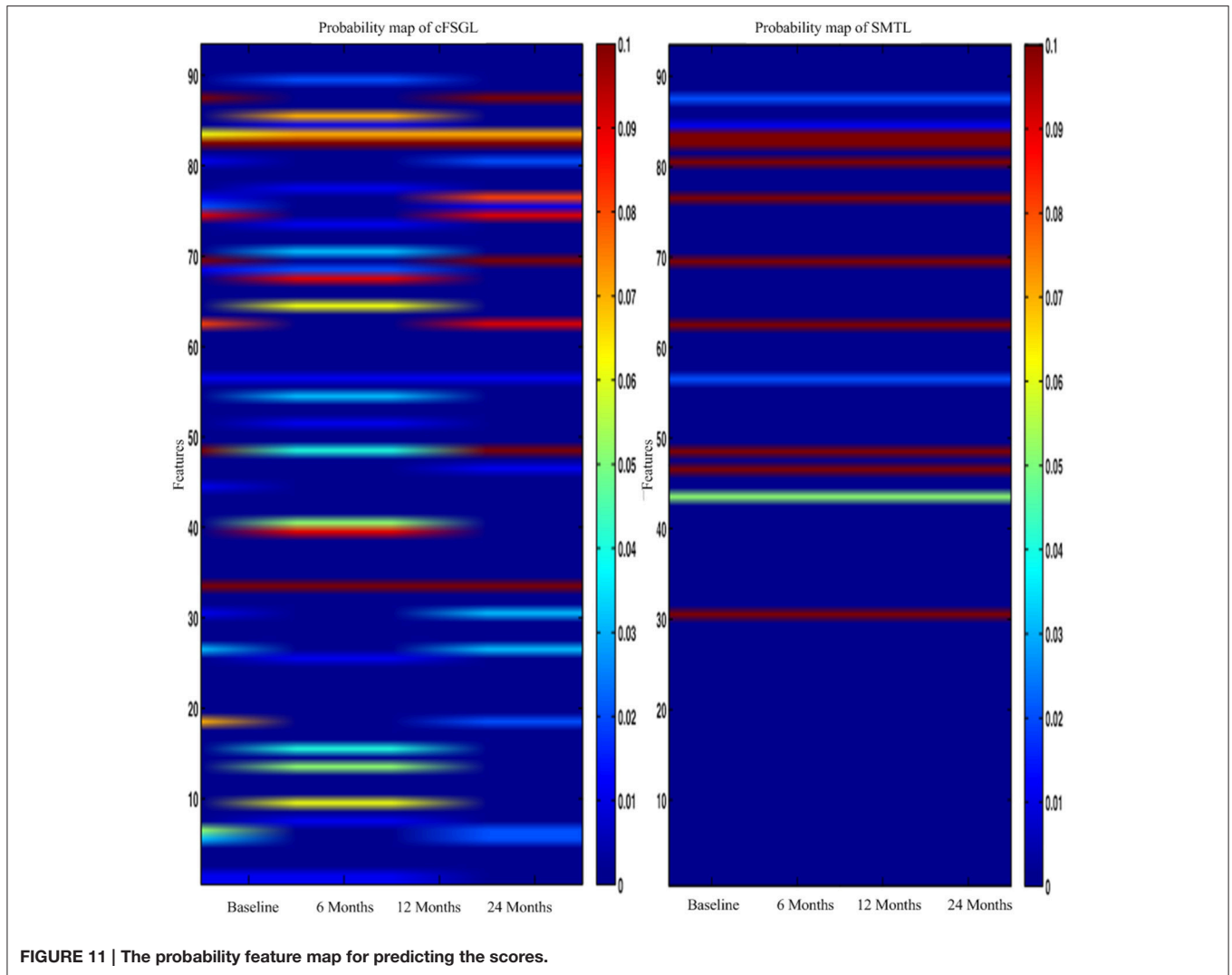


FIGURE 11 | The probability feature map for predicting the scores.

share common effective information. Other feature selection methods such as cFSGL demonstrate different patterns. As shown in the stable feature selection results, our method generally obtains more stable features for predicting both ADAS-Cog and MMSE scores at multiple time points. In general, our method outperforms the competing methods, which is also demonstrated in the probability map shown in **Figure 11**. The probability map further confirms that the hippocampus and amygdala are highly correlated with AD. The features extracted from these regions have high predictive power for the ADAS-Cog and MMSE prediction.

