

Original Research

Diagnosis of Alzheimer's disease by feature weighted-LSTM: a preliminary study of temporal features in brain resting-state fMRI

Jiyun Li¹, Binbin Song¹, Chen Qian^{1,*}¹School of Computer Science and Technology, Donghua University, 201620 Shanghai, China*Correspondence: qianchenemail@163.com (Chen Qian)

Academic Editor: François S. Roman

Submitted: 16 June 2021 Revised: 24 August 2021 Accepted: 31 August 2021 Published: 22 March 2022

Abstract

The long short-term memory network (LSTM) is widely used in time series data processing as a temporal recursive network. The resting-state functional magnetic resonance data shows that not only are there temporal variations in the resting state, but there are also interactions between brain regions. To integrate the temporal and spatial characteristics of brain regions, this paper proposes a model called feature weighted-LSTM (FW-LSTM). The feature weight is defined by spatial characteristics calculating the frequency of connectivity of each brain region and further integrated into the LSTM. Thus, it can comprehensively model both temporal and spatial changes in rs-fMRI brain regions. The FW-LSTM model on the Alzheimer's disease neuroimaging initiative (ADNI) dataset is used to extract the time-varying characteristics of 90 brain regions for Alzheimer's disease (AD) classification. The model performances are 77.80%, 76.41%, and 78.81% in accuracy, sensitivity, and specificity. It outperformed the one-dimensional convolutional neural networks (1D-CNN) model and LSTM model, which only used temporal features of brain regions.

Keywords: Rs-fMRI data; Temporal characteristics; Spatial characteristics; FW-LSTM

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease with a slow onset process that worsens over time [1]. The clinical manifestation of Alzheimer's disease is loss of memory, while behavior and language skills can be significantly affected. Therefore, AD often places a heavy burden on individuals and families. According to Alzheimer's Disease International, it is estimated that by 2050, 131.5 million people worldwide (one in every 85 people) will have the disease [2]. Unfortunately, to date, there are no drugs available to treat Alzheimer's disease. Therefore, it is crucial to detect Alzheimer's disease at the early stage, so that the interventions can be promoted as early as possible to slow down the progression.

With the rapid development of neuroimaging technology, Alzheimer's disease can be classified in a reliable manner [3]. Functional magnetic resonance imaging (fMRI), as a non-invasive neuroimaging technique, is increasingly used in the study of the human brain. It can be divided into task-state fMRI and resting-state fMRI (rs-fMRI). The latter refers to the hemodynamics of the brain at rest [4]. In addition, neuroimaging techniques in the resting state include electroencephalogram (EEG) and functional near-infrared spectroscopy (fNIRS) [5]. Nowadays, rs-fMRI is playing an essential role in the classification of AD [6,7]. fMRI spatially groups brain regions based on brain region templates, obtains the average time series of each brain region, and calculates connectivity between brain regions. It is worth noting that connectivity refers to the correlation, covariance, or mutual information between pairwise brain

region sequences. Chen *et al.* [8] used the Pearson correlation coefficient as a connectivity metric for Fisher Linear Discriminant Analysis (LDA)-based AD and Mild cognitive impairment (MCI) classification. Challis *et al.* [9] used covariance as a connectivity metric to achieve AD and MCI classification based on the Gaussian process, logistic regression models. It has also been suggested to construct brain networks from connectivity matrix graphs and calculate network metrics, e.g., Cui *et al.* [10] constructed a minimal spanning tree classification framework for brain functional connectivity networks with the aim of AD classification. Ju *et al.* [11] used deep learning of brain networks and clinically relevant textual information to classify AD. Wang *et al.* [12] calculated the brain functional connectivity matrix as features by selecting some brain regions, then projected the features onto a one-dimensional axis using regularized linear discriminant analysis, and finally completed the classification task using the AdaBoost classifier. Wang *et al.* [13] constructed a brain network based on the brain functional connectivity matrix, yet therefrom extracted relevant features, and used the Least absolute shrinkage and selection operator (LASSO) method for feature selection and an extreme learning machine to achieve the classification. Khazaei *et al.* [14] constructed a brain network based on the brain functional connectivity matrix, extracted graph-theoretic features and finally used support vector machines for classification. Jie *et al.* [15] extracted features based on global topology and local connectivity from the graph, performed feature selection by minimum absolute shrinkage and selection operators and finally used a multi-



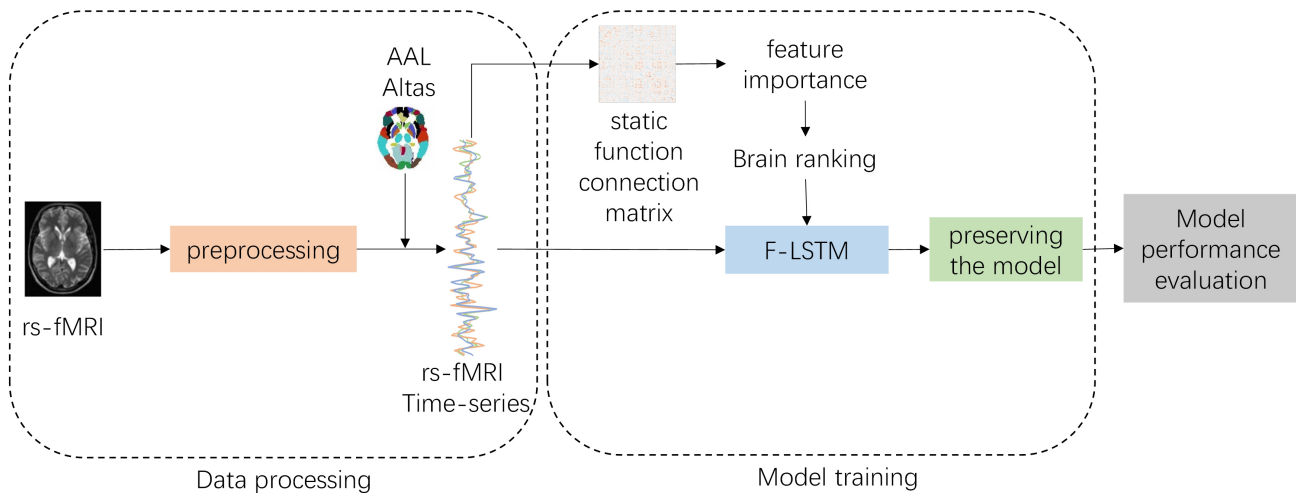


Fig. 1. The overall flow chart of this study. It includes a pre-processing data module, a training module and an evaluation module. The raw data were pre-processed and registered to the AAL template to obtain the regions of interest time series to train and evaluate the FW-LSTM model.

