# A Cascaded Multi-Modality Analysis in Mild Cognitive Impairment

Lu zhang[1], Akib Zaman[1], Dajiang Zhu[1*](dajiang.zhu@uta.edu)

[1] The University of Texas at Arlington

**Abstract.** Though reversing the pathology of Alzheimer's disease (AD) has so far not been possible, a more tractable goal may be the prevention or slowing of the disease when diagnosed in its earliest stage, such as mild cognitive impairment (MCI). Recent advances in deep modeling approaches trigger a new era for AD/MCI classification. However, it is still difficult to integrate multi-modal imaging data into a single deep model, to gain benefit from complementary datasets as much as possible. To address this challenge, we propose a cascaded deep model to capture both brain structural and functional characteristic for MCI classification. With diffusion tensor imaging (DTI) and functional magnetic resonance imaging (fMRI) data, a graph convolution network (GCN) is constructed based on brain structural connectome and it works with a one-layer recurrent neural network (RNN) which is responsible for inferring the temporal features from brain functional activities. We named this cascaded deep model as Graph Convolutional Recurrent Neural Network (GCRNN). Using Alzheimer's Disease Neuroimaging Initiative (ADNI-3) dataset as a test-bed, our method can achieve 97.3% accuracy between normal controls (NC) and MCI patients.

**Keywords:** Mild Cognitive Impairment, GCN, RNN

## 1 Introduction

Alzheimer's disease (AD) is the most common form of dementia and approximately 10% of people age 65 and older have AD [1]. Given the facts that AD is the only disorder that cannot be prevented or cured, a more tractable goal is to diagnose at its earlier stage - mild cognitive impairment (MCI), as more than 50% people with MCI will convert to AD eventually. To date, the most sensitive AD biomarkers include measures of brain Aβ deposition (amyloid PET)[2] and neurodegeneration (FDG PET)[3]. However, PET can be invasive and may not be feasible or readily available. Thus, non-invasive imaging modalities including MRI (T1 weighted), diffusion tensor imaging (DTI) and resting state fMRI (r-fMRI) are of great interest. There have been a variety of approaches for early diagnosis or classification of AD/MCI, such as voxel-based analysis[4,5], tract-based spatial statistics (TBSS for DTI)[6,7], and multivariate analysis/machine learning-based algorithms[8-10]. Recently, the advances in deep neural network approaches trigger a new era for AD/MCI studies. For example, Convolutional Neural Networks (CNN) [11] was applied on AD classification using structural MRI and PET data. The study in [12] conducted classification of MCI and regression of

cognitive scores simultaneously with a multi-channel CNN. Graph Convolutional Network (GCN) [13] was recently proposed to fulfil the MCI conversion with multiple datasets. Despite the advancements made by these methods, most proposed deep models only work with unstructured data in image space, such as voxel intensity. How to leverage the deep models over the brain network has been rarely studied. In addition, brain structure and function are closely related, but it is still difficult to integrate them, via multi-modal imaging data (i.e. DTI and fMRI), into a single deep model, to gain benefit from complementary imaging data as much as possible.

In this work, to capture both brain structural and functional characteristics in MCI patients, we proposed a cascaded GCN and Recurrent Neural Network (RNN) framework to classify MCI via modeling brain spatial and temporal information simultaneously. Brain regions are treated as graph nodes and structural connectome is used as the topological description of the graph. The representative fMRI signal of each brain region composes the feature vector associated with the node. Multiple RNN cells was connected to capture the temporal patterns contained in the structural network. We named this model as Graph Convolutional Recurrent Neural Network (GCRNN). **Fig.1** shows the major steps in this work. We use weighted adjacency matrix to store the fiber connectivity of different brain regions and a feature matrix to store the fMRI BOLD signals. During the training process, weighted adjacency matrix and feature matrix are firstly fed into GCN layers to learn the spatial features, then the RNN cells are built on the obtained features to learn temporal patterns in fMRI signals. Finally, a fully connected layer is used to perform the classification. Using Alzheimer's Disease Neuroimaging Initiative (ADNI-3) dataset (116NC/93MCI) as a test-bed, our method can achieve 97.3% accuracy between normal controls (NC) and MCI patients.

