# LONGITUDINAL ANALYSIS FOR MILD COGNITIVE IMPAIRMENT IDENTIFICATION VIA FUSED GROUP LEARNING WITH SMOOTH REGULARIZATION

*Baiying Lei，Xin Zhao, Dong Ni, Siping Chen*, and Tianfu Wang**
School of Biomedical Engineering, Shenzhen University,
National-Regional Key Technology Engineering Laboratory for Medical Ultrasound,
Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China.

## ABSTRACT

Alzheimer's disease (AD) and its early stage, mild cognitive impairment (MCI), have been widely analyzed by brain functional connectivity network (BFCN) due to its promising potential in identifying biomarkers for the neurodegenerative disorders and understanding the brain functions. The accurate construction of biologically meaningful brain network plays an essential role in these applications. Sparse learning has been widely applied for the network construction. However, the previous sparse learning studies often fail to consider the smoothness penalty from multiple time points. To address these problems, we propose a new multi-task learning method to integrate the fused penalty with smooth regularization. Specifically, a novel objective function is developed to consider fused learning of multiple time points via smoothness constraint. We evaluate our method for MCI identification via the resting-state functional magnetic resonating imaging (rs-fMRI) from Alzheimer's Disease Neuroimaging Initiative Phase-2 (ADNI-2) dataset. The extensive experimental results demonstrate that the proposed algorithm is quite effective for human brain connectivity modeling of MCI. In addition, our sparse BFCN modeling outperforms the conventional work and will benefit both basic and clinical neuroscience studies.

***Index Terms***—Mild cognitive impairment, longitudinal analysis, network connectivity, multitask fused learning.

## 1. INTRODUCTION

With a progressive decline of memory and cognitive function, Alzheimer's disease (AD) and its prodromal stage, mild cognitive impairment (MCI), are incurable neurodegenerative diseases[1]. Both AD and MCI are the main dementia leading to about 60-80% of dementia cases in the worldwide. MCI is convertible to AD with an average rate of 10-15%. Since MCI is misdiagnosed most of time due to explicit symptoms, the prompt treatment and monitor of AD progression before its onset is highly desirable . Currently, various imaging modalities have been widely applied for AD studies. Rs-fMRI is able to check functional integration and separation of brain networks disrupted by MCI and establish functional connectivity (FC) among brain regions for MCI network abnormalities characterization [2, 3]. Actually, FC is denoted as the temporal correlation of blood-oxygenation-level-dependent (BOLD) time series between two brain regions . The brain FC network (BFCN) study based on rs-fMRI has played an increasing important role in identifying biomarkers for neurological disorders . Hence, it is of great interest to develop early MCI diagnosis method to delay AD progression and treat this dementia via BFCN.

To this date, a myriad of FC modelling methods have been developed [4]. Namely, different regions-of-interest (ROIs) are parcellated from brain region to estimate the BOLD time series of ROI. For example, the pairwise Pearson's correlation (PC) among different brain regions is one of the widely applied FC modelling algorithm to construct brain regions for MCI topological properties revelation. However, PC focuses on pairwise relationship only, which fails to consider the interaction among multiple brain regions [5]. By contrast, another widely applied method is to establish FC network via sparse representation (SR) [6]. This sparse estimation is based on partial correlation via regularization to construct the relationship among certain ROIs while removing other ROIs' effects. SR network has been applied in AD and MCI by constructing brain networks [6-10]. However, the human is not only inherently sparse, but also has group structure, it is interesting to integrate both information.

It is known that machine learning techniques can make use of feature extracted from BFCN for MCI patient identification with a relatively high accuracy [7, 11-13]. Although the conventional studies mostly focused on single time point information from brain regions, one time point is limited due to lack of longitudinal analysis. To enhance the diagnostic performance, longitudinal analysis of multiple time point networks can model disease progression comprehensively and effectively [14, 15]. In the literature, longitudinal study for disease progression modeling has become a hot topic due to its effectiveness [14, 16]. For example, in [17], Zhou et al. proposed to model AD progression based on a novel designed convex fused learning for score prediction and achieved remarkable results. In [14], Huang et al. predicted the longitudinal score using weighted random forest and obtained superior results than the traditional study. In [16], a temporal smooth framework for longitudinal score prediction is proposed by Jie et al. In spite of these efforts, the previous studies mainly focused on the score prediction based on MRI or PET data only. It is argued that the FC network from rs-fMRI data can be effective for disease progression study. In view of this, our study concentrates on the longitudinal analysis for MCI

978-1-5090-1172-8/17/$31.00 ©2017 IEEE 

identification via rs-fMRI data. To our best knowledge, this is the first longitudinal analysis of MCI disease modelling based on FC network of rs-fMRI data.