core support vector machines (SVM) for AD classification. Khazaei *et al.* [16] computed integration and separation from the graph metrics, feature selection by Fisher scoring, and AD classification using SVM.

According to the papers abovementioned, the mission of AD classification was accomplished, while the correct average rate is 88.42%. However, they only used connectivity between brain regions (spatial features) to construct the brain network, without fully considering the dynamic changes in the region of interest. Such changes are obtained from the regionally averaged time series in fMRI, which would result in the lack of asynchronous information in the temporal dimension. Conversely, modeling that only considers temporal changes in brain region features would ignore the spatial features between brain regions. The LSTM model is able to deal with the temporal memory, which can extract the temporal change features of brain regions well. But it cannot realize the interactions between brain regions. Considering both cases, we hereby put forward a feature-weighted LSTM network using temporal and spatial features for integrated modeling.

2. Materials and methodology

2.1 Subjects

Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (<http://adni.loni.usc.edu/>) began operations in 2004 to work towards the early detection and tracking of AD. The data used in this paper are rs-fMRI data from ADNI, 189 in total. Subjects were AD and Normal control (CN) according to the diagnostic criteria for ADNI [17]. Table 1 shows the demographic information.

Table 1. Demographic information.

	AD	NC
Num	91	98
Age (mean \pm SD)	73.5 \pm 7.6	75.2 \pm 6.4
Sex (F/M)	52/39	57/41
Education/year (mean \pm SD)	15.4 \pm 2.6	16.7 \pm 2.1

AD, Alzheimer's disease; NC, Normal control; SD, standard deviation; M, Male; F, Female.

2.2 Data collection and pre-processing

Scanning images were acquired on a 3.0 Tesla MRI scanner from Philips Medical Systems (CommunityCare Inc., Latham, NY, USA). The acquisition parameters included: pulse sequence = GR, TR = 3000 ms, TE = 30 ms, matrix = 64×64 , slice thickness = 3.3 mm, slice number = 48, flip angle = 80° .

In this study, the Data Processing & Analysis for Brain Imaging (DPABI) toolbox was used to perform the pre-processing of rs-fMRI [18]. The main pre-processing steps include: removal of the first 10 acquired Magnetic Resonance Imaging (MRI) volumes for each subject; slice time correction; head motion correction; spatial normalization (nonlinear alignment of images to Echo Plane Imaging (EPI) templates in standard space by affine transformation); smooth (smooth kernel size is $6 \text{ mm} \times 6 \text{ mm} \times 6 \text{ mm}$); detrend; regression covariance (head movement parameters generated during head movement correction, whole-brain signal, white matter signal and cerebrospinal fluid signal) and filter (0.01–0.08 Hz).

After a standard pre-processing process for rs-fMRI, the brain was segmented into 90 regions of interest aligning well with the Automated Anatomical Labeling (AAL) template [19]. Finally, the average time series of the 90 regions of interest were obtained for each subject.

2.3 Research procedure

The model takes the time series of 90 brain regions as input and further predicts whether it is AD or not. The overall flow chart proposed is shown in Fig. 1. The basic process is as follows: rs-fMRI data are registered with the AAL template after pre-processing to obtain the time series of 90 brain regions, the Pearson correlation between each two brain regions is calculated to obtain the static functional connectivity matrix; the static functional connectivity matrix is assessed for feature importance by random forest, a threshold is set, and brain regions with feature importance greater than the threshold are selected. Next, brain region frequencies are counted and normalized to the weight matrix in the FW-LSTM model; finally, the model is trained and evaluated.

2.4 Static functional connection matrix

The correlation coefficient of pairwise brain regions is calculated as the static functional connectivity of brain regions. To obtain the correlation coefficient for pairwise brain regions, this paper uses the Pearson correlation coefficient, calculated as

$$\rho(R_i, R_j) = \frac{\text{cov}(r_i, r_j)}{\sigma_{r_i} \cdot \sigma_{r_j}} \quad (1)$$

In Eqn. 1, R_i denotes the i -th brain region, r_i represents the time series corresponding to R_i , $\text{cov}(r_i, r_j)$ denotes the covariate between r_i and r_j , and σ_r denotes the standard deviation of the time series r .

The static functional connectivity matrix is obtained for each subject by calculating the correlation coefficient for pairwise brain regions. Finally, the static functional connectivity matrix is transformed into normally distributed Z values by Fisher's-Z transformation.

2.5 Feature importance assessment

This paper uses the random forest (RF) algorithm for feature importance assessment of static functional connectivity matrices. RF is an integrated learning method based on decision trees, proposed by Breiman [20] in 2001. RF can analyze data with large dimensionality and compute variable importance scores, meanwhile keeps high prediction accuracy and tolerance to outliers and noises [21].

Feature importance assessment refers to the contribution made by each feature on each tree in the random forest. A feature's importance result is calculated by comparison of its contribution with the average one. Specific to the classification task in AD is assessing the contribution of connectivity into brain regions. The contribution of a feature can be measured using the Gini coefficient or the out-of-bag coefficient as an evaluation indicator.

