# Modeling and predicting AD progression by regression analysis of sequential clinical data

Qing Xie[a], Su Wang[a], Jia Zhu[b], Xiangliang Zhang[a,*], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Division of CEMSE, King Abdullah University of Science and Technology, Saudi Arabia
[b] School of Computer Science, South China Normal University, China

## ARTICLE INFO

## ABSTRACT

Alzheimer's Disease (AD) is currently attracting much attention in elders' care. As the increasing availability of massive clinical diagnosis data, especially the medical images of brain scan, it is highly significant to precisely identify and predict the potential AD's progression based on the knowledge in the diagnosis data. In this paper, we follow a novel sequential learning framework to model the disease progression for AD patients' care. Different from the conventional approaches using only initial or static diagnosis data to model the disease progression for different durations, we design a score-involved approach and make use of the sequential diagnosis information in different disease stages to jointly simulate the disease progression. The actual clinical scores are utilized in progress to make the prediction more pertinent and reliable. We examined our approach by extensive experiments on the clinical data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI). The results indicate that the proposed approach is more effective to simulate and predict the disease progression compared with the existing methods.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Alzheimer's Disease (AD) is a kind of presenile dementia, which is the common neurodegenerative disease appearing in the elders over 65 years old, with the symptoms of memory loss and disorder of central nervous system, and potentially resulting in death [1]. It has been reported that there are over 26 million AD patients all over the world by 2011, and this number will go beyond 114 million by 2050 [2,3]. Therefore, the timely AD diagnosis and treatment are of high significance and have attracted much concern recently, and researchers have taken efforts to simulate and predict the disease progression to benefit the elders' care.

With the increasing availability of medical diagnosis data [4] and the development of image processing [5], machine learning methods have been engaging the AD pattern analysis and progression prediction based on the massive diagnosis data, especially the medical images of brain scan, including Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET). These neuroimaging data are popularly used to understand the AD progression and identify the diagnosis of AD and its early stage, Mild Cognitive Impairment (MCI). Our work will focus on the challenges and innovations in neuroimaging analysis.

There are several research directions for analyzing the AD progression based on the medical image data. The first one starts from the image processing direction and focuses on the volume of different brain regions [6]. Guo et al. [7] discovered the relationship between the AD progression and the decrease of gray matter volume in the hippocampus, parahippocampal gyrus and insula and superior temporal gyrus. However, such methods only limit the applications within individual levels and it is hard to discover more general rules to identify the progression. Besides this, many works propose to develop classification models to distinguish different disease status [8–10], and determine the clinically defined categories of the subjects [11,12], such as AD, MCI and healthy Normal Control (NC). Survival model has also been applied to simulate the AD progression from the statistical point of view [13,14]. Recently, as the clinical scores are commonly accepted to indicate the disease status, regression model has been more popularly investigated to predict the AD patients' status in terms

of clinical scores, such as Mini-Mental State Examination (MMSE) or Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog), based on the original diagnosis data [15–17]. In the regression model research, since the high-dimensional neuro-image data are involved to generate the features, various methods have been proposed to effectively improve the performance, such as dimension reduction technique [18,19] and feature selection technique [20–23]. Specifically, Zhang et al. [24] proposed to address both disease diagnosis and clinical score prediction simultaneously, and combine these two tasks in a unified framework based on the correlated feature selection, which is also called Joint Regression and Classification (JRC) problem [25]. From the prediction effectiveness point of view, most methods of the regression model aim to predict the target score at a specific time point, such as one year [18], while more prediction scores at different time points are desired for a better prediction performance.

To address this problem, multi-task learning techniques [26] have been introduced into the regression model to simulate the disease progression and predict the clinical scores at different time points [27,28]. Multi-task learning aims to improve the performance of regression model building by utilizing the intrinsic commonality among different target tasks. The shared representation in parallel learning can help individual tasked be learned better. It has been demonstrated that multi-task learning is especially effective when the number of subjects is small and the number of input features is large, which is the case of AD simulation. The essential issue of multi-task learning is to discover how the tasks are related and identify the learning model. To achieve this, Zhou et al. [27,28] model the problem as longitudinal disease progression and the predictions of a patient's disease status at different time points are treated as regression tasks. These prediction tasks for different time points are performed simultaneously and the temporal smoothness across prediction models can be captured. Specifically, their work develop formulations that allow the simultaneous selection of a common set of biomarkers for multiple time points and specific sets of biomarkers for different time points, so as to capture the temporal patterns of the biomarkers in disease progression.

