

Multimodal Data Analysis of Alzheimer's Disease Based on Clustering Evolutionary Random Forest

Xia-an Bi*, *Member IEEE*, Xi Hu, Hao Wu and Yang Wang

Abstract—Alzheimer's disease (AD) has become a severe medical challenge. Advances in technologies produced high-dimensional data of different modalities including functional magnetic resonance imaging (fMRI) and single nucleotide polymorphism (SNP). Understanding the complex association patterns among these heterogeneous and complementary data is of benefit to the diagnosis and prevention of AD. In this paper, we apply the appropriate correlation analysis method to detect the relationships between brain regions and genes, and propose “brain region-gene pairs” as the multimodal features of the sample. In addition, we put forward a novel data analysis method from technology aspect, cluster evolutionary random forest (CERF), which is suitable for “brain region-gene pairs”. The idea of clustering evolution is introduced to improve the generalization performance of random forest which is constructed by randomly selecting samples and sample features. Through hierarchical clustering of decision trees in random forest, the decision trees with higher similarity are clustered into one class, and the decision trees with the best performance are retained to enhance the diversity between decision trees. Furthermore, based on CERF, we integrate feature construction, feature selection and sample classification to find the optimal combination of different methods, and design a comprehensive diagnostic framework for AD. The framework is validated by the samples with both fMRI and SNP data from ADNI. The results show that we can effectively identify AD patients and discover some brain regions and genes associated with AD significantly based on this framework. These findings are conducive to the clinical treatment and prevention of AD.

Index Terms—Alzheimer's disease, clustering evolutionary random forest, fMRI, multimodal data fusion, SNP.

I. INTRODUCTION

IT is well-established that Alzheimer's disease is defined as a typical form of neurodegenerative disorders in medicine. Generally, AD patients usually have mental retardation and memory impairment in clinical manifestations [1]. The latest research shows that the total prevalence of AD patients will

exceed 60 million in the next 50 years [2]. Therefore, a large quantity of studies on AD have been carried out to understand its pathogenesis, and prove that accurate early diagnosis is conducive to slow down the process of AD even though the process is irreversible [3].

Over the past two decades, advances in neuroimaging technologies have made great contributions to early diagnosis of AD [4], especially the application of fMRI technology which enables researchers to detect the brain activities of AD patients in real-time. Additionally, the genome-wide association study (GWAS) shows that genetic variations (e.g., SNP) are the intrinsic etiologies of AD through their abnormal expression in brain function and structure [5]. Therefore, multimodal fusion analysis to explore the correlation between neuroimaging data and gene data may be a potential breakthrough in AD research.

Multimodal data fusion analysis is an emerging field in brain research. Because the cost of obtaining labeled samples in the medical field is expensive, there is little public and credible multimodal data. A limitation that researchers often face is how to extract important information from high-dimensional fusion features in small samples [6]. In previous studies, some researchers used classical methods [7], such as principal component analysis (PCA) and independent component analysis (ICA), to solve the high-dimensional problem of multimodal fusion analysis. These methods achieved attribute reduction, but researchers need to devote lots of efforts if they want to analyze some significant fusion features separately. In recent researches, many improved methods have been proposed. For instance, Wang et al. [8] reduced the dimensionality by proposing sparse multiple canonical correlation analysis (SMCCA) for feature selection. Besides the methods mentioned above, other methods have also been proposed, such as gradient boosting decision tree [9] and low-rank dimensionality reduction [10]. However, it is worth noting that the complexity of the methods may also reduce its practicability and generalization.

In addition, most of the current researches mainly focus on a specific process in multimodal fusion, such as fusion feature construction [11], feature selection [12], or sample classification [13]. The lack of comprehensive framework is another challenge in the study of multimodal fusion between neuroimaging data and gene data. For example, Shi et al. [14] proposed the multimodal stacked deep polynomial networks (MM-SDPN) to construct fusion features. Zhu et al. [15]

This work was supported by the National Science Foundation of China (Grant No. 61502167), the Hunan Provincial Science and Technology Project Foundation (2018TP1018).

Xia-an Bi (bixiaan@hnu.edu.cn), Hao Wu, Xi Hu and Yang Wang are with Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, P.R. China, and also with College of Information Science and Engineering, Hunan Normal University, Changsha, P.R. China.