In this paper, the Gini index is used to evaluate the contribution of characteristics. For example, suppose there are m features: F_1, F_2, \dots, F_m , and calculate for each fea-

ture F_j of the Gini index score VIM_j and the Gini index is calculated as

$$GI_m = \sum_{k=1}^{|K|} \sum_{k' \neq k} P_{mk} P_{mk'} = 1 - \sum_{k=1}^{|K|} P_{mk}^2 \quad (2)$$

In Eqn. 2, the GI_m denotes the node m of the Gini index, the K denotes the number of categories in the sample set, k' denotes a category, and P_{mk} indicates the proportion of category k in a node m .

The importance of features F_j at the node m is the amount of change in the Gini index before and after the branching of the node m , calculated as

$$VIM_{jm} = GI_m - (GI_l + GI_r) \quad (3)$$

In Eqn. 3, the GI_l denotes the Gini index of the left node after branching, and GI_r denotes the Gini index of the right node after branching.

If the node where the feature F_j appears in the decision tree i is in the set M , then the importance of F_j in the tree i as

$$VIM_{ij} = \sum_{m \in M} VIM_{jm} \quad (4)$$

Let there be a total of random forest n trees, then

$$VIM_j = \sum_{i=1}^n VIM_{ij} \quad (5)$$

Finally, the importance scores obtained for the variables are normalized, then

$$VIM_j = \frac{VIM_j}{\sum_{i=1}^c VIM_i} \quad (6)$$

In Eqn. 6, c is the total number of trees.

2.6 Frequency of brain areas

The importance of features is calculated using random forests on a static functional connectivity matrix, followed by setting thresholds to select appropriate features. In this paper, a threshold (δ) is set to select the features. Selecting the functional connectivity of brain regions with importance scores greater than δ for the variables mentioned in the previous section and counting the number of occurrences of each brain region in the selected features can be expressed as

$$\text{Cul}(R_i) = \begin{cases} +1, & \text{score}(R_i, R_k) > \delta \\ 0, & \text{else} \end{cases} \quad (7)$$

2.7 FW-LSTM

LSTM is a temporal recurrent neural network, which is well suited for processing time-series data. Blood oxygen

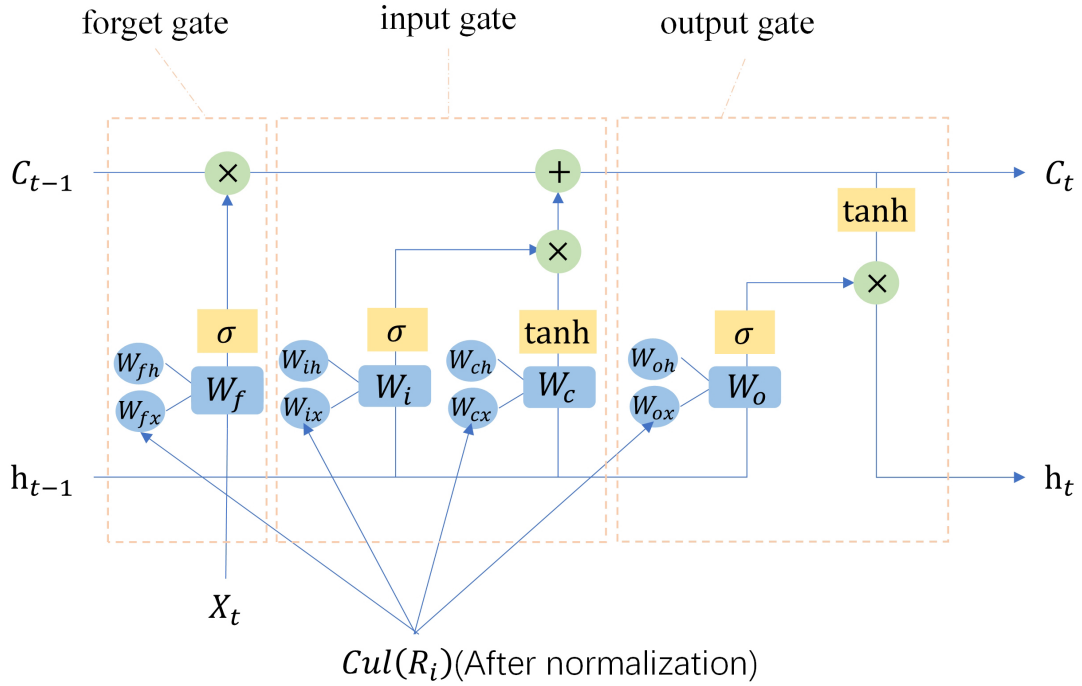


Fig. 2. Structure of the FW-LSTM unit. The weight terms include W_f , W_i , W_c and W_o , which can be specifically divided into W_{fh} , W_{fx} , W_{ih} , W_{ix} , W_{ch} , W_{oh} , W_{ox} . The normalization of $Cul(R_i)$ in the FW-LSTM model is combined with the corresponding weight matrix W_{fx} , W_{ix} , W_{cx} , W_{ox} .

level dependent (BOLD) fMRI signal data is time-series data, where the signal at the current moment is closely related to the signal in the past and future. The LSTM is mainly used to select valid information and the learning of long-term dependencies through forgetting gates, input gates and output gates. Therefore, this paper proposes the FW-LSTM model that uses temporal and spatial features for integrated modeling. Fig. 2 shows the FW-LSTM model.

In the model, the forget gate is used to determine what information should be discarded from the cell state. It can be expressed as

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (8)$$

In Eqn. 8, the input at the current moment is X_t with the output of the previous moment h_{t-1} through the activation function (sigmoid), the output is given a value between 0 and 1, representing whether it should be forgotten, with 0 indicating complete forgetting and 1 indicating complete remembering.

