Published in final edited form as:

Med Image Anal. 2019 April ; 53: 111–122. doi:10.1016/j.media.2019.01.007.

Multi-task Exclusive Relationship Learning for Alzheimer's Disease Progression Prediction with Longitudinal Data

Mingliang Wang[a], **Daoqiang Zhang**[a,†], **Dinggang Shen**[b,†], and **Mingxia Liu**[b,†]

[a]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

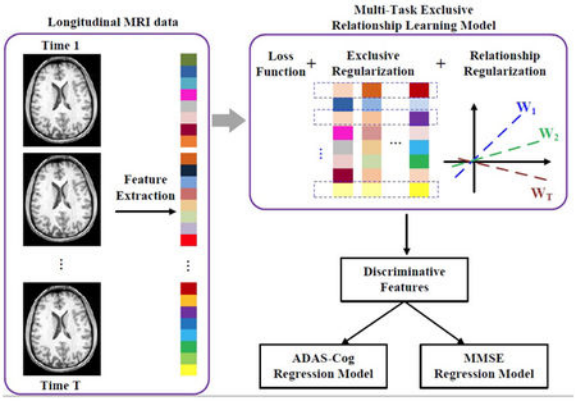
[b]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA

Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder characterized by progressive impairment of memory and other cognitive functions. Currently, many multi-task learning approaches have been proposed to predict the disease progression at the early stage using longitudinal data, with each task corresponding to a particular time point. However, the underlying association among different time points in disease progression is still under-explored in previous studies. To this end, we propose a multi-task exclusive relationship learning model to automatically capture the intrinsic relationship among tasks at different time points for estimating clinical measures based on longitudinal imaging data. The proposed method can select the most discriminative features for different tasks and also model the intrinsic relatedness among different time points, by utilizing an exclusive lasso regularization and a relationship induced regularization. Specifically, the exclusive lasso regularization enables partial group structure feature selection among the longitudinal data, while the relationship induced regularization efficiently introduces the relationship information from data to guide knowledge transfer. We further develop an efficient optimization algorithm to solve the proposed objective function. Extensive experiments on both synthetic and real datasets demonstrate the effectiveness of our proposed method. In comparison with several state-of-the-art methods, our proposed method can achieve promising performance for cognitive status prediction and also can help discover disease-related biomarkers.

Graphical Abstract

[†]Corresponding authors.: wml489@nuaa.edu.cn (Mingliang Wang), dqzhang@nuaa.edu.cn (Daoqiang Zhang), dgshen@med.unc.edu (Dinggang Shen), mxliu@med.unc.edu (Mingxia Liu).



Keywords

Alzheimer's disease; Longitudinal analysis; Exclusive lasso; Clinical status

1. Introduction

Alzheimer's disease (AD), characterized by progressive impairment of memory and cognitive functions, is the most common type of dementia in elderly people. As life expectancy increases, the number of AD patients will also increase correspondingly, resulting in a heavy socio-economic burden (Fan et al., 2008). It was reported that there are 26.6 million AD cases in the world in 2006, and about 56% of the cases are at the early stage (called mild cognitive impairment, MCI). In 2050, the AD population will increase to over 100 million (Brookmeyer et al., 2007). With the need for markers that can track the progress of the disease becomes increasingly urgent, many research groups have devoted their efforts to improve the understanding and monitoring of AD progression (Liu et al., 2015; Hinrichs et al., 2011; P et al., 2009).

Neuroimaging has proven to be a powerful tool for characterizing the neurodegenerative process of AD (Stonnington et al., 2010a; Wang et al., 2012a; Lian et al., 2019, 2018). In recent decades, neuroimaging-based longitudinal studies have been widely investigated to predict cognitive status, using data at multiple time points. Many cognitive measures have been designed to evaluate the cognitive decline, *e.g.*, the Mini Mental State Examination (MMSE) and the Alzheimer's Disease Assessment Scale Cognitive subscale (ADAS-Cog), which can be used to partially reveal AD progression. It has been reported that MMSE is correlated with the underlying AD pathology and neurodegenerative mechanism (Petrella et al., 2003), and ADAS-Cog is the gold standard in AD drug trial for cognitive function assessment (Rosen et al., 1984). However, to accurately predict the progression of AD/MCI remains a challenging task, since it is difficult to model the association between imaging features and a specific cognitive measure at multiple time points.

To address the above challenges, there already exist several feature selection efforts dedicated to AD disease progression modeling, which have been demonstrated as a useful way to reflect the association among prediction tasks at multiple time points. These efforts

generally fall into two categories:1) single-task learning, and 2) multi-task learning. In the first category, the disease progression at different time points is estimated separately. That is, neither the association between imaging features and a specific task nor the correlation among tasks is explored in single-task learning. For instance, Duchesne *et al.* (Duchesne et al., 2009) proposed to use robust linear regression model to predict one-year MMSE changes from baseline magnetic resonance imaging (MRI) data, and revealed that the baseline MRI data are highly associated with one-year cognitive changes. Wang *et al.* (Wang et al., 2010) developed a linear relevance vector regression model to predict continuous clinical variables. Since single-task learning treats each task as a stand-alone task (*i.e.*, without considering the intrinsic association among different tasks), the performance of these approaches could be suboptimal in progression predicting of brain diseases.

As another line of e orts, the multi-task learning (MTL) paradigm (Caruana, 1997; Zhou et al., 2013; Jie et al., 2017; Liu et al., 2016a; Nie et al., 2017; Liu and Zhang, 2016a) has been widely studied for the prediction of cognitive status in the domain of brain disease analysis. For instance, Zhou *et al.* (Zhou et al., 2013) proposed to integrate temporal smoothness into their multi-task learning model for predicting the disease progression and selecting predictive markers of the progression. Jie *et al.* (Jie et al., 2017) developed a temporally-constrained group sparse learning method for longitudinal analysis. Nie *et al.* (Nie et al., 2017) designed a novel multi-source multi-task learning method to evaluate the disease status using the prior knowledge of source consistency and temporal smoothness, with the effectiveness validated on the progression prediction of AD suffers. Most of the existing multi-task learning methods focus on learning multiple related tasks (*e.g.*, estimating multiple types of cognitive scores) simultaneously, and introducing particular regularization terms to model the prior knowledge of the longitudinal data (*e.g.*, different tasks share a common subset of features). For instance, in predicting cognitive scores, we usually expect to discover the commonality across different time points, thereby a group lasso regularizer is widely used to impose the structured sparsity on a parameter matrix in the longitudinal analysis (Zhou et al., 2013; Jie et al., 2017). The underlying assumption for group lasso regularizer is that these tasks are equally related to each other. That is, when one feature is important for several tasks, it is also expected to be important for the remaining tasks (Zhang et al., 2011; Liu et al., 2016b). We argue that such an assumption is too strong, because the true correlation among tasks for different clinical score predictions is actually unknown in practice. Also, if a specific feature is only discriminant to a particular task, conventional group lasso regularizer is prone to ignore this feature, because the group lasso regularizer mainly focuses on shared features among all tasks. Thus, more properly designed structured sparsity induced norms are desired in the longitudinal analysis for AD progression prediction.

In this paper, we propose a multi-task exclusive relationship learning (MTERL) approach to select predictive markers for disease progression prediction. In particular, our method can learn the intrinsic relatedness among multiple cognitive measures from data automatically, without assuming them to be known in advance. Specifically, we define a novel objective function to utilize the relationship information from adjacent time points. We first learn a least square regression model by using data from each time points, and further introduce the new mixed structured sparsity norms to overcome the above drawback in the existing sparse

learning based feature selection models. For accurate identification of effective imaging markers, we utilize the exclusive lasso regularization which can select a specific set of biomarkers at different time points for disease progression prediction. different from the group lasso regularization, if one feature in a group is given a large weight, the other features in the same group will tend to be assigned with small weights or even zero. In addition, to reflect the intrinsic relatedness from adjacent time points, we introduce the relationship induced regularization that can automatically learn the relatedness among the tasks from data. Unlike most previous multi-task based regression models (*e.g.*, (Zhou et al., 2013; Jie et al., 2017) that assumed the tasks shared a common set of features), we do not need to have prior knowledge about the longitudinal data and our model will learn the intrinsic relatedness from multiple-time-point data automatically, which is one of our major contributions and not investigated before.

To evaluate the efficiency of our proposed method, we perform experiments on both synthetic and real datasets. For real dataset, we conduct experiments on 445 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database with MRI data. Specifically, each subject has baseline MRI data and two types of clinical scores (*i.e.*, MMSE and ADAS-Cog) at 4 different time points (*i.e.*, baseline, 6-month, 12-month and 24-month). The experimental results show that the proposed method yields clearly improved prediction performance, compared with methods in previous studies.

The remainder of the paper is structured as follows. In Section 2, we describe the image pre-processing method and our proposed multi-task exclusive relationship learning approach. The experimental settings and experimental results are then described in Section 3 and Section 4, respectively. In Section 5, we investigate the influence of parameters, the significance of selected biomarkers, as well as the limitations of our method. Finally, a conclusion of this paper is presented in Section 6.