As illustrated in **Figure 12**, the most effective features for predicting ADAS-Cog and MMSE across all time points are, namely, angular R, right hippocampal formation, right temporal lobe WM, left uncus, left middle temporal, right perirhinal cortex, left perirhinal cortex, left inferior temporal, and left temporal. Our findings are consistent with many previous studies which determined the amygdala, hippocampal formation, angular, and uncus are the most predictive biomarkers for AD characterization (Convit et al., 2000; Laakso et al., 2000; Wolf

et al., 2003; Del Sole et al., 2008; Knafo et al., 2009; Derflinger et al., 2011; Poulin et al., 2011; Ota et al., 2014).

DISCUSSIONS AND CONCLUDING REMARKS

These relationships could provide inherent high-level information on AD progression, therefore modeling and utilizing such relationship information can enhance the learning performance of AD progression prediction. Although there are a myriad of progression models available, such as regression and survival model (Wang et al., 2010), we used a popular SVR model due to its desirable properties. The accurate prediction of ADAS-Cog and MMSE scores is essential in modeling disease progression of AD, and to assist clinical assessment. It is shown that clinical scores are highly correlated with ventricle change, shape, hippocampal, and gray matter volume loss. Because the medial temporal lobe and brain atrophy pattern are highly related to MMSE, the correlation should be utilized for clinical

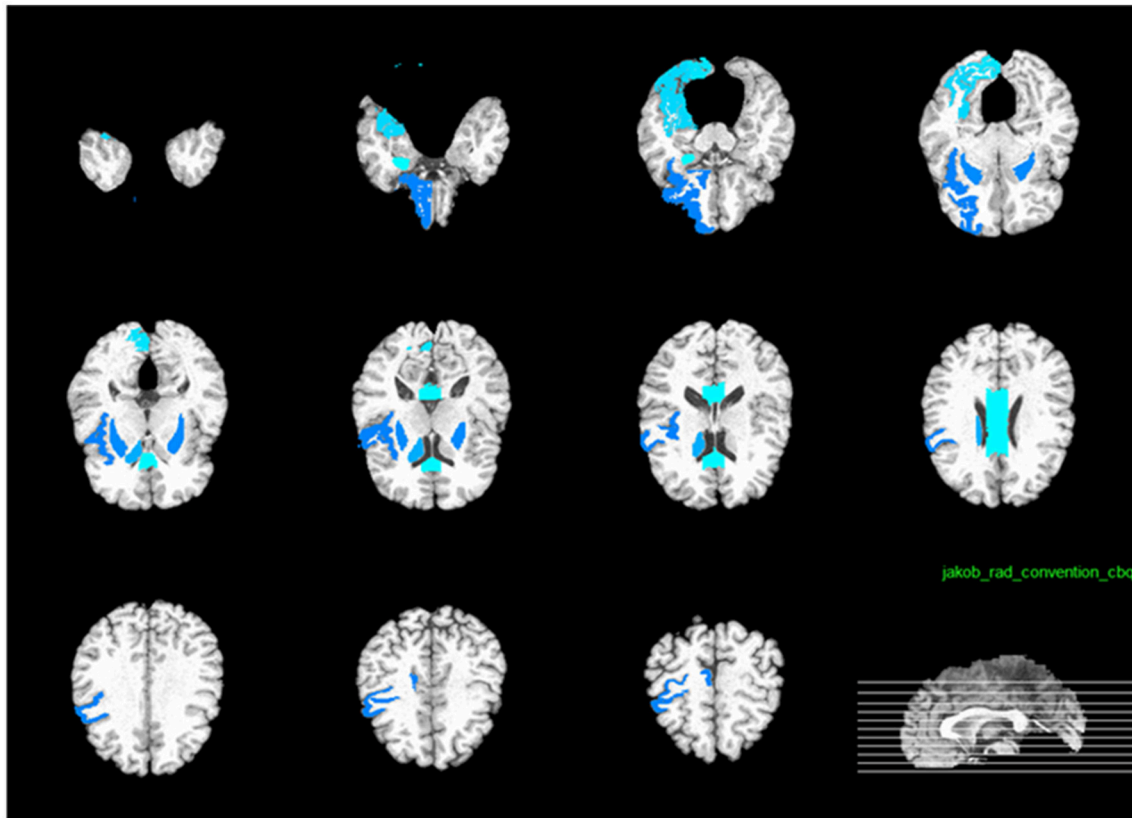


FIGURE 12 | Top 10 most predictive brain regions for predicting the cognitive measurements.