The input gate is used to memorize a selection of new information into the cell state C_t , which can be expressed as

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \\ C_t &= f_t \times C_{t-1} + i_t \times \tilde{C}_t \end{aligned} \quad (9)$$

In Eqn. 9, i_t represents the degree of information memorization obtained by the current moment X_t and the previous moment's output h_{t-1} through the sigmoid function. \tilde{C}_t denotes the new cell state, obtained by passing the input X_t and the previous moment's output h_{t-1} through a \tanh activation function. C_t is the updated cell state, selected by the forgetting gate to forget the previous moment's cell state C_{t-1} part of it, and then the input gate selects to add the new cell state \tilde{C}_t to complete the process.

The output gate output to the current output value of the LSTM through a fraction of the cell state h_t , which can be expressed as

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \\ h_t &= O_t \times \tanh(C_t) \end{aligned} \quad (10)$$

In Eqn. 10, O_t is calculated from h_{t-1} and X_t via sigmoid function, which is used to measure the output to be O_t how much output value there is.

In the model of this paper, the weight terms include W_{fh} , W_{fx} , W_{ih} , W_{ix} , W_{ch} , W_{oh} , W_{ox} . Where W_{fx} , W_{ix} , W_{cx} , W_{ox} dimension is (90, unit_num), which is acting on the input X_t at the current moment of the model. And W_{fh} , W_{ih} , W_{ch} , W_{oh} dimensions are (unit_num, unit_num) and are acting on the output h_{t-1} at the previous moment.

Therefore, we only adjust the four weight matrices acting on the input X_t , that is, the weights of the input time series of 90 brain regions. The four weight matrices

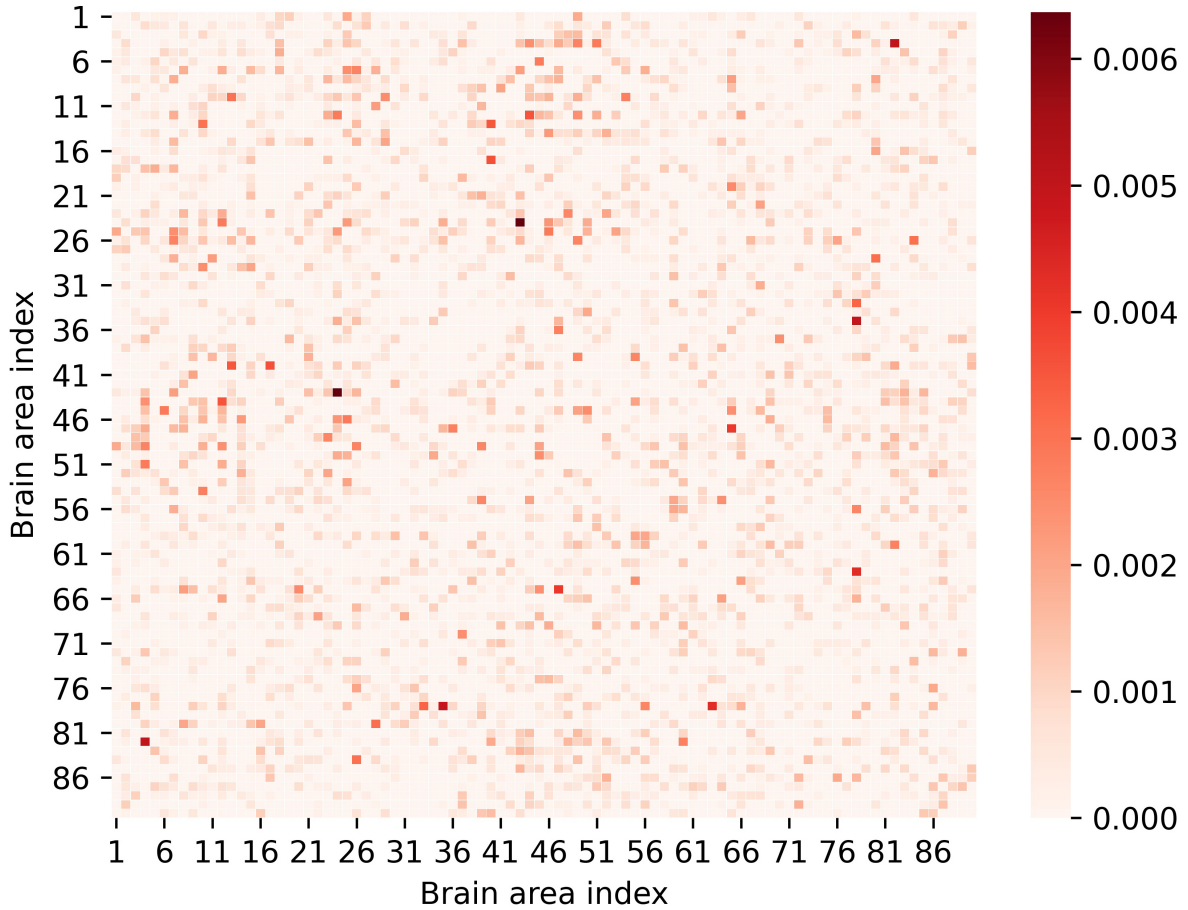


Fig. 3. The importance of the features of the static functional connectivity matrix. The brain region index represents numbering in the AAL template. The small colored blocks in the diagram represent the feature importance of each pair of brain region connections.

W_{fx} , W_{ix} , W_{cx} , and W_{ox} are adjusted using the normalized $Cul(R_i)(n_{-}Cul(R_i))$ for the respective importance of the 90 brain regions. The modifications are as follows.

$$\begin{aligned}
 W_{fx} &= W_{fx} \times n_{-}Cul(R_i) \\
 W_{ix} &= W_{ix} \times n_{-}Cul(R_i) \\
 W_{cx} &= W_{cx} \times n_{-}Cul(R_i) \\
 W_{ox} &= W_{ox} \times n_{-}Cul(R_i)
 \end{aligned} \quad (11)$$

A dropout layer follows the LSTM module with a loss rate of 0.5 and a dense layer with several neurons of 10. The dense layer extracts the fMRI time-varying features and finally uses the sigmoid function for the AD and Normal control (NC) binary classification task.