2. Materials and Method

2.1. Subjects

Data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.adni-info.org). ADNI is funded $60 million 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 in 2003. One goal of ADNI is to verify whether the serial MRI and positron emission tomography (PET), along with other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

In this study, a total of 445 ADNI subjects with all corresponding MRI data as well as two types of cognitive scores (*i.e.*, MMSE and ADAS-Cog) at 4 different time points (*i.e.*, baseline, 6-month, 12-month and 24-month) are used. Specifically, there are 91 AD, 152 NC, and 202 MCI subjects used in this study. Also, among 202 MCI subjects, there are 104 MCI converters (MCI-C) who progressed to AD in 36 months after the baseline time, and 98 MCI non-converters (MCI-NC) who did not progress to AD. In Table 1, we show the

detailed demographic and clinical information of the studied subjects. More details on acquiring MRI data from ADNI can be found at (Zhang et al., 2011; Zhang and Shen, 2012a). Briefly, the structural MR scan was obtained from each subject by using 1.5T scanners. We download the Raw Digital Imaging and Communications in Medicine (DICOM) MRI data from the public ADNI website (http://www.adni.loni.usc.edu), reviewed for quality, and automatically corrected for spatial distortion caused by gradient nonlinearity and B1 field inhomogeneity.

2.2. Image Pre-processing

In this study, image pre-processing is performed following the same procedures as in (Jie et al., 2017; Zhang et al., 2011; Zhang and Shen, 2012a). Specifically, the anterior commissure (AC) - posterior commissure (PC) correction was first performed for all images, followed by the N3 algorithm (Sled et al., 1998) to correct the intensity inhomogeneity. Subsequently, skull stripping was performed on structural MR images, after which the cerebellum was removed from the skull-stripping image. Then, the FAST method (Zhang et al., 2001) was applied to segment the brain into three tissues, including grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). After segmentation, all images were registered by HAMMER registration method (Shen et al., 2003) to a template with 93 manually-labeled Regions-Of-Interest (ROIs) (Kabani, 1998). Finally, the brain image of each subject was partitioned into 93 ROIs by atlas warping, and the total GM volume of each ROI was computed as the feature representation. In this study, we only use GM for feature extraction, because GM is the most affected tissue by AD and also widely used in the neuroimaging-based analysis (Lei et al., 2017; Zhang et al., 2011). Similar to previous studies (Wee et al., 2012; Zhou et al., 2013), we also normalize the obtained features to facilitate disease diagnosis and prognosis.

2.3. Multi-Task Exclusive Relationship Learning

In this paper, we focus on the problem of predicting longitudinal cognitive trajectories through time by using neuroimaging data. Since the cognitive scores from different time points can provide complementary perspectives on neuropsychological assessments, we aim to seek a jointly multi-task linear regression model to reveal the longitudinal associations between the phenotypic markers and the cognitive trajectories. In the following, we will elaborate the multi-task exclusive relationship learning (MTERL) model in detail.

For AD progression prediction using longitudinal phenotypic markers, we assume that the number of training subjects is N, and each subject has its corresponding imaging data at T *different* time points, represented as $\mathbf{X}_t = \left[\mathbf{x}_{t,1}; \mathbf{x}_{t,2}; \cdots; \mathbf{x}_{t,N}\right] \in \mathscr{R}^{N \times d}$ where $\mathbf{x}_{t,n} \in \mathscr{R}^{1 \times d}$ is a d-dimensional row vector at the t-th ($t = 1, \cdots, T$) time point. Denote $\mathbf{Y}_t = \left[y_{t,1}; y_{t,2}; \cdots; y_{t,N}\right] \in \mathscr{R}^{N \times 1}$ as the disease status at the t-th time point for all training subjects. We denote $f_t(x) = \mathbf{X}_t\mathbf{w}_t$ as the linear predictive function to estimate the clinical score from imaging data at the t-th time point, where $\mathbf{w}_T \in \mathscr{R}^{d \times 1}$ is the feature weight vector. Let $\mathbf{W} = \left[\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_T\right] \in \mathscr{R}^{d \times T}$ denote the weight matrix we aim to learn for all T

time points, with column vector corresponding to one specific task. We thus have the following standard exclusive lasso model:

$$\min_{\mathbf{W}} J(\mathbf{W}) = \frac{1}{2T}\sum_{t=1}^{T}\left\|\mathbf{Y}_t - \mathbf{X}_t\mathbf{w}_t\right\|_2^2 + \frac{\lambda}{2}\sum_{i=1}^{d}\|\mathbf{W}^i\|_1^2, \quad (1)$$

where the first term is a square loss function that measures the loss on the training data for all tasks. Let $h(\mathbf{W}) = \sum_{i=1}^{d}\|\mathbf{W}^i\|_1^2$ be the *exclusive lasso term* (Zhou et al., 2010), where $\mathbf{W}^i$ is the i-th row of the weight matrix $\mathbf{W}$. Different from the conventional $\ell_{2,1}$-norm regularization which encourages the common imaging biomarkers selected for all tasks, the exclusive lasso regularization assumes a competitive nature among the features shared by all the tasks. That is, if a feature is assigned a very large weight in one task, the weights for the same feature in the other tasks are expected to be small or even zero. As a result, it can capture the consecutive changes in the brain between adjacent time points, helping to reveal the unique information conveyed in MRI at a particular time point.

So far, the standard exclusive lasso model has been able to capture the partial group structures across all the cognitive scores, *i.e.,* each imaging marker has either small score or large score for each specific cognitive measure. Besides, to automatically learn the relationship of different prediction tasks, we propose to use the *relationship induced term* (Zhang and Yeung, 2010a) in our proposed model, defined as follows:

$$\mathrm{tr}\left(\mathbf{W}\,\mathbf{\Omega}^{-1}\,\mathbf{W}^{\mathrm{T}}\right), \quad (2)$$

with $\mathbf{\Omega} \geq 0$ and $\mathrm{tr}(\mathbf{\Omega}) = 1$. Here, $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix, $\mathbf{\Omega}^{-1}$ denotes the inverse of the matrix $\mathbf{\Omega}$, and $\mathbf{\Omega}$ is defined as a task covariance matrix that will benefit learning on weight matrix by inducing the correct relationship in longitudinal analysis. The relatedness among multiple clinical scores helps reflect the smooth changes between data derived from adjacent time points. The constraint term $\mathbf{\Omega} \geq 0$ is used to restrict $\mathbf{\Omega}$ as positive semi-definite matrix, and $\mathrm{tr}(\mathbf{\Omega}) = 1$ is used to penalize the complexity of $\mathbf{\Omega}$.

By incorporating the relationship induced term defined in Eq. (2) into Eq. (1), we formulate our objective function as follows:

$$\begin{aligned}\min_{\mathbf{W},\mathbf{\Omega}} J(\mathbf{W}) &= \frac{1}{2T}\sum_{t=1}^{T}\|\mathbf{Y}_t - \mathbf{X}_t\mathbf{w}_t\|_2^2 \quad (3)\\ &+ \frac{\lambda}{2}\sum_{i=1}^{d}\|\mathbf{W}^i\|_1^2 + \frac{\gamma}{2}\mathrm{tr}\left(\mathbf{W}\,\mathbf{\Omega}^{-1}\,\mathbf{W}^{\mathrm{T}}\right),\\ \text{s.t.}\quad \mathbf{\Omega} &\succcurlyeq 0, \mathrm{tr}(\mathbf{\Omega}) = 1,\end{aligned}$$

where λ and γ are the regularization parameters to control the balance between the exclusive lasso regularization and the relationship induced regularization. It is easy to know that when $\gamma = 0$, our method will degenerate to the standard exclusive lasso method (Zhou et al., 2010). In the following section, we will develop an efficient optimization algorithm to solve the objective function defined in Eq. (3).

2.4. Optimization

The objective function in Eq. (3) is a convex optimization problem, and we propose to solve it by using an alternating algorithm (see Algorithm 1). Specifically, we first optimize $\mathbf{W}$ with a fixed $\mathbf{\Omega}$, and then optimize $\mathbf{\Omega}$ when $\mathbf{W}$ is fixed.

2.4.1. Optimizing w.r.t. W when Ω is fixed—When $\mathbf{\Omega}$ is given and fixed, the objective function of Eq. (3) is convex but non-smooth, because of the non-smooth of exclusive lasso regularization term. Fortunately, according to (Chen et al., 2012, 2011), the smoothed approximation function of exclusive lasso penalty term can be rewritten as:

$$\nabla h(\mathbf{W}) = \left[q_\mu\left(\mathbf{W}^1\right)\mathbf{V}^1\left(\mathbf{W}^1\right), \cdots, q_\mu\left(\mathbf{W}^d\right)\mathbf{V}^d\left(\mathbf{W}^d\right)\right], \quad (4)$$

where $q_\mu\left(\mathbf{W}^i\right) = \max_{\|\mathbf{V}^i\|_\infty} \left\langle \mathbf{W}^i, \mathbf{V}^i \right\rangle - \frac{\mu}{2}\|\mathbf{V}^i\|_2^2$ and μ is the positive smoothness parameter. The $\mathbf{V}^i(\mathbf{W}^i)$ is the unique minimizer of $q_\mu(\mathbf{W}^i)$, for a fixed $\mathbf{W}^i$, it can be denoted as $\min\{1, \max\{-1, \frac{\mathbf{W}^i}{\mu}\}\}$. Moreover, according to Chen et al. (2012, 2011), $\nabla h(\mathbf{W})$ is Lipschitz-continuous with the constant $L_\mu = \|X^{\mathrm{T}}X\|_2 + (\frac{R}{\mu} + T)d$. Hence, the i-th column gradient of Eq. (3) can be computed as:

$$\nabla_{\mathbf{w}_t} J\left(\mathbf{W}_t\right) = \mathbf{X}_t^{\mathrm{T}}\left(\mathbf{X}_t\mathbf{w}_t - \mathbf{Y}_t\right) + \lambda \nabla h\left(\mathbf{w}_t\right) + \gamma \mathbf{w}_t \mathbf{\Omega}_t^{-1} \ . \quad (5)$$

