Knowledge-Based Systems 254 (2022) 109466

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Longitudinal study of early mild cognitive impairment via similarity-constrained group learning and self-attention based SBi-LSTM☆

Baiying Lei [a,b,1], Yuwen Zhang [a], Dongdong Liu [a], Yanwu Xu [d], Guanghui Yue [a], Jiuwen Cao [c], Huoyou Hu [e], Shuangzhi Yu [a], Peng Yang [g], Tianfu Wang [a], Yali Qiu [a], Xiaohua Xiao [e,*], Shuqiang Wang [f,*]

[a] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, 518060, China
[b] Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation, Southern University of Science and Technology, Shenzhen 518055, China
[c] Artificial Intelligence Institute, Hangzhou Dianzi University, Zhejiang, 310010, China
[d] Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo, China
[e] Affiliated Hospital of Shenzhen University, Health Science Center, Shenzhen University, Shenzhen Second People's Hospital, Shenzhen, China
[f] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Shenzhen, 518000, China
[g] Shenzhen University, China

ARTICLE INFO

Article history:
Received 11 November 2021
Received in revised form 20 June 2022
Accepted 13 July 2022
Available online 22 July 2022

Keywords:
Mild cognitive impairment
Similarity-constrained group learning
SBi-LSTM
Self-attention
Longitudinal study

ABSTRACT

Alzheimer's disease (AD) is an incurable neurodegenerative disease. Mild cognitive impairment (MCI) is often considered a critical time window for predicting early conversion to Alzheimer's disease (AD), with approximately 80% of amnestic MCI patients developing AD within 6 years. MCI can be further categorized into two stages (i.e., early MCI (EMCI) and late MCI (LMCI)). To identify EMCI effectively and understand how it changes brain function, the brain functional connectivity network (BFCN) has been widely used. However, the conventional methods mainly focused on detection from a single time-point data, which could not discover the changes during the disease progression without using multi-time points data. Therefore, in this work, we carry out a longitudinal study based on multi-time points data to detect EMCI and validate them on two public datasets. Specifically, we first construct a similarity-constrained group network (SGN) from the resting state functional magnetic resonance imaging (rs-fMRI) data at different time-points, and then use a stacked bidirectional long short term memory (SBi-LSTM) network to extract features for longitudinal analysis. Also, we use a self-attention mechanism to leverage high-level features to further improve the detection accuracy. Evaluated on the public Alzheimer's Disease Neuroimaging Initiative Phase II and III (ADNI-2 and ADNI-3) databases, the proposed method outperforms several state-of-the-art methods.

1. Introduction

Alzheimer's disease (AD) is one of the most severe dementia among elderlies. With the deepening of global aging, the number of AD patients is also growing. According to the latest statistics, the current total number of AD patients is about 50 million, which is expected to increase to about 152 million by 2050 and will cause a lot of burden to the economy and community [1]. As an aging disease, AD is an incurable disease [2] and its early stage is mild cognitive impairment (MCI). MCI contains early MCI (EMCI) and late MCI (LMCI), and its detection has been attracting wide attention [1,3]. If we can detect EMCI early and perform timely treatment and intervention, we can effectively delay the AD occurrence [4].

☆ This work was supported partly by National Natural Science Foundation of China (Nos. 61871274, 62101338, and U1909209), National Natural Science Foundation of Guangdong Province, China (No. 2019A1515111205), Guangdong Provincial Key Laboratory, China (No. 2020B121201001), Shenzhen Key Basic Research Project, China (Nos. KCXFZ20201221173213036, SGDX20201103095802007, JCYJ20180507184647636, JCYJ20190808155618806, GJHZ20190822 095414576, and JCYJ20190808145011259).

* Corresponding authors.
E-mail addresses: tu_xi8888@163.com (X. Xiao), sq.wang@siat.ac.cn (S. Wang).
[1] Senior Member, IEEE.

https://doi.org/10.1016/j.knosys.2022.109466
0950-7051/

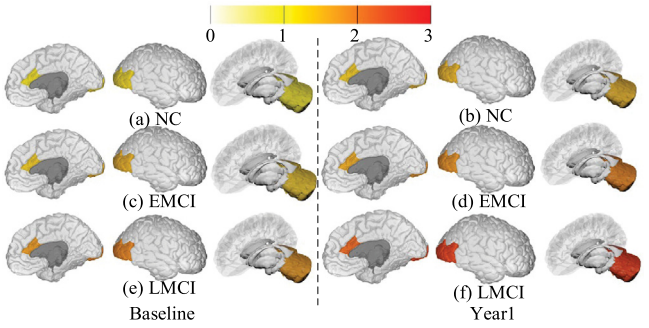


Fig. 1. Examples of brain regions' changes along time and different stages of disease.

With the further development of MCI, both structure and function of the brain region changes to a certain extent, but in the early stage of MCI, the brain structure has less change than brain function and it is hard to detect the slight changes of brain structures [5]. Moreover, the resting state functional magnetic resonance imaging (rs-fMRI) has been widely used in MCI detection [4,6]. With the development of MCI, the corresponding brain functions of relevant brain regions will also change [5]. Therefore, the longitudinal study of MCI can help us further identify the corresponding brain change as well as improve the accuracy of disease diagnosis [7]. In order to better represent the physiological function changes during different time points, we use BrainPainter [8] to further demonstrate the changes in Fig. 1, which shows the corresponding brain regions changes via the different stages of MCI and normal control (NC) [9]. We can find that the EMCI and NC only have a slight difference, which is difficult to detect them [10].

To address the above problems, establishing a new brain functional connectivity network (BFCN) among brain regions to detect the EMCI disease region is quite desirable. Recently, many researchers proposed various methods to construct an effective BFCN for MCI detection [11]. For example, the Pearson's correlation (PC) brain network is constructed to solve the problem of learning the relationship of different brain regions [12]. However, the BFCN of PC has lots of redundancy, which is undesirable for disease diagnosis. Hence, sparse representation (SR) is used to reduce the numerous useless features in BFCN for performance improvement [13]. However, these methods only consider the internal regional relationship, which fails to consider the external brain regional relationship of subjects. To improve it, the group-constrained sparse (GCS) brain network has been used [14]. However, the GCS method does not consider the similarity of brain regions during the BFCN construction. Also, the recent work has shown that multi-center data becomes more appealing in disease detection [15], because various scanning time points and parameters have been used on multiple centers and thus the combined data is more challenging to validate the generalization ability of the model. Hence, an effective method to construct the multi-center BFCN is quite desirable. For this reason, we devise a similarity-constrained group sparse network (SGN) for multi-center BFCN construction, which can learn the similarity among features and reduce the useless features simultaneously.

Apart from BFCN construction, discriminative feature learning plays an essential role in detecting EMCI. In the recent years, with the continuous development of artificial intelligence, various intelligent diagnosis methods are used to detect brain disease [16]. Recent work shows that constructing the effective BFCN and feature extraction/selection methods can improve the detection results via support vector machine (SVM) [17,18]. However, these feature extraction/selection methods are based on traditional machine learning methods [17], which cannot learn deep features for disease detection. At present, some deep learning method has achieved good results in the diagnosis of brain disease [19–21]. For example, Bakar et al. [22] proposed to use multilayer perceptrons (MLP) neural network for the detection of Parkinson's disease. Except MLP, long short term memory (LSTM) has also widely used for disease detection. For example, Hong et al. [23] used LSTM to learn the time-related biomarkers associated with disease status to predict the stages of AD. Wang et al. [24] proposed to use two layer stacked bidirectional LSTM (SBi-LSTM) for brain state recognition. However, the above disease detection methods mainly use magnetic resonance imaging (MRI) dataset, while the rs-fMRI has been proved more effective for detecting EMCI [4]. An more effective deep learning method is still needed for EMCI detection via rs-fMRI data. Therefore, we explore the SBi-LSTM to extract feature from BFCN for longitudinal study of EMCI.

Recently, various attention mechanisms have been proposed to learn the discriminative features for detecting disease and segmenting lesion areas [25,26]. For example, Song et al. proposed a novel attention guided method to improve their performance for action recognition [27]. Zhao et al. proposed to use spatial-channel attention based on U-net to find deeper features for gland segmentation and achieved promising result [28]. But the above attention mechanisms do not have similarity with the original feature, which has been proved important for classification [29]. Recently, the self-attention has achieved quite promising results on word detection [30] since it has the powerful ability to find similar features during feature learning. Therefore, we propose to use self-attention mechanism for EMCI detection.

Based on the above analysis, the BFCN constructed by SGN is integrated with self-attention via SBi-LSTM (SSBi-LSTM) to automatically detect and analyze EMCI longitudinally, which consists of the following three steps. Firstly, the longitudinal data is used to construct BFCN by our SGN method. Then, the brain network feature is sent to the SBi-LSTM for feature extraction. Finally, we propose to use self-attention to find more effective features during the feature learning. Experimental results on the public Alzheimer's Disease Neuroimaging Initiative Phase II and III (ADNI-2 and ADNI-3) (http://adni.loni.usc.edu/) databases show the promising performance for analyzing and detecting EMCI longitudinally. The main contributions of this work are summarized as below:

(1) We construct a similarity-constrained group brain network to learn representative features effectively.

(2) We design a SBi-LSTM framework to utilize the longitudinal information for EMCI detection.

(3) We explore a self-attention mechanism to find the most discriminative features to improve the detection performance.