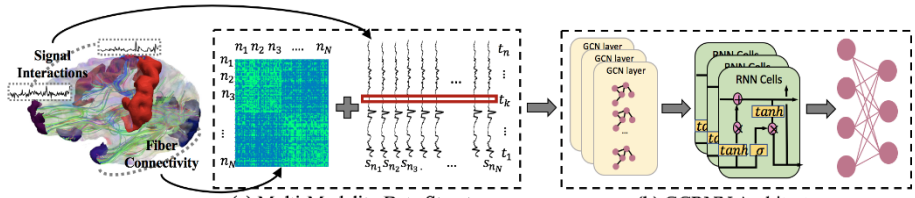


**Fig. 1.** The pipeline of the proposed work. **(a)** We generate two matrices as the input for GCN. Structural connectivity is represented as weighted adjacency matrix to depict the structure of the graph. fMRI signals are formed as feature matrix associated with the nodes. $n_i$ represents the $i^{\text{th}}$ brain region, $s_{n_i}$ is the averaged fMRI signal in region $n_i$, and $t_i$ is the value at $i^{\text{th}}$ time point. **(b)** Overview of the proposed GCRNN.

## 2 Method

### 2.1 Dataset

In this work, all the data used was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI 3) [14]. We began with 661 subjects which have both DTI and rs-

fMRI data. The DTI data is 2.0mm isotropic, TE=56ms, TR=7.2s, and the gradient directions is 54. For rsfMRI, the range of image resolution in X and Y dimensions was from 2.29mm to 3.31mm. The range of slice thickness was from 3.3mm to 3.4mm, TE=30ms, TR=3s, and there were 197 volumes (time points) for each subject.

We applied standard pre-processing procedures including skull removal for both modalities, spatial smoothing, slice time correction, temporal pre-whitening, global drift removal and band pass filtering (0.01-0.1 Hz) for rsfMRI, eddy current correction and fiber tracking via MedINRIA for DTI, registering rsfMRI to DTI space using FLIRT and adopt the Destrieux Atlas [15] for ROI labeling. The brain cortex is partitioned into 148 regions after removing two unknow areas. After the preprocessing and image quality check, we have 209 subjects (116 CN/93 MCI) for experiments.

## 2.2 Multi-Modality Data Structure.

### 2.2.1 Adjacency matrix and Feature matrix.

Following the basic brain network approach, we used the Destrieux atlas along with fiber tracking information to reconstruct the brain connectivity matrices, weighted by fiber count and ROI volume. For each brain region, we used the averaged fMRI BOLD signal as the temporal feature. Specifically, we used the adjacency matrix $A \in R^{N\times N}$ to represent the fiber connectivity of different brain regions, $N$=148 is the number of the brain regions. $A$ is a symmetric matrix, $A_{ij} \in A$ is the number of shared fibers of brain region $i$ and region $j$. The feature matrix $F \in R^{T\times N\times P}$ is used to store the averaged fMRI signal, $T$ is the number of time points, $P$ is the dimension of averaged fMRI signal of one brain region at one time point. By using the adjacency matrix $A$ and feature matrix $F$ simultaneously, we aim to integrate brain spatial and temporal information for MCI diagnosis with our cascaded deep model.

### 2.2.2 Transformation of adjacency matrix and feature matrix.

Before feeding the adjacency matrix and feature matrix into our deep model, we conducted normalization and transformations for two matrices separately. For feature matrix $F$, we normalize it by linear transformation:

$$\vec{Y} = \frac{\vec{X} - X_{min}}{X_{max} - X_{min}}, \tag{1}$$

where $\vec{X}$ is the vector before normalization, $X_{min}$ and $X_{max}$ are the minimum and maximum of $\vec{X}$, $\vec{Y}$ is the new vector.