score prediction. Apart from feature selection, clinical model and statistical analysis extract diagnostic information and learn large-scale medical imaging based features. The advantages of our models include discrete timestamps at irregular intervals for modeling the disease progression of AD, and the greater interpretability of our model due to the improvement in the selection of distinct features.

Even though promising prediction results were obtained in our study, there are still some limitations in the proposed feature selection method. First, only a single modality is used. If multiple modalities were available (such as PET, DTI, APOE, and CSF), the predictive power could be possibly enhanced. Second, all subjects should have complete feature values with no missing data. This would be challenging to achieve with larger experiments and more subjects. Future research can be conducted to investigate the issues associated with missing data. Third, we only investigate the ADAS-Cog and MMSE predictions. It is possible to extend this method to predict the sub-scores of clinical tests and other regression variables, such as the clinical dementia rating scale sum of boxes (CDR-SOB) and the auditory verbal learning test (AVLT) for cognitive measurement.

A problem that is often referred to as the curse of dimensionality, where the large size (i.e., large number of dimensions) of features for modeling AD progression makes it difficult to perform various numerical analyses on the data. This

problem led to increased difficulty to draw consistent conclusions from the dataset. The feature selection method addressed this problem by leveraging the sparse learning technique to model the AD progressive and predict AD progression using the temporal priors and hippocampus. Apart from feature selection, the clinical model and statistical analysis extracted diagnostic information and learned large-scale medical imaging based on features. By merging fused multi-task learning together with the temporal smoothing of the parametric hippocampus surface, clinical scores, and subjects, promising prediction results for future ADAS-Cog and MMSE prediction were obtained.

The main challenge is to identify task-specific features and significant biomarkers to model AD, as well as locating a common set of features using the learned model. Previous works showed that performance decreases with a smaller training set (Stonnington et al., 2010), but the trend and relative performance remain comparable. An interesting direction is to add constraints to find similar parametric surface to determine more similar and smoother weights, and hence better and more consistent results can be obtained. This work illustrated that the extraction of the volumetric information from parametric surface could aid in the prediction of AD progression, which could be also extended to fMRI studies. Because the dimensionality problem still exists when the number of voxel and vertex points increases, feature selection is generally a plausible approach to leverage prior

information and explore sparsity and smoothness as well. Sparse learning is a powerful tool to identify useful features and reduce feature dimension. Although there are a number of feature selection algorithms available, the features from these algorithms often lack biological meanings and reasonable interpretations. A method that is capable of locating desirable features with reasonable feature dimension is highly desirable.

Directions for future work include understanding the behavior of weights across the parametric surface space and time. Previous works shown that stability selection (Zhou et al., 2013) may be a good fit for analyzing the feature weights on the model. Future works including the stability analysis of weights which might provide more information on the relationship between the deformation of hippocampal subfields and other clinical indicators, such as AVLT during AD progression. The longitudinal information on brain structure is highly correlated to disease progression. Hippocampal atrophic rates and ventricular changes are assessed statistically with the surface change. The resulting maps are sensitive to longitudinal changes in brain structure as the disease progresses. Additional maps to localize atrophic change regions are linked to cognitive decline. Additional maps for hippocampal atrophy and clinical deterioration are also helpful for understanding AD progression. These quantitative, dynamic visualizations of hippocampal atrophy and ventricular expansion rates in aging and AD may provide a promising measure to track AD progression in drug trials. Furthermore, it would be interesting to investigate the feasibility of extending our joint learning method to model and predict other diseases, such as Parkinson's disease or Autism disorder.