3. Experimental results and analysis

3.1 Feature importance

The obtained static functional connectivity matrix is evaluated for feature importance. The static functional connectivity matrix is a matrix of 90×90 , where the upper right triangular matrix is symmetrical with the lower-left triangular matrix, and the upper right triangular matrix is

taken as the feature (i.e., 4005 dimensions) to evaluate its importance.

When the data is fed into the random forest, the grid-SearchCV is chosen to find the best parameters. The maximum number of iterations of the weak learner is searched in the range of [100, 150], which is finally determined to be 131, and the maximum number of features of the selected feature subset is searched in the range of [10, 100], which is finally determined to be 56. In order to avoid the chance of the results, the average of the ten feature importance assessment results is chosen as the feature importance assessment result.

As Fig. 3 shows, the feature of greatest importance is the functional connection between brain area 24 and 43, i.e., superior frontal gyrus (medial) and calcarine fissure and surrounding cortex; the second-ranked feature is the functional connection between brain area 36 and 47, i.e., posterior cingulate gyrus and lingual gyrus; the third-ranked feature is the functional connection between brain area 56 and 86, i.e., fusiform gyrus and middle temporal gyrus.

3.2 Frequency of brain areas

We compare the impact of the selected features on AD classification at different thresholds to select the most appropriate one. We use features larger than the threshold to classify AD again, and the classifier choose random forest. Here, we divide the dataset into a training set and a test set in the ratio of 4:1, and execute the model five times to acquire the accuracy evaluation metric. During the execution, the training set and the test set are randomly selected each time, Fig. 4 shows the results of the experiment. Moreover, we test the importance obtained from an average of 5 experiments with an accuracy of 65.04%.

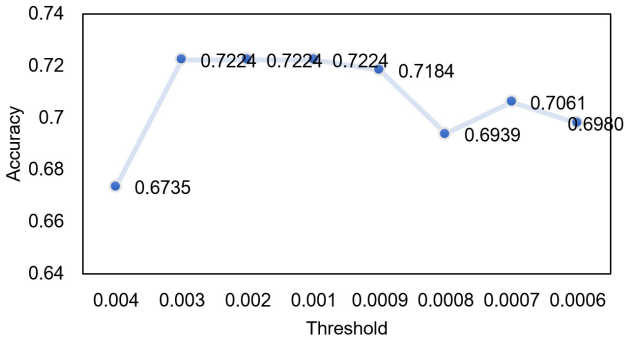


Fig. 4. Classification accuracy at different thresholds. We compared the accuracy of the model under eight different thresholds. It becomes obvious that the classification accuracy is highest at the threshold values of 0.003, 0.002, 0.001.

Now, we choose a threshold value of 0.001 for the subsequent experiment. A total of 303 functional connections to brain regions are obtained based on the calculated characteristic importance of the connections between brain regions and the threshold settings. The number of occurrences for each brain region is counted. The results are demonstrated in Fig. 5 (Ref. [22]).

3.3 FW-LSTM model training

The FW-LSTM model is implemented by Python 3.6 (Python Inc., Netherlands), while the deep learning development platform is the Keras framework (v2.1.6, Google Inc., USA) for Tensorflow (v1.7.0, Google Inc., USA). The development environment consists of Windows 10 as the operating system, Inter(R) Core(TM) i7-8565 CPU as the processor, and Nvidia GeForce MX250 (Nvidia Inc., USA) as the graphics card.

In this paper, the extracted time series of brain regions are input into the FW-LSTM model for temporal feature extraction and classification. The experimental data are time-series data with a dimension of 90×130 . The small amount of data could easily lead to overfitting the model, so we performed data segmentation. The 130-time points are segmented, and 10-time points are treated as one data, implying a dimensionality of 90×10 . Finally, a five-fold cross-

validation strategy is used to evaluate the model.

3.4 Analysis of results

To evaluate the effectiveness of the FW-LSTM model for temporal feature extraction in this paper and to compare it with other models, the performance of the model is evaluated using accuracy (Acc), sensitivity (Sen), and specificity (Spe).

Acc: Time-varying features are extracted for classification using the FW-LSTM model. Acc represents the proportion of correctly classified samples to the total samples (i.e., the full test set) and can be expressed as

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \quad (12)$$

In Eqn. 12, TP (True Positive) denotes positive cases correctly classified by the model; TN (True Negative) denotes negative cases correctly classified by the model; FN (False Negative) denotes positive cases misclassified as negative by the model, and FP (False Positive) denotes negative cases misclassified as positive by the model.

Sen: indicates the number of positive cases correctly predicted by the model as a proportion of positive cases and can be expressed as

$$Sen = \frac{TP}{TP+FN} \quad (13)$$

Spe: indicates the number of negative cases correctly predicted by the model as a proportion of negative cases and can be expressed as

$$Spe = \frac{TN}{TN+FP} \quad (14)$$

Sensitivity and specificity are often used to measure the diagnostic outcome of a disease, with higher sensitivity resulting in a lower rate of missed diagnoses and higher specificity resulting in a higher rate of confirmed diagnoses.

ROC curve: Receiver operating characteristic curve, a composite indicator of a continuous variable of sensitivity and specificity. On the ROC curve, if the corner is closer to the upper left, the better the result. AUC (Area under ROC curve) is a metric used to measure how good a classification model is, with a larger AUC representing better performance.

In the FW-LSTM model, we make changes to the four parameters that act on the input to the model: W_{fx} , W_{ix} , W_{cx} , and W_{ox} . We modify each of these four parameters, and the experimental results are shown in Table 2. Finally, we combine multiple parameters to view the effect. The experimental results are as follows.