2.4.2. Optimizing w.r.t. Ω when W is fixed—When $\mathbf{W}$ is given and fixed, the optimization problem for finding $\mathbf{\Omega}$ can be stated as:

$$\begin{aligned} &min_{\mathbf{\Omega}} \mathrm{tr}\left(\mathbf{W}\ \mathbf{\Omega}^{-1}\ \mathbf{W}^{\mathrm{T}}\right) \quad (6)\\ &\text{s.t. } \mathbf{\Omega} \succcurlyeq 0, \mathrm{tr}(\mathbf{\Omega}) = 1. \end{aligned}$$

According to (Zhang and Yeung, 2010a; Zhang et al., 2010; Liu et al., 2016a), by taking the partial derivative of the Eq. (6), we can get the closed-form solution for $\mathbf{\Omega}$[1]:

[1]The source code for this paper is available at http://ibrain.nuaa.edu.cn/code/.

$$\mathbf{\Omega} = \frac{(\mathbf{W}^{\mathrm{T}}\mathbf{W})^{\frac{1}{2}}}{\mathrm{tr}((\mathbf{W}^{\mathrm{T}}\mathbf{W})^{\frac{1}{2}})}. \quad (7)$$

Algorithm 1:

Alternating optimization for solving the proposed MTERL model

Input: $\mathbf{X}$,$\mathbf{Y}$, λ, γ, μ;

Output: $\mathbf{w}_t$, $\mathbf{\Omega}$;

1 Initialize *Calculate* L_μ, *set* $\mathbf{w}_0$, $\mathbf{v}_0$ *and let* $\alpha_0 = 1$, $t = 0$;

2 **repeat**

3 Given $\mathbf{\Omega}$, optimization of $\mathbf{w}_t$:

4 $\mathbf{u}_t = (1 - \alpha_t)\mathbf{w}_t + \alpha_t\mathbf{v}_t$;

5 Compute the $\nabla h(\mathbf{u}_t)$ according to Eq. (4);

6 $\mathbf{v}_{t+1} = \mathbf{v}_t - \frac{1}{\alpha_t L_\mu}\left(\mathbf{X}_t^{\mathrm{T}}\left(\mathbf{X}_t\mathbf{w}_t - \mathbf{Y}_t\right) + \lambda\nabla h\left(\mathbf{w}_t\right) + \gamma\mathbf{w}_t\,\mathbf{\Omega}_t^{-1}\right)$;

7 $\mathbf{w}_t+1 = (1 - \alpha_t)\mathbf{w}_t + \alpha_t\mathbf{v}_t+1$;

8 $\alpha_{t+1} = \frac{2}{t+1}, t = t + 1$;

9 Given $\mathbf{w}_t$, optimization of $\mathbf{\Omega}$:

10 Calculate the analytical solution for $\mathbf{\Omega}$, according to Eq. (7);

11 **until** *Converges*;

3. Validation

To evaluate the efficacy of the proposed disease progression model, we perform experiments on 445 ADNI MRI data with the corresponding cognitive scores (*i.e.*, MMSE and ADAS-Cog), tracked over four different time points (*i.e.*, baseline, 6-month, 12-month and 24-month). In the experiments, we use a 10-fold cross-validation strategy to evaluate the performance of our proposed method and those compared methods. Specifically, the whole set of subjects are randomly partitioned into 10 subsets (with roughly equal size), and the subjects within one subset are selected as the test data each time, while all other subjects in the remaining nine subsets are used to train the regression models. The support vector regressor (SVR) and support vector classifier (SVC) with a linear kernel with a default parameter (*i.e.*, $C = 1$) (Chang and Lin, 2012) are used for regression and classification, respectively. We perform another 5-fold cross-validation to choose the parameters using a line search strategy in the range $[10^{-4}, 10^{-3}, \cdots, 10^{2}]$ in each fold. After cross-validation, we choose the parameters with the best performance on the training data. To avoid randomness during dataset partitioning, we repeat the process ten times. We report the final performance after averaging the results of the repeated cross-validations.

[2] https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Also, to examine that the selected compact set of imaging markers by the proposed method are closely related to cognitive trajectories, we compare the following methods in disease progression domain for evaluation.

1. **Lasso** (Hastie et al., 2000). Lasso, being regularized by ℓ_1-norm, performs variable selection to enhance the prediction/regression performance. So it has been one of the most widely used methods to estimate the future disease status by modeling the cognitive measures at different time points separately.
2. **Multi-task Feature Learning (MTFL)** (Liu et al., 2009). MTFL is considered to be the $\ell_{2,1}$-norm regularized regression model for joint feature selection from multiple tasks. One appealing feature of this method is that it encourages multiple predictors to share similar sparsity patterns.
3. **Multi-task Exclusive Lasso (MTEL)** (Zhou et al., 2010). Essentially, the proposed method is an extension of MTEL work without considering the intrinsic relatedness across multiple time points. different from the $\ell_{2,1}$-norm regularization, the exclusive lasso regularization models the scenario when variables in the same group compete with each other.
4. **Multi-task Relationship Learning (MTRL)** (Zhang and Yeung, 2010a). MTRL learns task relationship matrix under a regularization framework. This model can be viewed as a novel generalization of the regularization framework for single-task learning. The task relationship matrix and model parameters (*i.e.*, ρ and η which control the model complexity and relationship measure separately), can be learned iteratively.

Both methods are implemented with a least squared loss function. For Lasso, MTFL and MTEL the trade-off parameter is chosen from the set $[10^{-4}, 10^{-3}, \cdots, 10^{2}]$. The candidate set for both trade-off parameters in MTRL (*i.e.*, ρ and η) and MTERL (*i.e.*, λ and γ) is $[10^{-4}, 10^{-3}, \cdots, 10^{2}]$. Note that unless otherwise specified, all the regularization parameters are tuned via another 5-fold cross-validation on the training data. We evaluate the performance of different methods via three criteria, *i.e.*, normalized mean squared error (nMSE), root mean square error (rMSE) and Pearson's correlation coefficient (Corr). More specifically, the nMSE is commonly used in multi-task learning literature (Argyriou et al., 2008; Zhang and Yeung, 2010b), which is the mean squared error divided by the variance of the ground truth. The rMSE and Corr are widely used in measuring performances of regression and association analysis between the predicted and the actual clinical scores (Stonnington et al., 2010a; Duchesne et al., 2009). It is worth noting that, for the nMSE and rMSE criteria, the smaller the value, the better performance. For the Corr criteria, the greater the value, the better performance.

4. Experiment

In this section, we first show results on the simulation data, and then present the disease prediction results on the ADNI database with longitudinal brain MRI data.

4.1. Results on Simulation Data

In this subsection, we use a toy example to show that our MTERL method would infer a correct relationship before we do experiments on real datasets. Specifically, we first generate the toy dataset as follows. There are three regression tasks defined as $y = 4x + 13$, $y = -4x - 6$ and $y = 1$, respectively. For each task, we randomly generate 15 points from a uniform distribution in [0, 10]. The corresponding function outputs are corrupted by a Gaussian noise process with zero mean and variance equal to 0.1. One example of the toy data is plotted in Fig. 1, with each color (and point type) corresponding to a particular task. Based on the coefficients of three regression functions, we expect the correlation between the first two tasks to be −1, and the correlation among other tasks to approach 0. For the proposed MTERL model, we empirically set λ_1 to 0.01 and λ_2 to 0.005.

The correlation matrix (*i.e.*, $\mathbf{\Omega}$) learned by MTERL is shown in Fig. 2. Also, the estimated regression functions for the three tasks are $y = 4.1807x + 9.3771$, $y = -3.7859x - 9.2830$ and $y = 0.0622x + 0.0114$. These results demonstrate that the relationship among three tasks learned by our proposed MTERL method is consistent with our expectation. Besides, we show the change of the objective function values on simulation data in Fig. 3. From Fig. 3, we can see that the values of the objective function decrease rapidly within twenty iterations and then levels off, showing the fast convergence of the proposed alternating optimization method in Algorithm 1.

4.2. Results on Real-world Longitudinal Data

In this section, to validate the effectiveness of our proposed MTERL method, we conduct experiments on the real ADNI dataset to study how the changes in the brain are associated with different clinical dementia scores (*i.e.*, ADAS-Cog, and MMSE) at four time points (*i.e.*, baseline, 6-month, 12-month and 24-month). Specifically, we treat the prediction of the cognitive scores at each time point as a learning task, so we have 4 regression tasks. We first partition the brain image of each subject into 93 ROIs with pre-defined template and compute the total GM volume of corresponding region as the feature. Thus, we obtained a 93 dimensional feature vector for representing each subject. We then use the 93 dimensional representation data of 445 subjects and the corresponding cognitive scores to jointly learn the linear predictive models as formulated in Eq. (3).

4.2.1. Learned Relationship among Tasks—On the real-world ADNI dataset, we show the learned relationship (*i.e.*, $\mathbf{\Omega}$ in Eq. (3)) of MTRL and MTERL methods among four regression tasks for estimating ADAS-Cog and MMSE, respectively. The experimental results are shown in Fig. 4 and Fig. 5, where yellow in the color bar denotes a high positive correlation coefficient. From Fig. 4 and Fig. 5, one can see that the correlation of regression tasks at four time points shows significant difference. Also, compared with the learned relationship by MTERL, the relationship learned by MTRL is lower at four time points with respect to both ADASCog and MMSE, which is a bit counter-intuitive that multiple time points tend to play similar roles. Hence, exploiting the real relationships among multiple tasks, as we do via the regularization term in Eq. (3), could help improve the learning performance of individual regressors.

4.2.2. Regression Results—In Fig. 6, we show the comparison of nMSE, rMSE and correlation coefficients achieved by different methods, including Lasso, MTFL, MTRL and MTEL, in estimating the clinical scores of ADAS-Cog and MMSE at four time points.