The approaches based on multi-task learning model tactfully fuse the regression tasks for different time durations into the uniformed framework. Although different kinds of loss functions have been employed to refine the prediction model, they still have the following two drawbacks. Firstly, most works only learn the regression model based on the subject features generated at the baseline time point, and all the prediction analysis for the following time points is derived from the baseline information. Therefore, if the subjects for model learning only cover limited status of AD progression, e.g., the early stage, the regression model will be inaccurate to describe the progression. Secondly, these approaches fail to use the evolving feature information in progress, which can be also utilized as feedback to enhance the regression model. For a long term task, which is the case of the engaged problem, the actual evolution of the subjects in progress can increase the uncertainty for prediction and should not be ignored.

Therefore, we propose a sequential data analysis mechanism to perform the regression model for simulating the AD progression, and the feedback concept is involved in the model to improve the prediction performance. In our work, we will effectively make use of the intermediate information during the AD patients' therapy, which includes the consecutive brain scan images and corresponding clinical scores. Based on these information, our work will jointly analyze the feature data in different time points, and the existing clinical scores will be embedded into the model to facilitate the disease simulation and the clinical score prediction. The relationship is established by building a fused sparse Lasso formulation [29], which incorporates the temporal smoothness.

Generally speaking, our framework focuses more on the solution suitable for the specific application in medical care. It explores the information and resources available in practical conditions, and suggests the regression model accordingly. The idea of sequential analysis is an effective simulating approach for the continuous medical monitoring, as well as other similar scenarios.

To evaluate the performance of our approach, we conducted extensive experimental studies on the clinical image data. Data used in the preparation of this paper were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). The results evidently demonstrate the effectiveness and accuracy of our proposed solutions.

## 2. Regression by sequential diagnosis data analysis

In this section, we will interpret our solutions in detail. We will introduce the basic concepts and notations first, and then explain the regression model building by sequential data analysis.

### 2.1. Preliminaries

In the AD caring, the target patients will receive regular MRI or PET scan in fixed time interval, and their cognitive scores will be measured accordingly. Based on the acquired medical image data, the regression object aims to predict the cognitive score at specific time points.

Assume that a patient's brain-scan image can be processed into $d$-dimensional feature data, and for this patient' brain-scan records at all time points, we can collect all the feature data $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$. Here, for each sample $\mathbf{x}_t \in \mathbb{R}^d$, there is a corresponding clinical score $y_t$ (MMSE or ADAS-Cog) measured at the same time. In our test corpora, we have the record collection for all patients under monitoring. The regression model simulates the relationship between the collected feature data and the corresponding target clinical measures, so as to predict the patient's potential clinical score at the specific data point in future.

### 2.2. Regression model building

In this part, we will introduce three different approaches to build the feasible regression model based on sequential analysis. We will focus on the analysis of one patient to interpret the model building.

#### 2.2.1. Baseline sequential prediction

The first model predicts the target clinical scores at different time points based on the baseline observation. When the patient takes the first brain scan, we can have his initial feature observation $\mathbf{x}_1$, and we predict his future clinical scores $y_t$ ($t > 1$) based on $\mathbf{x}_1$. Importantly, when the actual score $y_t$ is available, we embed it into the baseline feature data to facilitate the prediction of $y_{t+1}$.

More specifically, assume that $\tilde{y}_t$ is the prediction (or estimation) of $y_t$, and the regression model can be defined as given the baseline observation $\mathbf{x}_1$, and the actual clinical score $y_t$ ($t \geq 1$), predict the estimation for the clinical score at next time point $\tilde{y}_{t+1}$. Such model is called Baseline Sequential Prediction (BSP).