2. Related work

2.1. BFCN construction

Many studies in the literature focused on constructing an effective BFCN for brain disease detection. To reduce the noisy connection as well as the network complexity, previous methods focused on constructing a sparse brain network based on the kNN method (Zhang et al. [31]). Specifically, every node connects a subset of nodes (i.e., its k nearest neighbors, kNN graph for short) in the sparse brain networks. Moreover, the nearest neighbor is obtained based on the similarity measurement. For example, Yang et al. [32] proposed to first calculate the mean BFCN matrix of all training subjects within the same time-series block to construct a kNN graph. Yao et al. [33] proposed to

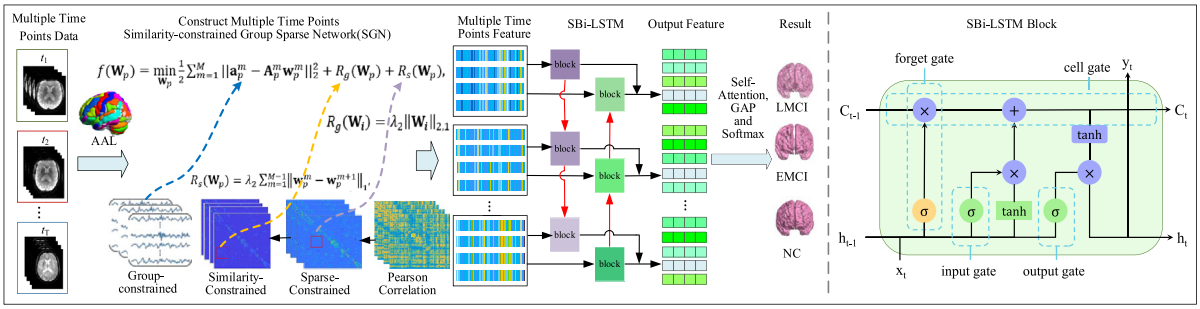


Fig. 2. Flowchart of the proposed SSBi-LSTM method for MCI detection from two time points rs-fMRI data.

first calculate PC among the nodes within the individual brain and then to connect each brain region with 8 neighbors for all subjects. In addition, Bi et al. [34] designed two deep learning methods for functional brain network classification, and also implemented an Extreme Learning Machine (ELM) augmentation structure to further improve the learning ability. Zhang et al. [35] proposed a multi-GCN based GAN (MGCN-GAN) to infer individual SC based on corresponding FC by automatically learning the complex associations between individual brain structural and functional networks. Gan et al. [36] proposed a framework for functional connectivity network (FCN) analysis, which conducted the brain disease diagnosis on the rs-fMRI data. Zhang et al. [37] developed a graph-based deep neural network to simultaneously model brain structure and function in MCI. Huang et al. [15] proposed to add both PC and modularity structure into sparse low-rank brain network (SLR) to get the PC-related SLR brain network features. Another fused sparse network (FSN) method was recommended by Yang et al. [5] to analyze the longitudinal features to detect MCI stages. However, these methods mainly focus on getting sparse brain network features but fail to consider the similarity characteristics while constructing BFCN, which are undesirable for improving detection performance. Therefore, we propose a similarity-constrained sparse brain network for BFCN construction.

2.2. Deep learning methods for longitudinal study

Most existing MCI detection methods are based on conventional machine learning, while EMCI detection from rs-fMRI data or using deep learning methods are quite rare. For example, Liu et al. proposed to use MRI data and the subjects' demographic information to establish the deep multi-task multi-channel learning model for clinical score regression [38]. Feng et al. proposed a new multi-task learning method to effectively detect brain disease via MRI and positron emission computed tomography (PET) data [39], where a 3D CNN is proposed to extract features and a specific SBi-LSTM network is designed to fuse data from different modalities. However, this method only focuses on brain structure changes caused by AD. Islam et al. [40] presented a MCI diagnosis method to learn the spatial–temporal dependency via rs-fMRI time signal directly and achieved good results based on LSTM. However, they may miss the effect of BFCN. Another novel method uses weight correlation kernel model for effectively constructing a dynamic BFCN from rs-fMRI, which obtained quite impressive results [41]. However, this method does not consider longitudinal study. Regarding the longitudinal study on AD, a novel deep polynomial network (DPN) is proposed to learn the longitudinal features for predicting the future development of AD and get good results [7]. Another work used parameter-free centralized multi-task learning to detect EMCI and LMCI [5].

Both constructing a meaningful BFCN and building a powerful feature learning model are beneficial to detect the brain regions longitudinally. Therefore, SBi-LSTM is utilized in this paper to obtain rich longitudinal information for EMCI detection.

3. Methodology

Fig. 2 shows the architecture of the proposed EMCI longitudinal analysis method using SSBi-LSTM model. Specifically, we first construct a SGN network to represent the longitudinal data of different centers. Then we combine the two center SGN features and feed them into the SBi-LSTM module to extract features. Thirdly, we add a self-attention mechanism to identify the most representative features. Finally, we use a Softmax function to detect EMCI.

3.1. Subjects and data acquisition

The datasets collected in this work are all from public ADNI-2 and ADNI-3 datasets, with 102 subjects in ADNI-2 and 51 subjects in ADNI-3 dataset. In ADNI-2 dataset, there are 30 LMCI, 39 EMCI, and 33 NC. In ADNI-3 dataset, there are 16 LMCI, 25 EMCI, and 10 NC. All subjects have two time-point (baseline and year1) rs-fMRI data with different scan parameters. Our multi-time point data is acquired with a 3.0 T Siemens MRI scanner. The rs-fMRI data collection uses a gradient echo planar imaging (EPI) sequence with specific parameters as follows: imaging matrix $= 64 \times 64$, field of view (FOV) $= 192 \times 192$ mm^2, voxel thickness $= 3.3$ mm, flip angle $= 80°$, time repetition (TR) $= 3000$ ms, echo time (TE) $= 30$ ms. Specifically, we collect data at 140 time points in ADNI-2, and gather data at 197 time points in ADNI-3. During the data collection, ADNI-2 subjects are asked to close their eyes, while ADNI-3 subjects are asked to open their eyes.

3.2. Data pre-processing

We follow the same preprocess parameters used in [5]. In this study, the collected rs-fMRI data is pre-processed by the Statistical Parametric Mapping toolbox (SPM12) and the Data Processing Assistant for Resting-State fMRI (DPARSFA) with a standardized data pre-processing method. Before pre-processing, we remove the first 10 time points of each subject's rs-fMRI data, which can make the magnetization equal. We utilize a staggered sequence of slice correction of the rest 170 time points. Also, the echo planar scan is utilized to make sure that each slice' data is corresponding to the unanimous point in time. We perform the following preprocessing steps: First, we set the interpolation time point to half TR, to minimize the relative error of each TR. Then, we perform time registration and spatial normalization after removing the head motion to reduce the influence of scan differences. The high pass filter parameters are set to [0.01, 0.08] to reduce low frequency noise. Finally, we divide the brain into 90 regions of interests (ROIs) by aligning to the anatomical automatic labeling (AAL) template [42].

3.3. Similarity-constrained group sparse network

After pre-processing, it is essential to establish an effective BFCN to improve the diagnosis performance [15,43]. Therefore, we propose a SGN model to construct BFCN. In this section, bold uppercase letters, bold lowercase letters, and italic letters are used to represent matrices, vectors, and scalars, respectively. Assuming that we have the input data $\mathbf{A}$ from M subjects, and a brain is separated into P ROIs by AAL template so that we can derive $\mathbf{A} = \left[\boldsymbol{a}_1, \ldots \boldsymbol{a}_p, \ldots \boldsymbol{a}_P\right] \in \mathbb{R}^{P\times M}$, where $\boldsymbol{a}_p^m = \left[a_{1p}^m, a_{2p}^m, \ldots, a_{Ep}^M\right]$ is a length E vector that denotes the regional average time series of the blood oxygen-dependent level in the pth ROI of subject k. We denote the signal matrix of all ROIs except $\mathbf{a}_p^k$ as $\mathbf{A}_p^k$, so that $\mathbf{A}_p^k = \left[\mathbf{a}_1^k, \ldots \mathbf{a}_{p-1}^k, \ldots \mathbf{a}_{p+1}^k, \ldots \mathbf{a}_P^k\right]$ is the column vector of weighted regression coefficients. Although the GCS network can guarantee both group constraint and sparseness, it may also ignore the similarity parameters among different subjects, and thus obtain less discriminative features. To solve it, we explore a model to learn the functional brain network of each subject via similarity-constrained group sparse learning. The objective function is formulated as

$$f(\mathbf{W}_p) = \min_{\mathbf{W}_p} \frac{1}{2}\sum_{m=1}^{M} \|\mathbf{a}_p^m - \mathbf{A}_p^m \mathbf{w}_p^m\|_2^2 + R_g(\mathbf{W}_p) + R_s(\mathbf{W}_p), \tag{1}$$

where $R_g(\mathbf{W}_p)$ and $R_s(\mathbf{W}_p)$ represent the group- and similarity-constraint, respectively, which are denoted as:

$$R_g(\mathbf{W}_{\boldsymbol{p}}) = \lambda_1 \left\|\mathbf{W}_{\boldsymbol{p}}\right\|_{2,1} = \lambda_1 \sum_{d=1}^{P-1} \left\|\mathbf{w}_p^d\right\|_2, \tag{2}$$

$$R_s(\mathbf{W}_p) = \lambda_2 \sum_{m=1}^{M-1} \left\|\mathbf{w}_p^m - \mathbf{w}_p^{m+1}\right\|_1, \tag{3}$$

where λ_1, λ_2 are two weighting parameters, $\mathbf{w}_p^d$ is the dth row vector of $\mathbf{W}_p$, $\left\|\mathbf{w}_p^d\right\|_2$ is the sum of l_2-normalized $\mathbf{w}_p^d$, and $\left\|\mathbf{w}_p^m - \mathbf{w}_p^{m+1}\right\|_1$ penalizes the two continuous weighted vectors from the same group to minimize their diversity.