In Table 2, for modifying individual parameters, we can see that in terms of accuracy and specificity, the best results can be obtained by modifying the W_{fx} . In terms of sensitivity, the best results can be obtained by modifying the

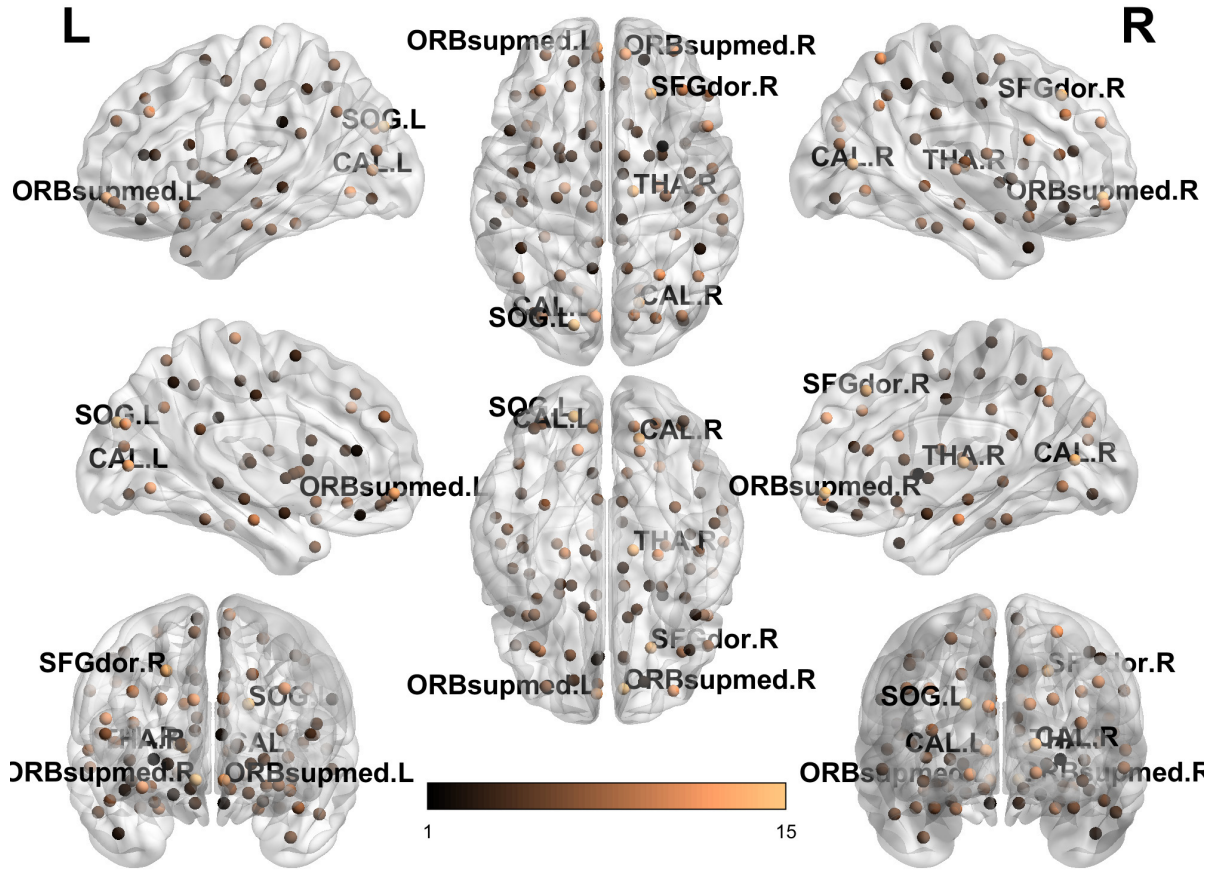


Fig. 5. Frequency map of brain areas. The brain networks are visualized with the BrainNet Viewer [22]. Because the brain area labels are so dense, only brain area labels with a frequency greater than 12 are expressed in this figure. Left superior occipital gyrus (SOG.L) appear 15 times, right superior frontal gyrus, medial orbital (ORBsupmed.R) appear 15 times, right thalamus (THA.R) appear 14 times, right calcarine fissure and surrounding cortex (CAL.R) appear 14 times, right superior frontal gyrus, dorsolateral (SFGdor.R) appear 14 times, left calcarine fissure and surrounding cortex (CAL.L) appear 13 times, left superior frontal gyrus, medial orbital (ORBsupmed.L) appear 13 times.

Table 2. Model comparison 1.

Method	Modify	Acc	Sen	Spe
	W_{ix}	67.56%	67.66%	67.46%
	W_{fx}	73.01%	72.81%	73.22%
	W_{cx}	72.76%	72.97%	72.54%
	W_{ox}	70.41%	69.22%	71.69%
	W_{fx}, W_{cx}	77.80%	76.41%	78.81%
	W_{fx}, W_{cx}, W_{ox}	77.72%	79.38%	75.93%
FW-LSTM	$W_{fx}, W_{ix}, W_{cx}, W_{ox}$	76.59%	75.31%	77.97%

W_{cx} . Modifying the W_{fx} in terms of accuracy compared to modifying the W_{ix} , the W_{cx} , and the W_{ox} increased by 5.45, 0.25 and 2.6 percentage points, respectively; modifying the W_{fx} in specificity compared to modifying the W_{ix} , the W_{cx} , and the W_{ox} increased by 5.76, 0.68 and 1.53 percentage points, respectively. Modifying the W_{cx} in specificity compared to modifying the W_{ix} , the W_{fx} , and the W_{ox} is improved by 5.31, 0.16, and 3.75 percentage points, respectively.

We can see that the effect of modifying the model with one parameter seems unsatisfactory. However, the effect of modifying multiple parameters is much better. Apparently, modifying the W_{fx} and the W_{cx} results in the highest accuracy and specificity; modifying the W_{fx} , the W_{cx} , and the W_{ox} in sensitivity can give the best results. Fig. 6 depicts the respective ROC curves.