The model of BSP is demonstrated in Fig. 1. From the figure, we can see that at each time point, the clinical score is predicted using
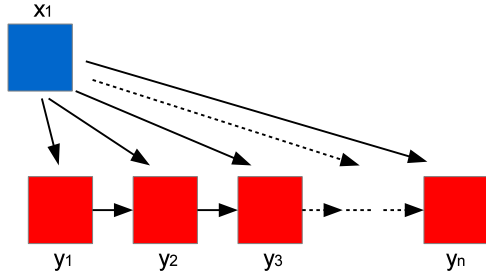
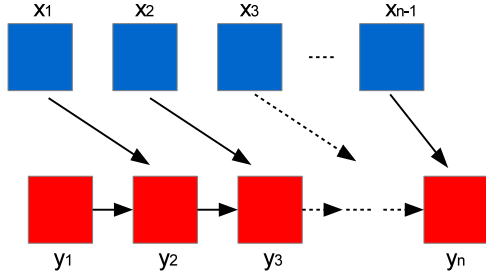**Fig. 1.** The demonstration of baseline sequential prediction.



**Fig. 2.** The demonstration of general sequential prediction.

the information of the previously recorded score and the initial observed sample. The motivation is, the existing clinical score can provide certain information about the current status of the patient, so it can facilitate to make more precise prediction for the following progression.

Formally, we denote the feature data matrix by $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^T \in \mathbb{R}^{N \times (d+1)}$. Here each $\mathbf{x}_i \in \mathbb{R}^{d+1}$ is the extension of baseline feature data with one more dimension by embedding the score record $y_i$. Note that we still use $\mathbf{x}_i$ to denote the feature data when no conflict occurs. We also maintain a target matrix $\mathbf{Y} = [y_1, ..., y_N]^T \in \mathbb{R}^N$. Here $y_i$ stands for the clinical score at the next time point of $\mathbf{x}_i$. We denote $\mathbf{W} = [w_1, ..., w_{d+1}]^T \in \mathbb{R}^{d+1}$ as the weight matrix, and the loss function can be defined as

$$L_{bsp}(W) = \| XW - Y \|_F^2. \tag{1}$$

Considering the regularization term, we apply the sparse group lasso [29,30], and the proposed formulation can be expressed as solving the following optimization problem:

$$\min_W L_{bsp}(W) + \lambda_1 \| W \|_1 + \lambda_2 \| W \|_{2,1}, \tag{2}$$

where $\lambda_1$ and $\lambda_2$ are regularization parameters. The $\ell_{2,1}$-norm is for feature selection, i.e., after finding the optimal solution, we will find some zero (close to zero) values in $W$, whose corresponding features are not relevant in predicting the clinical scores. It has been proved that the optimization problem can be solved by the accelerated gradient method (AGM) [31], and we employ the theoretical results of [28] to solve the result.

### 2.2.2. General sequential prediction

The next model has more general sense, which predicts the target clinical scores at different time points based on the instant observations in progress. Whenever the patient takes a brain scan, we can immediately obtain this feature observation $\mathbf{x}_t$, and we predict the clinical scores $y_{t+1}$ based on $\mathbf{x}_t$. Here we note that the corresponding score $y_t$ is also available, so we embed it into $\mathbf{x}_t$ to facilitate the prediction of $y_{t+1}$. By this mean, the regression model can be defined as given the feature observation $\mathbf{x}_t$, and the actual clinical score $y_t$, predict the estimation for the clinical score
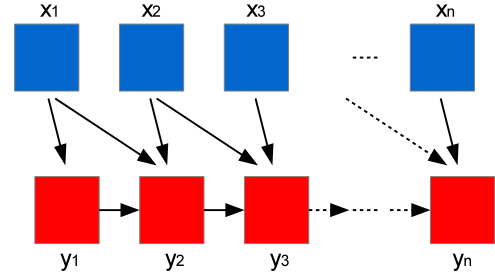
at next time point $\tilde{y}_{t+1}$. Such model is called General Sequential Prediction (GSP).

The model of GSP is demonstrated in Fig. 2. The figure shows that at each time point, the clinical score is predicted using the information of the previously observed sample and recorded score. Different from the BSP, GSP applies the instant observations that deliver the updated and inherent information reflecting the patient's current health status, which helps to make more precise prediction for the following progression.

The formulation for GSP model is similar to that of BSP, and the difference is that $\mathbf{x}_i \in \mathbb{R}^{d+1}$ is the extension of instant feature data other than the baseline data. We still apply sparse group lasso for regularization in this model.

