



Longitudinal score prediction for Alzheimer's disease based on ensemble correntropy and spatial–temporal constraint

Baiying Lei¹ · Wen Hou² · Wenbin Zou²  · Xia Li² · Cishen Zhang³ · Tianfu Wang¹

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Neuroimaging data has been widely used to predict clinical scores for automatic diagnosis of Alzheimer's disease (AD). For accurate clinical score prediction, one of the major challenges is high feature dimension of the imaging data. To address this issue, this paper presents an effective framework using a novel feature selection model via sparse learning. In contrast to previous approaches focusing on a single time point, this framework uses information at multiple time points. Specifically, a regularized correntropy with the spatial–temporal constraint is used to reduce the adverse effect of noise and outliers, and promote consistent and robust selection of features by exploring data characteristics. Furthermore, ensemble learning of support vector regression (SVR) is exploited to accurately predict AD scores based on the selected features. The proposed approach is extensively evaluated on the Alzheimer's disease neuroimaging initiative (ADNI) dataset. Our experiments demonstrate that the proposed approach not only achieves promising regression accuracy, but also successfully recognizes disease-related biomarkers.

Keywords Alzheimer's disease · Correntropy · Ensemble learning · Longitudinal score prediction · Spatial–temporal constraint

Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease and the leading cause of dementia, listed as the sixth-leading cause of death in the United States (Kochanek et al. 2012). In 2016, there are approximately 44 million people worldwide diagnosed with AD (Yin et al. 2017). The preclinical stage of AD is termed as mild cognitive impairment (MCI) (Shi et al. 2012; Liu et al. 2014). Although no treatment can reverse the progression of AD, early diagnosis of MCI (Wang et al. 2017) followed by appropriate treatment

can greatly defer the disease progression (Alzheimer's Association 2015).

It is known that the progression of AD can be clinically indicated by cognitive scores at multiple time points. A variety of clinical assessment scores have been developed, such as the mini-mental state examination (MMSE) (Folstein et al. 1975), Alzheimer's disease assessment scale-cognitive subscale (ADAS-Cog) (Rosen et al. 1984), and the clinical dementia rating-global and the sum of boxes (CDR-GLOB and CDR-SOB) (Morris 1993). They are considered as quantifiable measurements of the disease progression and have been used in clinical investigations extensively (Yau et al. 2015). Accurate prediction of clinical scores is desirable so that appropriate treatment plans can be initiated and adjusted based on the prediction results of disease progression.

To analyze AD progression via clinical score prediction, various pattern recognition and regression methods have been devised based on imaging data such as magnetic resonance imaging (MRI) and positron emission topography (PET) obtained through neuroimaging tools (Hao et al. 2016). In general, these methods have provided insights into the neurophysiological characteristics of AD (Eskildsen et al. 2015). In neuroimaging-based AD studies, one of the

✉ Wenbin Zou
zouszu@sina.com

¹ National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China

² Shenzhen Key Lab of Advanced Telecommunication and Information Processing, College of Information Engineering, Shenzhen University, Shenzhen 518060, China

³ Faculty of Science, Engineering & Technology, Swinburne University of Technology, Melbourne, Australia

major challenges for computer-aided AD diagnosis is the high feature dimensionality (Zhu et al. 2015). To address this issue, feature selection has become the core step due to its effectiveness (Ota et al. 2015). Currently, a great amount of work is devoted to the cross-sectional analysis using only the data collected at a single time point. It is argued that AD is a progressive disease and the imaging data of multiple time points are often available. Hence, the longitudinal analysis with multiple time point is more beneficial for revealing the progression patterns than single time point, and is also an active research field (Hao et al. 2017). In view of this, a spatial–temporal group LASSO method (STGL) is developed in this paper. Specifically, two regularization terms are imposed into the sparse least square regression framework. The first regularization term minimizes the difference between successive weights to simulate the smooth change of adjacent time points (Zhang et al. 2012b; Jie et al. 2017). The second regularization term is a locality-preserving-projection (LPP) based term, which retains the neighborhood structure of each sample and reduces the adverse effects of noise (Zhu et al. 2012, 2016a). To remove outliers and select robust features, correntropy is introduced to boost the performance of STGL model (CSTGL). Since ensemble classifier is usually considered to be more accurate and robust than individual classifier (Che et al. 2011; Kuncheva et al. 2010; Liu et al. 2012), the ensemble strategy is also adopted in the proposed framework to further improve the regression performance (ESTGL and ECSTGL). Due to good generalization ability of these models, our proposed method can achieve the balance between optimal feature selection and potential data over-fitting to specific training data. Furthermore, the optimal learning problem is solved by an efficient optimization algorithm.

The proposed models are evaluated by predicting clinical scores, MMSE and ADAS-Cog, based on the MRI data of 445 subjects from the Alzheimer’s disease neuroimaging initiative (ADNI) database with the scores collected at four different time points. Experimental results show that the proposed STGL model outperforms other sparse learning methods, and the adaption of correntropy formulation and ensemble strategies by ECSTGL model further improves the prediction accuracy. The disease-related biomarkers are successfully identified through the proposed models, which demonstrates the effectiveness of the proposed models as well. Overall, the contributions of this paper are summarized as follows:

- 1) To our best knowledge, this is the first approach that integrates spatial–temporal information in the feature selection model for prediction of longitudinal AD scores.
- 2) Correntropy is successfully integrated into the STGL model to extract informative features while removing outliers from training data.

- 3) Efficient optimization algorithms are developed to solve the proposed STGL and CSTGL models so that AD-related biomarkers can be discovered for subsequent regression tasks.
- 4) Ensemble learning of support vector regression (SVR) is exploited to predict longitudinal AD scores more accurately from the selected features.

Materials and methods

Related work

The main focus of this paper is on the dimensionality reduction of neuroimaging-based features, and relevant methods can be divided into two categories.

The first category is feature selection using statistical t-test and Chi-squared methods (Liu et al. 2013) or sparse linear models (Gallego-Jutglà et al. 2015; Zhu et al. 2016b). Among those methods, sparse learning techniques have attracted increasing attention due to their outstanding performance (Suk et al. 2016). To start with, the widely used least absolute shrinkage and selection operator (LASSO) has been proposed to diagnose AD and MCI from MRI data (Tibshirani 1996; Hinrichs et al. 2009) and further extended. For example, one LASSO variant uses an elastic net (Shen et al. 2011), and another one incorporates domain-specific information (Ng and Abugharbieh 2011). Moreover, the recently developed group LASSO with an $\ell_{2,1}$ -norm regularization has been applied for joint learning tasks and demonstrated its effectiveness in enhancing the mechanistic understanding of AD (Zhang et al. 2012a; Zhu et al. 2017).

The second category is subspace learning that projects original features into a lower-dimensional space (Zhang et al. 2012c). The classical methods include principle component analysis (PCA), independent component analysis (ICA), linear discriminant analysis (LDA), and canonical correlation analysis (CCA) (Zhu et al. 2012). Subspace learning methods have gained extensive applications in medical image analysis. For example, local linear embedding has been employed to process multivariate MRI data (Roweis and Saul 2000).

Both categories have their own advantages. On the one hand, feature selection methods are preferred in terms of interpretability to identify the biomarkers of AD, while subspace learning fails in this aspect because all features are involved to find a lower-space feature representation. On the other hand, it is reported (Zhu et al. 2012) that subspace learning obtains superior performance in revealing the disease status of a subject compared with feature selection methods. Hence, it will be beneficial to combine the two categories of approaches to overcome limitations of each category and achieve optimal performance.