From the results of longitudinal cognitive scores prediction in Fig. 6, we can derive several interesting observations. First, the joint learning methods considering the cognitive progression information from multiple time points generally achieve significantly better performance, compared with the single task learning method (*i.e.*, Lasso). This demonstrates that, the relation information among multiple time points is beneficial to the predicting of disease progression. Second, our proposed MTERL method, which models the intrinsic relation from adjacent time points automatically, consistently outperforms four competing methods regarding nMSE, rMSE and Corr. Specifically, MTERL achieves the average (*i.e.*, across four time points) nMSE of 0.56 and 0.61 for estimating ADAS-Cog and MMSE scores, respectively. Also, MTERL achieves the average rMSE of 0.76 and 0.79 for estimating ADAS-Cog and MMSE scores, respectively. These results validate the efficacy of our proposed method, which not only considers the partial group structures within phenotypic markers but also models the correlations among cognitive measures in jointly estimating the clinical scores based on the longitudinal analysis. Finally, it is worth noting that, Fig. 6 indicates that estimating later time point (*i.e.*, M24) scores often achieves better performance than estimating earlier time point scores. The possible reason for high performance at the late time point is that the relationship between imaging features and clinical scores becomes much stronger with the 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. The McNemar test (Dietterich, 1998) is used to evaluate the statistical significance of the difference between regression performances of two methods. We report the p-values in Fig. 6 and mark statistically significant differences ($p < 0.05$) with the asterisk (*).

We also show the scatter plots for the predicted values versus the actual values for ADAS-Cog and MMSE on the testing data in Fig. 7 and Fig. 8, respectively. Specifically, for each time point, we learn a linear SVR on training subjects and estimate ADAS-Cog and MMSE scores for testing subjects, respectively. From Fig. 7 and Fig. 8, we can observe that the predicted values achieved by our MTERL method are highly correlated to the actual clinical scores, manifested as high Corr values. Also, Fig. 7–8 show that the overall performance for estimating ADAS-Cog scores is better than that for estimating MMSE scores, which is similar to the previous studies (Zhou et al., 2013; Nie et al., 2017). These results further validate the effectiveness of the proposed MTERL method in estimating clinical scores using longitudinal data.

4.2.3. Results of MCI Conversion Prediction using Informative Biomarkers—In this set of experiments, we conduct the classification task of predicting the future conversion of MCI patients based on only the baseline MRI data, where the informative biomarkers discovered in the regression experiments are fed into a particular classifier. Specifically, for each competing method, we first build a prediction model using longitudinal training MRI data at four time points to select the important brain regions (with respect to ADAS-Cog/MMSE scores). Then, we choose the top cumulative absolute weight

across different time points as the important brain regions for the multi-task-based classification methods. It is worth noting that, for the Lasso method that learns at four time points separately, we select only the brain regions with higher absolute weights at the baseline for subsequent classification. Finally, we train the corresponding support vector machine (SVM) classifiers with different numbers of above-selected features (*w.r.t.*, brain regions) from the baseline training MRI data, respectively. In Figs. 9–10, we show the classification accuracy using different numbers of selected features on the baseline MRI data.

From Figs. 9–10, we can clearly see that the prediction results of our MTERL method consistently outperform the competing methods in nearly all the test cases. Also, MTERL can achieve the best performance with a smaller number (*i.e.*, 15) of selected features, in comparison to other methods. The underlying reason for the superiority of our MTERL method over four competing methods could be listed as follows. Lasso independently estimates the cognitive measures at different time points, which neglects the correlations across different time points. Although the MTFL method can learn the underlying relationship among the prediction tasks at different time points, it simply assumes that these tasks are closed to each other by using the $l_{2,1}$-norm, while such assumption is too strong and may not hold true in practice. Also, unlike MTRL and MTEL which consider one of the partial group structures within longitudinal phenotypic markers or the correlations shared among different tasks, the proposed MTERL method captures all the valuable information (*i.e.*, partial group structures and correlation shared among different tasks) for AD progression prediction.

We also summarize the performance of different methods in Table 2. We further use the McNemar's test (Dietterich, 1998) (with the significance level at 0.05) to assess whether the difference in AUC values between our proposed method and each competing method is significant, with the corresponding p-values reported in Table 2. From Table 2, we can see that the proposed MTERL method consistently outperforms the compared methods in all performance measures. For instance, the ACC value obtained by our method is 71.98% in the conversion of MCI classification, which is better than the second-best result 69.28% yielded by MTFL method with guidance from ADAS-Cog. Also, our method achieves a classification accuracy of 73.04%, a sensitivity of 78.29%, and a specificity of 68.89% with the guidance from MMSE. These results are consistently better than other methods on each performance measure. In addition, Table 2 also indicates that the joint analysis of longitudinal data would be beneficial to use the complete information for the identification of relevant imaging markers (Zhou et al., 2011; Wang et al., 2012b).

4.2.4. Informative Brain Regions—It is meaningful to identify a subset of biomarkers which are highly correlated to the AD progression. Therefore, we investigate the top ten brain regions identified by our proposed method in this section. It is worth noting that, because the selected brain regions are different in each 10-fold cross-validation, we choose the cumulative absolute weight (Yahata et al., 2016) as an indicator of the brain region contribution in the regression task. Fig. 11 shows the top 10 identified important brain regions selected by the proposed method and those selected brain region names are listed in Table 3.

From Table 3, we can see that the most informative brain regions that are commonly selected in two regression tasks include, the *corpus callosum*, *hippocampus* and *amygdala*. In previous studies (Lei et al., 2017; Jie et al., 2017; Wang et al., 2012c; Zhang and Shen, 2012a; Zhou et al., 2013; Pan et al., 2018), these regions are reported to be highly associated with the Alzheimer's disease. Meanwhile, the *hippocampus* plays an important role in identifying brain conditions through AD modeling and measuring the cognitive outcomes such as ADAS-Cog and MMSE, which has been confirmed by existing works (Derflinger et al., 2011; Knafo et al., 2009; Convit et al., 2000). In addition, the selected brain regions such as *uncus* and *middle temporal regions* are also sensitive to AD diagnosis, which also has been observed in previous studies (Zhou et al., 2013; Zhang et al., 2012; Braak and Braak, 1991; Delacourte et al., 1999). To sum up, the identified imaging markers are highly suggestive and effective for tracking the progression of AD, since it strongly agrees with the existing research findings.

4.2.5. Comparison with State-of-the-Art Methods—Furthermore, we compare the results achieved by our method with several recent state-of-the-art results reported in the literature studying the relationship between cognitive scores and imaging markers (*i.e.*, clinical score regression based on MRI data). We list the details of each method and the corresponding results in Table 4. From Table 4, we can see that our method obtains competitive results in most cases. Specifically, our method achieves the average correlation coefficient of longitudinal analysis with respect to ADAS-Cog and MMSE scores are 0.664 and 0.625, respectively. It is worth noting that, although a previous study in (Zhang and Shen, 2012a) reported a higher correlation coefficient, a relatively smaller number of subjects with multi-modality data are used in their work.

5. Discussion

In this section, we first investigate the influence of parameters. Then we analyze the effectiveness of the biomarkers identified by the proposed method. Finally, we further discuss the limitations of the proposed method.

5.1. Influence of Parameters

There are two important parameters (*i.e.*, λ and γ) in our proposed method, which is used to balance the relative contribution of two regularization terms. Now we investigate the influence of these two parameters on the performances of our method in the regression tasks with respect to ADAS-Cog and MMSE scores. Specifically, we vary the value of λ and γ from the range $\{10^{-4}, 10^{-3}, \cdots, 10^{2}\}$, and record results achieved by our method in both tasks of ADAS-Cog and MMSE score regression in Fig. 12 and Fig. 13, respectively. From Fig. 12 and Fig. 13, we can see that our method achieves relatively stable results using different parameters, demonstrating that our method is not very sensitive to parameters.

5.2. effectiveness of MRI Biomarkers

The correlation analysis between biomarkers and the clinical scores reveals the effective for tacking the progression of AD, which has been widely used in medical imaging analysis (Morgan et al., 2017). In Fig. 14 and Fig. 15, we show the relationship between the MMSE

score and the top two biomarkers listed in Table 3. From Fig. 14 and Fig. 15, we can see that the selected biomarkers by the proposed method are highly associated with the MMSE scores at multiple time points. Specifically, the values of MRI imaging in the region of *corpus callosum* have a significant negative correlation with MMSE scores, while the values of *hippocampal formation left* show a positive correlation with MMSE scores. This correlation indicates that these biomarkers are sensitive to the clinical status of AD, and have the potential to be used as a diagnostic indicator.

5.3. Limitations

There are still several limitations to be considered in this study. *First*, the current study requires each subject has complete data, (*i.e.*, subjects with MRI data and clinical scores at each time point), which limits the size of subjects that can be used. For example, while there are more than 400 MCI subjects in the ADNI dataset, there are only 202 MCI subjects having complete MRI data and the corresponding ADAS-Cog and MMSE scores at four time points (*i.e.*, BL, M06, M12, and M24). *Second*, in the current study, we performed experiments to estimate the clinical scores based on only MRI data. Actually, there exist many other data modalities (*e.g.*, PET and CSF) for subjects in ADNI. It is interesting to apply our method to disease progression prediction using multimodality data for further performance improvement. *Third*, in the practical prediction of AD progression, besides ADAS-Cog and MMSE scores, the sub-scores of clinical variables or other clinical variables are generally acquired, *e.g.*, the clinical dementia rating scale sum of boxes (CDR-SOB) and the auditory verbal learning test (AVLT). Since the clinical variables are helpful to reflect the status of AD progression, using more continuous clinical values may help discover more robust disease-relevant markers, which will also be one of our future works.