For adjacency matrix $A$, besides linear transformation using formula (1), we also performed Laplace transformation which has three different forms:

$$L_1 = D - \hat{A} \tag{2}$$

$$L_2 = D^{-1/2}(D - \hat{A})\, D^{-1/2} \tag{3}$$

$$L_3 = D^{-1}(D - \hat{A}) \ . \tag{4}$$

$\hat{A}$ is the adjacency matrix after linear transformation of $A$, and $D$ is the degree matrix of $\hat{A}$ which is a diagonal matrix: $D_{ii} = \sum_j \hat{A}_{ij}$. The results from Laplace transformation have good properties including: 1) all of them are positive semi-definite 2) 0 is

an eigenvalue and the corresponding eigenvectors are constant one vector $1^c$, $D^{-1}1^c$ and $1^c$ respectively 3) they all have n non-negative real-valued eigenvalues. We designed a comparative experiment in section 3.3 to show different performance by using $\widehat{A}$ , $L_1$, $L_2$, $L_3$ as the adjacency matrix respectively. Here we use the notation $\tilde{A}$ and $\tilde{F}$ represent the adjacency matrix and feature matrix after the normalization and transformation. The input data is $\{X_1, X_2, \ldots, X_T\}$, where $X_t = \{\tilde{A}, \tilde{F}_t\}$ and $\tilde{F} = \{\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_T\}$, $\tilde{F}_t \in R^{N \times P}$.

### 2.3 GCRNN Architecture.

The motivation of this work is to design a deep model to integrate brain spatial and temporal information for MCI classification. Graph Convolutional Neural Network (GCN) and Recurrent Neural Network (RNN) have shown their effectiveness in modeling structured data, such as graph and sequence [9,14]. Here, we used GCN to model brain structural connectivity and RNN to capture temporal pattern of fMRI data. We proposed a Graph Convolutional Recurrent Neural Network (GCRNN) to combine these two powerful models for MCI classification. Because RNN model has two different variants: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), we will consider both of them and develop two different GCRNN models: GCN-GRU and GCN-LSTM. The overall architectures of the two different GCRNN and the implementation details are shown in **Fig. 2**.

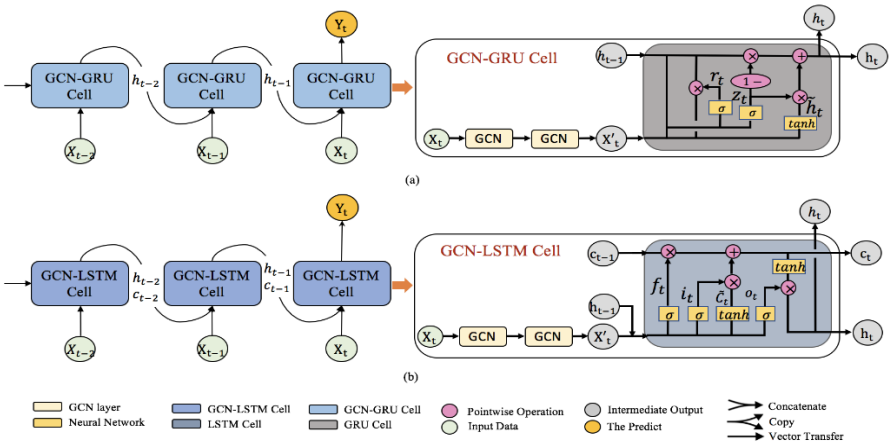


**Fig. 2. (a)** GCRNN with GRU. **(b)** GCRNN with LSTM. Both of the two GCRNN have the same two-layer GCN. Training process: input data $X_t$ is fitted into GCN layers and obtains the output of GCN, $X'_t$, then $X'_t$ together with $h_{t-1}$ and $c_{t-1}$ are used as input data to fit into a one-layer RNN. According to different RNN, data will be processed differently.

**Table 1** is the mathematical definition of GCN-GRU and GCN-LSTM models. Two-layers GCN is defined by formula (5), $W_0$ and $W_1$ are the learnable weights of GCN layer 1 and layer 2 respectively, $\tilde{A}$ and $\tilde{F}$ are the adjacency matrix and feature matrix mentioned in section 2.2.3. $g(\tilde{A}, \tilde{F}_t)$ is the output of the second GCN layer and is also

the input of two different RNN models. As the output of GCN, $g(\tilde{A}, \tilde{F}_t)$ contains both topological structure and functional features. LSTMis defined by formula (6) - (11) and GRUis defined by formula (12) - (15). In these formulae, $W_*$ and $b_*$ are the weight and bias of RNN model. Note that GRU and LSTM have different performance when applied to different problems. In order to examine which RNN model works better in GCRNN architecture, we conducted a comparative experiment to compare the performance of the two structures in **Table 2**.