* indicates the corresponding author

designed a novel relational regularization feature selection approach to extract the most effective fusion features. In addition to the above two aspects, sample classification has attracted more attention of researchers, and the deep learning methods are becoming increasingly popular. Chaddad et al. [16] extracted the radial features from MRI images, and used the convolutional neural network (CNN) model to classify 100 AD and 135 healthy control (HC). Choi et al. [17] input 3-dimensional positron emission tomography (PET) volumes into deep CNN to recognize 139 AD and 182 HC, and the classification accuracy reached 84%. However, the deep learning model is more similar to the black box, and the interpretation of classification results is easier to make the problem more complex. Therefore, the design of a comprehensive framework integrating multimodal fusion feature construction, feature selection and sample classification is of great significance for the early diagnosis of AD.

Considering the challenges mentioned above, we carry out a multimodal fusion study of AD based on fMRI data and gene data. Firstly, we design the clustering evolutionary random forest to solve the small samples problem. The method constructs the initial random forest by selecting samples and fusing features randomly, and introduces the idea of clustering evolution to delete irrelevant and redundant decision trees dynamically. Secondly, we test the different construction methods of fusion features, and integrate the fusion features construction, feature selection and sample classification into a framework. The multimodal data comprehensive analysis framework for AD is showed in Fig. 1. The overall description of the model is as follows: 1) The preprocessing of multimodal

data. 2) The construction of multimodal fusion features. 3) The establishment of CERF and the classification of samples. 4) The extraction of features with strong classification ability. 5) The detection of abnormal brain regions and pathogenic genes. Finally, we use multimodal data from ADNI database to evaluate our method, and its reliability is verified.

The rest of the article is arranged as follows. Section II is the description of methods. Section III is the experimental results. Section IV is the related discussions, and Section V is the conclusion of our paper.

II. MATERIAL AND METHODOLOGY

A. Data Acquisition and Preprocessing

The data is acquired from the ADNI database, which serves as a large public cohort for Alzheimer's disease (<http://adni.loni.usc.edu/>). Substantial PET data, MRI data and gene data of AD patients are available in the cohort. In this work, 37 AD patients and 35 HC are collected from ADNI as experimental samples, and each sample has resting fMRI and SNP data.

In order to avoid the influences of age and sex on the follow-up experiments, this study conducts statistical tests for the AD and HC groups. As shown in Table I, there is no significant difference in age and sex between the two groups.

TABLE I
BASIC INFORMATION OF AD AND HC

Variables(Mean ± SD)	AD(n=37)	HC(n=35)	P value
Gender(M/F)	19/18	13/22	0.324*
Age(years)	75.35±7.949	77.14±6.175	0.291**

Notes: Values are mean ± SD;

* The P value was obtained by the chi-square test.

** The P value was obtained by the two-sample t-test.

For quality control, it is necessary to preprocess fMRI data and SNP data respectively. Standard preprocessing steps for fMRI data are applied using DPARSF such as slice timing correction and head-motion adjustment. The specific steps of fMRI data preprocessing are as follows:

- Transforming the original data into NIFTI format file.
- Deleting the first 10 time volumes of all samples to ensure magnetic gradient field stabilization of the scanner.
- Performing slice timing correction on the remaining volumes.
- Adjusting the head-motion to guarantee that the brain of each sample is in the same position.
- Normalizing the image with the EPI template.
- Smoothing the noises of image using Gaussian kernel with the full width at half maximum (FWHM) of 6 mm.
- Using the linear model to remove covariate signal interference including subject motion, white matter and global signal.
- Filtering the functional time series by a 0.01Hz to 0.08Hz frequency range.