Taking these indicators together, modifying the W_{fx} and the W_{cx} is the best. Therefore, our FW-LSTM model used modifications of these two parameters.

We chose Random Forest to classify and the static functional connectivity matrix as features to compare the accuracy, specificity and sensitivity obtained using only spatial features. 1D-CNN can be well applied to temporal sequence analysis. Thereby, we use it as a comparison model. The 1D-CNN model we use contains two 1-dimensional convolutional layers, followed by a dropout layer, then a pooling layer, then a fully connected layer, before reaching the output layer for prediction. For this

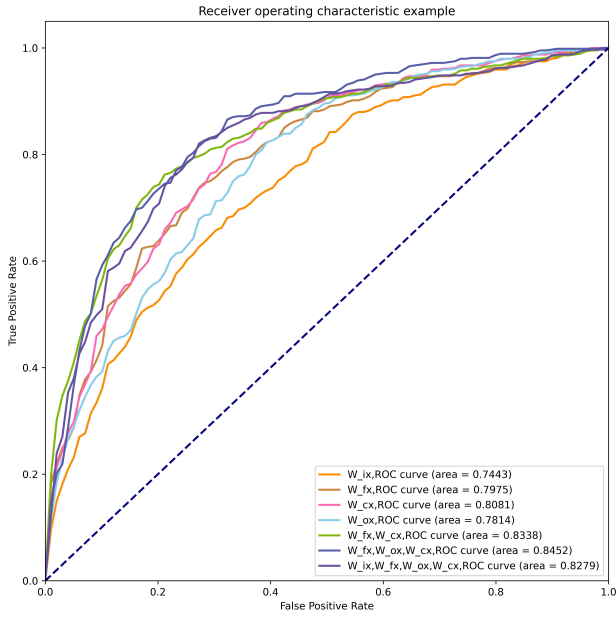


Fig. 6. ROC curves were obtained. As seen from modifying individual parameter, modifying the W_{ix} has the worst result and modifying W_{cx} gives the best result. The area under the curve of the classifier obtained by modifying the W_{fx} , W_{ix} , W_{cx} , and W_{ox} is 0.7975, 0.7443, 0.8081, 0.7814, respectively. As can be seen from modifying multiple parameters, it is not clear which classifier is the best. The area under the curve of the classifier obtained by modifying W_{fx} and W_{cx} is 0.8338, modifying W_{fx} , W_{ox} , and W_{cx} is 0.8452, modifying W_{fx} , W_{ix} , W_{cx} , and W_{ox} is 0.8279.

model, we used 16 feature maps with a kernel of size 3. We used the efficient Adam algorithm to optimize the network, using the binary crossentropy for classification as the loss function. For the other comparison model, the LSTM, we used the same training configuration as the FW-LSTM. For each of the comparison models, we used a five-fold cross-validation strategy and repeating five-fold cross-validation assignments from resampling the data five times at random.

The model comparison in Table 3 shows that the FW-LSTM model proposed in this paper outperforms the random forest, 1D-CNN and LSTM models in terms of accuracy, sensitivity and specificity. Compared with the random forest, the FW-LSTM model proposed in this paper has 12.5 percentage points better accuracy, 22.49 percentage points better sensitivity and 10.19 percentage points in specificity; Compared with the 1D-CNN model, the FW-LSTM model proposed in this paper has improved 10.61 percentage points in accuracy, 10.01 percentage points in sensitivity and 11.02 percentage points in specificity; compared with the LSTM model, the FW-LSTM model proposed in this paper has improved 7.35 percentage points in accuracy and 5.48 percentage points in sensitivity and 9.16 percentage points in specificity. Next, we look at the respective ROC curves of the FW-LSTM, 1D-CNN, and

Table 3. Model comparison 2.

Models	Acc	Sen	Spe
Random Forest	65.71%	55.41%	68.12%
1D-CNN	67.60%	67.89%	67.29%
LSTM	70.86%	72.42%	69.15%
FW-LSTM	78.21%	77.90%	78.31%

LSTM models in Fig. 7.

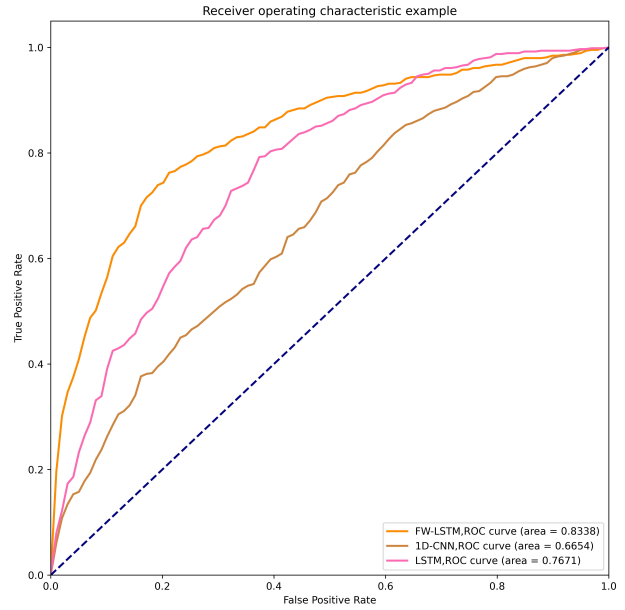


Fig. 7. These models obtained ROC curves. We can clearly see from the ROC curves that the FW-LSTM model has the best performance, followed by the LSTM model and finally the 1D-CNN model. The area under the curve for the FW-LSTM model, LSTM model, and 1D-CNN model is 0.8338, 0.7671, 0.6654, respectively.

4. Discussion

In this paper, FW-LSTM model is introduced for the time-varying characteristics unique to fMRI data in Alzheimer's and the interactions between the 90 brain regions. First, the Pearson correlation between pairwise brain regions is calculated, i.e., the static functional connectivity matrix is calculated; then, the feature importance of the static functional connectivity matrix is calculated, and those features with feature importance greater than a threshold are counted to obtain the brain region frequencies. Finally, the weight matrices in the FW-LSTM unit are combined to extract time-varying features for classification based on the values of the frequency of brain areas.