To characterize the complex MCI disease and boost the discrimination performance, we propose to develop a novel brain network model based on multi-task fused learning with smoothness constraint. Specifically, we devise a multiple time points network to take advantage of relationship of successive time points. A novel framework harnessing multiple time point information for longitudinal functional analysis of MCI disease is developed for network construction. Moreover, we perform feature selection via the least absolute shrinkage and selection operator (LASSO) to identify the most informative features, and the final selected features are fed into support vector machine (SVM) for MCI identification. We evaluate our proposed method based on the Alzheimer's Disease Neuroimaging Initiative Phase-2 (ADNI-2) database. Our experiments confirm that our method outperforms the traditional methods for MCI diagnosis.

## 2. METHODOLOGY

### 2.1. Data and post-processing

We used the ADNI-2 dataset in our study, which contains 24 MCI patients and 23 normal controls (NCs). All subjects were scanned using 3.0T Philips Achieva scanners with matched age and gender. The slice thickness is 3.3 mm. We removed the first 4 volumes of each subject to keep magnetization equilibrium and applied slice timing and motion correction to correct the rs-fMRI images. We also employed a rigid-body transformation for head correction and the subjects with head motion above 2mm are removed. Rs-fMRI images are normalized to the standard Montreal Neurological Institute (MNI) space and applied Gaussian spatial smoothing using full width. The rs-fMRI was parcellated into 116 brain regions based on the automated anatomical labelling (AAL) template. FSL software is used for pre-processing in our experiment. The mean rs-fMRI time series of each brain region was also high pass filtered. In addition, we regressed out head motion parameters, mean BOLD time series of the white matter and the cerebrospinal fluid. The mean of BOLD signal in each region of interest (ROI) is used as features. Accordingly, the original rs-fMRI signal is denoted by 116 ROIs (i.e.,116 nodes) and connections between each pair of 116 ROIs (i.e., the edges connecting them). PC of two mean time series between a pair of ROIs is computed to measure the connection strength.

### 2.2. Method

We have multiple time points rs-fMRI as our input. After pre-processing of rs-fMRI data, multiple FC networks of various time points are constructed by the proposed method. Then the feature selection method is applied to identify the most informative features and the selected features are fed to the classifier for MCI identification.

To model rs-fMRI dataset, the dimension curse is always an issue. It is recommended that group lasso is an effective way to address these problems. The group lasso is able to identify a small set of features of specific groups to construct connectivity map using non-zero weights of all predictor. The connectional differences of MCI and NC groups are appealing to investigation as suggested by [6]. Assuming each brain is parcellated into R ROIs using AAL template, a response vector with M length regional mean time series of the r-th ROI is represented as: $\boldsymbol{y} = [y_{1r}, y_{2r}, \dots, y_{Mr}] \in \mathbb{R}^M$, where M is the number of time points in the time series, and $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_R] \in \mathbb{R}^{M\times(R-1)}$ represents a predictor data matrix of a subject, $A_r^n = [y_{1r}^1, \dots y_{2r}^n, \dots, y_{Mr}^n]$is data matrix of r-th ROI (the whole BOLD time series except for r-th ROI), $w_r^n \in \mathbb{R}^{R-1}$is weighting regression coefficient vector, $W_r = [w_r^1, \dots w_r^n, \dots, w_r^N]$, where N is the total number of subjects. Then the key step of constructing the BFCN for this subject is to estimate the FC matrix W$\in \mathbb{R}^{R\times R}$, considering $R$ nodes (i.e., $\boldsymbol{x}_i$, i = 1, 2, ..., $R$) denote all ROIs. There are many researches to construct a sparse network to model the brain region connectivity, and the typical group lasso sparse learning network is formulated as below:

$$J(W_r) = \min_{W_r} \frac{1}{2}\sum_{n=1}^{N} ||y_r^n - A_r^n w_r^n||_2^2 + \lambda||W_r||_{2,1}, \quad (1)$$