In this paper, we proposed a novel longitudinal prediction model which incorporates multiple relation information of data in a unified objective function. We applied our proposed model for AD progression prediction at multiple future time points, using the baseline data only. Specifically, we developed a novel multi-task sparse feature selection model by considering the relationships between features, subjects, and clinical scores. The

feature selection procedure selects the most relevant features for the task of clinical scores prediction at multiple future time points, followed by the use of regression models for predictions. Our experimental results of the proposed method based on the ADNI database demonstrated promising results in estimating the clinical cognitive scores at multiple future time points.

AUTHOR CONTRIBUTIONS

BL, FJ, SC, DN, and TW designed the experiments; BL and DN performed the data analysis and experiments; BL wrote the main manuscript text; SC and TW provided scientific interpretation; BL, FJ, SC, DN, and TW reviewed the manuscript, provided final approval of the version to be published and ensured the accuracy and integrity of the work.

ACKNOWLEDGMENTS

This work was supported partly by National Natural Science Foundation of China (Nos. 61402296, 61571304, 81571758, 61501305, 61571304, and 61427806), National Key Research and Develop Program (No. 2016YFC0104703), (Key) Project of Department of Education of Guangdong Province (No. 2014GKXM052), Shenzhen Key Basic Research Project (Nos. JCYJ20150525092940986, JCYJ20150525092940982, and JCYJ20140509172609164), Guangdong Medical Grant (No. B2016094), and the Open Fund Project of Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (No. MJUKF201711), the National Natural Science Foundation of Shenzhen University (No. 2016077). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering.

REFERENCES

- Alzheimer's Association (2014). 2014 Alzheimer's disease facts and figures. *Alzheimers Dement.* 10, e47–e92. doi: 10.1016/j.jalz.2014.02.001
- Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 15, 1373–1396. doi: 10.1162/089976603321780317
- Brookmeyer, R., Johnson, E., Ziegler-Graham, K., and Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement.* 3, 186–191. doi: 10.1016/j.jalz.2007.04.381
- Caroli, A., and Frisoni, G. B. (2010). The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease neuroimaging initiative cohort. *Neurobiol. Aging* 31, 1263–1274. doi: 10.1016/j.neurobiolaging.2010.04.024
- Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199
- Convit, A., de Asis, J., de Leon, M. J., Tarshish, C. Y., De Santi, S., and Rusinek, H. (2000). Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiol. Aging* 21, 19–26. doi: 10.1016/S0197-4580(99)00107-4
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.-O., et al. (2011). Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage* 56, 766–781. doi: 10.1016/j.neuroimage.2010.06.013
- Daianu, M., Mendez, M. F., Baboyan, V. G., Jin, J., Melrose, R. J., Jimenez, E. E., et al. (2016). An advanced white matter tract analysis in frontotemporal dementia and early-onset Alzheimer's disease. *Brain Imaging Behav.* 10, 1038–1053. doi: 10.1007/s11682-015-9458-5
- Davatzikos, C., Bhatt, P., Shaw, L. M., Batmanghelich, K. N., and Trojanowski, J. Q. (2010). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* 32, 2322.e19–2322.e27. doi: 10.1016/j.neurobiolaging.2010.05.023
- Davatzikos, C., Fan, Y., Wu, X., Shen, D., and Resnick, S. M. (2008). Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiol. Aging* 29, 514–523. doi: 10.1016/j.neurobiolaging.2006.11.010
- Davatzikos, C., Genc, A., Xu, D., and Resnick, S. M. (2001). Voxel-based morphometry using the RAVENS Maps: methods and validation