5. Conclusions

The experimental results prove that our model can extract time-varying features from 90 brain regions, which is of great significance for the classification of AD. During the diagnosis of AD, rs-fMRI should not only consider either time-varying characteristics or spatial characteristic. In our experiments, not only we can classify AD, but also tell from the frequency ranking of brain areas that the main brain areas contributing to the classification are Superior occipital gyrus and Superior frontal gyrus, medial orbital. Regardless, our work is not flawless: the dataset is small; only static correlation is considered; model performance still has room to be improved.

Abbreviations

LSTM, long short-term memory network; rs-fMRI, resting-state functional magnetic resonance; FW-LSTM, Feature-weighted long short-term memory networks; AD, Alzheimer disease; sMRI, structural magnetic resonance imaging; fMRI, functional magnetic resonance imaging; PET, positron emission tomography imaging; DTI, diffusion tensor imaging; AAL, Automated Anatomical Labeling; RF, random forest; VIM, variable importance scores.

Author contributions

BBS and JYL conceived the essay, BBS designed the experiment and wrote the paper, JYL and CQ made constructive changes to the paper.

Ethics approval and consent to participate

The dataset we used was the public dataset of ADNI, the ethical approval was not required. In addition, informed written consent was obtained from all participants at every center.

Acknowledgment

We thank two/three anonymous reviewers for excellent criticism of the article.

Funding

This work has been supported by the National Key R&D Program of China under Grant 2019YFE0190500, the Fundamental Research Funds for the Central Universities of Ministry of Education of China (Grant No.2232021D-22), Shanghai Engineering Research Center on Big Data Management System, and the Initial Research Funds for Young Teachers of Donghua University.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Burns A, Iliffe S. Alzheimer's disease. *British Medical Journal*. 2009; 338: b158.
- [2] Prince M, Comas-Herrera A, Knapp M, Guerchet M, Karagianidou M. Improving Healthcare for People Living with Dementia: Coverage, Quality and Costs Now and in the Future. *Alzheimer's Disease International*. 2016.
- [3] Zhang F, Li Z, Zhang B, Du H, Wang B, Zhang X. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing*. 2019; 361: 185–195.
- [4] Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience*. 2002; 3: 142–151.
- [5] Chiarelli AM, Perpetuini D, Croce P, Filippini C, Cardone D, Rotunno L, *et al*. Evidence of Neurovascular Un-Coupling in Mild Alzheimer's Disease through Multimodal EEG-fNIRS and Multivariate Analysis of Resting-State Data. *Biomedicine*. 2021; 9: 337.
- [6] Dennis EL, Thompson PM. Functional brain connectivity using fMRI in aging and Alzheimer's disease. *Neuropsychology Review*. 2014; 24: 49–62.
- [7] Li W, Lin X, Chen X. Detecting Alzheimer's disease Based on 4D fMRI: an exploration under deep learning framework. *Neurocomputing*. 2020; 388: 280–287.
- [8] Chen G, Ward BD, Xie C, Li W, Wu Z, Jones JL, *et al*. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology*. 2011; 259: 213–221.
- [9] Challis E, Hurley P, Serra L, Bozzali M, Oliver S, Cercignani M. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *NeuroImage*. 2015; 112: 232–243.
- [10] Cui X, Xiang J, Guo H, Yin G, Zhang H, Lan F, *et al*. Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls with Subnetwork Selection and Graph Kernel Principal Component Analysis Based on Minimum Spanning Tree Brain Functional Network. *Frontiers in Computational Neuroscience*. 2018; 12: 31.
- [11] Ju R, Hu C, Zhou P, Li Q. Early Diagnosis of Alzheimer's Disease Based on Resting-State Brain Networks and Deep Learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2019; 16: 244–257.
- [12] Wang Z, Zheng Y, Zhu DC, Bozoki AC, Li T. Classification of Alzheimer's Disease, Mild Cognitive Impairment and Normal Control Subjects Using Resting-State fMRI Based Network Connectivity Analysis. *IEEE Journal of Translational Engineering in Health and Medicine*. 2018; 6: 1–9.
- [13] Wang Z, Jiang W, Liu B, Chen S. Computer-aided diagnosis of mild cognitive impairment based on extreme learning machine. *Journal of Harbin Engineering University*. 2021; 1–7.
- [14] Khazae A, Ebrahimzadeh A, Babajani-Feremi A. Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. *Brain Imaging and Behavior*. 2016; 10: 799–817.
- [15] Jie B, Zhang D, Gao W, Wang Q, Wee C, Shen D. Integration of network topological and connectivity properties for neuroimaging classification. *IEEE Transactions on Bio-Medical Engineering*. 2014; 61: 576–589.
- [16] Khazae A, Ebrahimzadeh A, Babajani-Feremi A. Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory. *Clinical Neurophysiology*. 2015; 126: 2132–2141.
- [17] Petersen R, Weiner MW, Albert M, Salmon D, Morris J, Shaw LM, *et al*. Alzheimer's Disease Neuroimaging Initiative 2. Available at: <https://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf> (Accessed: 1 July 2008).
- [18] Yan C, Wang X, Zuo X, Zang Y. DPABI: Data Processing &

- Analysis for (Resting-State) Brain Imaging. *Neuroinformatics*. 2016; 14: 339–351.
- [19] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*. 2002; 15: 273–289.
- [20] Breiman L. Random forests. *Machine Learning*. 2001; 45: 5–32.
- [21] Kai Y, Yan H, Kang L. Variable Importance Measure of Random Forest and Its Progress. *Sciencepaper Online*. 2015. (In Chinese)
- [22] Xia MR, Wang JH, He Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*. 2013; 8: e68910.