6. Conclusion

In this paper, we proposed a novel multi-task exclusive relationship learning model, aiming to reveal the relationship among prediction tasks for cognitive measures at different time points based on longitudinal neuroimaging markers. Specifically, an exclusive lasso regularization and a relationship induced regularization are used to encourage the partial group structure feature selection among the longitudinal data and to capture the relationship among tasks, respectively. To efficiently solve the proposed objective function, we develop an iterative algorithm with a closed form solution in each iteration. We have proved the convergence of our algorithm on simulation data. Also, experimental results on the real ADNI dataset with MRI data demonstrate that our method outperforms those state-of-the-art approaches in both tasks of estimating clinical scores and brain disease classification.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61876082, 61861130366, 61703301, 61473149), and the NIH grants (EB006733, EB008374, EB009634, MH100217, AG041721, AG042599, AG010129, AG030514). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The investigators within the ADNI did not participate in analysis or writing of this study. A complete listing of ADNI investigators can be found online[2].

References

Argyriou A, Evgeniou T, Pontil M, 2008 Convex multi-task feature learning. Machine Learning 73, 243–272.

Braak H, Braak E, 1991 Neuropathological stageing of Alzheimer-related changes. Acta Neuropathologica 82, 239–259. [PubMed: 1759558]

Brookmeyer R, Johnson E, Zieglergraham K, Arrighi HM, 2007 Forecasting the global burden of Alzheimer's disease. Alzheimer's and Dementia 3, 186–191.

Caruana R, 1997 Multitask learning. Machine Learning 28, 41–75.

Chang CC, Lin CJ, 2012 LIBSVM: A library for support vector machines. Acm Transactions on Intelligent Systems and Technology 3, 1–27.

Chen X, Yuan X, Yan S, Tang J, Rui Y, Chua TS, 2011 Towards multi-semantic image annotation with graph regularized exclusive group lasso, in: ACM International Conference on Multimedia, pp. 263–272.

Chen X, Yuan XT, Chen Q, Yan S, Chua TS, 2012 Multi-label visual classification with label exclusive context, in: IEEE International Conference on Computer Vision, pp. 834–841.

Convit A, De AJ, de Leon MJ, Tarshish CY, De SS, Rusinek H, 2000 Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiology of Aging 21, 19–26. [PubMed: 10794844]

Delacourte A, David JP, Sergeant N, Buee L, Wattez A, Vermersch P, Ghozali F, Falletbianco C, Pasquier F, Lebert F, 1999 The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease. Neurology 54, 1158–1165.

Derflinger S, Sorg C, Gaser C, Myers N, Arsic M, Kurz A, Zimmer C, Wohlschlager A, Muhlau M, 2011 Grey-matter atrophy in Alzheimer's disease is asymmetric but not lateralized. Journal of Alzheimers Disease Jad 25, 347–357. [PubMed: 21422522]

Dietterich TG, 1998 Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation 10, 1895–1923. [PubMed: 9744903]

Duchesne S, Caroli A, Geroldi C, Collins DL, Frisoni GB, 2009 Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. NeuroImage 47, 1363–1370. [PubMed: 19371783]

Fan Y, Kaufer D, Shen D, 2010 Joint estimation of multiple clinical variables of neurological diseases from imaging patterns, in: IEEE International Symposium on Biomedical Imaging: From Nano To Macro, pp. 852–855.

Fan Y, Resnick SM, Wu X, Davatzikos C, 2008 Structural and functional biomarkers of prodromal Alzheimer's disease: A high-dimensional pattern classification study. NeuroImage 41, 277–285. [PubMed: 18400519]

Hastie T, Tibshirani R, Friedman J, 2000 The Elements of Statistical Learning. Springer-Verlag, New York.

Hinrichs C, Singh V, Xu G, Johnson SC, 2011 Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. NeuroImage 55, 574–589. [PubMed: 21146621]

Jie B, Liu M, Liu J, Zhang D, Shen D, 2017 Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease. IEEE Transactions on Biomedical Engineering 64, 238–249. [PubMed: 27093313]

Kabani N, M.D.H.C.E.A., 1998 A 3D atlas of the human brain. NeuroImage 7, S717.

Knafo S, Venero C, Merinoserrais P, Fernaudespinosa I, Gonzalezsoriano J, Ferrer I, Santpere G, Defelipe J, 2009 Morphological alterations to neurons of the amygdala and impaired fear conditioning in a transgenic mouse model of Alzheimer's disease. Journal of Pathology 219, 41–51. [PubMed: 19449368]

Lei B, Feng J, Chen S, Dong N, Wang T, 2017 Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning. Frontiers in Aging Neuroscience 9, 1–12. [PubMed: 28174533]

Lian C, Liu M, Zhang J, Shen D, 2019 Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Lian C, Zhang J, Liu M, Zong X, Hung SC, Lin W, Shen D, 2018 Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images. Medical Image Analysis 46, 106–117. [PubMed: 29518675]

Liu J, Ji S, Ye J, 2009 Multi-task feature learning via efficient $\ell_{2,1}$-norm minimization, in: Conference on Uncertainty in Artificial Intelligence, pp. 339–348.

Liu M, Zhang D, 2016a Feature selection with effective distance. Neurocomputing 215, 100–109.

Liu M, Zhang D, 2016b Pairwise constraint-guided sparse learning for feature selection. IEEE Transactions on Cybernetics 46, 298–310. [PubMed: 26151948]

Liu M, Zhang D, Chen S, Xue H, 2016a Joint binary classifier learning for ECOC-based multi-class classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 38, 2335–2341.

Liu M, Zhang D, Shen D, 2015 View-centralized multi-atlas classification for Alzheimer's disease diagnosis. Human Brain Mapping 36, 1847–1865. [PubMed: 25624081]

Liu M, Zhang D, Shen D, 2016b Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. IEEE Transactions on Medical Imaging 35, 1463–1474. [PubMed: 26742127]

Morgan AR, Touchard S, O' Hagan C, Sims R, Majounie E, Escott-Price V, Jones L, Williams J, Morgan BP, 2017 The correlation between inflammatory biomarkers and polygenic risk score in Alzheimer's disease. Journal of Alzheimer's Disease 56, 25–36.

Nie L, Zhang L, Meng L, Song X, Chang X, Li X, 2017 Modeling disease progression via multisource multitask learners: A case study with Alzheimer's disease. IEEE Transactions on Neural Networks and Learning Systems 28, 1508–1519. [PubMed: 26929064]

V. P, W. HJ, W. SD, S. LM, T. JQ, W. MW, K. DS, P. RC, C. J, Jr, 2009 MRI and CSF biomarkers in normal, MCI, and AD subjects predicting future clinical change. Neurology 73, 294–301. [PubMed: 19636049]

Pan Y, Liu M, Lian C, Zhou T, Xia Y, Shen D, 2018 Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 455–463.

Petrella JR, Coleman RE, Doraiswamy PM, 2003 Neuroimaging and early diagnosis of Alzheimer disease: A look to the future. Radiology 226, 315–336. [PubMed: 12563122]

Rosen WG, Mohs RC, Davis KL, 1984 A new rating scale for Alzheimer's disease. American Journal of Psychiatry 141, 1356–1364. [PubMed: 6496779]

Shen D, Resnick SM, Davatzikos C, 2003 4D hammer image registration method for longitudinal study of brain changes. Proceedings of the Human Brain Mapping, 1–8.

Sled JG, Zijdenbos AP, Evans AC, 1998 A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97. [PubMed: 9617910]

Stonnington CM, Chu C, Kloppel S, Jr CRJ, Ashburner J, Frackowiak RSJ, 2010a Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. NeuroiImage 51, 1405–1413.

Stonnington CM, Chu C, Klppel S, Jack CR, Ashburner J, Frackowiak RS, 2010b Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. NeuroImage 51, 1405–1413. [PubMed: 20347044]

Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L, 2012a Identifying quantitative trait loci via group-sparse multitask regression and feature selection: An imaging genetics study of the ADNI cohort. Bioinformatics 28, 229–237. [PubMed: 22155867]

Wang H, Nie F, Huang H, Yan J, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L, 2012b From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. Bioinformatics 28, i619–i625. [PubMed: 22962490]

Wang H, Nie F, Huang H, Yan J, Kim S, Risacher SL, Saykin AJ, Shen L, 2012c High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction. Advances in Neural Information Processing Systems 2, 1286–1294.

Wang Y, Fan Y, Bhatt P, Davatzikos C, 2010 High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. NeuroImage 50, 1519–1535. [PubMed: 20056158]

Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-bohmer KA, Wang L, Shen D, 2012 Identification of MCI individuals using structural and functional connectivity networks. NeuroImage 59, 2045–56. [PubMed: 22019883]

Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y, Kuwabara H, Kuroda M, Yamada T, Megumi F, 2016 A small number of abnormal brain connections predicts adult autism spectrum disorder. Nature Communications 7, 1–12.

Yan J, Li T, Wang H, Huang H, Wan J, Nho K, Kim S, Risacher SL, Saykin AJ, Shen L, 2015 Cortical surface biomarkers for predicting cognitive outcomes using group $\ell_{2,1}$ norm. Neurobiology of Aging 36, S185–S193. [PubMed: 25444599]

Zhang D, Liu J, Shen D, 2012 Temporally-constrained group sparse learning for longitudinal data analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 264–271.