Proposed regression framework

Figure 1 illustrates the framework of the proposed regression approach, which contains three major steps: image preprocessing and feature extraction, feature selection, and SVR based prediction. The details of each step are described in the following.

Subjects and data acquisition

The data used in the preparation of this paper has been obtained from the ADNI database (<http://www.adni-info.org>). The MRI data and corresponding cognitive scores (MMSE and ADAS-Cog) at four different time points (i.e., baseline, 6-month, 12-month and 24-month) of 445 ADNI subjects (including 91 AD, 202 MCI, and 152 NC) are used in this study. Particularly for the MCI cohort, there are 104 MCI converters (MCI-C) and 98 MCI non-converters (MCI-NC). The demographic information of the studied subjects can be found in (Jie et al. 2017).

Structural MR scans in ADNI have been acquired from 1.5T scanners. The raw digital imaging and communications in medicine (DICOM) MRI scans are downloaded from the public ADNI site (<http://adni.loni.usc.edu>). These scans are reviewed for quality and spatial distortion caused by gradient nonlinearity and B1 field inhomogeneity are corrected.

Image preprocessing and feature extraction

Pre-processing is carried out for all MR images in a similar way as previous works (Zhang et al. 2011). The pre-processing consists of five main steps: correction, skull stripping, segmentation, registration, and feature extraction. Details are as follows.

- 1) Correction: Anterior commissure (AC) - posterior commissure (PC) correction is realized via MIPAV software (<http://mipav.cit.nih.gov/index.php>) and correction of intensity inhomogeneity is realized via N3 algorithm (Sled et al. 1998).
- 2) Skull-stripping: The skull-stripping on structural MR images is performed with a learning based method proposed in (Shi et al. 2012) that includes both brain surface extractor (BSE) and brain extraction tool (BET). The stripping results are further manually reviewed to ensure clean skull and dura removal.
- 3) Segmentation: Structural MR images are segmented into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) via FMRIB software library (FSL) (Zhang et al. 2001).
- 4) Registration: All different time-point images of each subject are registered to a template with 93 manually-labeled regions of interest (ROI) via 4D HAMMER

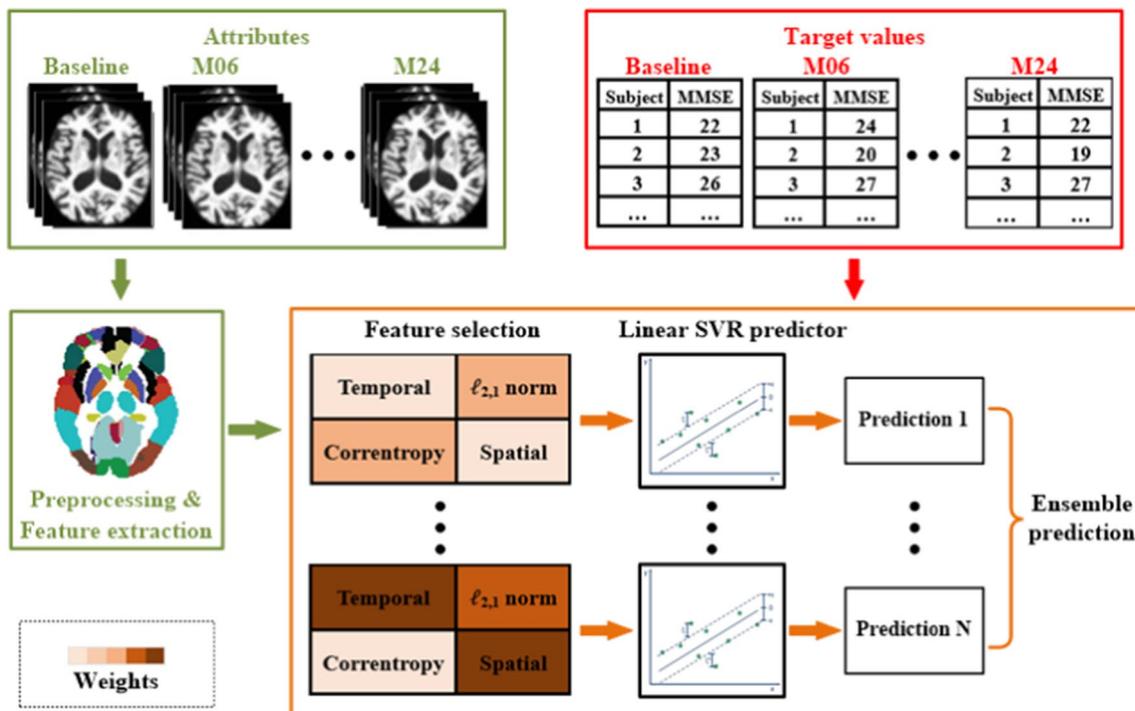


Fig. 1 Schematic illustration of the proposed regression framework. Feature selection model incorporates group sparsity, correntropy, and spatial-temporal information, where darker color indicates larger weights

(Shen et al. 2003), which is a fully automatic 4-dimensional atlas warping method.

- 5) Feature extraction: For each of the 93 ROIs in the labeled MR image, the total GM volume of each region is calculated and used as the feature. Only grey matter (GM) volume is used as features because it is the most affected by AD and widely used in the literature (Zhang et al. 2012a). If there is no GM in a specific region, the feature value for this region will be 0.

Feature selection

Supposing there are n training subjects, and for each subject, the imaging data is derived from t different time points, $\mathbb{X} = \{X_1, X_2, \dots, X_t\}$ where $X_i (i = 1, 2, \dots, t) \in \mathbb{R}^{n \times d}$ with d as the feature number. The corresponding clinical scores are denoted by $\mathbb{Y} = \{Y_1, Y_2, \dots, Y_t\}$ with $Y_i \in \mathbb{R}^{n \times 1}$. A linear model is used to estimate the clinical score from the imaging data at the i -th time point as $Y_i = X_i W_i$, where the feature weight vector $W_i \in \mathbb{R}^{d \times 1}$. Let $W = [W_1, W_2, \dots, W_t] \in \mathbb{R}^{d \times t}$ denote the weight matrix for all t learning tasks. The commonly used group LASSO model across the regression tasks is formulated as

$$\min_w \frac{1}{2} \sum_{i=1}^t \|Y_i - X_i W_i\|_F^2 + \rho_0 \|W\|_{2,1}, \quad (1)$$

where ρ_0 is the group regularization parameter, $\|W\|_{2,1}$ is the $\ell_{2,1}$ -norm term, which penalizes the coefficients in the same row of W together for joint selection. Group LASSO model encourages the weights corresponding to the same feature across multiple time-points to be grouped together.