where $\lambda$ is a group regularization parameter, the regularization term is denoted as: $||W_r||_{2,1} = \sum_{g=1}^{G} ||w_{rg}||_2$(i.e., $l_1$norm of $||w_{rg}||_2$ ), $w_{rg}$ represents the connectivity coefficients of g-th predictor. The utilization of $l_2$-norm on row vectors imposes the weights related with the g-th feature across multiple time points to be grouped together, and the further adoption of $l_1$-norm jointly select features via the weights of R time points. The group lasso regularization is the traditional sparse regression network, which makes sure that all of the regression models in different groups have the shared set of connections. The $l_2$ norm group penalty imposes every representation coefficient using the same weight. Namely, this $l_2$ norm treats each ROI in the same way to reconstruct a target ROI. Accordingly, SR model with this objective function is able to reconstruct the target ROI by the ROIs different from the target ROI. In addition, each ROI reconstruction is independent from other ROI reconstruction.

The main goal in brain disease diagnosis is to enhance the diagnostic performance between NC and MCI, but the group lasso regression model with penalty fails to consider the smooth properties of different time points in the framework. For this reason, we devise a novel framework to jointly learn shared functional brain networks of each subject by the group sparse regularization and fused smoothness information with the devised regularization terms as below:

$$J(W_r) = \min_{W_r} \frac{1}{2}\sum_{n=1}^{N} ||y_r^n - A_r^n w_r^n||_2^2 + \lambda||W_r||_{2,1} +$$
$$\lambda_1 \sum_{r=1}^{R-1} ||A_r^n w_r^n - A_{r-1}^n w_{r-1}^n||_2^2 + \lambda_2 \sum_{r=1}^{R-1} ||w_r^n - w_{r-1}^n||_1, \quad (2)$$

where $\lambda_1$ and $\lambda_2$ are the smoothness regularization parameters. The last term, $||w_r^n - w_{r-1}^n||_1$ , is the regularization penalty derived from fused LASSO, which

ensures that adjacent weight vectors of the continuous time points are as small as possible. Since $l_1$ norm is used in this fused smoothness term to have sparsity of weighting difference, zeros will occur in the difference vectors of weighting. Namely, adjacent weighting vectors will be the same due to the fused smoothness regularization. Only the features with non-zero weights for our classification task will be selected. $||A_r^n w_r^n - A_{r-1}^n w_{r-1}^n||_2^2$ is the target smoothness, which encourages two successive models of continuous time points as small as possible. The smoothness regularization terms tradeoff the two terms' contribution. When both smoothness terms are zeros, the proposed method is the conventional group lasso method. By introducing the fused smoothness terms, the connectivity coefficients for the subjects in different time points are smoothed. Furthermore, the integration of lasso and group lasso penalty imposes high penalty via regularization terms. We name this new sparse learning model as multitask fused sparse regression model (MFSR).

### 2.3. Selective feature fusion

Since multiple time point feature is used for rs-fMRI data, the feature dimension of FC is too high to cause overfitting problem. To address this issue and reduce feature dimension to a certain extent, we adopt the feature selection to benefit from the discriminative feature learning and redundancy reduction algorithm. Specifically, we utilize a sparse regression framework with the wrapper-based feature selection , which can identify and select the most predictive features. Namely, given a current set of feature blocks, a new feature block can be selected to obtain a discriminative subset. The LASSO feature selection is utilized on the training data to find a small feature subset to benefit our classification results. Then the selected features from the training data are utilized to train the LibSVM model. We select the one yielding optimal accuracy finally.

## 3. EXPERIMENTS AND RESULTS

### 3.1. Experimental setup

In our experiment, we implement our algorithm using Matlab 2015a software. The sparse regression and classification are implemented by SLEP and LibSVM toolboxes, respectively. Since our data size is small, we adopt the leave-one-out cross validation to evaluate the proposed performance. For the hyper-parameter in each method, we utilize the greedy strategy to select the best parameters of our proposed method. For instance, we obtain the optimal values of $\lambda, \lambda_1$, and $\lambda_2$ in our method by exhaustive search ranging from $10^{-5}$ to $10^5$ via a grid search strategy. To measure performance of different methods, we use the following metrics: accuracy (ACC), area under receiver operating characteristic (ROC) curve (AUC), sensitivity (SEN), specificity (SPEC), Youden's Index (Youden), F1-score (F1), and balanced accuracy (BAC). To show the effectiveness of our proposed method, we compare the proposed MFSR network with the closely related networks including:1) Baseline PC network; 2) Baseline SR network; 3) Baseline MFSR network; 4) Year 1 PC network; 5) Year 1 SR network; and 6) Year 1 MFSR network.