3.4. Optimization algorithm

Since our objective function includes both similarity and group regularization, we propose to use the iterative projected gradient descent algorithm to solve this problem. Specifically, the objective function is divided into similarity constraints and non-similarity constraints, which is defined in Eqs. (4) and (5):

$$s(\mathbf{W}_p) = \min_{\mathbf{W}_p} \frac{1}{2}\sum_{m=1}^{M} \|\mathbf{a}_p^m - \mathbf{A}_p^m \mathbf{w}_p^m\|_2^2, \tag{4}$$

$$m(\mathbf{W}_p) = \lambda_1 \|\mathbf{W}_p\|_{2,1} + \lambda_2 \sum_{k=1}^{M-1} \left\|\mathbf{w}_p^m - \mathbf{w}_p^{m+1}\right\|_1, \tag{5}$$

In specific projection gradient descent algorithm, we use two steps to complete our optimization process in the first step in the first n iterations. Specifically, the first step uses $g(\mathbf{W}_p)$ and γ_n to represent $g'\left(\mathbf{W}_p^n\right)$ in $\mathbf{W}_p^n$'s gradient and step length, respectively. Then the first and second steps are denoted in Eqs. (6) and (7), respectively.

$$\mathbf{V}_p^n = \mathbf{W}_p^n - \gamma_n g'\left(\mathbf{W}_p^n\right), \tag{6}$$

$$\mathbf{W}_p^{n+1} = \arg min \frac{1}{2}\|\mathbf{W}_p - \mathbf{V}_p^n\|_2^2 + m(\mathbf{W}_p). \tag{7}$$

After the optimization of the similarity constraint, we optimize the non-sparse constraint $n(\mathbf{W}_p)$ in Eq. (5) by cyclic calculation of the proximal operators related to the Lasso group and Lasso constraint [44,45]. To obtain our approximate solution faster, we use accelerated gradient descent method to further accelerate the above gradient. Specifically, our gradient descent calculation is based on $\mathbf{S}_p^n$ rather than $\mathbf{W}_p^n$.

$$\mathbf{S}_p^n = \mathbf{W}_p^n + \alpha_n\left(\mathbf{W}_p^n - \mathbf{W}_p^{n-1}\right), \tag{8}$$

$$\mathbf{V}_p^n = \mathbf{S}_p^n - \gamma_j g'\left(\mathbf{S}_p^n\right), \tag{9}$$

where α_n is a pre-defined variable and $g'(\mathbf{S}_p^n)$ and γ_p represent $g'\left(\boldsymbol{W}_{\mathrm{p}}^{\mathrm{n}}\right)$ in $\boldsymbol{W}_{\mathrm{p}}^{\mathrm{n}}$'s gradient and step length, respectively. Finally, we gain a new approximate solution via our optimization *algorithm. The specific solutions are listed in Algorithm 1.*

3.5. Feature learning by SBi-LSTM

The basic LSTM cell has the input, output gate to get and output the features after learning [46], which are defined as:

$$i_t = \sigma\left(\mathbf{Z}_{xi}x_t + \mathbf{Z}_{hi}h_{t-1} + b_i\right), \tag{10}$$

$$o_t = \sigma\left(\mathbf{Z}_{xo}x_t + \mathbf{Z}_{ho}h_{t-1} + b_o\right), \tag{11}$$

where i_t, o_t represent the input gate, output gate, respectively. $\mathbf{Z}_{xi}$ and $\mathbf{Z}_{hi}$ represent the weights parameters of input features x_t and the output features of former cell h_{t-1}, b_i and b_o are the bias of input gate and output gate, respectively. σ is the activation function. The forget gate in basic LSTM cell is also significant, which can be given as:

$$f_t = \sigma\left(\mathbf{Z}_{xf}x_t + \mathbf{Z}_{hf}h_{t-1} + b_f\right), \tag{12}$$

where $\mathbf{Z}_{xf}$, $\mathbf{Z}_{hf}$ are used to express the weight of x_t and h_{t-1}, respectively, and b_f is used to define the bias. Then the basic LSTM will use a special mechanism to select which unit to remember and forget, which are defined as:

$$g_t = \tanh\left(\mathbf{Z}_{\boldsymbol{xc}}x_t + \mathbf{Z}_{\mathrm{hc}}h_{t-1} + b_c\right), \tag{13}$$

$$c_t = i_t \odot g_t + f_t \odot c_{t-1}, \tag{14}$$

where g_t is the current state of neurons, c_t is a long-term memory unit controlled by i_t, g_t, f_t and c_{t-1}, and c_{t-1} is the parameter of the last unit, when the parameter is bigger than c_{t-1}, the former unit will be remembered, otherwise, it will be forgotten. $\mathbf{Z}_{xc}$ is the weight of input data x_t, and $\mathbf{Z}_{hc}$ is the weight of the last output neuron h_{t-1}, $\odot$ is the element-wise multiplication term. After interpreting the basic LSTM cell, we introduce the bidirectional LSTM, and D(■) is used to define a LSTM cell. The forward and backward structure are both needed in the bidirectional LSTM's feature learning. Therefore, $\vec{f}$ and $\overleftarrow{b}$ are used to define the forward and backward LSTM, respectively, which are given as follow:

$$\vec{f}_t = D\left(\mathbf{Z}_{x\vec{f}}x_t + \mathbf{Z}_{d\vec{f}}\overrightarrow{f_{t-1}} + \vec{b}_f\right), \tag{15}$$

$$\overleftarrow{b}_t = D\left(\mathbf{Z}_{\underset{x\,b}{\leftarrow}}x_t + \mathbf{Z}_{\underset{d\,b}{\leftarrow}}\overleftarrow{h_{t-1}} + \overleftarrow{b}_d\right), \tag{16}$$

where $\mathbf{Z}_{x\vec{f}}$ and $\mathbf{Z}_{d\vec{f}}$ are the weights of the forward LSTM, $\mathbf{Z}_{\underset{x\,b}{\leftarrow}}$ and $\mathrm{Z}_{\underset{d\,b}{\leftarrow}}$ are the weights of the backward LSTM, $\vec{b}_f$ and $\overleftarrow{b}_d$ are the biases of the forward and backward LSTM layers, respectively. Then the final output of each direction is connected as the final output.

$$y_t = D\left(\mathbf{Z}_{\vec{f}y}\vec{f}_t + \mathbf{Z}_{\underset{b\,y}{\leftarrow}}\overleftarrow{b}_t + b_y\right), \tag{17}$$

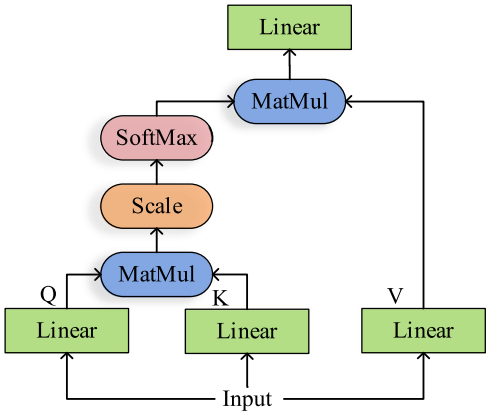


Fig. 3. The interpretation of self-attention model.

Algorithm 1: The optimization algorithm of similarity term and non-similarity term

Input: $\boldsymbol{a}_p^m \in \mathbb{R}^{E\times 1}$, $\mathbf{A}_p^k \in \mathbb{R}^{E\times(P-1)}$, λ_1, λ_2
Output: $\mathbf{W}_p \in \mathbb{R}^{(P-1)\times M}$
Initialize $n = 0$, $\mathbf{S}_p^n$, $\mathbf{W}_p^n$
While not converged
  Repeat
    update $\mathbf{V}_p^n$ in Eq. (6)
    update $\mathbf{S}_p^n$ in Eq. (8)
    update $\mathbf{W}_p^n$ in Eq. (7)
    $n = n + 1$
  End
Until convergence or stop when criterion is met
End
Return $\mathbf{W}_p$

Finally, $\mathbf{Z}_{\vec{f}y}$ and $\mathbf{Z}_{\underset{b}{\leftarrow}y}$ are the weights of the forward and backward hidden state weights, respectively.

3.6. Self-attention based SBi-LSTM

Although SBi-LSTM has shown good results for disease detection [27], it is still feasible to be improved via attention. Therefore, we explore a novel self-attention mechanism that can capture the most representative part of the disease in response to a given aspect, which is demonstrated in Fig. 3.

Specifically, we get the output feature y_t from the previous section, and then we take the output feature y_t as input, and create a query vector ($\mathrm{Q} = y_t \mathrm{w_Q}$), a value vector ($\mathrm{V} = y_t \mathrm{w_V}$), a key vector ($\mathrm{K} = y_t \mathrm{w_K}$) based on the input features, and then we score the input through dot-product of Q and V. Next, we divide the score by $8\sqrt{d_k}$, so that the output is normalized by the Softmax function, and the last column list adds up to 1. The next step is to multiply each value vector by Softmax. Finally, we get the cumulative weighted vector via combining them together. This produces output from the constrained attention layer at this location. The self-attention can be expressed as:

$$\text{Attention (Q, K, V)} = \text{Softmax}\left(\frac{\mathrm{QK^T}}{8\sqrt{d_k}}\right)\mathrm{V}, \tag{18}$$

where Q, K, V are the query vector, key vector and value vector, respectively.

4. Experimental results

4.1. Experimental setup

In this paper, two public datasets, ADNI-2 and ADNI-3 are used in the experiments. The details of the dataset are shown in Table 1. All subjects have two time-point (baseline and year1) rs-fMRI data with different scan parameters. In detail, the length of time series is 140 and 197 in ADNI-2 and ADNI-3, respectively. The subjects close their eyes in ADNI-2 while scan but open their eyes in ADNI-3. The SLEPT toolbox is used for constructing our BFCN. The Keras library with Tensorflow as backend is used to extract features for detection. All the detection tasks have three different scenarios: baseline, year1, and their fusion. We can learn the relationship across different time points and study them longitudinally. We use two constrained parameters in our SGN to get the feature maps from multi-time points data, group-constrained and similarity-constrained parameters, both are set from 2^{-5} to 2^5. For the SBi-LSTM, we choose two-layer Bi-LSTM, and the first layer has 128 cells while the second layer has 32 cells.

In the self-attention mechanism, we set d_k to 32. In order to facilitate training, we add a fully-connected (FC) layer and softmax as the classifier. We choose the RMS optimizer [47] for optimization and categorical cross entropy as the loss function. In addition, the batch size is 20, epoch numbers are 200, the learning rate is 10^{-3}. To avoid the overfitting, the fuzz factor, rho, and the learning rate decay are 10^{-8}, 0.9, and 10^{-5}, respectively.

4.2. Classification performance of our method

To evaluate the diagnosis performance of our model, four quantitative metrics are used, namely accuracy (ACC), sensitivity (SEN), specificity (SPEC), area under receiver operating characteristic (ROC) curve (AUC). We also draw the ROC curves of the ADNI-2 and ADNI-3 datasets combine together in Fig. 4. Since the PC based BFCN get significant lower performance, we remove its ROC curves in the figures. To represent the excellent performance of our proposed method, we utilize the radar chart in Fig. 5 to show the different performances among five different methods.