The spatial-temporal constraint is then introduced to the group LASSO model to explore relevant characteristics of longitudinal MR data. On the one hand, a fused smoothness term $\sum_{i=1}^{t-1} \|W_i - W_{i+1}\|_F^2$ is incorporated into the group LASSO frame. Inspired by the fused LASSO proposed in (Tibshirani et al. 2005). Equation (1) ignores the temporal dependence among data, while the imposition of the fused smoothness term encourages the sparsity in the weight difference and thus the longitudinal similarity is fully utilized. On the other hand, locality-preserving-projection (LPP) is shown to alleviate the sensitivity of noise or outlier in the training samples (He et al. 2005). The preservation of the neighborhood structure of data points can be realized to reflect the similarity of data in the same class by adding the LPP based regularization term to the objective function, and a graph Laplacian matrix is constructed using the local topological relation. A weighted graph $X_i W_i$ is constructed based on the given dataset X_i , and the similarity matrix S_i between the p -th and q -th node of X_i is given by

$$S_{i,pq} = \exp\left(-\frac{\|X_{i,p} - X_{i,q}\|^2}{\sigma_3}\right), \quad (2)$$

where $X_{i,p}, X_{i,q}$ are the p -th and q -th row of X_i and σ_3 is a free parameter to be tuned empirically. Big $S_{i,pq}$ indicates that the difference between the p -th and q -th node is small, thus a locality preserving feature can be chosen by minimizing

$$\begin{aligned} & \sum_{pq} (f_{i,rp} - f_{i,rq})^2 S_{i,pq} \\ &= \sum_{pq} (f_{i,rp}^2 + f_{i,rq}^2 - 2f_{i,rp}f_{i,rq}) S_{i,pq} \\ &= \sum_{pq} (2f_{i,rp}^2 S_{i,pq} - 2f_{i,rp}f_{i,rq} S_{i,pq}) \\ &= 2(X_i W_i)^T (D_i - S_i) (X_i W_i), \end{aligned} \quad (3)$$

where $f_{i,rp}$ and $f_{i,rq}$ denote the p -th and q -th sample of the r -th feature for X_i , and D_i is the diagonal matrix defined as $D_i = \sum_{p,q=1}^n S_{i,pq}$. $L_i = D_i - S_i$ represents the Laplacian matrix for task i . Based on the discussions above, the objective function of the proposed *spatial-temporal group sparse model (STGL)* is defined as

$$\begin{aligned} \min_w \frac{1}{2} \sum_{i=1}^t \|Y_i - X_i W_i\|_F^2 + \rho_0 \|W\|_{2,1} + \rho_1 \sum_{i=1}^{t-1} \|W_i - W_{i+1}\|_F^2 \\ + \rho_2 \sum_{i=1}^t (X_i W_i)^T L_i (X_i W_i), \end{aligned} \quad (4)$$

where ρ_1, ρ_2 are the temporal and spatial regularization parameters, respectively.

To remove the potential outliers in the training data, the correntropy algorithm is studied as it is effectively against non-Gaussian noise and impulsive noise (Liu et al. 2007), and has seen numerous applications in computer vision and signal processing field (He et al. 2012). Inspired by its effectiveness, we include *correntropy* in the STGL (CSTGL) for robust feature selection, and the new CSTGL model is formulated as

$$\begin{aligned} \min_w \frac{1}{2} \sum_{i=1}^t \exp\left(-\frac{\|Y_i - X_i W_i\|_F^2}{\sigma_4}\right) + \rho_0 \sum_{i=1}^{t-1} \|W_i - W_{i+1}\|_F^2 \\ + \rho_2 \sum_{i=1}^t (X_i W_i)^T L_i (X_i W_i), \end{aligned} \quad (5)$$

where σ_4 denotes the kernel size and controls all properties of correntropy. By removing outliers via correntropy, the CSTGL model is expected to select robust and informative features for subsequent regression.

We start with solving the STGL model. To the best of our knowledge, the objective function defined in Eq. (4) is the first to simultaneously include the group sparsity as well as the spatial-temporal regularizations. This objective function cannot be solved by the existing sparse learning models. After careful analysis, we observe that the objective function $f(W)$ is the sum of the smooth part $f_s(W)$ and the non-smooth part $f_n(W)$, where

$$\begin{aligned} f_s(W) = \frac{1}{2} \sum_{i=1}^t \|Y_i - X_i W_i\|_F^2 + \rho_1 \sum_{i=1}^{t-1} \|W_i - W_{i+1}\|_F^2 \\ + \rho_2 \sum_{i=1}^t (X_i W_i)^T L_i (X_i W_i), \end{aligned} \quad (6)$$

$$f_n(W) = \rho_0 \|W\|_{2,1} \quad (7)$$

By applying the first-order Taylor expansion at the point \mathbf{W}_k for $f_s(\mathbf{W})$, the following model $g(\bullet)$ is constructed to approximate the composite function (Nesterov 2013; Boyd et al. 2003)

$$g_{L, \mathbf{W}_k}(\mathbf{W}) = f_s(\mathbf{W}_k) + \langle f'_s(\mathbf{W}_k), \mathbf{W} - \mathbf{W}_k \rangle + \frac{L}{2} \|\mathbf{W} - \mathbf{W}_k\|_2^2 + f_{ns}(\mathbf{W}), \quad (8)$$

where $L > 0$, and the regularization term $\frac{L}{2} \|\mathbf{W} - \mathbf{W}_k\|_2^2$ keeps \mathbf{W}_k in the neighborhood of \mathbf{W} .

The optimization algorithms are implemented via the accelerated gradient methods (AGM). Differing from the traditional gradient method, the search point \mathbf{U}_k of AGM at every iteration is the affine combination of previous two points \mathbf{W}_k and \mathbf{W}_{k-1} , written as:

$$\mathbf{U}_k = \mathbf{W}_k + \beta_k (\mathbf{W}_k - \mathbf{W}_{k-1}), \quad (9)$$

where β_k is a properly chosen coefficient.

The approximate solution \mathbf{W}_{k+1} can be obtained by minimizing $g_{L_k, \mathbf{U}_k}(\mathbf{W})$:

$$\mathbf{W}_{k+1} = \arg(\min)_{\mathbf{W}} g_{L_k, \mathbf{U}_k}(\mathbf{W}), \quad (10)$$

where L_k is found through line search following the Armijo-Goldstein rule. The update of \mathbf{W} is to solve the $\ell_{2,1}$ -norm regularized Euclidean projection problem, which belongs to the composite absolute penalties (CAP) family.

$$\begin{aligned} \mathbf{W}_{k+1} &= \arg \min_{\mathbf{W}} \frac{1}{2} \|\mathbf{W} - \mathbf{V}\|_F^2 + \frac{1}{L} f_{ns} \\ (\mathbf{W}) &= \arg \min_{\mathbf{w}_1, \dots, \mathbf{w}_d} \frac{1}{2} \sum_{j=1}^d \left(\|\mathbf{w}_j - \mathbf{v}_j\|_2^2 \right) + \frac{\rho_0}{L} \|\mathbf{w}_j\|_2, \end{aligned} \quad (11)$$

where \mathbf{w}_j and \mathbf{v}_j denote the j -th row of \mathbf{W} and \mathbf{V} , respectively, and $\mathbf{V} = \mathbf{W}_k - \frac{1}{L} f'_s(\mathbf{W}_k)$. Therefore, this problem is decomposed into d separate subproblems through Eq. (11.) It is crucial to solve the update step efficiently and we adopt the algorithm in (Liu and Ye 2010; Chen et al. 2009) to compute \mathbf{W}_{k+1} :

$$\mathbf{w}_j^* = \begin{cases} \left(1 - \frac{\rho_0}{L \|\mathbf{v}_j\|_2}\right) \mathbf{v}_j & \text{if } \|\mathbf{v}_j\|_2 > \frac{\rho_0}{L} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

This algorithm falls into the category of exact ADM and thus inherits its convergence rate of $O(1/K^2)$ for the convex loss, where K is the maximum iteration. The full optimization procedure is summarized in Algorithm 1.

The CSTGL model in Eq. (5) is solved in a similar manner with the exception for $g_{L, \mathbf{U}_k}(\mathbf{W})$, where $f_s(\mathbf{U}_k)$ and $f'_s(\mathbf{U}_k)$ are calculated in a different way.