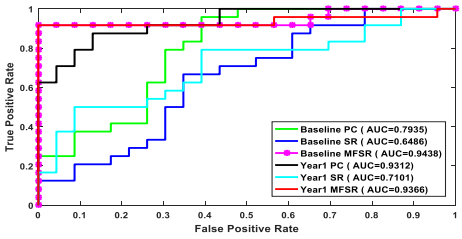


Fig. 1. ROC curves of various methods.

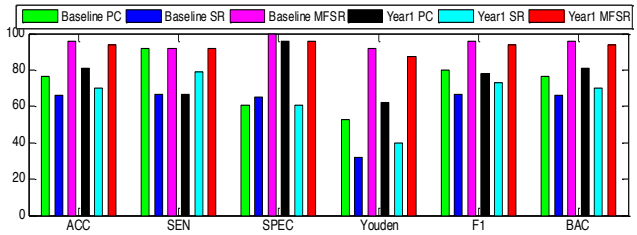


Fig. 2. Classification results via various metrics.

### 3.2. Experimental results

Figs.1 and 2 show the ROC curves and classification results and of all the listed networks. It is clear that our proposed network beats other competing classification networks. The proposed network overcomes the previous network's drawbacks, and indeed it achieves better classification results in terms of various metrics. As shown in Fig. 2, the proposed network achieves the best classification performance with an accuracy of 95.74% and 93.64% for baseline and year 1 rs-fMRI data, respectively. Upon close examination, it is obvious that multi-task fused learning is better than each individual task since it can unclose the underlying relationship among multiple time points than the simple averaging of them. By considering the relationship of the successive ROIs, we can observe that MFSR model outperforms the traditional SR and PC models in both baseline and year 1 data, which confirms multiple time points networks are promising for MCI identification. Another encouraging observation is that classification performance is less sensitive to time change, namely, a similar pattern is observed in both baseline and year 1 cases, where year 1 performs slightly poorer than baseline case. By comparing with those models, the effectiveness of the introduced fused and group structure via smooth constraints is justified.

To investigate our algorithm further, three NC and MCI patients are randomly selected in our database. Fig. 3 plots the PC, SR and MFSR networks in two time points, baseline and year 1, respectively. We can see that MFSR networks reveal more block-like structure and clear layout compared with conventional networks. Additionally, the conventional networks are similar between NC and MCI networks, while

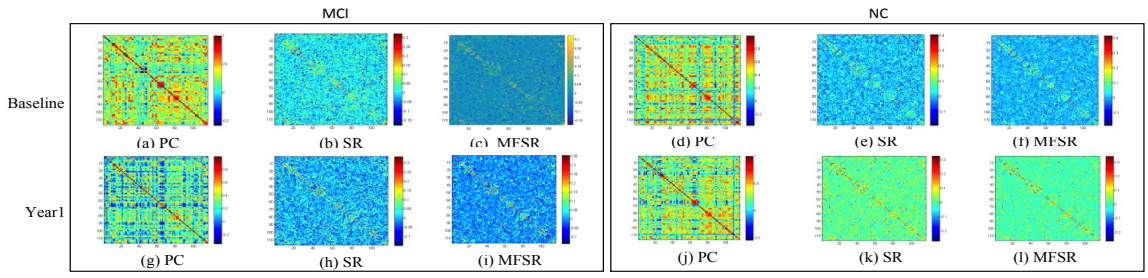


Fig.3. Sampled MCI and NC connectivity networks of baseline and year 1 data.

MFSR networks show obvious differences between MCI and NC groups.