**Table 1.** The mathematical definition of GCN-GRUand GCN-LSTM models.

| **GCN-LSTM** | **GCN-GRU** |
|---|---|
| $\boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}}) = ReLU(\tilde{A}ReLU(\tilde{A}\tilde{F}_t W_0)W_1)$ (5) | |
| $f_t = \sigma(W_f \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})] + b_f)$ (6) | $z_t = \sigma(W_z \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})])$ (12) |
| $i_t = \sigma(W_i \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})] + b_i)$ (7) | $r_t = \sigma(W_r \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})])$ (13) |
| $\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})] + b_C)$ (8) | $\tilde{h}_t = tanh(W \cdot [r_t * h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})])$ (14) |
| $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ (9) | $h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$ (15) |
| $o_t = \sigma(W_o \cdot [h_{t-1}, \boldsymbol{g}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{F}}_{\boldsymbol{t}})] + b_o)$ (10) | |
| $h_t = o_t * tanh(C_t)$ (11) | |

## 3 Results

We perform binary classification (CN vs. MCI) using GCRNN models and multi-modality data. In order to show the reproducibility of the results, we repeated each experiment for 10 times with the same parameter setting. The performance was evaluated by the best accuracy (Best), worst accuracy (Worst), average accuracy (Ave) and the standard deviation (Stdev) of the 10 runs. To be specific, we have three comparative experiments: 1) classification accuracy 2) influence of brain structural connectivity 3) performance of using different Laplace transformation.

### 3.1 Classification performance of GCRNN model

For demonstrating the effectiveness of the proposed GCRNN model, we performed 540 experiments under 27 different model settings with both GCN-LSTM and GCN-GRU. The model setting is defined as: GCN = (number of layers, [feature number of each layer]), RNN = (number of layers, [hidden nodes number of each layer]) and model name. **Table 2** displays the results of 220 experiments including two-layer GCN / one-layer RNN, one-layer GCN / one-layer RNN and three-layer GCN / one-layer RNN.

In **Table 2**, the best results and the corresponding model settings are highlighted in bold. Besides the best results, some items are marked with "*", because except the average accuracy is not the best, they have relatively good performance and stability at the same time. Note that using more hidden nodes in RNN leads to higher accuracy.. With the same number of hidden nodes in RNN, the model with two-layer GCN has

the best performance comparing to one-layer and three-layer GCN. This suggests that our data might sufficient for current RNN settings. For GCN training, three layers might need more data to retain the accuracy.

**Table 2.** Results of different model settings.

| Model settings | GCN-LSTM-Cell | | | | GCN-GRU-Cell | | | |
|---|---|---|---|---|---|---|---|---|
| | Best | Worst | Ave | Stdev | Best | Worst | Ave | Stdev |
| GCN = (2, [64, 128]) RNN = (1, 1024) | 95.3 | 87.3 | 90.9 | 2.9 | 94.1 | 80.1 | 87.8 | 5.2 |
| **GCN = (2, [32, 64]) RNN = (1, 1024)** | **97.3** | **80.6** | **93.5** | **5.2** | 96.1 | 81.4 | 89.7 | 5.0 |
| GCN = (2, [16, 32]) RNN = (1, 1024) | 96.9 | 72.7 | 87.8 | 7.6 | 91.1 | 76.6 | 83.8 | 4.5 |
| GCN = (2, [64, 128]) RNN = (1, 512) | 98.8 | 58.1 | 77.8 | 11.9 | 93.4 | 61.3 | 75.8 | 11.4 |
| GCN = (2, [32, 64]) RNN = (1, 512) | 94.0 | 73.7 | 82.7 | 5.7 | 93.2 | 56.1 | 76.2 | 10.6 |
| GCN = (2, [16, 32]) RNN = (1, 512) | 96.9 | 53.2 | 84.3 | 13.7 | 86.3 | 58.4 | 68.8 | 8.3 |
| GCN = (2, [64, 128]) RNN = (1, 2048) | 97.2 | 84.5 | 91.3 | 4.1 | 96.5 | 80.9 | 90.9 | 6.5 |
| **GCN = (2, [32, 64]) RNN = (1, 2048)** * | 96.1* | 88.1* | 91.4* | 2.8* | **97.7** | **83.6** | **91.5** | **4.4** |
| GCN = (2, [16, 32]) RNN = (1, 2048) | 92.1 | 83.4 | 87.3 | 2.4 | 91.4 | 81.5 | 87.7 | 3.1 |
| GCN = (1, 64) RNN = (1, 1024) | 91.1 | 82.5 | 87.6 | 2.7 | 93.8 | 83.3 | 88.4 | 4.1 |
| GCN = (3, [32, 64, 128]) RNN = (1, 1024) | 76.1 | 59.6 | 67.6 | 5.3 | 76.1 | 53.1 | 65.0 | 8.6 |