### 2.2.3. Incrementally sequential prediction

Finally, we introduce the concept of incrementally sequential prediction (ISP). For the time sequence $\{1, 2, ..., n\}$, $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ are the series of consecutive feature observations. The incrementally sequential prediction model can be defined as given consecutive observations $\mathbf{x}_{t-1}$, $\mathbf{x}_t$ and the previous clinical score $y_{t-1}$, predict the estimation for current clinical score $\tilde{y}_t$.

The incrementally sequential prediction model is demonstrated in Fig. 3. From the figure, we can see that at each time point, the clinical score is predicted using the information of the previously observed sample and the current observation. The motivation of such a model is based on the following consideration: usually speaking, the initial status of a patient is uncertain, and also the time interval between two consecutive observations can be various, so the incremental prediction is more suitable in such situation. According to the model, the prediction in health caring will not be limited for single time point.

We collectively use $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_N]^T \in \mathbb{R}^{N \times 2d}$ to denote all feature pairs of all patient samples as data matrix. Here each $\mathbf{z}_i \in \mathbb{R}^{2d}$ is the fusion of a consecutive feature pair $\{\mathbf{x}_{t-1}, \mathbf{x}_t\}$, with $\mathbf{z}_i(2j-1) = \mathbf{x}_{t-1}(j)$, and $\mathbf{z}_i(2j) = \mathbf{x}_t(j)$, $\forall j \in [1, d]$. We also maintain a target matrix $\mathbf{Y} = [y_1, ..., y_N]^T \in \mathbb{R}^N$, and a reference score matrix $\mathbf{V} = [v_1, ..., v_N]^T \in \mathbb{R}^N$. Here $v_i$ and $y_i$ stand for the clinical scores of $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$, respectively. We denote $\mathbf{W} = [w_1, ..., w_{2d}]^T \in \mathbb{R}^{2d}$ as the weight matrix, and the loss function can be defined as

$$L_{isp}(W) = \| V + ZW - Y \|_F^2. \tag{3}$$

For the regularization term, in our model, the prediction is made based on the information of consecutively observed feature data, so we can assume that the difference between the target prediction and the previously recorded reference score is small, which is so-called temporal smoothness [27]. Therefore, besides group sparse lasso, we employ the fused Lasso penalty to incorporate the temporal smoothness. Formally, the proposed formulation can be expressed as solving the following optimization problem:

$$\min_W L_{isp}(W) + \lambda_1 \| W \|_1 + \lambda_2 \| RU^T \|_1 + \lambda_3 \| W \|_{2,1}, \tag{4}$$



**Fig. 3.** The demonstration of incrementally sequential prediction.

where $U \in \mathbb{R}^{d \times 2}$ is derived from $W$ with $U_{i,1} = W_{2i-1}$ and $U_{i,2} = W_{2i}, \forall i \in [1, d]; R = [1, -1]$.

By employing ISP analysis, the incrementally temporal information can be incorporated into the model, and the relationship among consecutive observations can be discovered to improve the prediction performance. With our proposed model, we can continuously predict a patient's disease progress and take the relevant medical treat accordingly.

## 3. Experiments

In this section, we will report the empirical evaluation of our proposed framework in modeling AD progression. The experiments will be conducted on the dataset collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We will compare the performance of simulating and predicting the disease progression with the baseline approach and the latest solution based on multi-task learning framework.

### 3.1. Dataset description

The ADNI project is designed to collect the serial of MRI, PET and other clinical assessment scores to measure the progression of selected subjects, including Alzheimers Disease patients (AD), Mild Cognitive Impairment patients (MCI) and normal controls (NC), and the subjects will be observed repeatedly and continuously over a 6-month or 1-year interval. For each observation, the MRI and PET scans will be collected, as well as other corresponding measurements, e.g., clinical scores such as MMSE and ADAS-Cog. In our work, we use the MRI scans to generate feature data, which are obtained from over 700 subjects. Five types of MRI features are used in the work: white matter parcellation volume (Vol.WM.), cortical parcellation volume (Vol.C.), surface area (Surf. Area), cortical thickness average (CTA) and cortical thickness standard deviation (CTStd). The date when the patient receives the MRI scans for the first time is called baseline, and the time points of the following observations are denoted by the duration starting from the baseline. For example,"M06" means the screening taken at the time point 6 months after the first visit. All the subjects are under the repeated observations for up to 48 months. Table 1 records the dimensionality of MRI feature, as well as the sample sizes available at different time points. Notice that some patients may quit the monitoring and examination after a certain period, so there are some missing records, so the missing value estimation techniques [32] can be applied here. However, we are not expecting all patients have complete records, but every consecutive partial records can contribute the model building.