- using simulated longitudinal atrophy. *Neuroimage* 14, 1361–1369. doi: 10.1006/nimg.2001.0937
- Del Sole, A., Clerici, F., Chiti, A., Lecchi, M., Mariani, C., Maggiore, L., et al. (2008). Individual cerebral metabolic deficits in Alzheimer's disease and amnesic mild cognitive impairment: an FDG PET study. *Eur. J. Nucl. Med. Mol. Imaging* 35, 1357–1366. doi: 10.1007/s00259-008-0773-6
- Derflinger, S., Sorg, C., Gaser, C., Myers, N., Arsic, M., Kurz, A., et al. (2011). Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized. *J. Alzheimers Dis.* 25, 347–357. doi: 10.3233/JAD-2011-110041
- Dickerson, B. C., Goncharova, I., Sullivan, M. P., Forchetti, C., Wilson, R. S., Bennett, D. A., et al. (2001). MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiol. Aging* 22, 747–754. doi: 10.1016/S0197-4580(01)00271-8
- Duchesne, S., Caroli, A., Geroldi, C., Collins, D. L., and Frisoni, G. B. (2009). Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. *Neuroimage* 47, 1363–1370. doi: 10.1016/j.neuroimage.2009.04.023
- Fan, Y., Resnick, S. M., Wu, X., and Davatzikos, C. (2008). Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage* 41, 277–285. doi: 10.1016/j.neuroimage.2008.02.043
- Frisoni, G. B., Fox, N. C., Jack C. R. Jr., Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 6, 67–77. doi: 10.1038/nrneuro.2009.215
- Gaser, C., Nenadic, I., Buchsbaum, B. R., Hazlett, E. A., and Buchsbaum, M. S. (2001). Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. *Neuroimage* 13, 1140–1145. doi: 10.1006/nimg.2001.0771
- Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182. doi: 10.1162/15324430322753616
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797
- Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., and Johnson, S. C. (2009). Spatially augmented LBoosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48, 138–149. doi: 10.1016/j.neuroimage.2009.05.056
- Hinrichs, C., Singh, V., Xu, G., and Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 55, 574–589. doi: 10.1016/j.neuroimage.2010.10.081
- Huang, L., Gao, Y., Jin, J., Thung, K.-H., Shen, D., and Alzheimer's Disease Neuroimaging Initiative. (2015). “Sparse regression based random forest for predicting future clinical scores of Alzheimer's disease,” in *6th International Conference on Machine Learning in Medical Imaging, LNCS 9532* (Munich), 246–254.
- Huang, L., Jin, Y., Gao, Y., Thung, K.-H., and Shen, D. (2016). Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiol. Aging* 46, 180–191. doi: 10.1016/j.neurobiolaging.2016.07.005
- Ito, K., Ahadi, S., Corrigan, B., French, J., Fullerton, T., and Tensfeldt, T. (2010). Disease progression meta-analysis model in Alzheimer's disease. *Alzheimer's Dement.* 6, 39–53. doi: 10.1016/j.jalz.2009.05.665
- Jack, C. R. Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049
- Jack C. R. Jr., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 9, 119–128. doi: 10.1016/S1474-4422(09)70299-6
- Jie, B., Zhang, D., Cheng, B., and Shen, D. (2015). Manifold regularized multitask feature learning for multimodality disease classification. *Hum. Brain Mapp.* 36, 489–507. doi: 10.1002/hbm.22642
- Jin, Y., Huang, C., Daianu, M., Zhan, L., Dennis, E. L., Reid, R. I., et al. (2017). 3D tract-specific local and global analysis of white matter integrity in Alzheimer's disease. *Hum. Brain Mapp.* doi: 10.1002/hbm.23448. [Epub ahead of print].
- Jin, Y., Shi, Y., Zhan, L., and Thompson, P. M. (2015). “Automated multi-atlas labeling of the fornix and its integrity in Alzheimer's disease,” in *IEEE 12th International Symposium on Biomedical Imaging* (Brooklyn Bridge, NY), 140–143.