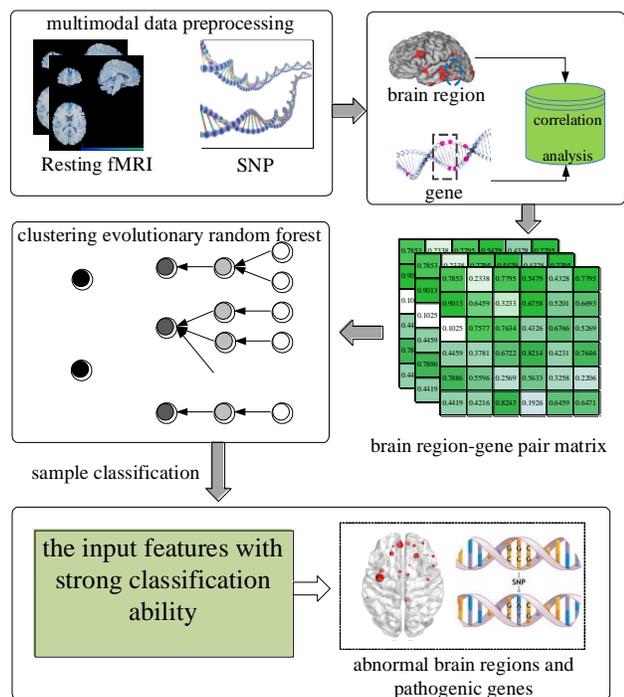


Fig. 1. The framework of multimodal data analysis. The framework of multimodal data analysis includes five parts: data preprocessing, correlation analysis, feature matrix construction, sample classification and pathogenic factors extraction.

The SNP information of the sample is obtained through Illumina Omni 2.5M BeadChip. PLINK is used to perform preprocessing for SNP data, and detailed steps are as follows:

- The threshold of sample call rate is set to 95% to evaluate the overall quality of gene data.
- The thresholds of genotyping, minimum allele frequency and Hardy-Weinberg equilibrium test are set to 99.9%, 4% and 1E-4 respectively to eliminate the SNPs with poor quality.

B. Construction of Fusion Features

The first aim of our study is to detect the correlations between genes and brain regions, and construct the fusion features. Although there are a few existing studies trying to carry out the fusion analyze of brain region and gene [8], but multimodal data fusion is still a challenging work in the field of brain science. This paper designs an attractive fusion scheme for multimodal fusion research. We use a more practical correlation analysis method to detect the correlations between specific brain regions and genes from sequence coding aspect. These correlations are sample fusion features named “brain region-gene pairs”. Compared with the classical method, it has better construction efficiency and interpretability. The specific construction method is as follows.

After preprocessing of the multimodal data, the resting fMRI and gene data of samples are standardized. Firstly, fMRI data is separated into 90 brain regions based on the Automatic Anatomical Labeling (AAL) template, and time series in brain regions are obtained. Secondly, the SNPs are grouped according to their corresponding genes, and M genes with SNP count greater than threshold len are kept. Thirdly, four types of bases (e.g., A, T, C and G) in SNPs are converted into different digits (e.g., 1, 2, 3 and 4), and then we obtain the digital sequences of genes by the transformation method. Finally, we adjust the length of gene sequences and brain time series to 2len respectively, and use Pearson correlation analysis, canonical correlation analysis (CCA) and distance correlation analysis (DCA) as candidate methods to construct “brain region-gene pairs”. We will compare the applicability of different correlation analysis methods in the proposed comprehensive framework to select the optimal fusion feature construction method.

C. Construction of Clustering Evolutionary Random Forest

The function of CERF is to process high-dimensional fusion features. It is generally known that ensemble learning has advantageous superiority in high-dimensional data processing. Random forest is the representative of ensemble learning technology, and has desirable processing ability for high-dimensional data in some cases. However, in the field of multimodal brain science data fusion research, it is often faced with the challenge of high dimension and small sample. Therefore, the CERF is proposed in this paper. Clustering evolution and random forest are combined to realize an adaptive ensemble learner. Through hierarchical clustering of decision trees in random forest, the most recognizable features between HC and AD are gradually selected from the high-dimensional features. The design idea of CERF is

displayed in Fig. 2, and the concrete realization of CERF is as follows.

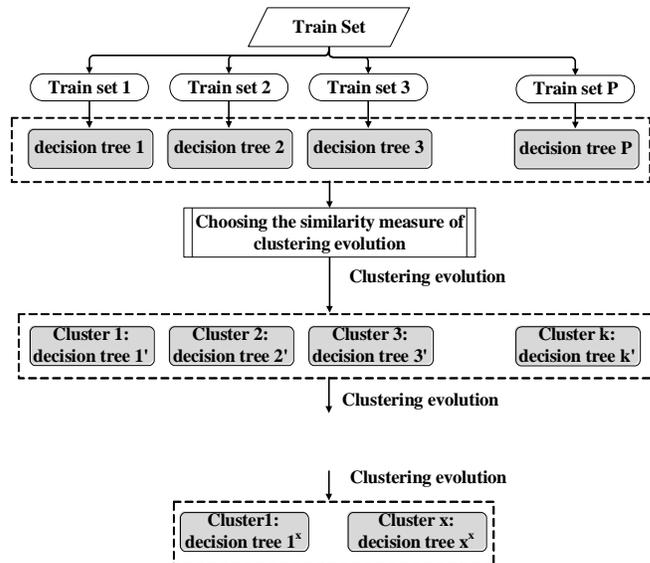


Fig. 2. The design of CERF. The idea of clustering evolution is introduced to improve the diversity and accuracy of base classifiers through repeated clustering evolutions.