### 3.2. Prediction performance

In the experiments, we will compare our proposed models with some existing approaches. The baseline approach is the basic regression model simulating the relationship between MRI feature and its corresponding clinical score. The second approach in comparison is the Convex Fused Sparse Group Lasso (cFSGL) based on multi-task learning [28]. The performance is evaluated by the

prediction of MMSE and ADAS-Cog scores. For each target score, we build the regression model, and the model parameters are determined by 10-fold cross validation. To measure the regression performance, we employ the mean squared error (MSE) calculated for the score prediction at different time points, and weighted correlation coefficient (R-value) as suggested in [18,28].

The comparison results between different approaches are listed in Table 2. Here our approaches are not applicable for "M06MSE" of MMSE score, because there are no baseline score records. From the table we can observe that our approaches can significantly improve the prediction results over those achieved by basic model and cFSGL, and the correlation can be improved over 0.85. Generally speaking, the prediction error of cFSGL is lower than basic model because the shared representation in parallel learning can help individual tasked be learned better. However, the sequential prediction can further improve the performance due to the feedback from the intermediate information. More specifically, the prediction error will increase for the records at "M24", and this is because the status of patients in monitoring will become significantly worse after "M12", so the prediction error will increase accordingly, since the basic assumption for all models is the temporal smoothness. The performance for MMSE and ADAS-Cog are also different because of the different distribution of the patients records. We can observe the obvious increase of prediction error when processing ADAS-Cog scores. When comparing the distributions of two different scores, we can observe that the ADAS-Cog has much wider range and higher fluctuation along the recording period, and this is the reason bringing the high prediction error. We will provide further error analysis in the following part.

### 3.3. Error analysis: effect of patients' status

Based on the overall prediction comparison, we further provide an error analysis to discover the reason that influences the prediction performance. For each time point, we divide the patient objects into different ranges (AD, MCI and NC) based on their actual clinical scores (target scores), and compare the prediction errors of different approaches, so as to examine the effect of the patients' actual scores, i.e., the patients' actual status in AD progression.

Figs. 4 and 5 show the error analysis for MMSE and ADAS-Cog prediction respectively, and the absolute prediction errors for patients of different cognitive scores are summarized at different time points. From the figures, we can conclude that, at different time points, the major prediction error is produced when predicting those patients with bad cognitive measurement (For MMSE score, [0, 20] means AD patients; for ADAS-Cog score, [20, 70] means AD patients). For those with better cognitive scores (MCI or Normal patients), since their status is relatively stable, the prediction error is significantly improved. Overall speaking, our approaches can achieve better performance at almost all score ranges and all time periods. At last, the same statement can be concluded that the prediction error for MMSE score is much lower than ADAS-Cog score, and the major cause is those records of AD patients.

## 4. Conclusions

This work addresses the significant problem of simulating and predicting AD progression for elders' care. Based on the existing works using multi-task learning technique to model the AD progression, we further introduce a sequential data analysis mechanism for continuous clinical score prediction to improve the modeling and regression performance. The basic disease progression model is built

**Table 1**
The feature dimensionality of MRI scans and the sample size available at different time points in the experiments.

| Target score | BS | M06 | M12 | M24 | M36 | M48 | Dim |
|---|---|---|---|---|---|---|---|
| MMSE | 740 | 641 | 588 | 452 | 236 | 57 | 327 |
| ADAS-Cog | 740 | 641 | 588 | 452 | 236 | 57 | 327 |

**Table 2**
Comparison of different approaches on MMSE and ADAS-Cog prediction using MRI features, in terms of average correlation coefficient (R) and mean squared error (MSE) for each time point.