Algorithm 1: Group sparsity learning by encoding spatial-temporal information

Input: $\rho_0 \geq 0, \rho_1 \geq 0, \rho_2 \geq 0, L_0 > 0, \mathbf{W}_0, K$
Output: \mathbf{W}_{K+1}
 1: Initialize $\mathbf{W}_1 = \mathbf{W}_0, \alpha_{-1} = 0, \alpha_0 = 1$, and $L = L_0$.
 2: **for** $k=1$ to K **do**
 3: Set $\beta_k = \frac{\alpha_{k-2}-1}{\alpha_{k-1}}, \mathbf{U}_k = \mathbf{W}_k + \beta_k(\mathbf{W}_k - \mathbf{W}_{k-1})$
 4: Find the smallest $L = L_{k-1}, 2L_{k-1}, \dots$ such that
 $f(\mathbf{W}_{k+1}) \leq g_{L, \mathbf{U}_k}(\mathbf{W}_{k+1})$,
 where \mathbf{W}_{k+1} is updated as in Eq. (8).
 5: Set $L_k = L$ and $\alpha_k = \frac{1 + \sqrt{1 + 4\alpha_k^2}}{2}$
 6: **end for**

Ensemble SVR prediction

Using the selected features, SVR is adopted for score prediction. SVR uses the same principles as SVM, which is to minimize error by individualizing the hyperplane that maximizes the margin for classification and regression (Vapnik and Lerner 1963). The main difference lies on the fact that the output of SVR is a real number, which has infinite possibilities and thus makes the prediction task challenging. In the case of regression, a margin of tolerance is set and the algorithm is more complicated.

The feature selection algorithm might cause overfitting problem because it is dependent on the training data, and thus it is not optimal for the test data. To address this problem, an ensemble classification strategy is adopted by averaging the top predicted scores obtained by the proposed feature selection model with various sets of parameters, which is different from building a single classifier with an optimal subset of features. Improvement of the generalization ability, robustness of individual classifiers, and regression accuracy can be also achieved via ensemble learning.

Experimental results

Two regression variables, i.e. MMSE and ADAS-Cog, at four time points are estimated, respectively. Performance evaluation is performed on a subset of the ADNI dataset composed of 91 AD, 104 MCI-C, 98 MCI-NC and 152 NC with the four proposed models: STGL, Ensemble STGL (ESTGL), Correntropy STGL (CSTGL) and Ensemble CSTGL (ECSTGL). The following methods are chosen for comparison: LASSO, group LASSO (GL), temporally-constrained group LASSO (TGL) (Jie et al. 2017)

and subspace-regularized group LASSO (SGL) (Zhu et al. 2016a). Four regression performance metrics are adopted, which are given by

$$\text{rMSE}(Y, \hat{Y}) = \sqrt{\frac{\|Y - \hat{Y}\|_2^2}{m}}, \quad (13)$$

$$\text{Corr}(Y, \hat{Y}) = \frac{\text{cov}(\hat{Y}, Y)}{\sigma(Y)\sigma(\hat{Y})}, \quad (14)$$

$$\text{nMSE}(Y, \hat{Y}) = \frac{\sum_{i=1}^t \|Y_i - \hat{Y}_i\|_2^2 / \sigma(Y_i)}{\sum_{i=1}^t m_i}, \quad (15)$$

$$\text{wR}(Y, \hat{Y}) = \frac{\sum_{i=1}^t \text{Corr}(Y_i, \hat{Y}_i) m_i}{\sum_{i=1}^t m_i}, \quad (16)$$

where Y and \hat{Y} stand for the actual and predicted clinical scores at a single time point with m objects for rMSE and Corr. In the case of nMSE and wR, Y_i and \hat{Y}_i denote the actual and predicted clinical scores at time point i with m_i objects.

Due to limited number of samples, a 10-fold cross-validation technique is applied. Specifically, the whole dataset is randomly divided into 10 subsets. One subset is used for testing and the remaining nine subsets are used for training. This process is repeated 10 times to avoid any possible bias caused by dataset partitioning for cross-validation. The final result is obtained by averaging the results from all experiments. To be specific, it is guaranteed that optimal parameters are chosen from the overall 10-fold cross validation, and for the implementation of ensemble, top sets of optimal parameters are consistent throughout the 10-fold cross validation.

Estimating clinical scores

As time goes by, the increasing standard deviation of longitudinal scores indicates that the data spreads out over a wide range of values, which makes prediction even more challenging. The results of nMSE and wR are illustrated in Table 1 and rMSE and Corr results are plotted in Fig. 2. We have the following observations.

- 1) The proposed method STGL consistently outperforms other methods (i.e. LASSO, GL, TGL and SGL) in estimating clinical scores. To be specific, it achieves the smallest nMSE and biggest wR among the competing methods.
- 2) The ensemble strategy moderately improves the prediction performance, which is indicated by the slight advantage of ESTGL over STGL and that of ECSTGL over CSTGL.
- 3) The correntropy model yields remarkably more accurate results by enabling robust feature selection, as revealed by the comparisons between STGL and CSTGL.

Since the two terms of correntropy and LPP both deal with spatial outliers/noise, it would be meaningful to find out whether these two terms have a functional overlap. Another simulation is run by using Eq. (5) without the last LPP term, and the corresponding model is denoted as CTGL. For MMSE and ADAS-Cog prediction, the nMSEs obtained by CTGL are 2.397 and 4.692, respectively. Compared with Table 1, it can be seen that CTGL obtains more accurate predictions than STGL, which indicates that for this dataset and for the proposed model, correntropy is superior to LPP. Additionally, the ultimately integrated model CSTGL still has the best results, showing that the two terms of correntropy and LPP promote the advantage of each other, and thus the combination of them is beneficial.

Further analysis of the proposed ECSTGL method is conducted using the scatter plots (see Fig. 3) of predicted values versus the ground truth of MMSE and ADAS_Cog on the testing data, respectively. Two lines are drawn in

Table 1 Comparison of nMSEs and wRs of different methods in predicting longitudinal MMSE and ADAS-Cog scores

MMSE	LASSO	GL	TGL	SGL	STGL	ESTGL	CSTGL	ECSTGL
nMSE	2.948 ± 0.511	2.679 ± 0.448	2.558 ± 0.334	2.560 ± 0.347	2.425 ± 0.307	2.418 ± 0.289	2.344 ± 0.275	2.323 ± 0.279
wR	0.467 ± 0.126	0.545 ± 0.127	0.580 ± 0.080	0.578 ± 0.093	0.604 ± 0.076	0.606 ± 0.074	0.622 ± 0.067	0.629 ± 0.062
ADAS-Cog	LASSO	GL	TGL	SGL	STGL	ESTGL	CSTGL	ECSTGL
nMSE	6.277 ± 0.707	5.391 ± 0.651	5.263 ± 0.633	5.184 ± 0.695	4.718 ± 0.444	4.690 ± 0.436	4.628 ± 0.439	4.595 ± 0.446
wR	0.383 ± 0.161	0.574 ± 0.061	0.587 ± 0.070	0.604 ± 0.073	0.648 ± 0.060	0.647 ± 0.059	0.655 ± 0.058	0.655 ± 0.054