We suppose that the sample set is $S = \{x, y\}_{n=1}^N$. Then, the train set $A = \{x_a^A, y_a^A\}_{a=1}^{n_A}$, validation set $B = \{x_b^B, y_b^B\}_{b=1}^{n_B}$ and test set $C = \{x_c^C, y_c^C\}_{c=1}^{n_C}$ are randomly selected from the sample set according to the ratio of 5:3:2. Specifically, $x_a^A = \{BG_a^1, BG_a^2, \dots, BG_a^m\}$ and $y_a^A = \{-1, +1\}$, where x_a^A represents a sample in the train set A, BG_a^m indicates the m-th feature of the sample, y_a^A indicates the corresponding class label of the sample, “+1” is HC and “-1” is AD.

According to the splitting strategy for data set, the train set A and the corresponding validation set B are obtained. The G features and H samples are randomly extracted from the train set A, the value of G is defined as follows.

$$G = \text{fix}(\sqrt{m}) \quad (1)$$

where m represents the total quantity of sample feature, $\text{fix}(x)$ is rounding function. Then, we use the Gini index to find the optimal classification points of different features and construct a decision tree (please see Eq.2).

$$\text{GINI}(A) = 1 - \sum P_k^2 \quad (2)$$

where P_k represents the probability of which the classification result is k. Specifically, we use the BG^m to represent the m-th feature of all samples. When the value of feature BG^m is j, the Gini index is calculated by Eq.3

$$\text{Gain_GINI}_{BG^m, j}(A) = \frac{n_1}{N} \text{GINI}(A1) + \frac{n_2}{N} \text{GINI}(A2) \quad (3)$$

where N is the total quantity of samples in the train set A, n_1 and n_2 are the number of samples in the sample subset A1 and A2 respectively. Next, we calculate the Gini coefficients of all values of feature BG^m , and select the value corresponding to the minimum Gini coefficient as the optimal classification point. Furthermore, the optimal binary classification points of all features are calculated according to the Eq.2-3, a decision

tree is constructed.

The construction of a decision tree mentioned above is repeated for P times. P decision trees are obtained, and assembled into the initial random forest.

The similarities between decision trees in the initial random forest are taken as clustering criteria to construct the CERF. We try to use disagreement measure, relevance measure, and kappa similarity measure to detect the similarities. By comparing the classification performance of CERF based on different similarity measures on the test set, the study will select the best one as the similarity measure of decision trees. Here are the formulas for the three indicators.

The formula for calculating disagreement measure is Eq.4. The smaller value of $DM_{i,j}$ is, the greater the similarity is.

$$DM_{i,j} = \frac{R_{i-j} + R_{j-i}}{T_{ij} + R_{i-j} + R_{j-i} + F_{ij}} \quad (4)$$

The expression for calculating the kappa similarity measure is Eq.5.

$$\text{kappa}_{i,j} = \frac{2(T_{ij}F_{ij} - R_{i-j}R_{j-i})}{(T_{ij} + R_{i-j})(T_{ij} + R_{j-i})(F_{ij} + R_{i-j})(F_{ij} + R_{j-i})} \quad (5)$$

Finally, the relevance measure is defined as

$$RM_{i,j} = \frac{T_{ij}F_{ij} - R_{i-j}R_{j-i}}{\sqrt{(T_{ij} + R_{i-j})(T_{ij} + R_{j-i})(F_{ij} + R_{i-j})(F_{ij} + R_{j-i})}} \quad (6)$$

The parameters of the above formulas are explained as follows. We hypothesize that dt_i and dt_j represent two decision trees. T_{ij} indicates the quantity of samples that can be correctly classified by dt_i and dt_j in the train set. R_{i-j} indicates the quantity of samples that is only correctly classified by dt_i in the train set. R_{j-i} indicates the quantity of samples that is only correctly classified by dt_j in the train set. F_{ij} indicates the quantity of samples that is classified incorrectly by dt_i and dt_j in the train set.