| Comparison item | Basic | cFSGL | BSP | GSP | ISP |
|---|---|---|---|---|---|
| Target:MMSE | | | | | |
| R | 0.690 ± 0.037 | 0.719 ± 0.039 | 0.848 ± 0.040 | 0.853 ± 0.034 | **0.862 ± 0.022** |
| M06 MSE | 7.847 ± 1.731 | 6.843 ± 1.400 | NA | NA | NA |
| M12 MSE | 10.033 ± 2.235 | 8.876 ± 2.371 | 5.780 ± 1.330 | 5.847 ± 1.577 | **5.214 ± 1.016** |
| M24 MSE | 11.520 ± 3.279 | 12.733 ± 6.243 | 6.601 ± 2.121 | **6.497 ± 2.386** | 6.536 ± 2.585 |
| M36 MSE | 9.900 ± 3.710 | 9.331 ± 2.582 | 5.015 ± 1.682 | 5.015 ± 2.878 | **4.754 ± 2.838** |
| Target:ADAS-Cog | | | | | |
| R | 0.707 ± 0.040 | 0.776 ± 0.025 | 0.872 ± 0.017 | 0.873 ± 0.024 | **0.874 ± 0.016** |
| M06 MSE | 33.05 ± 9.52 | 30.15 ± 8.20 | 15.92 ± 3.24 | 15.99 ± 4.05 | **15.81 ± 3.48** |
| M12 MSE | 36.67 ± 13.61 | 31.92 ± 6.30 | **17.01 ± 2.96** | 17.14 ± 3.75 | 17.20 ± 3.40 |
| M24 MSE | 50.02 ± 17.02 | 48.63 ± 16.40 | 23.83 ± 4.73 | 23.62 ± 7.40 | **22.95 ± 7.98** |
| M36 MSE | 35.81 ± 13.33 | 38.19 ± 13.80 | 16.75 ± 8.45 | **16.09 ± 5.37** | 16.43 ± 4.54 |
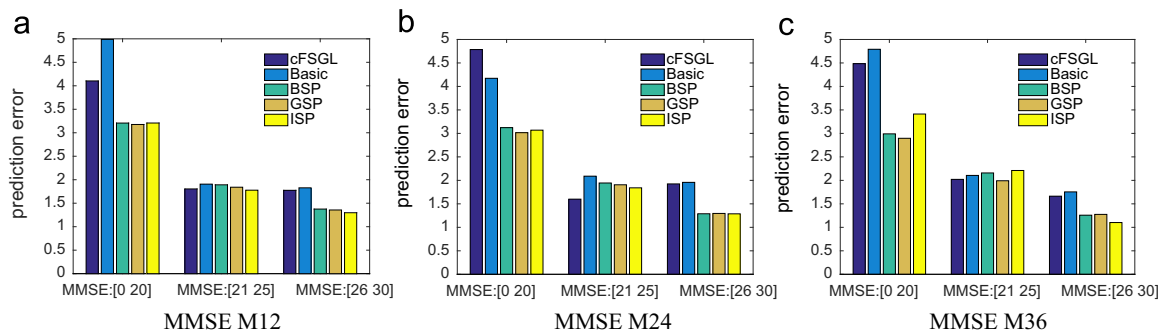