### 3.2 Influence of brain structural connectivity

One major advantage of GCRNN is to use structural connectivity, constructed by counting DTI derived fibers between brain regions, to describe the topology of the graph for GCN. In order to verify the influence of using brain structural network, we make a comparison with the variant that uses all ONE matrix as the adjacency matrix and see how the topology of the graph affects the performance. Note that the all ONE matrix has the same size as brain structural connectivity matrix and it is a complete graph. That is, we assume each pair of brain regions is connected, hence it is unbiased and doesn't introduce any spatial information of brain network during the training. The comparison is under the setting of GCN = (2, [32, 64]) RNN = (1, 1024) with a LSTM and repeated for 10 times. The results are shown in **Fig. 3**.

From the **Fig. 3** we can see that using brain structural connectivity as adjacency matrix has an overwhelming advantage compared to all ONE matrix, which means that the structural connectivity is an effective representation of brain spatial information during GCN training. Moreover, the classification results also suggest that the functional features along with individual brain structures have the potential to capture the intrinsic alterations in MCI patients.

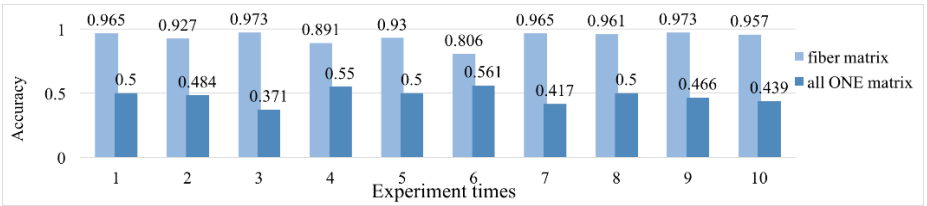


**Fig. 3.** Results of testing influence of brain structural connectivity

### 3.3 Influence of Laplace Transformation

Though Laplace transformation has been applied to a wide range of applications as well as GCN constructions, it is still largely unclear whether it can improve the performance in our binary classification. In this comparative experiment, we used $\widehat{A}$ (structural connectivity matrix after linear transformation) $L_1$ $L_2$ and $L_3$ ($\widehat{A}$ after three forms of Laplace transformation) which are defined in section 2.2.2, respectively. Our experiment is under the model setting with GCN = (2, [32, 64]) RNN = (1, 1024) with LSTM and repeated for 10 times. The results are shown in **Fig. 4**.

The results in **Fig. 4** show that among three forms Laplace transformation and original matrix $\widehat{A}$, the performance of all three forms of Laplace transformation have lower standard deviation (Stdev), thus is relatively stable, and there is no significant difference in best accuracy (Best), worst accuracy (Worst) and average accuracy (Ave) between Laplace transformation and original matrix. So using Laplace transformation will increase the stability of the results in the experiments, but there is no significant improvement in the accuracy.