Due to the small dataset, the leave one out cross validation (LOOCV) is used in the experiments. Supposing we have S data in total, we use $S-1$ subjects for training and the left one for testing. Then we loop the procedure for S times, and each subject is tested. The regularization parameters with the best performance are selected as the optimal parameters. $S-1$ classifiers are used to classify the completely unknown test subject. A majority voting strategy is used for the final classification decision. Each subject in the dataset will be picked out for test and the process is repeated S times, which computes the overall cross-validation classification results [48]. To verify the efficacy of our model, we set up the experiments of EMCI vs. NC. Also, we add the LMCI vs. NC, and EMCI vs. LMCI tasks to verify the effectiveness of our method.

Tables 2 and 3 show the experiment results of ADNI-2 and ADNI-3 datasets. From the above two tables, we can verify the effectiveness of the proposed method. First, we compare five methods including SBi-LSTM, LSTM, SVM and MLP to verify the effectiveness of SSBi-LSTM, which are verified on ADNI-2 and ADNI-3 datasets, we can find the proposed method always achieves the best performance. In addition, we also verify the improvement of multi-time data from the fusion of baseline and year1. From the detection results of the EMCI vs. NC classification on ADNI-2, we can find that year1 has made some improvement compared with baseline, which are 86.11% and 84.53%, while Fusion gets the performance of 87.50%, which achieved an improvement over

Table 1
Demographic information of the subjects used in our study.

Time	ADNI-2 dataset			ADNI-3 dataset		
	NC	EMCI	LMCI	NC	EMCI	LMCI
Baseline	33(15/18)	39(18/21)	30(19/11)	16(8/8)	25(13/12)	10(6/4)
Baseline	74.8 ± 6.43	71.0 ± 6.63	71.6 ± 8.34	76.4 ± 7.92	75.6 ± 6.96	76.9 ± 8.41
Year1	33(15/18)	39(18/21)	30(19/11)	16(8/8)	25(13/12)	10(6/4)
Year1	75.9 ± 6.40	72.0 ± 6.60	72.6 ± 8.34	77.2 ± 8.10	77.3 ± 7.07	77.5 ± 8.46

Table 2
Algorithm comparison of different scenarios on ADNI-2 dataset.

DATA	BFCN	Method	LMCI vs. NC				EMCI vs. LMCI				EMCI vs. NC			
			ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
Baseline	PC	SVM	50.52	50.87	48.56	**52.53**	50.78	50.85	48.67	33.56	50.00	50.08	56.41	42.42
		MLP	50.80	50.67	49.40	52.28	50.78	50.58	56.41	26.67	50.61	50.69	46.15	45.45
		SBi-LSTM	51.24	52.08	**56.67**	46.28	52.05	52.14	56.41	**46.67**	51.60	**52.31**	52.40	**46.82**
		SSBI-LSTM	**51.90**	**52.49**	51.51	50.00	**52.61**	**52.94**	**58.97**	**46.67**	**52.08**	51.37	**58.97**	44.90
	GCS	SVM	66.67	71.62	64.93	68.55	66.67	71.25	71.79	60.00	62.50	63.25	64.10	58.97
		MLP	69.84	75.56	66.67	69.70	66.67	72.39	66.73	**76.67**	63.88	65.97	51.28	**78.79**
		SBi-LSTM	**74.60**	**81.62**	81.82	66.67	71.04	75.38	67.61	**76.67**	65.28	71.72	66.67	69.70
		SSBI-LSTM	72.14	80.10	**83.33**	**72.74**	**72.46**	**75.47**	**74.36**	73.38	**68.05**	**76.61**	**69.23**	69.70
	SGN	SVM	75.36	78.25	73.92	81.44	75.36	82.22	74.36	76.67	75.00	80.81	85.74	67.46
		MLP	79.36	78.28	**83.33**	75.76	79.71	82.74	87.18	70.00	75.00	86.48	**89.74**	57.58
		SBi-LSTM	80.95	93.33	70.00	**90.91**	85.50	90.26	**94.87**	**83.33**	82.22	87.18	75.27	82.90
		SSBI-LSTM	**87.30**	**93.54**	**83.33**	87.88	**86.95**	**93.08**	89.74	**83.33**	**84.53**	**91.45**	82.05	**84.85**
Year1	PC	SVM	50.80	50.84	47.78	**54.55**	50.78	47.11	56.41	26.67	51.38	50.65	46.15	**57.58**
		MLP	50.80	51.07	40.00	**54.55**	51.42	47.26	53.85	33.33	51.22	50.89	**64.10**	27.27
		SBi-LSTM	52.50	52.73	50.00	**54.55**	52.17	50.27	**58.97**	43.33	51.96	51.79	58.97	45.09
		SSBI-LSTM	**53.33**	**55.56**	**60.00**	45.45	**53.62**	**51.79**	56.41	**50.00**	**52.08**	**52.07**	56.41	45.45
	GCS	SVM	68.25	73.64	66.67	69.70	66.67	71.54	74.36	60.00	62.50	66.50	69.23	54.55
		MLP	71.42	75.56	60.93	52.83	68.11	70.16	76.92	63.33	63.88	66.67	71.79	54.55
		SBi-LSTM	74.60	81.61	**73.33**	75.76	72.46	77.52	**79.49**	**66.67**	66.67	71.72	**74.36**	57.58
		SSBI-LSTM	**76.19**	**86.33**	66.67	**78.79**	**73.91**	**78.38**	**79.49**	**66.67**	**68.05**	**74.51**	69.23	**66.67**
	SGN	SVM	77.78	75.85	72.92	81.44	76.81	85.47	76.92	76.67	75.00	80.34	76.92	72.72
		MLP	77.78	79.49	80.00	76.67	79.71	88.97	74.36	**86.67**	76.38	89.51	**89.74**	60.61
		SBi-LSTM	85.71	90.02	**90.00**	81.81	85.50	90.26	87.18	83.33	**86.11**	91.92	**89.74**	81.82
		SSBI-LSTM	**90.47**	**93.74**	**90.00**	**91.84**	**89.85**	**93.47**	**94.87**	83.33	**86.11**	**92.35**	87.18	**84.85**
Fusion	PC	SVM	51.28	51.24	50.00	53.54	51.44	51.98	50.22	43.74	51.38	51.48	**64.10**	27.27
		MLP	51.28	51.41	49.40	54.55	51.95	52.68	56.41	46.38	51.61	52.65	53.85	45.45
		SBi-LSTM	55.56	53.54	53.33	57.58	54.04	54.55	58.97	43.33	53.47	53.15	56.41	48.48
		SSBI-LSTM	**57.66**	**59.70**	**56.67**	**63.64**	**57.61**	**55.56**	**61.54**	**53.33**	**55.56**	**56.41**	**64.10**	**49.19**
	GCS	SVM	69.84	73.64	70.00	67.70	68.11	73.42	74.36	60.00	63.88	69.85	56.41	72.72
		MLP	71.42	81.31	60.93	58.55	69.56	76.75	71.79	60.00	65.28	72.56	51.28	**78.79**
		SBi-LSTM	76.19	83.13	66.67	84.85	73.91	79.23	79.49	**66.67**	69.44	77.23	69.23	60.61
		SSBI-LSTM	**77.78**	**90.56**	**72.50**	**91.84**	**75.46**	**81.45**	**84.62**	63.33	**72.22**	**83.99**	**79.49**	63.64
	SGN	SVM	79.36	81.31	**90.00**	57.58	79.71	88.97	74.36	86.67	76.38	84.85	79.49	75.75
		MLP	80.95	89.90	73.33	87.88	84.05	91.79	87.18	80.00	79.17	86.40	79.49	78.79
		SBi-LSTM	87.30	93.54	**90.00**	84.84	89.85	94.02	83.33	**97.44**	86.11	90.37	**87.18**	84.85
		SSBI-LSTM	**92.06**	**94.75**	86.67	**96.97**	**91.30**	**95.47**	**94.87**	86.67	**87.50**	**92.61**	**87.18**	**87.88**

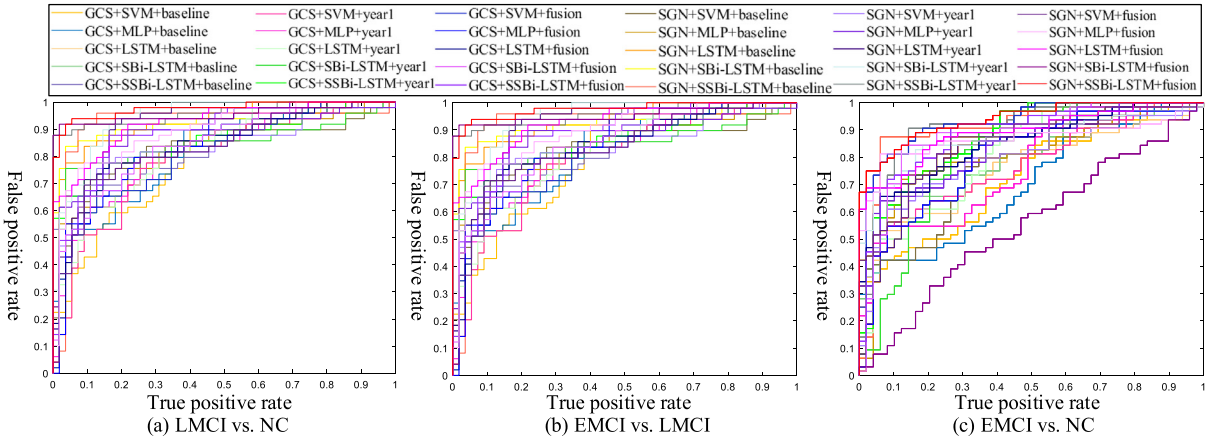
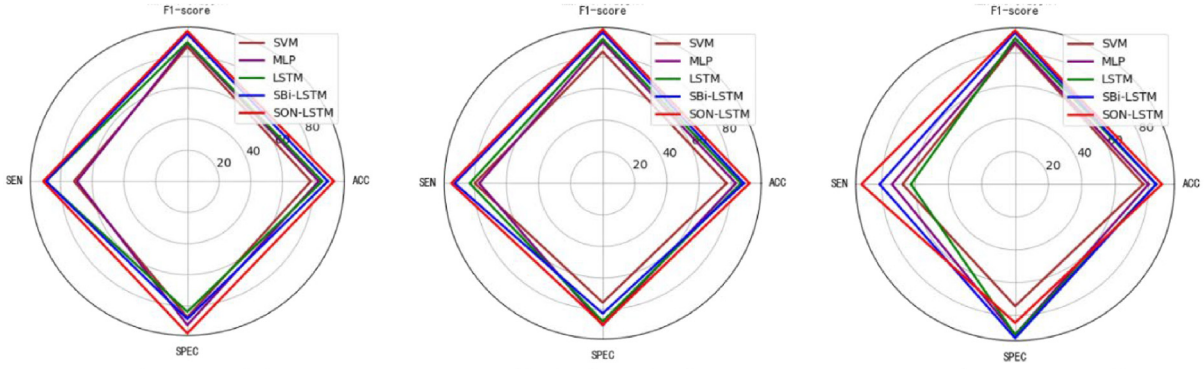


Fig. 4. The ROC curves of different methods on the ADNI-2&3dataset.