**Fig. 4.** Error analysis for MMSE prediction at different time points.
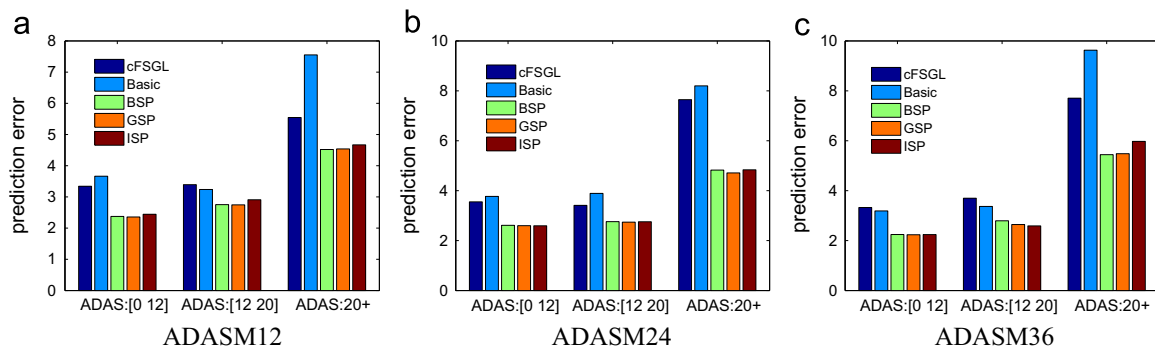


**Fig. 5.** Error analysis for ADAS-Cog prediction at different time points.

by a fused sparse group lasso formulation, which embeds the intermediate feature observation and actual score information for incremental model training, and also incorporates the temporal smoothness. In the simulation, we continuously obtain the actual diagnosis data and the clinical scores of selected subjects at different time points to contribute the model building, so as to enhance the model learning and increase the accuracy of prediction. The effectiveness of the proposed progression model is evaluated by the experimental studies on the famous data sets collected from the Alzheimers Disease Neuroimaging Initiative (ADNI). The results show that the proposed regression model is more effective than the baseline regression model and the one based on multi-task learning formulation for simulating the disease progression.

## References

[1] Z.S. Khachaturian, Diagnosis of alzheimer's disease, Arch. Neurol. 42 (1985) 1097–1105.

[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H.M. Arrighi, Forecasting the global burden of alzheimer's disease, Alzheimer's Dement. 3 (2007) 186–191.

[3] A. Wimo, B. Winblad, H. Aguero-Torres, E. von Strauss, The magnitude of dementia occurrence in the world, Alzheimer Dis. Associat. Disord. 17 (2003) 63–67.

[4] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3d object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (2012) 4290–4303.

[5] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, X. Wu, Visual-textual joint relevance learning for tag-based social image search, IEEE Trans. Image Process. 22 (2013) 363–376.

[6] D. Holland, J.B. Brewer, D.J. Hagler, C. Fennema-Notestine, A.M. Dale, M. Weiner, L. Thal, R. Petersen, C.R. Jack, W. Jagust, et al., Subregional neuroanatomical change as a biomarker for alzheimer's disease, Proc. Natl. Acad. Sci. 106 (2009) 20954–20959.

[7] X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, L. Yao, K. Chen, Voxel-based assessment of gray and white matter volumes in alzheimer's disease, Neurosci. Lett. 468 (2010) 146–150.

[8] K. Ito, B. Corrigan, Q. Zhao, J. French, R. Miller, H. Soares, E. Katz, T. Nicholas, B. Billing, R. Anziano, et al., Disease progression model for cognitive deterioration from alzheimer's disease neuroimaging initiative database, Alzheimer's Dement. 7 (2011) 151–160.

[9] X. Zhu, L. Zhang, Z. Huang, A sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process. 23 (2014) 3737–3750.

[10] X. Zhu, X. Li, S. Zhang, Block-row sparse multiview multilabel learning for image classification, IEEE Trans. Cybern. 46 (2016) 450–461.

[11] R.S. Desikan, H.J. Cabral, F. Settecase, C.P. Hess, W.P. Dillon, C.M. Glastonbury, M.W. Weiner, N.J. Schmansky, D.H. Salat, B. Fischl, et al., Automated mri measures predict progression to alzheimer's disease, Neurobiol. Aging 31 (2010) 1364–1374.

[12] C. Misra, Y. Fan, C. Davatzikos, 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 (2009) 1415–1422.

[13] R.K. Pearson, R.J. Kingan, A. Hochberg, Disease progression modeling from historical clinical databases, in: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, 2005, pp. 788–793.

[14] P. Vemuri, H. Wiste, S. Weigand, L. Shaw, J. Trojanowski, M. Weiner, D. Knopman, R. Petersen, C. Jack, et al., Mri and csf biomarkers in normal, mci, and ad subjects: predicting future clinical change, Neurology 73 (2009) 294–301.

[15] B. Cheng, D. Zhang, S. Chen, D.I. Kaufer, D. Shen, A.D.N. Initiative, et al., Neuroinformatics, Semi-supervised multimodal relevance vector regression improves cognitive performance estimation from imaging and biological biomarkers 11 (2013) 339–353.

[16] K. Franke, G. Ziegler, S. Klöppel, C. Gaser, A.D.N. Initiative, et al., Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: exploring the influence of various parameters, Neuroimage 50 (2010) 883–892.