In order to carry out the clustering evolution of the initial random forest, we calculate the similarities between decision trees according to the above methods, and form the similarity matrix M_{dt} which is defined as

$$M_{dt} = \begin{bmatrix} SM_{1,1} & \cdots & SM_{1,q} \\ \vdots & \ddots & \vdots \\ SM_{q,1} & \cdots & SM_{q,q} \end{bmatrix} \quad (7)$$

where $SM_{q,1}$ indicates the similarities between decision tree dt_1 and decision tree dt_q , and is calculated by disagreement measure, relevance measure or kappa similarity measure. If the similarity value $SM_{q,1}$ between dt_1 and dt_q is the highest, the two decision trees will be considered as a cluster. According to the strategy, we carry out the agglomerative hierarchical clustering from bottom to top, the linkage algorithm is used to cluster most similar decision trees into one cluster. Then, the decision trees in the initial random forest are clustered into several clusters. We test the classification accuracies of decision trees in each cluster by the Eq.8

$$\text{Acc}_z = \frac{W_z}{W} \quad (8)$$

where Acc_z indicates the accuracy of the decision tree z , W_z indicates the quantity of samples that are classified correctly by the decision tree z in the validation set, and W indicates the sample size of the validation set. The decision tree with the highest accuracy in a cluster is selected as the representative of the cluster to construct an improved random forest. The above process is called as clustering evolution. The process of clustering evolution is repeated until the generalization performance of random forest reaches its peak. Then we stop clustering evolution and get the final random forest. In clustering evolution process, the decision trees with the highest similarity will be clustered into one cluster, and the number of decision trees in random forest will be reduced. By setting the step size of clustering evolution E , the number of decision trees reduced is E in each clustering evolution. Its main function is to control the clustering evolution process of random forest, avoid the clustering process too fast, to ensure that the decision trees with large similarity will be the one cluster, and then retain the decision trees with the best classification performance. After several clustering evolutions, the amount of decision trees in the CERF can be defined as

$$K = D - iE \quad (i = 1, 2, 3 \dots n) \quad (9)$$

where D is the amount of decision trees in the initial random forest, i is the clustering evolution times, E is the step size of clustering evolution.

The construction process is summarized in Algorithm 1. The input of the algorithm is the sample set, and the output is a clustering evolutionary random forest.

Algorithm 1 CERF learning process

Input: experimental data set $\{X, Y\}$
Output: The clustering evolutionary random forest

- 1: Initialize $\{X, Y\}$, D , E , i .
- 2: $\{X, Y\}$ is experimental data set,
- 3: D is the number of initial decision trees
- 4: E is the step size of clustering evolution
- 5: i is the clustering evolution times.
- 6: Partitioned the $\{X, Y\}$ into $\{X, Y\}_{\text{tra}_1}, \{X, Y\}_{\text{val}_1}, \dots$,
- 7: $\{X, Y\}_{\text{tra}_n}, \{X, Y\}_{\text{val}_n}$ and $\{X, Y\}_{\text{test}}$
- 8: $i=1$
- 9: **for** $k = 1$ to D :
 10: select $\{X, Y\}_{\text{tra}_k}$
 11: Randomly select a subset of features as $\{\text{Features}\}_{\text{tra}_k}$
 12: $\{X, Y\}_{\text{tra}_k}$ and $\{\text{Features}\}_{\text{tra}_k} \rightarrow$ decision tree $\{Tb_k\}$
- 13: $\{X, Y\}_{\text{val}_k} \rightarrow$ test the classification accuracy of decision tree $_k$
- 14: **end for**
- 15: Random Forest = Ensemble of decision trees $\{Tb_1, \dots, Tb_D\}$
- 16: Choose a similarity measure from Eq. (4-6)
- 17: Computing the similarities between decision trees
- 18: **Do**
 19: Clustering the decision trees in the Random Forest into several clusters
 20: Retain the decision trees with the highest accuracy in each cluster \rightarrow remove the inefficient trees
- 21: $D = n - iE$, $i=i+1$
- 22: $\{\text{Random Forest}\}_{\text{new}} =$ Ensemble of the retained decision trees
- 23: **until** the accuracy of $\{\text{Random Forest}\}_{\text{new}}$ reaches the peak

D. Classification Method of Clustering Evolutionary Random Forest

Let us assume that “+1” denotes HC and “-1” denotes AD. We classify the unlabeled sample T with the final random forest after clustering evolutions, and the plurality voting method is used to determine the final classification result. The specific voting method is shown in Eq.10-11. Firstly, we count the voting results of the decision trees in the final random forest using Eq.10.

$$R_l = \sum_{i=1}^n dt_{l,T}(T) \quad (10)$$

where R_l denotes the total votes of the class label l , and $dt_{l,T}(T)$ denotes the voting result of a decision tree in the final random forest. If the voting result is l , the value of $dt_{l,T}(T)$ is “1”, otherwise the value is “0”. Subsequently, according to Eq.11, we choose the label with the largest number of votes as the class label of the sample.

$$\text{Result} = \text{Arg max}(R) \quad (11)$$

More concretely, for an unlabeled sample T , if more than half of the decision trees in the final random forest consider the sample is positive, the final classification result is normal. Otherwise, the result of the sample classification is patient.