Table 3
Algorithm comparison of different scenarios on ADNI-3 dataset.

Data	BFCN	Method	LMCI vs. NC				EMCI vs. LMCI				EMCI vs. NC			
			ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
Baseline	PC	SVM	51.11	51.11	53.33	47.22	50.85	51.08	51.02	50.23	50.90	50.25	47.62	53.89
		MLP	50.85	52.32	52.00	48.48	50.85	51.28	50.33	52.94	51.22	52.74	**53.27**	48.26
		SBi-LSTM	**57.69**	55.56	**58.77**	56.73	51.42	53.20	55.56	50.69	52.34	53.24	52.00	52.86
		SSBi-LSTM	**57.69**	**58.36**	56.58	**59.21**	**53.71**	**56.40**	**60.00**	**54.00**	**53.25**	**58.74**	52.62	**55.35**
	GCS	SVM	66.67	71.62	64.93	68.55	66.67	71.25	71.79	60.00	62.50	63.25	64.10	58.97
		MLP	69.84	75.56	66.67	69.70	65.85	64.20	71.79	50.00	63.88	65.97	51.28	**78.79**
		SBi-LSTM	**74.60**	**81.62**	81.82	66.67	71.04	75.38	67.61	**76.67**	70.73	71.25	**76.00**	62.50
		SSBi-LSTM	72.14	80.10	**83.33**	**72.74**	**72.46**	**75.47**	**74.36**	73.38	**73.17**	**74.50**	72.00	75.00
	SGN	SVM	70.73	73.85	33.33	41.11	74.49	41.11	73.85	75.33	67.22	70.00	68.48	63.48
		MLP	76.53	81.25	78.75	70.00	78.55	84.67	71.25	**81.79**	68.11	71.54	76.92	63.33
		SBi-LSTM	84.61	86.88	81.25	**90.00**	**82.86**	**90.40**	**84.00**	80.00	78.04	84.00	80.00	75.00
		SSBi-LSTM	**88.46**	**87.50**	**87.50**	80.00	**82.86**	87.20	**84.00**	80.00	**82.93**	**85.32**	**84.00**	**87.50**
Year1	PC	SVM	51.68	51.27	49.87	53.33	51.57	51.65	52.60	50.68	51.78	50.79	52.38	48.78
		MLP	51.68	54.95	48.33	54.55	51.22	52.24	48.89	**56.27**	52.34	52.23	52.38	49.00
		SBi-LSTM	**57.69**	60.00	56.00	54.17	**54.29**	53.80	56.60	52.37	53.22	56.81	58.00	49.30
		SSBi-LSTM	**57.69**	**61.00**	**58.46**	**56.00**	**54.29**	**55.40**	**60.00**	50.00	**56.09**	**59.62**	**60.68**	**52.08**
	GCS	SVM	68.25	73.64	66.67	69.70	66.67	71.54	74.36	60.00	62.50	66.50	69.23	54.55
		MLP	71.42	75.56	60.93	52.83	68.57	67.20	74.50	64.00	63.88	66.67	71.79	54.55
		SBi-LSTM	74.60	81.61	**73.33**	75.76	72.46	77.52	**79.49**	**66.67**	70.73	74.25	72.00	**68.75**
		SSBi-LSTM	**76.19**	**86.33**	66.67	**78.79**	**73.91**	**78.38**	**79.49**	**66.67**	**73.17**	**78.75**	**76.00**	**68.75**
	SGN	SVM	71.11	73.85	73.33	67.22	76.00	78.48	73.48	**80.00**	70.68	80.00	78.48	63.33
		MLP	79.23	74.35	68.75	70.00	78.55	80.00	84.00	60.00	70.73	80.75	68.00	**75.00**
		SBi-LSTM	84.61	88.12	**87.50**	**80.00**	85.71	**92.80**	88.00	**80.00**	82.92	86.38	88.00	**75.00**
		SSBi-LSTM	**88.46**	**89.38**	81.25	**80.00**	**88.57**	91.20	**92.00**	**80.00**	**85.36**	**88.25**	**92.00**	**75.00**
Fusion	PC	SVM	51.49	52.23	53.85	49.33	52.71	53.00	54.00	50.00	51.78	50.69	56.25	48.27
		MLP	52.85	53.65	50.22	53.74	52.71	53.28	52.27	52.96	52.34	53.19	57.14	49.21
		SBi-LSTM	57.69	58.75	55.67	58.33	57.14	60.80	60.00	53.56	56.09	58.75	64.00	52.00
		SSBi-LSTM	**61.53**	**67.32**	**57.50**	**68.75**	**60.00**	**62.90**	**64.00**	**60.00**	**60.97**	**60.25**	**68.00**	**55.09**
	GCS	SVM	69.84	73.64	70.00	67.70	68.11	73.42	74.36	60.00	63.88	69.85	56.41	72.72
		MLP	71.42	81.31	60.93	58.55	71.42	80.00	71.79	**70.00**	65.28	72.56	51.28	**78.79**
		SBi-LSTM	76.19	83.13	66.67	84.85	73.91	79.23	79.49	66.67	73.17	79.50	**80.00**	62.50
		SSBi-LSTM	**77.78**	**90.56**	**72.50**	**91.84**	**75.46**	**81.45**	**84.62**	63.33	**75.60**	**80.00**	**80.00**	68.75
	SGN	SVM	80.73	80.75	80.00	75.00	72.50	91.84	75.46	81.45	70.62	83.33	72.22	73.99
		MLP	83.07	83.75	81.25	80.00	71.43	75.20	76.00	60.00	73.17	77.00	**84.00**	56.25
		SBi-LSTM	88.46	87.35	87.50	**90.00**	88.57	93.60	88.00	**90.00**	85.36	95.50	**84.00**	**87.50**
		SSBi-LSTM	**92.30**	**92.30**	**93.75**	**90.00**	**91.42**	**96.40**	**96.00**	80.00	**87.80**	**93.25**	**84.00**	**87.50**



(a) LMCI vs. NC (b) EMCI vs. LMCI (c) EMCI vs. NC

Fig. 5. The ROC curves of different methods on the ADNI-2&3dataset.

baseline and year1. The same findings can be found for the performance of the other two detection tasks of LMCI vs. NC and EMCI vs. LMCI. Then, we also demonstrate the good performance of our longitudinal study on ADNI-3, which can be seen in Table 3. In addition, we also show the results on the combination of ADNI-2 and ADNI-3 (hereinafter to be referred as ADNI-2&3 dataset) dataset in Table 4, to show the effectiveness of the proposed method. Compared with the above two tables, the results on the combined dataset ADNI-2&3 decrease relatively, comparing to the results on the two datasets individually, which are from 87.50% to 86.11%, 91.30% to 89.42% and 92.06% to 91.42% in the three detection tasks of EMCI vs. NC, EMCI vs. LMCI and LMCI vs. NC, respectively. However, this can be expected because of the differences existed in the scanning time series length and scanning mode between the two datasets.

5. Discussion

5.1. Influence of the SBi-LSTM depth

To evaluate the influence on the depth of SBi-LSTM, we test our model with varied number of layers while fixing the other

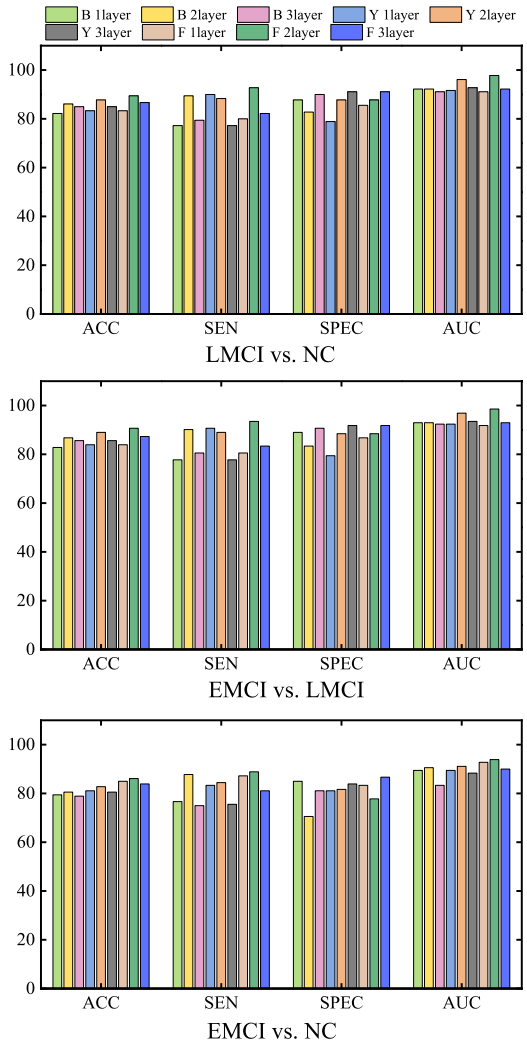


Fig. 6. The influence of LSTM's layer.

parameters. Namely, we test the following models: baseline of 1 layer SBi-LSTM (B 1layer), year1 of 1 layer SBi-LSTM (Y 1layer), fusion of 1 layer SBi-LSTM (F 1layer), baseline of 2 layer SBi-LSTM (B 2layer), year1 of 2 layer SBi-LSTM (Y 2layer), fusion of 2 layer SBi-LSTM (F 2layer), and baseline of 3 layer SBi-LSTM (B 3layer), year1 of 3 layer SBi-LSTM (Y 3layer), fusion of 3 layer SBi-LSTM (F 3layer), we use the same SGN output feature maps to test these configurations. Then we analyze the significance of the obtained results and compare them in Fig. 6. The final results show that we get better results when the SBi-LSTM has only 2 layers. The results are consistent with Wang et al. [24] since the LSTM's layer can improve the detection results. Only one layer is too shallow to discover representative information for detection, while too many layers may omit discriminative features during feature learning.