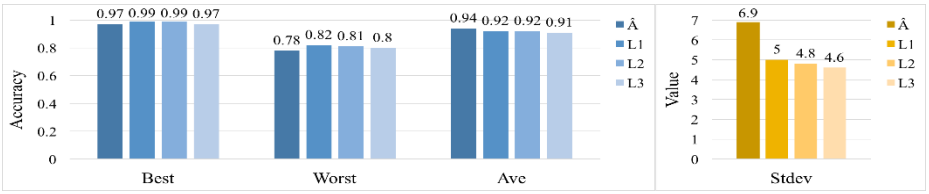


**Fig. 4.** Results of testing influence of Laplace Transformation

## 4 Conclusion

In this work, we proposed a cascaded GCN and RNN framework, named as GCRNN, to classify MCI via modeling brain spatial and temporal information simultaneously. Using brain structural connectivity as the graph and fMRI signals as the feature, our cascaded deep model effectively combines GCN and RNN together and achieve 97.7% accuracy for classification of MCI and NC. Note that, besides the considerations of using structural connectivity and different deep architectures, another factor that might influence our result is the length of the fMRI signal used. Previous studies [16] suggested that for resting state fMRI 14 time points (when TR=2s) are sufficient to capture functional dynamic patterns. Here, we also examined the influence when adopting different length of fMRI signals. The results are shown in **Fig. 5**. and it suggests that there is no obvious relationship between length and accuracy in our settings. In the future, this need to be further examined with more samples.

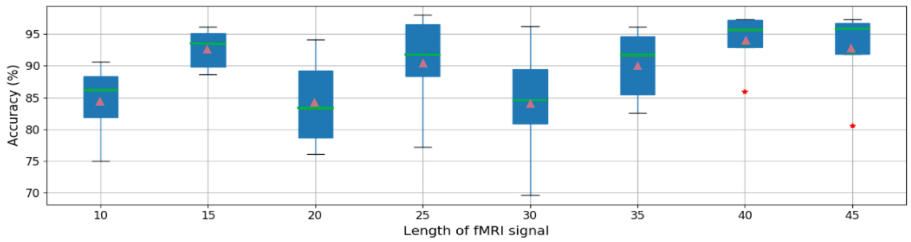


**Fig. 5.** Results of testing length of fMRI signal

## References


1. Alzheimer's Association 2019 Alzheimer's Disease Facts and Figures Report, https://www.alz.org/alzheimers-dementia/facts-figures.
2. Rowe, C. C. et al.: Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol. Aging 31, 1275–1283 (2010)
3. Jagust, W. J. et al.: The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. Alzheimers. Dement. 6, 221–9 (2010)
4. Ashburner, J. & Friston, K. J.: Voxel-Based Morphometry—The Methods. Neuroimage 11, 805–821 (2000).
5. THOMPSON, P. M., APOSTOLOVA, L. G.: Computational anatomical methods as applied to ageing and dementia. Br. J. Radiol. 80, S78–S91 (2007)
6. Vemuri, P. et al.: Accelerated vs. unaccelerated serial MRI based TBM-SyN measurements for clinical trials in Alzheimer's disease. Neuroimage 113, 61–69 (2015).
7. Smith, S. M. et al.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505 (2006)
8. Jiang, X., et al.: Intrinsic Functional Component Analysis via Sparse Representation on Alzheimer's Disease Neuroimaging Initiative Database. Brain Connect. 4, 575–586 (2014)
9. Tong, Tong, et al.: Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. Pattern recognition 63, 171-181 (2017).
10. Jack Jr, Clifford R., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine27(4), 685-691 (2008).
11. Liu, Mingxia, et al.: Deep multi-task multi-channel learning for joint classification and regression of brain status. In International conference on medical image computing and computer-assisted intervention, Springer, Cham, pp. (3-11). (2017).
12. Liu, Mingxia, et al.: Deep multi-task multi-channel learning for joint classification and regression of brain status. In International conference on medical image computing and computer-assisted intervention, Springer, Cham, pp. (3-11). (2017).
13. Parisot, S., et al.: Spectral graph convolutions for population-based disease prediction. In International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, Cham. pp. (177-185), (2017).
14. ADNI | Alzheimer's Disease Neuroimaging Initiative, http://adni.loni.usc.edu/.
15. Destrieux, Christophe, et al.: Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Nuroimage 53(1), 1-15 (2010).
16. Zhang, Xin, et al. "Characterization of task-free and task-performance brain states via functional connectome patterns." Medical image analysis 17(8) 1106-1122. (2013)