[17] Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables, Neuroimage 50 (2010) 1519–1535.

[18] S. Duchesne, A. Caroli, C. Geroldi, D.L. Collins, G.B. Frisoni, Relating one-year cognitive change in mild cognitive impairment to baseline mri features, Neuroimage 47 (2009) 1363–1370.

[19] X. Zhu, Z. Huang, Y. Yang, H.T. Shen, C. Xu, J. Luo, Self-taught dimensionality reduction on the high-dimensional small-sized data, Pattern Recognit. 46 (2013) 215–229.

[20] C.M. Stonnington, C. Chu, S. Klöppel, C.R. Jack, J. Ashburner, R.S. Frackowiak, A. D.N. Initiative, et al., Predicting clinical scores from magnetic resonance scans in alzheimer's disease, Neuroimage 51 (2010) 1405–1413.

[21] X. Zhu, H.-I. Suk, D. Shen, Matrix-similarity based loss function and feature selection for alzheimer's disease diagnosis, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3089–3096.

[22] Z. Li, J. Liu, Y. Yang, X. Zhou, H. Lu, Clustering-guided sparse structural learning for unsupervised feature selection, IEEE Trans. Knowl. Data Eng. 26 (2014) 2138–2150.

[23] Z. Li, J. Liu, J. Tang, H. Lu, Robust structured subspace learning for data representation, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 2085–2098.

[24] D. Zhang, D. Shen, A.D.N. Initiative, et al., Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease, Neuroimage 59 (2012) 895–907.

[25] X. Zhu, H.-I. Suk, D. Shen, A novel matrix-similarity based loss function for joint regression and classification in ad diagnosis, Neuroimage 100 (2014) 91–105.

[26] L. Jacob, J.-p. Vert, F.R. Bach, Clustered multi-task learning: a convex formulation, in: Advances in Neural Information Processing Systems, 2009.

[27] J. Zhou, L. Yuan, J. Liu, J. Ye, A multi-task learning formulation for predicting disease progression, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, pp. 814–822.

[28] J. Zhou, J. Liu, V.A. Narayan, J. Ye, Modeling disease progression via fused sparse group lasso, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 1095–1103.

[29] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 67 (2005) 91–108.

[30] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 68 (2006) 49–67.

[31] A. Nemirovski, Efficient methods in convex programming.

[32] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, Z. Xu, Missing value estimation for mixed-attribute data sets, IEEE Trans. Knowl. Data Eng. 23 (2011) 110–121.

**Qing Xie** received the BE degree in information science from University of Science and Technology of China in 2008, and the PhD degree in computer science from the University of Queensland, Australia, in 2013. He is currently a postdoctoral research fellow in Division of CEMSE, King Abdullah University of Science and Technology. His research interests include stream data mining, time series databases, continuous query optimization and e-Health.



**Su Wang** is a master student in computer science at King Abdullah University of Science and Technology. He received his Bachelor degree from Zhejiang University in 2014. He has a broad research interest in machine learning and data mining.



**Jia Zhu** is currently an associate professor in the School of Computer Science at South China Normal University after finished his Postdoctoral fellow at United Nations University. Prior to that, he received his PhD degree from the University of Queensland in 2013, and his BS and MS degree from Bond University, Australia in 2004 and 2006 respectively. His research interests are Big Data Applications, Machine Learning and Information Retrieval. He published several papers on top conferences and journals, such as Information Sciences and WWW.
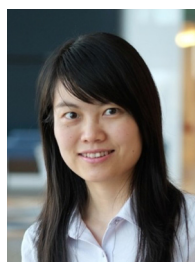


**Xiangliang Zhang** is currently Assistant Professor and directs the Machine Intelligence and kNowledge Engineering (MINE) Laboratory in King Abdullah University of Science and Technology (KAUST), Saudi Arabia. She was an European ERCIM research fellow in Norwegian University of Science and Technology, Norway, in 2010. She earned her PhD degree in computer science from INRIA-Universite Paris-Sud, France, in July 2010. She has authored or co- authored over 50 refereed papers in various journals and conferences. Her main research interests and experiences are in machine learning, data mining, and cloud computing.