5.2. Influence of the similarity and sparse constraint

Regarding the BFCN, we study the influence of the sparse and similarity constraint. Specially, both parameters are set from 2^{-5} to 2^{5}. Then we compare different methods with all the possible combinations of the two parameters in Fig. 6. Instead of applying LOOCV that takes too much time, we choose 10-fold cross-validation to perform the comparison. From Fig. 7, we can find that the optimal similarity and sparsity parameter combination is $(2^{-4}, 2^{-3})$ for the task of EMCI vs. NC classification, while achieving the best accuracy of 86.11%.

The optimal parameters for classification on EMCI vs. NC and EMCI vs. LMCI are $(2^{-3}, 2^{-3})$ and $(2^{-3}, 2^{-4})$, while the corresponding best results are 91.42% and 89.42%. We can conclude that: (1) When the similarity constraint and sparse constraint parameters are larger than 1, the accuracy does not improve greatly; (2) The accuracy improves greatly when the constrained parameters are smaller than 1. The experiment results also show that the pattern of decline can already be seen even the high detection results are achieved. Specifically, the similarity parameter is set to 0 in the GCS method.

5.3. Influence of the batch-size

We conduct experiments on different datasets and batch-sizes to understand their impact on our method. As can be seen from Table 5, for the same dataset, with the increase of batch-size, the accuracy rate tends to increase, but it will fluctuate up and down. The results on the ADNI-3 dataset are slightly worse than those on ADNI-2, which should be due to the slightly larger size of the ADNI-2 dataset.

5.4. Visualization analysis

We visualize the t-SNE features of the final outputs of LSTM, SBi-LSTM, and SSBi-LSTM in Fig. 8, which demonstrates that the proposed method can learn more discriminative features than the other two methods, especially for the classification of EMCI and NC. It is well-known that the brain is the central hub responsible for the daily activities, and different brain regions correspond to different tasks, not all of them are closely associated with MCI. Therefore, we use our proposed approach to detect these associated ROIs to find abnormal brain regions. Specifically, we use a brain region that blocks a network of functional connections to systematically eliminate other ROIs. We put the BFCN to zeros associated with these ROIs if these brain region connections to systematically eliminate other ROIs. We put the BFCN to zeros associated with these ROIs if these brain region connections to systematically eliminate other ROIs. We put the BFCN to zeros associated with these ROIs, and if these brain regions have significant influence on area that is related to EMCI, the correct class prediction will significantly drop. We use the whole brain as a control group. We then shield 90 brain regions in turn and get 90 classification results after shielding. We also choose the top 10 brain regions, which have the greatest impacts on the diagnosis of LMCI and EMCI, respectively. Fig. 9 shows the sagittal plane visualization of the most important brain region we have selected. Also, we draw the BFCN of the top ten ROIs in Fig. 10.

From the above results, we conclude that the top ten ROIs with the greatest impact on the detections of EMCI are Lingual gyrus left, Temporal pole, middle temporal gyrus left, Inferior frontal gyrus, triangular part left, Middle occipital gyrus right, Middle frontal gyrus left, Angular gyrus, Thalamus left, Middle frontal gyrus left, Angular gyrus right, Superior occipital gyrus right, and Posterior cingulate gyrus left. The top ten ROIs with the greatest impact on the detections of LMCI are Gyrus rectus left, Postcentral gyrus right, Superior temporal gyrus left, Gyrus rectus right, Orbital part, Gyrus rectus, Middle frontal gyrus, Orbital part right, Posterior cingulate gyrus left, Posterior cingulate gyrus right, Superior temporal gyrus right, Angular gyrus left, and Middle occipital gyrus right. These brain regions have been proved to have high correlations with EMCI and LMCI [4]. In addition, we also illustrate the top ten brain region connectivity in Fig. 11.

Table 4
Algorithm comparison of different scenarios on ADNI-2&3 dataset.

DATA	BFCN	Method	LMCI vs. NC				EMCI vs. LMCI				EMCI vs. NC			
			ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
Baseline	PC	SVM	51.68	51.92	49.17	**51.70**	51.92	51.37	**55.76**	48.61	50.44	50.69	57.81	46.26
		MLP	51.42	52.14	55.00	46.94	52.14	50.98	50.20	45.85	51.39	**53.99**	62.50	38.78
		SBi-LSTM	**54.02**	54.03	**63.27**	47.50	52.09	53.68	46.94	56.36	54.77	52.87	55.29	**53.08**
		SSBi-LSTM	53.87	**57.76**	55.00	51.65	**52.81**	**54.77**	42.68	**61.82**	**55.45**	**53.99**	**64.06**	50.77
	SGC	SVM	70.00	72.70	71.79	69.39	70.19	78.59	62.25	78.79	70.83	72.26	60.93	58.55
		MLP	71.42	75.56	62.50	77.51	72.57	81.37	69.39	**80.00**	70.83	69.93	67.69	**73.47**
		SBi-LSTM	**74.28**	**82.19**	68.77	**77.69**	**72.85**	81.25	**77.65**	67.64	71.16	80.00	69.23	**73.47**
		SSBi-LSTM	**74.28**	80.10	**75.00**	74.76	**72.85**	**84.90**	73.74	70.67	**72.91**	**80.13**	**77.50**	55.83
	SGN	SVM	75.00	79.14	77.55	72.50	75.00	82.56	73.74	80.68	73.93	77.74	67.19	81.63
		MLP	75.00	81.38	79.59	70.00	77.85	82.75	80.00	73.00	75.00	80.36	67.19	**83.67**
		SBi-LSTM	85.17	90.41	89.13	83.72	84.61	**92.50**	85.00	**84.37**	80.56	84.06	85.71	75.00
		SSBi-LSTM	**87.14**	**94.54**	**89.13**	**90.70**	**85.71**	91.95	**89.06**	82.50	**82.63**	**90.69**	**87.50**	70.83
Year1	PC	SVM	51.68	52.50	47.50	55.10	51.92	53.68	53.94	46.94	51.32	52.65	53.67	48.69
		MLP	52.85	53.15	57.50	47.06	52.62	50.71	**58.37**	42.73	51.68	54.38	54.67	47.96
		SBi-LSTM	55.13	54.49	50.00	**59.18**	**54.90**	53.87	44.90	**63.64**	55.75	56.92	**62.50**	53.09
		SSBi-LSTM	**56.38**	**54.85**	**62.54**	52.00	**54.90**	**57.66**	53.06	56.36	**56.33**	**57.08**	59.38	**54.75**
	SGC	SVM	71.28	76.86	62.50	78.79	71.15	79.11	60.93	58.55	71.16	76.37	66.67	**75.00**
		MLP	72.14	76.73	65.00	78.79	73.57	81.41	70.76	77.55	71.92	75.98	74.36	67.33
		SBi-LSTM	75.71	84.23	71.55	79.67	75.43	84.23	**76.92**	71.43	75.43	80.13	76.92	71.43
		SSBi-LSTM	**78.68**	**86.33**	**77.35**	**80.67**	**78.57**	**86.09**	76.72	**81.25**	**76.31**	**86.88**	**78.13**	57.14
	SGN	SVM	78.46	82.75	75.43	83.33	75.42	83.38	70.42	80.68	74.89	80.55	70.31	80.67
		MLP	79.28	80.15	69.38	**88.98**	77.85	87.12	71.43	87.27	77.14	82.94	79.69	79.59
		SBi-LSTM	86.07	**94.44**	85.68	87.50	86.42	94.24	**89.06**	85.00	81.57	87.40	76.92	**87.75**
		SSBi-LSTM	**89.28**	93.42	**91.30**	88.37	**87.85**	**95.96**	87.96	**87.50**	**84.53**	**90.91**	**84.38**	81.63
Fusion	PC	SVM	52.80	53.99	46.67	54.42	52.88	52.78	49.66	58.82	52.21	53.62	**60.94**	47.62
		MLP	52.85	54.08	**59.18**	40.00	53.00	52.21	48.24	55.64	52.08	52.07	56.41	45.45
		SBi-LSTM	55.13	54.49	50.00	59.18	**56.36**	57.22	51.02	**60.00**	57.72	58.07	58.65	**57.14**
		SSBi-LSTM	**60.91**	**58.01**	55.00	**65.39**	**56.36**	**58.56**	**61.82**	51.02	**58.33**	**59.28**	59.38	**57.14**
	SGC	SVM	72.14	78.52	**77.55**	69.70	72.31	81.15	66.73	78.61	72.91	81.95	68.25	78.79
		MLP	74.28	83.88	70.00	74.68	75.00	84.27	69.23	75.00	72.91	80.01	76.56	71.43
		SBi-LSTM	77.86	88.11	72.50	81.63	76.83	84.60	69.23	81.63	**76.83**	88.17	69.23	**81.63**
		SSBi-LSTM	**80.85**	**90.50**	74.40	**87.32**	**79.16**	**89.09**	**75.00**	**87.76**	**76.83**	**88.62**	**78.12**	63.27
	SGN	SVM	77.86	86.42	71.43	87.27	77.14	83.42	79.69	75.59	76.83	85.71	67.50	74.37
		MLP	82.85	88.57	69.57	92.30	82.85	89.53	76.92	**90.00**	80.00	86.73	73.88	91.83
		SBi-LSTM	88.57	94.69	**89.13**	88.37	87.85	95.70	**92.19**	82.50	84.53	92.64	81.25	**93.88**
		SSBi-LSTM	**91.42**	**96.43**	86.96	**97.67**	**89.42**	**97.59**	**92.19**	87.50	**86.11**	**93.90**	**89.06**	77.55