- Knafo, S., Venero, C., Merino-Serrais, P., Fernaud-Espinosa, I., Gonzalez-Soriano, J., Ferrer, I., et al. (2009). Morphological alterations to neurons of the amygdala and impaired fear conditioning in a transgenic mouse model of Alzheimer's disease. *J. Pathol.* 219, 41–51. doi: 10.1002/path.2565
- Laakso, M. P., Frisoni, G. B., Könönen, M., Mikkonen, M., Beltramello, A., Geroldi, C., et al. (2000). Hippocampus and entorhinal cortex in frontotemporal dementia and Alzheimer's disease: a morphometric MRI study. *Biol. Psychiatry* 47, 1056–1063. doi: 10.1016/S0006-3223(99)00306-6
- Lei, B., Chen, S., Ni, D., and Wang, T. (2016). Discriminative learning for Alzheimer's Disease diagnosis via canonical correlation analysis and multimodal fusion. *Front. Aging Neurosci.* 8:77. doi: 10.3389/fnagi.2016.00077
- Lei, B., Ni, D., Chen, S., and Wang, T. (2015a). “Joint learning of multiple longitudinal prediction models by exploring internal relations,” in *Machine Learning in Medical Imaging*, eds L. Zhou, L. Wang, Q. Wang, and Y. Shi (Springer International Publishing), 9352, 330–337. doi: 10.1007/978-3-319-24888-2_40
- Lei, B., Yao, Y., Chen, S., Li, S., Li, W., Ni, D., et al. (2015b). Discriminative learning for automatic staging of placental maturity via multi-layer fisher vector. *Sci. Rep.* 5:12818. doi: 10.1038/srep12818
- Leow, A. D., Klunder, A. D., Jack, C. R. Jr., Toga, A. W., Dale, A. M., Bernstein, M. A., et al. (2006). Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage* 31, 627–640. doi: 10.1016/j.neuroimage.2005.12.013
- Liu, F., Wee, C.-Y., Chen, H., and Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. *Neuroimage* 84, 466–475. doi: 10.1016/j.neuroimage.2013.09.015
- Liu, M., Zhang, D., and Shen, D. (2012). Ensemble sparse classification of Alzheimer's disease. *Neuroimage* 60, 1106–1116. doi: 10.1016/j.neuroimage.2012.01.055
- Liu, M., Zhang, D., and Shen, D. (2014). Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. *Hum. Brain Mapp.* 35, 1305–1319. doi: 10.1002/hbm.22254
- Liu, M., Zhang, D., and Shen, D. (2015). View-centralized multi-atlas classification for Alzheimer's disease diagnosis. *Hum. Brain Mapp.* 36, 1847–1865. doi: 10.1002/hbm.22741
- Misra, C., Fan, Y., and Davatzikos, C. (2009). 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* 44, 1415–1422. doi: 10.1016/j.neuroimage.2008.10.031
- Ota, K., Oishi, N., Ito, K., and Fukuyama, H. (2014). A comparison of three brain atlases for MCI prediction. *J. Neurosci. Methods* 221, 139–150. doi: 10.1016/j.jneumeth.2013.10.003
- Poulin, S. P., Dautoff, R., Morris, J. C., Barrett, L. F., and Dickerson, B. C. (2011). Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res.* 194, 7–13. doi: 10.1016/j.psychres.2011.06.014
- Shen, D., and Davatzikos, C. (2002). HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21, 1421–1439. doi: 10.1109/TMI.2002.803111
- Shi, L., Zhao, L., Wong, A., Wang, D., and Mok, V. (2015). Mapping the relationship of contributing factors for preclinical Alzheimer's Disease. *Sci. Rep.* 5:11259. doi: 10.1038/srep11259
- Stanciu, S. G., Xu, S., Peng, Q., Yan, J., Stanciu, G. A., Welsch, R. E., et al. (2014). Experimenting liver fibrosis diagnostic by two photon excitation microscopy and bag-of-features image classification. *Sci. Rep.* 4:4636. doi: 10.1038/srep04636
- Stonnington, C. M., Chu, C., Klöppel, S., Jack, C. R. Jr., Ashburner, J., and Frackowiak, R. S. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. *Neuroimage* 51, 1405–1413. doi: 10.1016/j.neuroimage.2010.03.051
- Teipel, S. J., Born, C., Ewers, M., Bokde, A. L., Reiser, M. F., Möller H.-J., et al. (2007). Multivariate deformation-based analysis of brain atrophy to predict