As the selected features by Lasso in each validation might be different, we record all selected features during the training process. The top 10 most frequently selected brain regions of baseline and year 1 data are visualized in Fig.5. Note that the colors of arcs are randomly generated to differentiate ROIs and connectivity for clear visualization. We can see that several brain regions such as frontal and temporal features are frequently identified are jointly selected as important features for MCI identification for baseline and year 1 data. The identified brain regions such as temporal inferior frontal gyrus, supplementary motor area, insula, frontal middle gyrus, middle temporal gyrus, superior gyrus are promising for clinical diagnosis as the potential biomarker, which are in line with findings of the most selected regions in MCI in previous studies [6-8]. Overall, the top selected most discriminative brain regions are closely related with MCI pathology and consistent with previous clinical findings as well [6-8].

## 4. CONCLUSIONS

In this study, we have developed a new multi-task fused sparse learning framework for MCI disease identification, which combines the sparse learning, longitudinal analysis, and network modeling together. Accordingly, we can model the complicated brain network more accurately compared with other widely used methods. We validate our work to distinguish MCI from NC and our findings show that the longitudinal analysis is quite effective and powerful for MCI identification. Our proposed method obtains the superior results than the common network modeling algorithms. In our future work, we will work on more time points and smooth constraints to further boost the MCI diagnosis results. The global graph theory and high order statistics (mean clustering coefficient, covariance of the clustering) can be integrated in our framework as well.

## REFERENCES

[1] A. Alzheimer's, "2015 Alzheimer's disease facts and figures," Alzheimers Dement, vol. 11, pp. 332-384, 2015.

[2] X. Yang, Y. Jin, X. Chen, H. Zhang, G. Li, and D. Shen, "Functional Connectivity Network Fusion with Dynamic Thresholding for MCI Diagnosis," in MLMI, vol. 10019, pp. 246-253, 2016.

[3] Y. Jin, C. Huang, M. Daianu, L. Zhan, E. L. Dennis, R. I. Reid, et al., "3D tract-specific local and global analysis of white matter integrity in Alzheimer's disease," Human Brain Mapping, pp. n/a-n/a, 2016.

[4] S. M. Smith, K. L. Miller, S. Moeller, J. Xu, E. J. Auerbach, M. W. Woolrich, et al., "Temporally-independent functional modes of spontaneous brain activity," Proceedings of the National Academy of Sciences of the United States of America, vol. 109, pp. 3131-3136, 2012.

[5] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, et al., "Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation," NeuroImage, vol. 50, pp. 935-949, 2010.

[6] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, and D. Shen, "Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification," Brain Structure and Function, vol. 219, pp. 641-656, 2014.

[7] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "Supervised discriminative group sparse representation for mild cognitive impairment diagnosis," Neuroinformatics, vol. 13, pp. 277-295, 2015.

[8] B. Jie, D. Zhang, C. Y. Wee, and D. Shen, "Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification," Human brain mapping, vol. 35, pp. 2876-2897, 2014.

[9] L. Huang, Y. Gao, Y. Jin, K.-H. Thung, and D. Shen, "Soft-Split Sparse Regression Based Random Forest for Predicting Future Clinical Scores of Alzheimer's Disease," in MLMI, vol. 9532, pp. 246-254, 2015.

[10] T. Wang, F. Shi, Y. Jin, W. Jiang, D. Shen, and S. Xiao, "Abnormal Changes of Brain Cortical Anatomy and the Association with Plasma MicroRNA107 Level in Amnestic Mild Cognitive Impairment," Frontiers in Aging Neuroscience, vol. 8, 2016.

[11] B. Lei, S. Chen, D. Ni, and T. Wang, "Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion," Front. Aging Neurosci., vol. 8, pp. 1-17, 2016.

[12] B. Jie, C.-Y. Wee, D. Shen, and D. Zhang, "Hyper-connectivity of functional networks for brain disease diagnosis," Medical image analysis, vol. 32, pp. 84-100, 2016.

[13] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," NeuroImage, vol. 129, pp. 292-307, 2016.

[14] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen, "Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest," Neurobiol Aging, vol. 46, pp. 180-191, 2016.

[15] B. Lei, S. Chen, D. Ni, and T. Wang, "Joint Learning of Multiple Longitudinal Prediction Models by Exploring Internal Relations," in MLMI, pp. 330-337, 2015.

[16] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, "Temporally-Constrained Group Sparse Learning for Longitudinal Data Analysis in Alzheimer's Disease," IEEE Trans Biomed Eng, vol. 64, pp. 238-249, 2017.

[17] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via multi-task learning," NeuroImage, vol. 78, pp. 233-248, 2013.