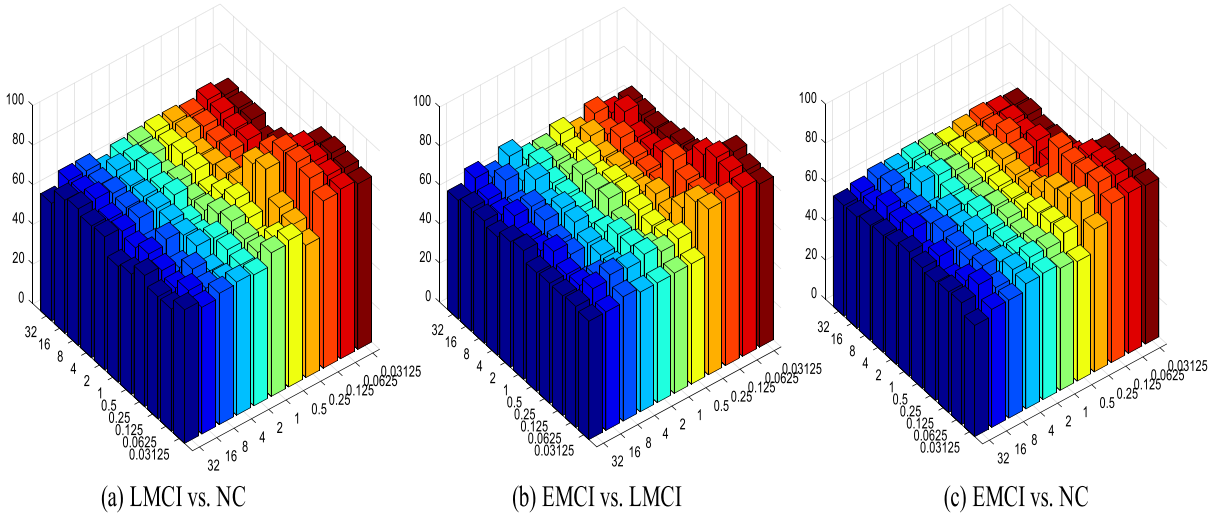


Fig. 7. The influence of sparse and similarity constrained parameters in ADNI-2&3 dataset.

5.5. Comparison with related works

Apart from comparing the proposed model with different parameters, we also compare it with several state-of-the-art methods, and we report the results in Table 6. Specifically, we present the number of subjects, method of BFCN construction, classifier and obtained results, where S denotes single time-point and M denotes multiple time-points. We can draw a conclusion that our model achieves the best result in terms of all metrics and the longitudinal data can improve the detect results of EMCI. With the better prediction performance, our proposed method also provides an effective way to detect other brain diseases through BFCN construction and longitudinal analysis.

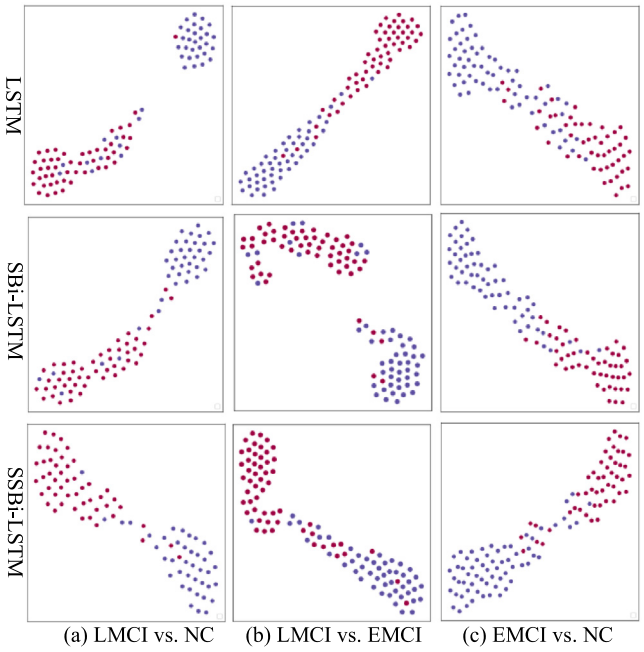


Fig. 8. The t-SNE visualization the last layer in our model of different detection tasks.

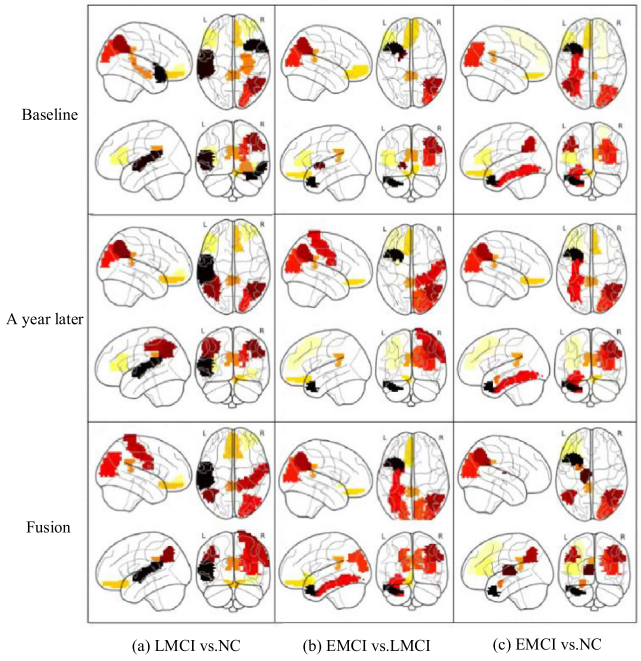


Fig. 9. The top 10 brain regions of different classification tasks.

6. Conclusion

In this paper, we propose a novel deep learning model to detect EMCI from longitudinal data, which comprises a SGN for effective BFCN construction, a SBi-LSTM network for discriminative feature learning and longitudinal analysis, and a self-attention module for high-level feature extraction from two time-points. We test the model on two public datasets individually and jointly. The experimental results show that our model outperforms several state-of-the-art methods. However, several issues need to be improved in our future study. First, only LSTM is applied to fuse the longitudinal data in our current study, other deep fusion methods will be introduced to further improve the overall performance. Second, the experiments are conducted on two datasets independently and jointly. However, a comprehensive multi-center study with more data varieties is needed to verify the robustness of the proposed method, in which the data from each center is collected with different scan parameters. Third, the datasets used in the experiments are relatively small, larger datasets should be introduced to validate the generalization ability of the proposed method.

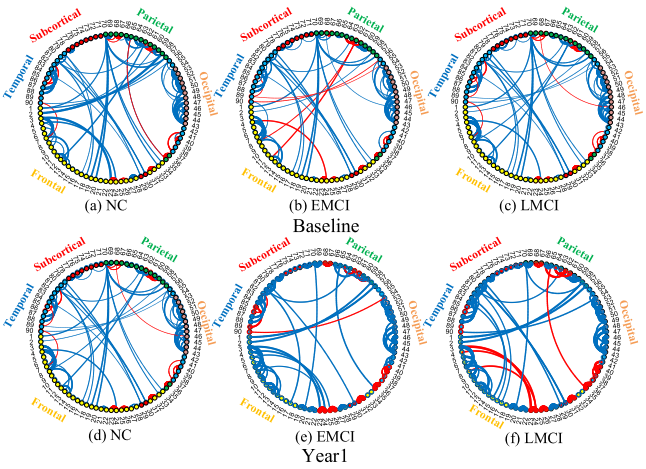


Fig. 10. Key brain connective change over different time points.

Table 5
Performance comparison of different batch-sizes (%).

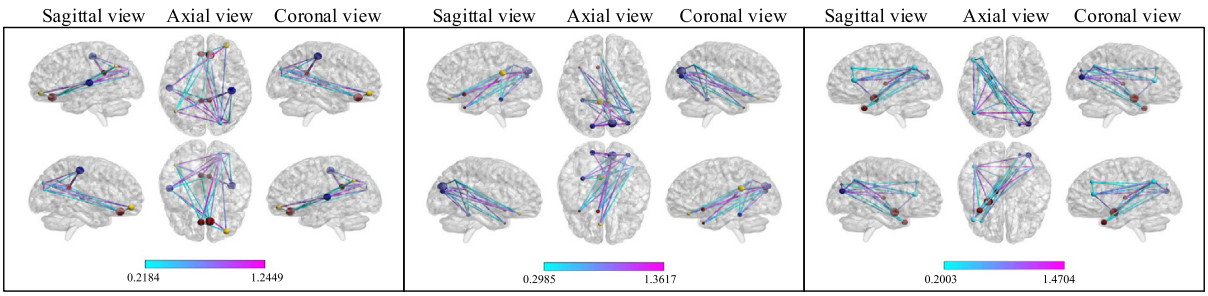
Dataset	Dataset size	Batch-size	ACC	AUC	SEN	SPE
ADNI-2	102	2	86.3	92.2	83.03	86.17
ADNI-2	102	4	88.42	92.45	84.31	88.24
ADNI-2	102	8	87.11	92.81	85.89	89.38
ADNI-2	102	16	90.86	93.06	88.56	90.72
ADNI-2	102	32	89.28	93.37	90.22	92.84
ADNI-2	102	64	91.05	93.94	89.73	95.03
ADNI-2	102	128	92.18	94.75	89.16	93.72
ADNI-3	51	2	86.11	88.61	85.48	86.31
ADNI-3	51	4	87.24	89.03	86.26	88.28
ADNI-3	51	8	87.56	89.74	88.3	89.64
ADNI-3	51	16	88.67	90.18	89.71	90.63
ADNI-3	51	32	90.48	90.84	90.32	91.49
ADNI-3	51	64	89.96	91.15	93.75	92.75
ADNI-3	51	128	91.35	92.3	91.03	91.62

Table 6
Performance comparison of different methods under various scenarios (%).

Method	Subject	Time	BFCN	Classifier	Acc
Guo et al. [18]	33 EMCI+28 NC	S	PC	SVM	72.10
Guo et al. [18]	32 LMCI+28 NC	S	PC	SVM	78.63
Wee et al. [14]	29 EMCI+30 NC	S	GCS	SVM	79.66
Yang et al. [5]	18 LMCI+29 NC	M	FSN	SVM	87.23
Yang et al. [5]	29 EMCI+18 LMCI	M	FSN	SVM	80.35
Yang et al. [5]	29 EMCI+29 NC	M	FSN	SVM	82.76
Proposed	40 LMCI+49 NC	M	SGN	softmax	91.42
Proposed	64 EMCI+40 LMCI	M	SGN	softmax	89.42
Proposed	64 EMCI+49 NC	M	SGN	softmax	86.11

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



(a) LMCI vs.NC (b) EMCI vs.LMCI (c) EMCI vs.NC

Fig. 11. The top 10 brain regions and the corresponding connective network of different classification tasks.