Bold numbers signifies the best results of ECSTGL

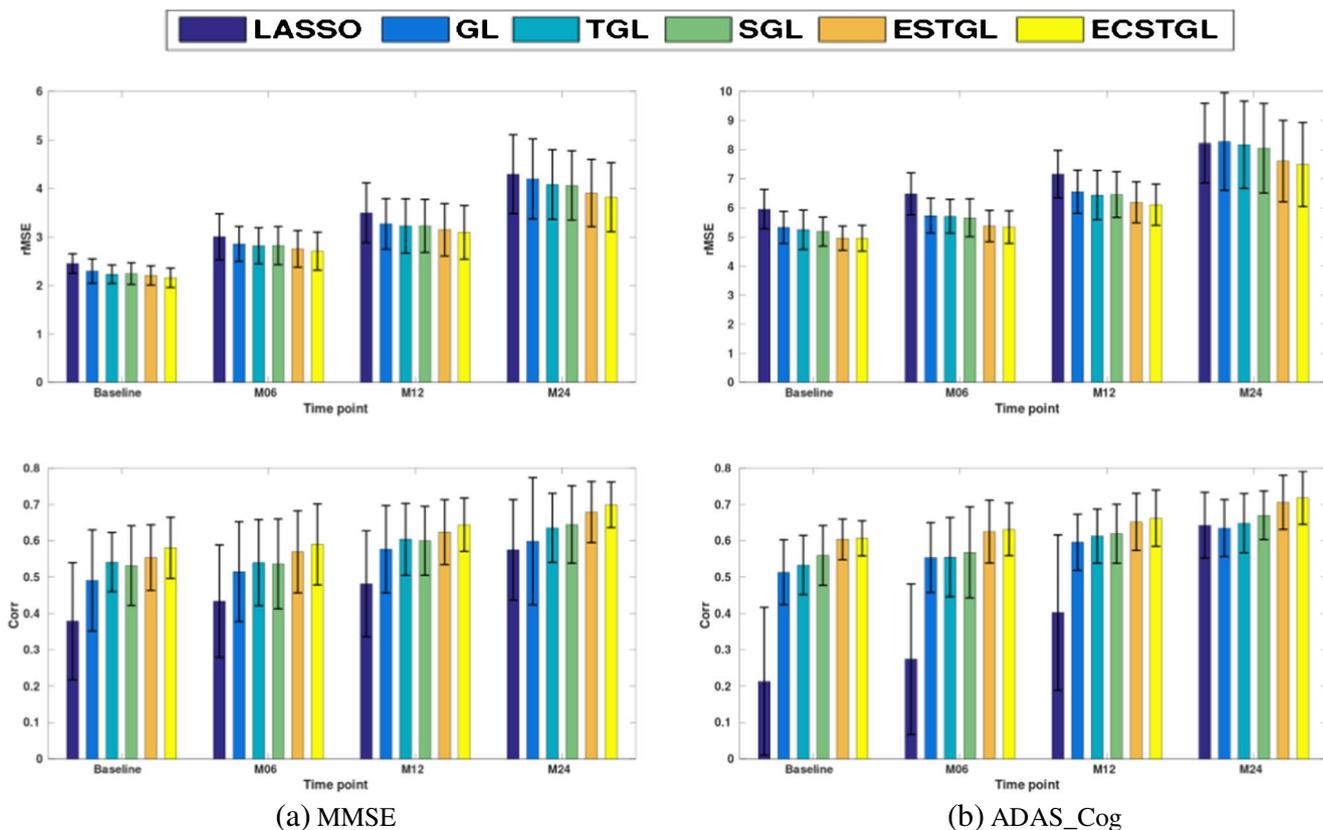


Fig. 2 Mean values and standard deviations of rMSEs and Corrs between actual clinical scores and predicted scores by different methods. Proposed models generally have the best performance, as shown as the smallest rMSE and the biggest Corr

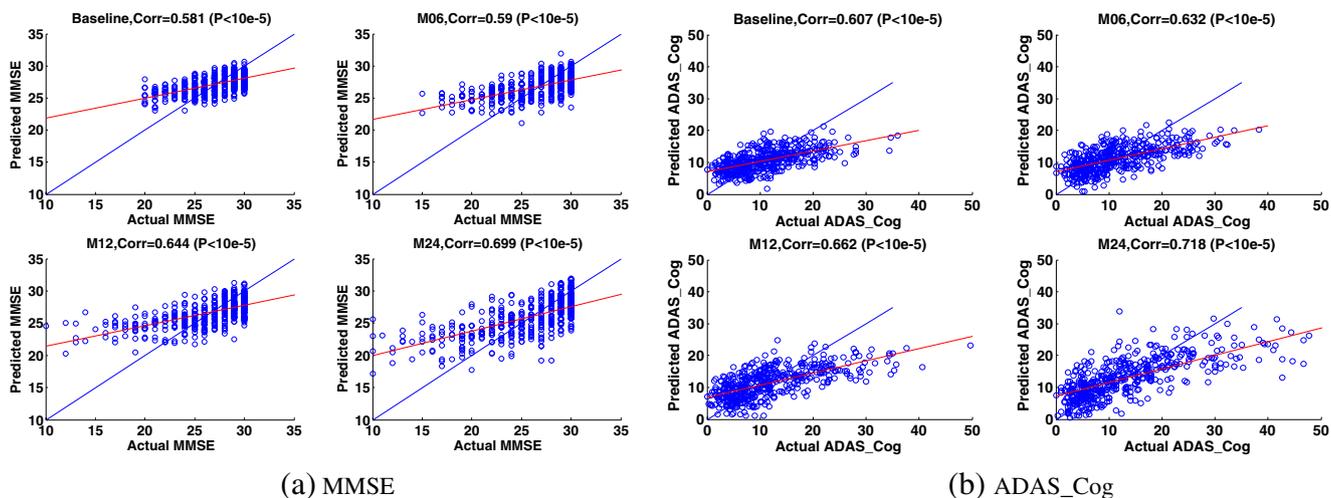


Fig. 3 Scatter plots of ground truth versus predicted scores predicted by ECSTGL method. Blue line represents reference line and red line represents regression line. Regression accuracy is higher when two lines are closer

each scatter plot. The reference line indicates the perfect correlation where the predicted value exactly equals to the actual value. The other is the regression line, obtained by performing least squares regression on the points shown

in the scatter plots. The closer the reference line is to the regression line, the better the predictions. As seen from the figures, predicted values and actual clinical scores have a high correlation, and patients with high

actual MMSE and low actual ADAS_Cog values are more predictable.

Most discriminative regions

The potential of using certain regions of the brain as biomarkers in AD diagnosis based on the selected frequency of ROIs is investigated. Due to the adoption of the 10-fold cross-validation and the ensemble technique, the brain regions with top occurrence frequency in all tests are marked as the most important ones. In MMSE prediction, there are 17 brain regions that are selected throughout the whole selection process, while in the case of ADAS_Cog, nine regions are selected all the time and 11 regions are selected one time fewer. These regions are shown in Fig. 4 and listed in Table 2. There are nine common regions, which indicate that the clinical scores, MMSE, and ADAS_Cog are highly related. The brain regions selected by the proposed ECSTGL such as amygdala, medial occipitotemporal gyrus, inferior temporal gyrus, temporal pole, and uncus are known to be related to AD severity in previous literatures (Misra