Zhang D, Shen D, 2012a Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59, 895–907. [PubMed: 21992749]

Zhang D, Shen D, 2012b Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. Plos One 7, e33182. [PubMed: 22457741]

Zhang D, Wang Y, Zhou L, Yuan H, Shen D, 2011 Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55, 856–867. [PubMed: 21236349]

Zhang Y, Brady M, Smith S, 2001 Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging 20, 45–57. [PubMed: 11293691]

Zhang Y, Yeung DY, 2010a A convex formulation for learning task relationships in multi-task learning, in: Twenty-sixth Conference on Uncertainty in Artificial Intelligence, pp. 733–742.

Zhang Y, Yeung DY, 2010b Multi-task learning using generalized t process. Journal of Machine Learning Research 9, 964–971.

Zhang Y, Yeung DY, Xu Q, 2010 Probabilistic multi-task feature selection, in: International Conference on Neural Information Processing Systems, pp. 2559–2567.

Zhou J, Liu J, Narayan VA, Ye J, 2013 Modeling disease progression via multi-task learning. NeuroImage 78, 233–248. [PubMed: 23583359]

Zhou J, Yuan L, Liu J, Ye J, 2011 A multi-task learning formulation for predicting disease progression, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, 8, pp. 814–822.

Zhou Y, Jin R, Hoi SCH, 2010 Exclusive lasso for multitask feature selection. Journal Machine Learning Research 9, 988–995.

Highlights

- A relationship induced regularization models the relationship among tasks at different time points for estimating clinical measures based on longitudinal imaging data;
- An exclusive lasso regularization enables partial group structure feature selection among tasks;
- Extensive experiments on both synthetic and real datasets demonstrate the effectiveness of the proposed method.

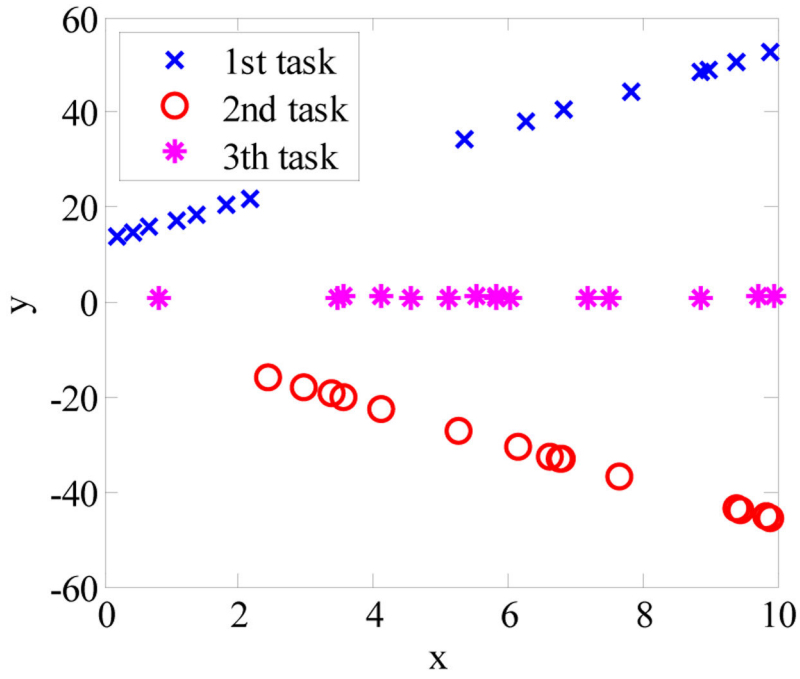


Figure 1:
Illustration of the simulation data, with each color denoting data points in a specific task.

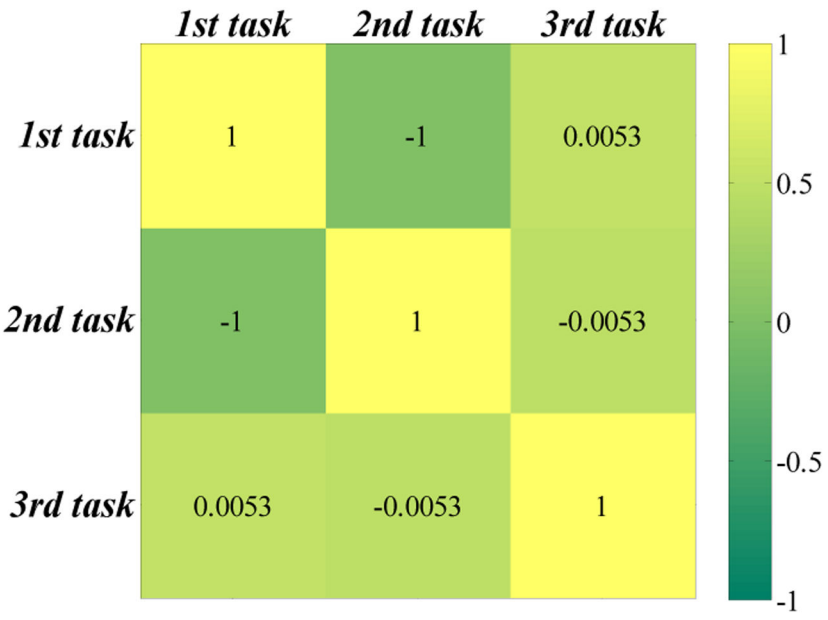


Figure 2:
The learned task correlation matrix.

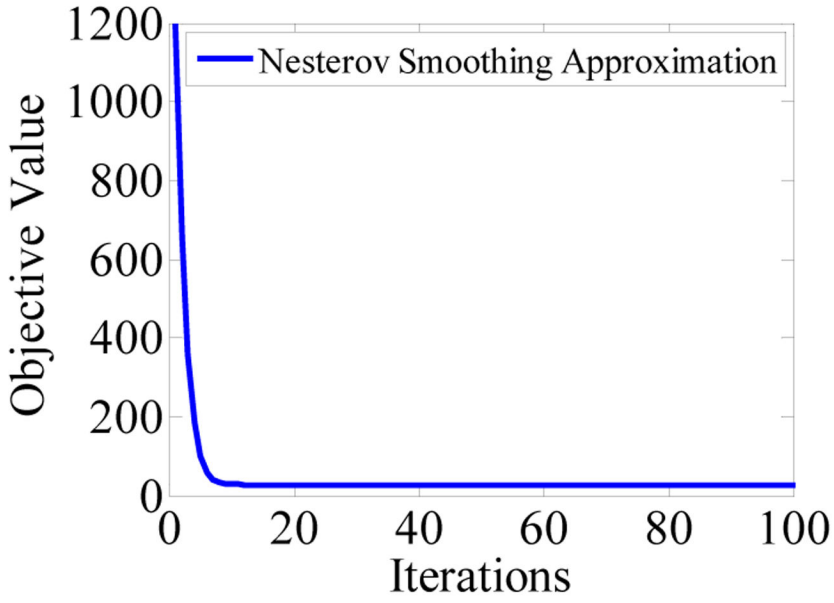
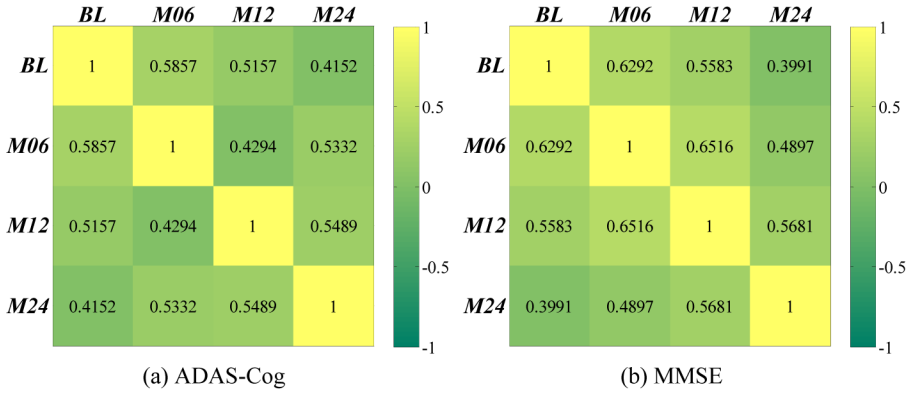


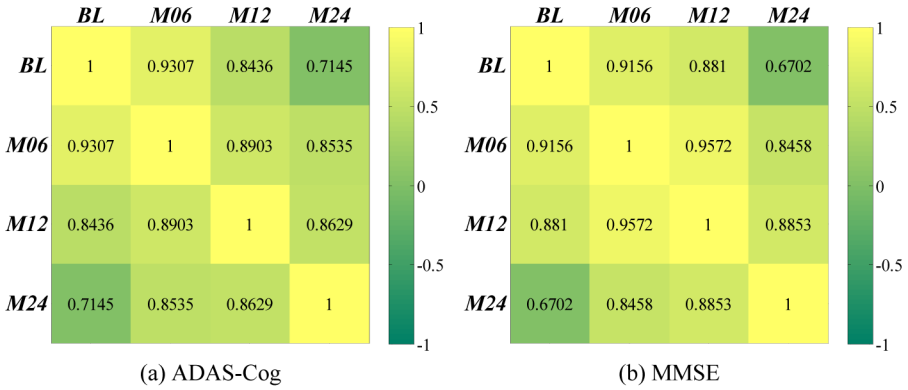
Figure 3:
Convergence of the objective function on simulation data.



(a) ADAS-Cog (b) MMSE

Figure 4:

The correlation matrix learned by our MTRL method among four regression tasks for estimating (a) ADAS-Cog and (b) MMSE, respectively. Note that each regression task is corresponding to a particular time point. BL: baseline; M06: 6 months after baseline; M12: 12 months after baseline; M24: 24 months after baseline.