References

[1] B. Lei, N. Cheng, A.F. Frangi, E.-L. Tan, J. Cao, P. Yang, A. Elazab, J. Du, Y. Xu, T. Wang, Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease, Med. Image Anal. 61 (2020) 101652.

[2] E. Eruysal, L. Ravdin, H. Kamel, C. Iadecola, M. Ishii, Plasma lipocalin-2 levels in the preclinical stage of Alzheimer's disease, in: Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, Vol. 11, 2019, pp. 646–653.

[3] S.J. Teipel, A. Wohlert, C. Metzger, T. Grimmer, C. Sorg, M. Ewers, E. Meisenzahl, S. Klöppel, V. Borchardt, M. Grothe, Multicenter stability of resting state fMRI in the detection of alzheimer's disease and amnestic MCI, NeuroImage: Clin. 14 (2017) 183–194.

[4] B. Lei, S. Yu, X. Zhao, A.F. Frangi, E.-L. Tan, A. Elazab, T. Wang, S. Wang, Diagnosis of early Alzheimer's disease based on dynamic high order networks, Brain Imaging Behav. 15 (1) (2021) 276–287.

[5] P. Yang, F. Zhou, D. Ni, Y. Xu, S. Chen, T. Wang, B. Lei, Fused sparse network learning for longitudinal analysis of mild cognitive impairment, IEEE Trans. Cybern. 51 (1) (2019) 233–246.

[6] L. Qian, L. Zheng, Y. Shang, Y. Zhang, Y. Zhang, A. s. d. N. Initiative, Intrinsic frequency specific brain networks for identification of MCI individuals using resting-state fMRI, Neurosci. Lett. 664 (2018) 7–14.

[7] B. Lei, M. Yang, P. Yang, F. Zhou, W. Hou, W. Zou, X. Li, T. Wang, X. Xiao, S. Wang, Deep and joint learning of longitudinal data for Alzheimer's disease prediction, Pattern Recognit. 102 (2020) 107247.

[8] R.V. Marinescu, A. Eshaghi, D.C. Alexander, P. Golland, BrainPainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes, in: Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy, Springer, 2019, pp. 112–120.

[9] Y. Lin, X. Liang, Y. Yao, H. Xiao, Y. Shi, J. Yang, Osthole attenuates APP-induced Alzheimer's disease through up-regulating miRNA-101a-3p, Life Sci. 225 (2019) 117–131.

[10] Y. Zhang, H. Zhang, X. Chen, S.-W. Lee, D. Shen, Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis, Sci. Rep. 7 (1) (2017) 1–15.

[11] J. Göttler, C. Preibisch, I. Riederer, L. Pasquini, P. Alexopoulos, K.P. Bohn, I. Yakushev, E. Beller, S. Kaczmarz, C. Zimmer, Reduced blood oxygenation level dependent connectivity is related to hypoperfusion in Alzheimer's Disease, J. Cereb. Blood Flow Metab. 39 (7) (2019) 1314–1325.

[12] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, M.W. Woolrich, Network modelling methods for FMRI, NeuroImage 54 (2) (2011) 875–891.

[13] L. Qiao, H. Zhang, M. Kim, S. Teng, L. Zhang, D. Shen, Estimating functional brain networks by incorporating a modularity prior, NeuroImage 141 (2016) 399–407.

[14] C.-Y. Wee, P.-T. Yap, D. Zhang, L. Wang, D. Shen, Group-constrained sparse fMRI connectivity modeling for mild cognitive impairment identification, Brain Struct. Funct. 219 (2) (2014) 641–656.

[15] F. Huang, E.-L. Tan, P. Yang, S. Huang, L. Ou-Yang, J. Cao, T. Wang, B. Lei, Self-weighted adaptive structure learning for ASD diagnosis via multi-template multi-center representation, Med. Image Anal. 63 (2020) 101662.

[16] M. Liu, J. Zhang, C. Lian, D. Shen, Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores, IEEE Trans. Cybern. 50 (7) (2019) 3381–3392.

[17] X.-A. Bi, X. Hu, H. Wu, Y. Wang, Multimodal data analysis of alzheimer's disease based on clustering evolutionary random forest, IEEE J. Biomed. Health Inf. 24 (2020) 2973–2983.

[18] H. Guo, F. Zhang, J. Chen, Y. Xu, J. Xiang, Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer's disease, Front. NeuroSci. 11 (2017) 615.

[19] F.J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, D. Castillo-Barnes, Studying the manifold structure of Alzheimer's Disease: a deep learning approach using convolutional autoencoders, IEEE J. Biomed. Health Inf. 24 (1) (2019) 17–26.

[20] S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A.s.D.N. Initiative, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, NeuroImage 189 (2019) 276–287.

[21] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A.s.D.N. Initiative, Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, NeuroImage: Clin. 21 (2019) 101645.

[22] Z.A. Bakar, D.I. Ispawi, N.F. Ibrahim, N.M. Tahir, Classification of Parkinson's disease based on Multilayer Perceptrons (MLPs) Neural Network and ANOVA as a feature extraction. 63–67.

[23] X. Hong, R. Lin, C. Yang, N. Zeng, C. Cai, J. Gou, J. Yang, Predicting Alzheimer's disease using LSTM, IEEE Access 7 (2019) 80893–80901.

[24] H. Wang, S. Zhao, Q. Dong, Y. Cui, Y. Chen, J. Han, L. Xie, T. Liu, Recognizing brain states using deep sparse recurrent neural network, IEEE Trans. Med. Imaging 38 (4) (2018) 1058–1068.

[25] L. Gao, Z. Guo, H. Zhang, X. Xu, H.T. Shen, Video captioning with attention-based LSTM and semantic consistency, IEEE Trans. Multimed. 19 (9) (2017) 2045–2055.

[26] R. Gu, G. Wang, T. Song, R. Huang, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, S. Zhang, CA-Net: Comprehensive attention convolutional neural networks for explainable medical image segmentation, IEEE Trans. Med. Imaging 40 (2) (2020) 699–711.

[27] S. Song, C. Lan, J. Xing, W. Zeng, J. Liu, Spatio-temporal attention-based lstm networks for 3d action recognition and detection, IEEE Trans. Image Process. 27 (7) (2018) 3459–3471.

[28] P. Zhao, J. Zhang, W. Fang, S. Deng, SCAU-Net: Spatial-channel attention U-Net for gland segmentation, Front. Bioeng. Biotechnol. 8 (2020) 670.

[29] H. Zhao, J. Jia, V. Koltun, Exploring self-attention for image recognition. 10076–10085.

[30] W. Li, F. Qi, M. Tang, Z. Yu, Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification, Neurocomputing 387 (2020) 63–77.

[31] S. Zhang, X. Li, M. Zong, X. Zhu, R. Wang, Efficient kNN classification with different numbers of nearest neighbors, IEEE Trans. Neural Netw. Learn. Syst. 29 (5) (2018) 1774–1785.

[32] X. Yang, Y. Jin, X. Chen, H. Zhang, G. Li, D. Shen, Functional connectivity network fusion with dynamic thresholding for MCI diagnosis, in: International Workshop on Machine Learning in Medical Imaging, 2016, pp. 246–253.

[33] D. Yao, J. Sui, E. Yang, P.-T. Yap, D. Shen, M. Liu, Temporal-adaptive graph convolutional network for automated identification of major depressive disorder using resting-state fMRI, in: International Workshop on Machine Learning in Medical Imaging, 2020, pp. 1–10.

[34] X. Bi, X. Zhao, H. Huang, D. Chen, Y. Ma, Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine, Cogn. Comput. 12 (3) (2020) 513–527.

[35] L. Zhang, L. Wang, D. Zhu, A. s. D. N. Initiative, Predicting brain structural network using functional connectivity, Med. Image Anal. 79 (2022) 102463.

[36] J. Gan, Z. Peng, X. Zhu, R. Hu, J. Ma, G. Wu, Brain functional connectivity analysis based on multi-graph fusion, Med. Image Anal. 71 (2021) 102057.

[37] L. Zhang, L. Wang, J. Gao, S.L. Risacher, J. Yan, G. Li, T. Liu, D. Zhu, f.t.A.s.D.N. Initiative, Deep fusion of brain structure-function in mild cognitive impairment, Med. Image Anal. 72 (2021) 102082.

[38] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.

[39] C. Feng, A. Elazab, P. Yang, T. Wang, F. Zhou, H. Hu, X. Xiao, B. Lei, Deep learning framework for alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM, IEEE Access 7 (2019) 63605–63618.

[40] M. Wang, C. Lian, D. Yao, D. Zhang, M. Liu, D. Shen, Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network, IEEE Trans. Biomed. Eng. 68 (8) (2019) 2241–2252.
[41] B. Jie, M. Liu, C. Lian, F. Shi, D. Shen, Designing weighted correlation kernels in convolutional neural networks for functional connectivity based brain disease diagnosis, Med. Image Anal. 63 (2020) 101709.
[42] R.C. Craddock, G.A. James, P.E. Holtzheimer III, X.P. Hu, H.S. Mayberg, A whole brain fMRI atlas generated via spatially constrained spectral clustering, Human Brain Mapp. 33 (8) (2012) 1914–1928.
[43] B. Lei, Y. Zhao, Z. Huang, X. Hao, F. Zhou, A. Elazab, J. Qin, H. Lei, Adaptive sparse learning using multi-template for neurodegenerative disease diagnosis, Med. Image Anal. 61 (2020) 101632.
[44] J. Liu, L. Yuan, J. Ye, An efficient algorithm for a class of fused lasso problems. 323–332.
[45] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (1) (2006) 49–67.
[46] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610.
[47] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[48] R. Yu, Z. Han, A. Le, X. Chen, Z. Wei, D. Shen, Connectivity strength–weighted sparse group representation-based brain network construction for MCI classification, Human Brain Mapp. 38 (5) (2017) 2370–2383.