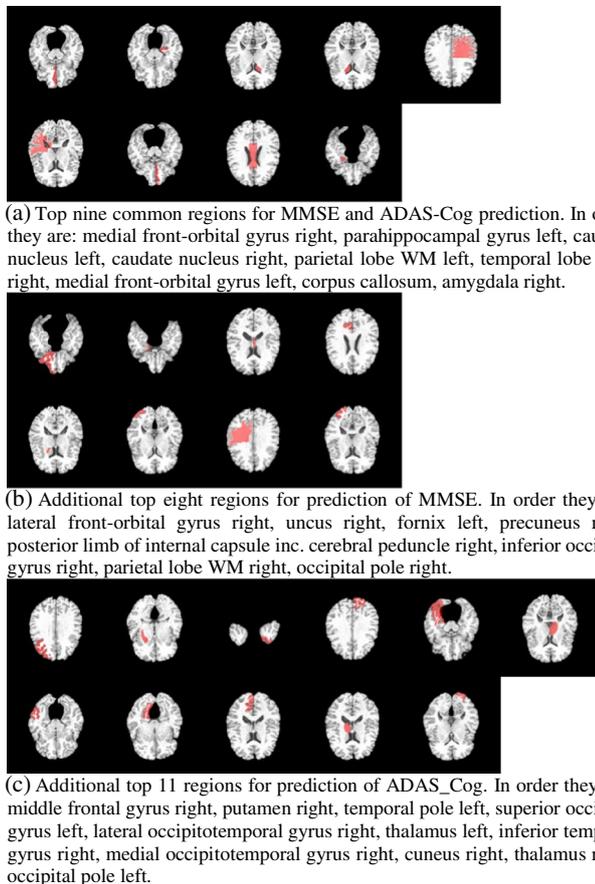


Fig. 4 Important brain regions detected by proposed ECSTGL model

Table 2 Most important brain regions selected by ECSTGL model

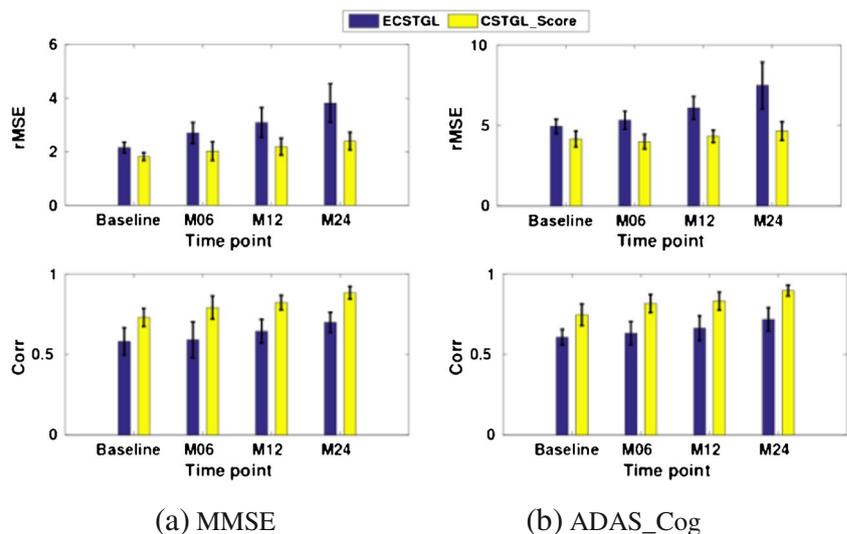
MMSE	ADAS_Cog
Amygdala right	Amygdala right
Caudate nucleus left	Caudate nucleus left
Caudate nucleus right	Caudate nucleus right
Corpus callosum	Corpus callosum
Fornix left	Cuneus right
Inferior occipital gyrus right	Inferior temporal gyrus right
Lateral front-orbital gyrus right	Lateral occipitotemporal gyrus right
Medial front-orbital gyrus left	Medial front-orbital gyrus left
Medial front-orbital gyrus right	Medial front-orbital gyrus right
Occipital pole right	Medial occipitotemporal gyrus right
	Middle frontal gyrus right
	Occipital pole left
Parahippocampal gyrus left	Parahippocampal gyrus left
Parietal lobe WM left	Parietal lobe WM left
Precuneus right	Putamen right
Posterior limb of internal capsule inc. cerebral peduncle right	Superior occipital gyrus left
Temporal lobe WM right	Temporal lobe WM right
Parietal lobe WM right	Temporal pole left
Uncus right	Thalamus left
	Thalamus right

et al. 2009; Convit et al. 2000). It might be noticed that the well-known hippocampus areas do not make to our list. The reason is that here we list the areas that are picked 100% and 99% through the whole process, while hippocampus areas are selected 98% or 97%. In addition, the proposed approach successfully identifies three pairs of ROIs that are not important for the score prediction of AD: lateral ventricle left/right, subthalamic nucleus left/right and anterior limb of internal capsule left/right.

Using clinical scores as attributes

In the last subsection, it is indicated that ADAS-Cog and MMSE are highly related, therefore it would be beneficial to use them as the attributes for the prediction of each other. In this subsection, we run several simulations for validation. For the prediction of MMSE, CSTGL model is first used to select features. Subsequently, ADAS-Cog is incorporated as the attributes for SVR. Similar procedures are followed by the prediction of ADAS-Cog. The rMSE and Corr results are compared with those of ECSTGL, as shown in Fig. 5. It can be observed that the incorporation of scores as attributes remarkably improves the prediction accuracy due to the high correlation between the scores.

Fig. 5 Mean values and standard deviations of rMSEs and Corrs between actual clinical scores and predicted scores predicted by ECSTGL and CSTGL_Score



Discussions

Significance of results

Though sparse learning models have been widely applied to the diagnosis of AD, commonly available information at multiple time points has not been fully utilized. In this paper, we explored the potential of the longitudinal data to improve the performance of automatic AD diagnosis. It is worth noting that the proposed method is the first attempt to simultaneously include the group sparsity as well as the temporal and spatial information for longitudinal data analysis. The regularization term incorporating the temporal smoothness reflects well the changes of the longitudinal patterns of the brain with the disease progression, and the LPP regularization term is beneficial for the prediction of clinical scores with big variance. Another main novelty lies in the formulation of correntropy into the STGL model. The correntropy induced model successfully extracts informative features, while removes outliers from training data. In addition, the adoption of ensemble strategy further improves the prediction accuracy. Extensive experiments demonstrated that the proposed approach outperforms state-of-the-art

methods in terms of regression performance. In addition, the proposed model successfully discovers more biomarkers than other models reported in the literature. For example, precuneus, posterior and parahippocampal regions were missed by TGL (Jie et al. 2017), but were detected by the proposed ECSTGL.

Regression performance and prediction of MCI conversion

Various methods were proposed to predict clinical scores of AD based on neuroimaging data. We compared the correlation coefficients of several recently proposed methods with the proposed ECSTGL model in Table 3, which shows that ECSTGL achieves the best results.

Clinical score prediction is not only useful for monitoring disease progression, but also aids patients and family to prepare for the possible progression from MCI to AD. Hence, the accuracy rate for predicting MCI progression is very important as well. The rate of relevant methods is summarized in Table 4, and the proposed ECSTGL model shows outstanding performance.