(a) ADAS-Cog (b) MMSE

Figure 5:

The correlation matrix learned by our MTERL method among four regression tasks for estimating (a) ADAS-Cog and (b) MMSE, respectively. Note that each regression task is corresponding to a particular time point. BL: baseline; M06: 6 months after baseline; M12: 12 months after baseline; M24: 24 months after baseline.

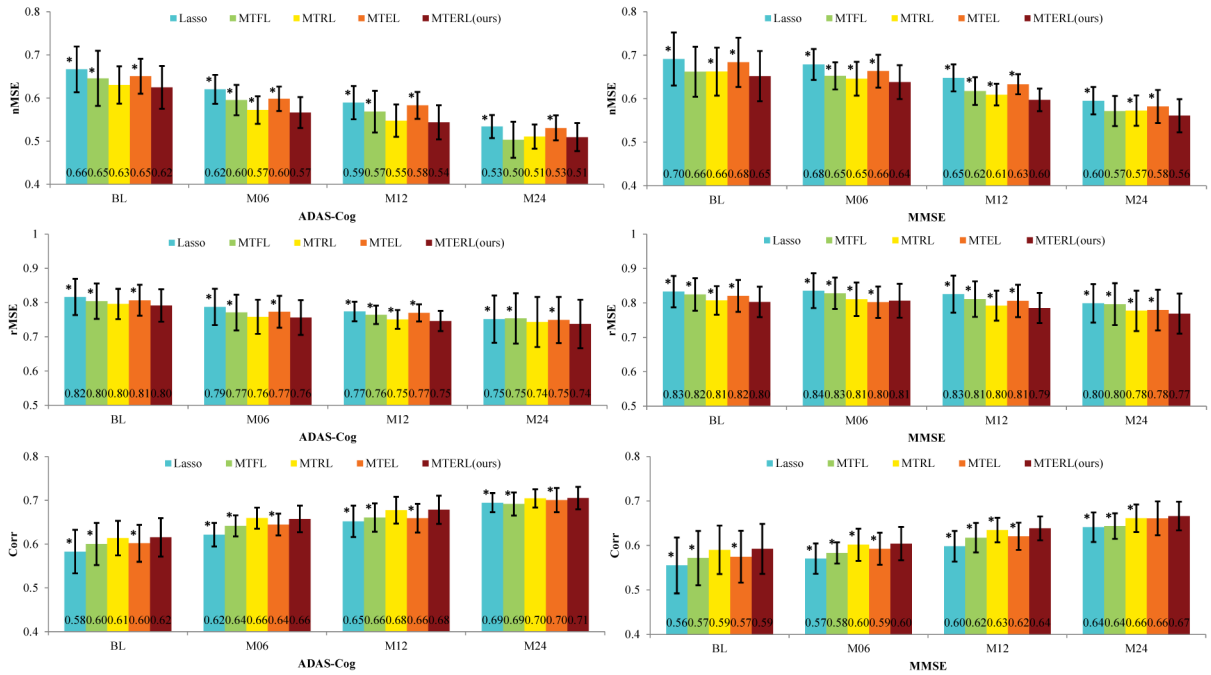


Figure 6:

Comparison of different methods on longitudinal ADAS-Cog and MMSE prediction in terms of (a) normalized mean square error (nMSE), (b) root mean square error (rMSE) and (c) Pearson's correlation coefficient (Corr). BL: baseline; M06: 6 months after baseline; M12: 12 months after baseline; M24: 24 months after baseline.

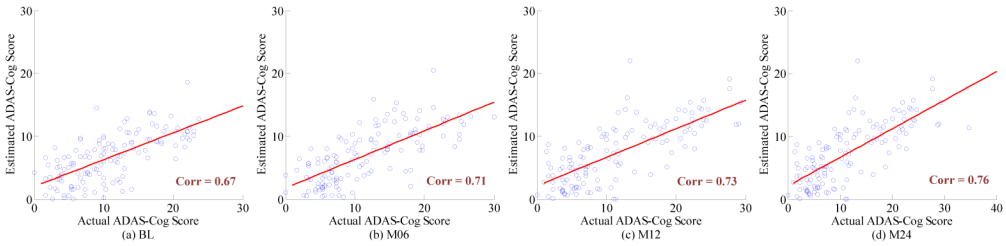


Figure 7:

Scatter plots of actual ADAS-Cog scores versus predicted values on testing data achieved by our proposed method based on MRI features at different time points. (a) Baseline (BL), (b) 6 months (M06), (c) 12 months (M12) and (d) 24 months (M24). Corr: Correlation coefficient between estimated scores and actual scores.

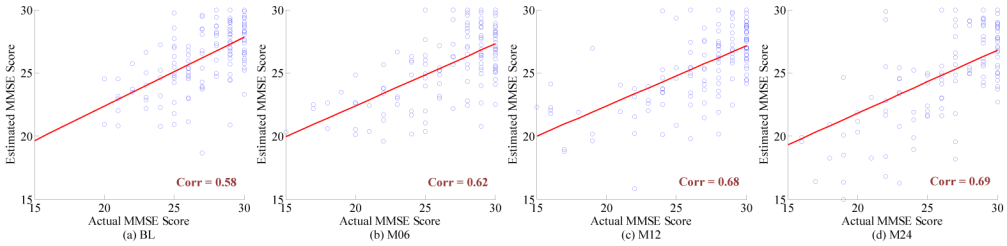


Figure 8:

Scatter plots of actual MMSE scores versus predicted values on testing data achieved by our proposed method based on MRI features at different time points. (a) Baseline (BL), (b) 6 months (M06), (c) 12 months (M12) and (d) 24 months (M24). Corr: Correlation coefficient between estimated scores and actual scores.

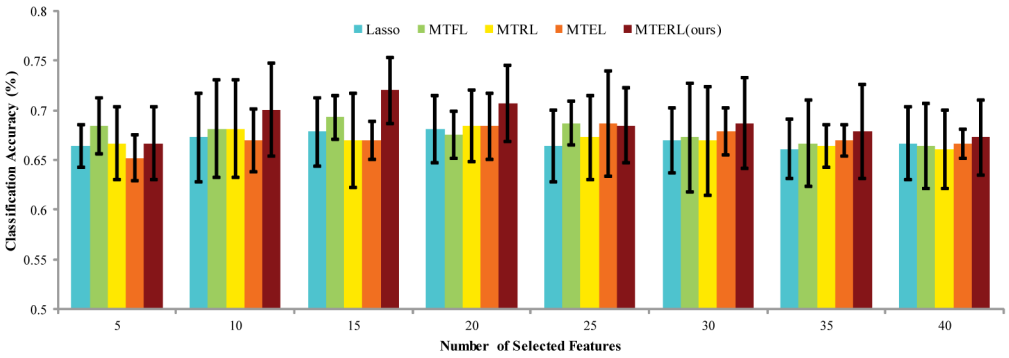


Figure 9:

Classification results achieved by different methods using features selected from the corresponding regression model for ADAS-Cog estimation.

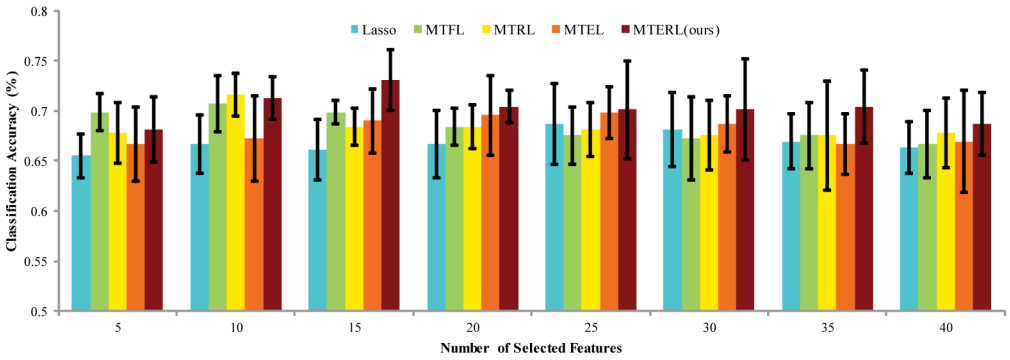
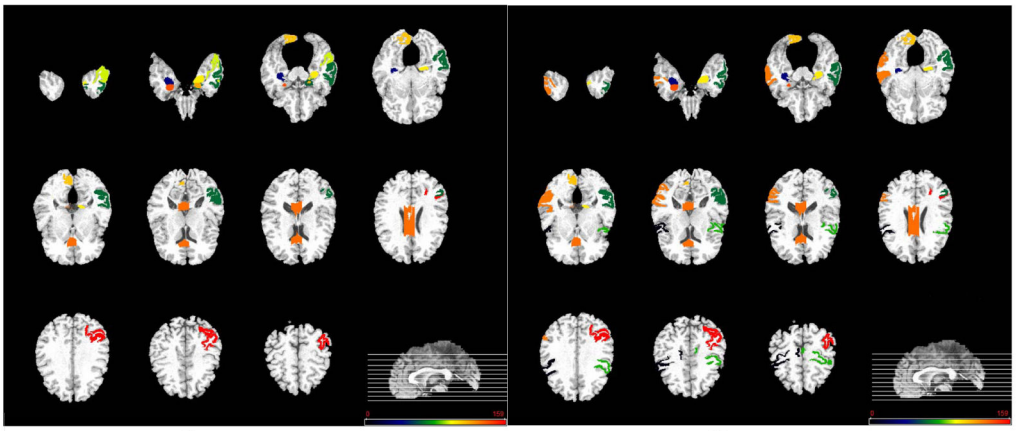


Figure 10:

Classification results achieved by different methods using features selected from the corresponding regression model for MMSE estimation.