- Alzheimer's disease in mild cognitive impairment. *Neuroimage* 38, 13–24. doi: 10.1016/j.neuroimage.2007.07.008
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* 58, 267–288.
- Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., et al. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage* 39, 1186–1197. doi: 10.1016/j.neuroimage.2007.09.073
- Vemuri, P., Wiste, H. J., Weigand, S. D., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., et al. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. *Neurol* 73, 294–301. doi: 10.1212/WNL.0b013e3181af79fb
- Wang, H., Nie, F., Huang, H., Yan, J., Kim, S., Risacher, S., et al. (2012). High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. *Adv. Neural Inf. Process. Syst.* 2, 1277–1285.
- Wang, Y., Fan, Y., Bhatt, P., and Davatzikos, C. (2010). High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. *Neuroimage* 50, 1519–1535. doi: 10.1016/j.neuroimage.2009.12.092
- Wang, T., Shi, F., Jin, Y., Jiang, W., Shen, D., and Xiao, S. (2016a). Abnormal changes of brain cortical anatomy and the association with plasma microRNA107 level in amnesic mild cognitive impairment. *Front. Aging Neurosci.* 8:112. doi: 10.3389/fnagi.2016.00112
- Wang, T., Shi, F., Jin, Y., Yap, P.-T., Wee, C.-Y., Zhang, J., et al. (2016b). Multilevel deficiency of white matter connectivity networks in Alzheimer's disease: a diffusion MRI study with DTI and HARDI models. *Neural Plast.* 2016:2947136. doi: 10.1155/2016/2947136
- Wang, Y., Nie, J., Yap, P.-T., Li, G., Shi, F., Geng, X., et al. (2014). Alzheimer's Disease neuroimaging initiative. Knowledge-guided robust mri brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. *PLoS ONE* 9:e77810. doi: 10.1371/journal.pone.0077810
- Wee, C. Y., Yap, P. T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., et al. (2012). Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59, 2045–2056. doi: 10.1016/j.neuroimage.2011.10.015
- Wolf, H., Jelic, V., Gertz, H. J., Nordberg, A., Julin, P., and Wahlund, L. O. (2003). A critical discussion of the role of neuroimaging in mild cognitive impairment. *Acta Neurol. Scand.* 107, 52–76. doi: 10.1034/j.1600-0404.107.s179.10.x
- Wu, H., Chen, G., Jin, Y., Shen, D., and Yap, P.-T. (2016). Embarrassingly parallel acceleration of global tractography via dynamic domain partitioning. *Front. Neuroinform.* 10:25. doi: 10.3389/fninf.2016.00025
- Yang, X., Jin, Y., Chen, X., Zhang, H., Li, G., and Shen, D. (2016). "Functional connectivity network fusion with dynamic thresholding for MCI diagnosis," in *7th International Conference on Machine Learning in Medical Imaging, LNCS 10019* (Athens), 246–253.
- Yuan, M., and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. B* 68, 49–67. doi: 10.1111/j.1467-9868.2005.00532.x
- Zhang, D., Liu, J., and Shen, D. (2012). "Temporally-constrained group sparse learning for longitudinal data analysis," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*, eds N. Ayache, H. Delingette, P. Golland, and K. Mori (Berlin; Heidelberg: Springer), 7512, 264–271. doi: 10.1007/978-3-642-33454-2_33
- Zhang, D., and Shen, D. (2012). Alzheimer's Disease neuroimaging initiative. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS ONE* 7:e33182. doi: 10.1371/journal.pone.0033182
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424
- Zhou, F., Jiao, J., and Lei, B. (2015). A linear threshold-hurdle model for product adoption prediction incorporating social network effects. *Inf. Sci.* 307, 95–109. doi: 10.1016/j.ins.2015.02.027
- Zhou, J., Liu, J., Narayan, V. A., and Ye, J. (2013). Modeling disease progression via multi-task learning. *Neuroimage* 78, 233–248. doi: 10.1016/j.neuroimage.2013.03.073
- Zhu, X., Suk, H. I., and Shen, D. (2014b). A novel multi-relation regularization method for regression and classification in AD diagnosis. *Med. Image Comput. Comput. Assist. Intervent.* 17, 401–408. doi: 10.1007/978-3-319-10443-0_51
- Zhu, X., Suk, H. I., and Shen, D. (2014a). A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *Neuroimage* 100, 91–105. doi: 10.1016/j.neuroimage.2014.05.078
- Zhu, X., Wu, X., Ding, W., and Zhang, S. (2013). "Feature selection by joint graph sparse coding," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, 803–811. doi: 10.1137/1.9781611972832.89
- Zhuo, Y., Ni, D., Chen, S., Lei, B., and Wang, T. (2016). "Adaptive ensemble manifold learning for neuroimaging retrieval," in *13th International Symposium on Biomedical Imaging (Prague)*, 859–862.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lei, Jiang, Chen, Ni and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.