Table 3 Comparisons of correlation coefficients of state-of-the-art methods and proposed ECSTGL method

Method	Subject	Time span	MMSE	ADAS-Cog
Principal component analysis (Duchesne et al. 2009)	20 MCI-C + 29 MCI-NC	12 months	0.31	/
Joint Bayesian classifier (Fan et al. 2010)	52 AD + 148 MCI + 64 NC	6 moths	0.57	0.52
Group sparse learning method (Yan et al. 2015)	172 AD + 349 MCI + 197 NC	Single time point	0.555	0.644
Temporally constrained group sparse learning (Jie et al. 2017)	91 AD + 202 MCI + 152 NC	24 months	0.613	0.639
Proposed method ECSTGL	91 AD + 202 MCI + 152 NC	24 months	0.629	0.655

Table 4 Comparisons of accuracy rate of state-of-the-art methods and proposed ECSTGL method for predicting MCI conversion

Method	Subject	Time span	Accuracy
Four supervised learning methods (Aguilar et al. 2013)	21 MCI-C+98 MCI-NC	12 months	67.4% – 74.7%
Local linear embedding (Liu et al. 2013)	97 MCI-C+93 MCI-NC	36 months	68%
Temporally constrained group sparse learning (Jie et al. 2017)	104 MCI-C+98 MCI-NC	24 months	74.7% – 75.7%
Proposed method ECSTGL	104 MCI-C+98 MCI-NC	24 months	78%

Effect of parameters

There are five parameters that can be freely tuned in the proposed ECSTGL model, where ρ_0 , ρ_1 and ρ_2 are the weights of the group sparsity, temporal and spatial regularization terms, respectively. These parameters balance the relative contributions of these terms. σ_3 is the width of the heat kernel in the spatial regularization term and σ_4 the kernel size that controls all properties of correntropy. For simplicity, σ_4 is set to one in all simulations. The larger ρ_0 emphasizes the group sparsity constraint imposed by the $\ell_{2,1}$ -norm, hence, fewer features are selected for estimating the clinical scores.

For the prediction of MMSE, the optimal parameters of the STGL method are $\rho_0 = 0.35$, $\rho_1 = 0.55$, $\rho_2 = 0.25$, $\sigma_3 = 0.5$, while those of CSTGL are $\rho_0 = 0.005$, $\rho_1 = 0.1$, $\rho_2 = 0.35$, $\sigma_3 = 1$. For the prediction of ADAS-Cog, the optimal parameters of the STGL method are $\rho_0 = 0.6$, $\rho_1 = 0.4$, $\rho_2 = 0.3$, $\sigma_3 = 1$, while those of CSTGL are $\rho_0 = 0.005$, $\rho_1 = 0.005$, $\rho_2 = 0.15$, $\sigma_3 = 0.1$.

In this subsection, we mainly focus on the CSTGL model. ρ_0 is fixed to be the optimal value 0.005, and ρ_1 and ρ_2 are chosen from [0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5], σ_3 from [0.05, 0.1, 0.5, 1]. The average of rMSEs across four time points is plotted in Fig. 6 with different parameter settings. Big rMSE is more likely to appear when ρ_1 and ρ_2 are smaller, which indicates the efficacy of the spatial–temporal regularization terms.

Subsequently, the effect of ρ_0 on the prediction accuracy is explored with $[\rho_1 \rho_2 \sigma_3]$ set to the optimal values. Similarly, the average of rMSEs across four time points is plotted in Fig. 7 with respect to ρ_0 . The prediction error generally goes up with bigger ρ_0 , where fewer features are selected.

Conclusions

This paper carried out the longitudinal data analysis for the computer-aided diagnosis of AD. A novel feature selection method was proposed by integrating correntropy and spatial–temporal information in a sparse linear learning framework, followed by the ensemble SVR regression. Experimental results of estimating clinical scores from MRI data at multiple time points demonstrated the advantage of the proposed method over existing sparse methods in both prediction accuracy and the ability to

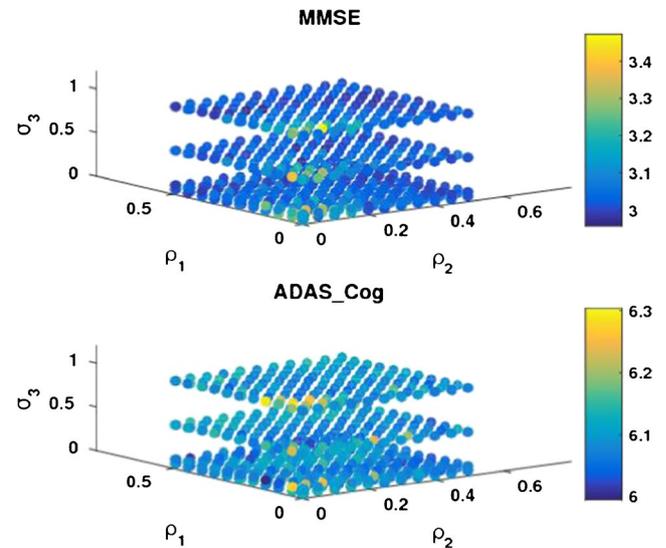


Fig. 6 Prediction accuracy with different settings of parameters $[\rho_1 \rho_2 \sigma_3]$ when ρ_0 is set to be an optimal value of 0.005

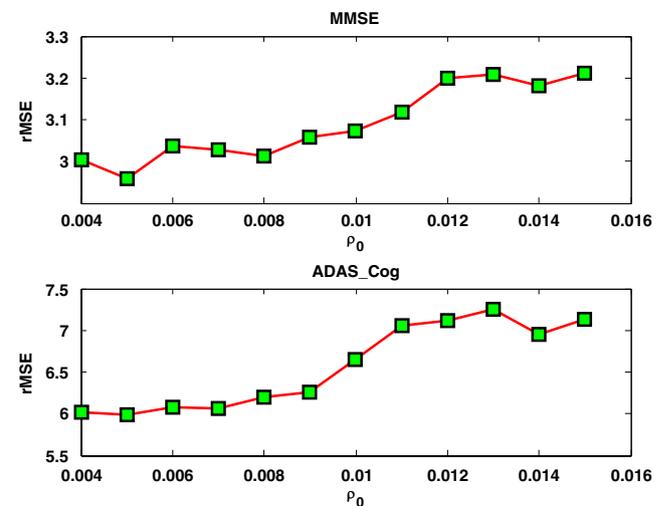


Fig. 7 Prediction accuracy with different settings of parameter ρ_0 when $[\rho_1 \rho_2 \sigma_3]$ are set to optimal values

discover disease-related biomarkers. In future, we would like to explore multimodal information as (Lei et al. 2016, 2017) for more comprehensive analysis of AD/MCI data.