(a) ADAS-Cog score (b) MMSE score

Figure 11:
The important brain regions identified by the proposed method when estimating clinical scores. (a) ADAS-Cog score, (b) MMSE score.

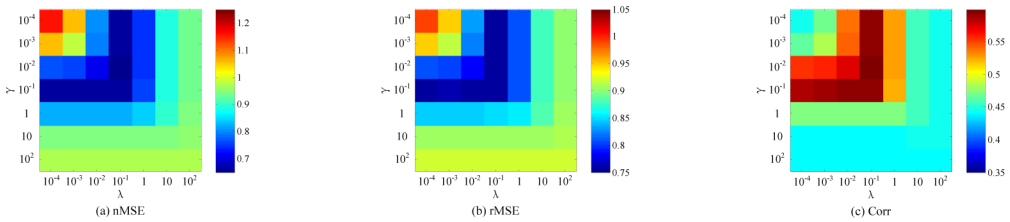


Figure 12:

Regression performance achieved by our method when estimating ADAS-Cog score with different criterias. (a) nMSE, (b) rMSE and (c) Corr.

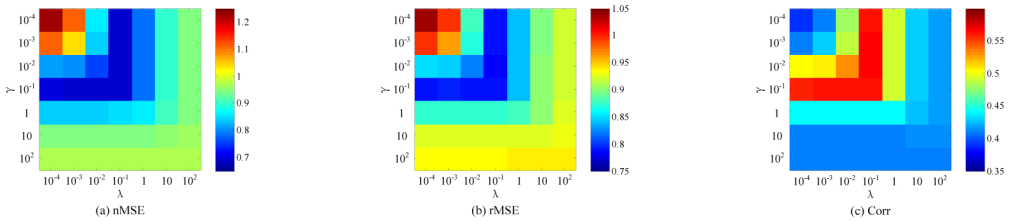


Figure 13:

Regression performance achieved by our method when estimating MMSE score with different criterias. (a) nMSE, (b) rMSE and (c) Corr.

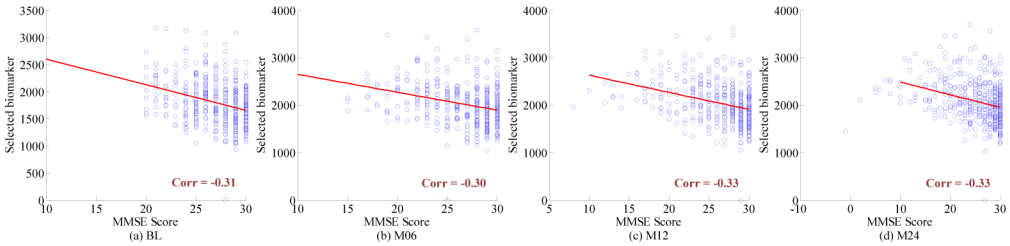


Figure 14:

Correlation between selected biomarker (*i.e.*, corpus callosum) and MMSE score at different time points. (a) Baseline (BL), (b) 6 months (M06), (c) 12 months (M12) and (d) 24 months (M24).

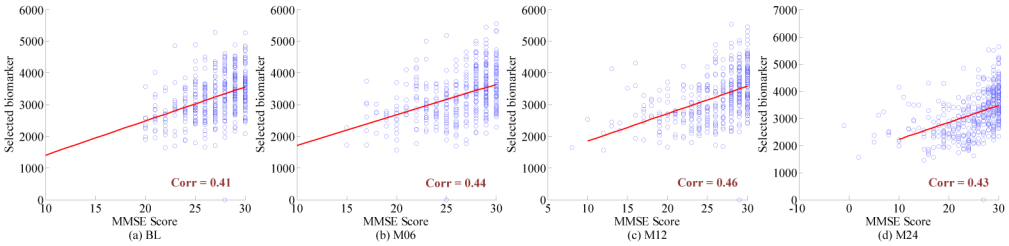


Figure 15:

Correlation between selected biomarker (*i.e.*, hippocampal formation left) and MMSE score at different time points. (a) Baseline (BL), (b) 6 months (M06), (c) 12 months (M12) and (d) 24 months (M24).

Table 1:

Demographic information of the studied subjects at different time points. The values are denoted as Mean ±Stand Deviation (SD). BL: Baseline; M06: 6 month after baseline; M12: 12 month after baseline; M24: 24 month after baseline.

	AD (*n* = 91)	**MCI-C (*n* = 104)**	**MCI-NC (*n* = 98)**	**NC (*n* = 152)**
Male/Female	53/38	66/38	68/30	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

Table 2:

The average performance of different methods in predicting the conversion of MCI patients at their best dimensions. ACC=Accuracy; SEN=Sensitivity; SPE=Specificity; AUC= Area Under Curve.

Method		Lasso	MTFL	MTRL	MTEL	MTERL(ours)
ADAS-Cog	ACC(%)	67.83 ± 3.14	69.28 ± 2.15	68.41 ± 3.61	68.69 ± 5.29	**71.98 ± 3.35**
	SEN(%)	73.53 ± 8.32	72.94 ± 10.27	74.71 ± 16.97	72.94 ± 16.69	**80.39 ± 11.14**
	SPE(%)	61.67 ± 11.35	65.00 ± 10.87	60.56 ± 13.66	62.78 ± 14.65	**65.04 ± 14.25**
	AUC(%)	77.31 ± 7.54	79.58 ± 7.80	79.75 ± 5.28	78.57 ± 6.27	**81.20 ± 5.07**
	p-value	0.0305*	0.0062*	0.0415*	0.0416*	–
MMSE	ACC(%)	68.69 ± 4.05	70.72 ± 2.79	70.73 ± 2.15	69.57 ± 3.97	**73.04 ± 3.01**
	SEN(%)	71.18 ± 8.92	77.65 ± 12.75	77.06 ± 9.84	75.88 ± 12.20	**78.29 ± 10.73**
	SPE(%)	65.00 ± 8.70	62.22 ± 13.69	62.78 ± 9.74	61.67 ± 12.33	**68.89 ± 7.71**
	AUC(%)	75.38 ± 4.96	79.30 ± 2.77	79.30 ± 2.77	76.16 ± 3.07	**80.81 ± 2.95**
	p-value	0.0105*	0.0193*	0.0193*	0.0386*	–

Table 3:

The most important ten ROIs identified by the proposed method.

ADAS-Cog		
Index	ROI Name	Related Studies
82	corpus callosum	(Lei et al., 2017; Jie et al., 2017)
69	hippocampal formation left	(Zhang and Shen, 2012a; Zhou et al., 2013)
87	angular gyrus left	(Wang et al., 2012c; Jie et al., 2017)
48	middle temporal gyrus left	(Lei et al., 2017; Zhou et al., 2013)
30	hippocampal formation right	(Wang et al., 2012c; Zhang and Shen, 2012a)
83	amygdala right	(Zhou et al., 2013; Wang et al., 2012c)
76	amygdala left	(Wang et al., 2012c; Lei et al., 2017)
62	inferior temporal gyrus left	(Lei et al., 2017; Zhou et al., 2013)
74	lingual gyrus right	(Zhou et al., 2013; Wang et al., 2012c)
46	uncus left	(Zhang and Shen, 2012a; Jie et al., 2017)
MMSE		
Index	ROI Name	Related Studies
82	corpus callosum	(Lei et al., 2017; Jie et al., 2017)
69	hippocampal formation left	(Wang et al., 2012c; Zhang and Shen, 2012a)
74	lingual gyrus right	(Zhou et al., 2013; Lei et al., 2017)
48	middle temporal gyrus left	(Lei et al., 2017; Jie et al., 2017)
83	amygdala right	(Wang et al., 2012c; Zhou et al., 2013)
87	angular gyrus left	(Zhang and Shen, 2012a; Wang et al., 2012c)
80	middle temporal gyrus right	(Wang et al., 2012c; Jie et al., 2017)
30	hippocampal formation right	(Zhou et al., 2013; Lei et al., 2017)
5	precentral gyrus right	(Lei et al., 2017; Zhou et al., 2013)
55	precentral gyrus left	(Zhou et al., 2013; Jie et al., 2017)

Table 4:

Comparison with the state-of-the-art methods in terms of correlation coefficient in clinical score regression.

Method	Features	Subjects	ADAS-Cog	MMSE
(Duchesne et al., 2009)	Baseline MRI	*75AD* + *49MCI* + 75*NC*	–	0.310
(Fan et al., 2010)	Baseline MRI	*52AD* + *148MCI* + 64*NC*	0.520	0.570
(Yan et al., 2015)	Baseline MRI	*172AD* + 349*MCI* + *197NC*	0.644 ± 0.026	0.555 ± 0.008
(Zhang and Shen, 2012a)	Baseline MRI	*45AD* + *91MCI* + 50*NC*	0.609 ± 0.014	0.504 ± 0.038
(Stonnington et al., 2010b)	Baseline MRI, CSF	*113AD* + *351MCI* + *122NC*	0.48	0.57
(Liu and Zhang, 2016b)	MRI, Age, Gender, Education (Years)	*143AD* + *271MCI* + *185NC*	0.59	0.57
(Zhang and Shen, 2012b)	Baseline, M06, M12, M18, M24 MRI, FDG-PET	*38MCI-C* + *50MCI-NC*	0.777 ± 0.027	0.786 ± 0.013
(Jie et al., 2017)	Baseline, M06, M12, M24 MRI	*91AD* + *202MCI* + 152*NC*	0.639 ± 0.008	0.613 ± 0.010
MTERL(ours)	Baseline, M06, M12, M24 MRI	*91AD* + *202MCI* + 152*NC*	0.664 ± 0.025	0.625 ± 0.032