Acknowledgements Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Funding This study was funded partly by National Natural Science Foundation of China (Nos. 61771321, 61501305 and 81771922), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Key Basic Research Project (Nos. KQJSCX20170327151357330, JCYJ20170302153337765, JCYJ20160307154003475 JCYJ20150525092940982 and 201502007), Shenzhen Peacock Plan (NO. KQTD2016053112051497), the Interdisciplinary Innovation Team of Shenzhen University, and the National Natural Science Foundation of Shenzhen University (Nos. 827-000152, 2016077 and 201565 and 2016089). 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, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; MesoScale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (<http://www.fnih.org>). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Aguilar, C., Westman, E., Muehlboeck, J. S., et al. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research*, *212*(2), 89–98.
- Alzheimer's Association. (2015). 2015 Alzheimer's disease facts and figures. *Alzheimers Dement*, *11*(3), 332–384.
- Boyd, S., Xiao, L., & Mutapcic, A. (2003). Subgradient Methods.
- Che, D., Liu, Q., Rasheed, K., & Tao, X. (2011). Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Advances in Experimental Medicine and Biology*, *696*, 191–199.
- Chen, X., Pan, W., Kwok, J. T., & Carbonell, J. G. (2009). Accelerated gradient method for multi-task sparse learning problem. In *9th IEEE International Conference on Data Mining* (pp. 746–751).
- Convit, A., De Asis, J., De Leon, M. J., Tarshish, C. Y., De Santi, S., & 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*(1), 19–26.
- Duchesne, S., Caroli, A., Geroldi, C., Collins, D. L., & Frisoni, G. B. (2009). Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. *Neuroimage*, *47*(4), 1363–1370.
- Eskildsen, S. F., Coupé, P., Fonov, V. S., Pruessner, J. C., & Collins, D. L. (2015). Structural imaging biomarkers of Alzheimer's disease: predicting disease progression. *Neurobiology of Aging*, *36*, S23–S31.
- Fan, Y., Kaufer, D., & Shen, D. (2010). Joint estimation of multiple clinical variables of neurological diseases from imaging patterns. *ISBI* (pp. 852–855).
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.
- Gallego-Jutglà, E., Solé-Casals, J., Vialatte, F.-B., Elgendi, M., Cichocki, A., & Dauwels, J. (2015). A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. *Journal of Neural Engineering*, *12*(1), 16018.
- Hao, X., Yao, X., Yan, J., et al. (2016). Identifying Multimodal intermediate phenotypes between genetic risk factors and disease status in Alzheimer's disease. *Neuroinformatics*, *14*(4), 1–14.
- Hao, X., Li, C., Yan, J., et al. (2017). Identification of associations between genotypes and longitudinal phenotypes via temporally-constrained group sparse canonical correlation analysis. In *Bioinformatics*, *33*(14), i341–i349.
- He, X., Cai, D., & Niyogi, P. (2005). Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, *18*, 507–514.
- He, R., Tan, T., Wang, L., & Zheng, W. (2012). L2,1 regularized corentropy for robust feature selection. *CVPR* (pp. 2504–2511).
- Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., Johnson, S. C., & Alzheimer's Disease Neuroimaging Initiative. (2009). Spatially augmented LP boosting for AD classification with evaluations on the ADNI dataset. *Neuroimage*, *48*(1), 138–149.
- 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*(1), 238–249.
- Kochanek, K. D., Xu, J., Murphy, S. L., Minino, A. M., & Kung, H. C. (2012). National vital statistics reports deaths: final data for 2009. *National Center for Health Statistics*, *60*(3), 1–117.
- Kuncheva, L. I., Rodriguez, J. J., Plumpton, C. O., Linden, D. E. J., & Johnston, S. J. (2010). Random subspace ensembles for FMRI classification. *IEEE Transactions on Medical Imaging*, *29*(2), 531–542.
- Lei, B., Chen, S., Ni, D., Wang, T., & Alzheimer's Disease Neuroimaging Initiative. (2016). Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion. *Frontiers in Aging Neuroscience*, *8*, 77.
- Lei, B., Chen, S., Ni, D., & Wang, T. (2017). Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. *IEEE Transactions on Cybernetics*, *47*(4), 1102–1113.
- Liu, J., & Ye, J. (2010). Efficient l1/lq norm regularization. *arXiv Prepr. arXiv*, *1009.4766*, 1–19.

- Liu, W., Pokharel, P. P., & Principe, J. C. (2007). Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Transactions on Signal Processing*, *55*(11), 5286–5298.
- Liu, M., Zhang, D., Shen, D. & Alzheimer's Disease Neuroimaging Initiative. (2012). Ensemble sparse classification of Alzheimer's disease. *Neuroimage*, *60*(2), 1106–1116.
- Liu, X., Tosun, D., Weiner, M. W., Schuff, N. & Alzheimer's Disease Neuroimaging Initiative. (2013). Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. *Neuroimage*, *83*, 148–157.
- Liu, F., Wee, C. Y., Chen, H., & 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.
- Misra, C., Fan, Y., & 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*(4), 1415–1422.
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, *43*(11), 2412–2414.
- Nesterov, Y. (2013). Gradient methods for minimizing composite function. *Mathematical Programming*, *140*(1), 125–161.
- Ng, B., & Abugharbieh, R. (2011). Generalized sparse regularization with application to fMRI brain decoding. *Information Processing in Medical Imaging*, *22*, 612–623.
- Ota, K., Oishi, N., Ito, K., & Fukuyama, H. (2015). Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease. *Journal of Neuroscience Methods*, *256*, 168–183.
- Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*(11), 1356–1364.
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326.
- Shen, D., Resnick, S. M., & Davatzikos, C. (2003). 4D HAMMER image registration method for longitudinal study of brain changes. *Human Brain Mapping*, *1*–8.
- Shen, L., Kim, S., Qi, Y., et al. (2011). Identifying neuroimaging and proteomic biomarkers for MCI and AD via the elastic net. *Multimodal Brain Image Analysis*, *7012*, 27–34.
- Shi, F., Wang, L., Dai, Y., Gilmore, J. H., Lin, W., & Shen, D. (2012). LABEL: pediatric brain extraction using learning-based meta-algorithm. *Neuroimage*, *62*(3), 1975–1986.
- Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97.
- Suk, H. I., Lee, S.-W., Shen, D. & Alzheimer's Disease Neuroimaging Initiative. (2016). Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis. *Brain Structure & Function*, *221*(5), 2569–2587.
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO: a retrospective. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *58*(1), 267–288.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused LASSO. *Journal of the Royal Statistical Society. Series B, Methodological*, *67*(1), 91–108.
- Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, *24*, 774–780.
- Wang, Z., Zhu, X., Adeli, E., et al. (2017). Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning. *Medical Image Analysis*, *39*, 218–230.
- Yan, J., Li, T., Wang, H., et al. (2015). Cortical surface biomarkers for predicting cognitive outcomes using group l2,1 norm. *Neurobiology of Aging*, *36*(1), S185–S193.
- Yau, W. Y. W., Tudorascu, D. L., McDade, E. M., Ikonomic, S., James, J. A., Minhas, D., et al. (2015). Longitudinal assessment of neuroimaging and clinical markers in autosomal dominant Alzheimer's disease: a prospective cohort study. *Lancet Neurology*, *14*(8), 804–813.
- Yin, S., Wen, Z., Shi, J., Peng, Y., Peng, J., et al. (2017). Manifold preserving: an intrinsic approach for semisupervised distance metric learning. *IEEE Transactions on Neural Networks and Learning Systems*, *99*, 1–12.
- 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*(1), 45–57.
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D. & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *55*(3), 856–867.
- Zhang, D., Shen, D. & Alzheimer's Disease Neuroimaging Initiative. (2012a). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage*, *59*(2), 895–907.
- Zhang, D., Liu, J., & Shen, D. (2012b). Temporally-Constrained Group Sparse Learning for Longitudinal Data Analysis. MIC-CAI (pp. 264–271).
- Zhang, L., Wang, L., & Lin, W. (2012c). Conjunctive patches subspace learning with side information for collaborative image retrieval. *IEEE Transactions on Image Processing*, *21*(8), 3707–3720.
- Zhu, X., Huang, Z., Shen, H. T., Cheng, J., & Xu, C. (2012). Dimensionality reduction by mixed kernel canonical correlation analysis. *Pattern Recognition*, *45*(8), 3003–3016.
- Zhu, X., Suk, H. I., Wang, L., Lee, S. W., & Shen, D. (2015). A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. *Medical Image Analysis*, *75*(6), 570–577.
- Zhu, X., Suk, H. I., Lee, S. W., & Shen, D. (2016a). Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. *IEEE Transactions on Biomedical Engineering*, *63*(3), 607–618.
- Zhu, X., Suk, H. I., Lee, S. W., & Shen, D. (2016b). Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. *Brain Imaging and Behavior*, *10*(3), 818–828.
- Zhu, X., Suk, H. I., Lee, S. W., & Shen, D. (2017). Discriminative self-representation sparse regression for neuroimaging-based Alzheimer's disease diagnosis. *Brain Imaging and Behavior*. <https://doi.org/10.1007/s11682-017-9731-x>.