E. Abnormal “Brain Region-Gene Pairs” Analysis

Besides the classification, the CERF also can find out the discriminative “brain region-gene pairs” between HC and AD, and extract abnormal brain regions and pathogenic genes, which is the other aim of our study. The detailed analysis process is as follows.

Initially, the frequencies of “brain region-gene pairs” selected by the decision trees in the final random forest are counted, and the r features with higher frequencies are regarded as “important brain region-gene pairs”. The formula for calculating frequency is defined as

$$\text{FBG}_j = \sum_{i=1}^m \text{FBG}_{i,j} \quad (12)$$

where FBG_j is the total frequency of the j -th “brain region-gene pair” and $\text{FBG}_{i,j}$ is the frequency of the j -th dimension feature in the i -th decision tree. Although important features have a great contribution to classification, there may still be some redundant or irrelevant features. Subsequently, different subsets are extracted from “important brain region-gene pairs” and taken as the input features of traditional random forest. The classification abilities of these subsets are reflected by the classification accuracy of random forest. We get the subset with the strongest classification ability which is the “optimal brain region-gene pairs”. The difference in optimal features between AD and HC is more obvious, which means that the brain regions contained in optimal features are more likely to have functional or structural lesions, and the genes contained in optimal features are more likely to have abnormal expression. Through the analysis of these optimal fusion features, it is more likely to determine the abnormal brain regions and risk genes. Finally, we further count the frequencies of brain regions and genes in the “optimal brain region-gene pairs”, and select brain

regions and genes with higher frequencies as abnormal brain regions and pathogenic genes.

F. Parameters Optimization

There are two free parameters in the proposed CERF method, which are the number of initial decision trees and the clustering evolution times. The two parameters should be optimized for achieving the best performance of CERF. The specific optimization methods are as follows.

Firstly, we hypothesize that the number of initial decision trees and the clustering evolution times are in large interval $[a, b]$ and $[c, d]$ respectively. Then, we carry out multiple clustering evolutions to make the classification performance of random forest with different number of initial decision trees reach the peak value, and calculate the optimal clustering evolution times corresponding to the different amounts of initial decision trees. Finally, the optimal combination of the two parameters is determined by the following criterion. When the number of initial decision trees is D , the clustering evolution times required for the random forest performance to reach the peak value are the least, which helps ensure the efficiency of CERF construction.

III. EXPERIMENT RESULTS

A. The Result of Fusion Feature Construction

After preprocessing of multimodal data, the fMRI voxels were clustered to 90 brain regions by AAL template, and the 82400 SNPs were retained. According to the fusion feature construction method mentioned in Section II, we extracted the average time series of each brain region, and extracted 36 genes with more than 30 SNPs from 82400 SNPs. Then, the first 60 points of gene sequence were intercepted, and the similar process exist for the brain regions. Finally, three types of correlation coefficients between the gene sequences and brain regions were calculated, including Pearson correlation coefficients, canonical correlation coefficients and distance correlation coefficients. Based on each correlation analysis method mentioned above, 3240 “brain region-gene pairs” were obtained from a sample.

The following contents of Section III are organized as follows: In part B and part C, we take the fusion features based on Pearson correlation analysis as an example to illustrate the construction of CERF and the extraction of optimal features. Part D is the comparison of different correlation analysis methods and the construction of AD comprehensive diagnosis framework. Part E is the extraction of brain regions and genes based on the comprehensive framework.

B. Optimal Clustering Evolutionary Random Forest

By comparing the performances of different similarity measures mentioned in Section II, our study selected the disagreement measure to construct the optimal CERF. The specific process of CERF construction was as follows.

Initially, 36 samples were abstracted randomly from the train set, and $\sqrt{3240} \approx 57$ features were selected randomly from the “brain region-gene pairs” to construct a single decision tree. By

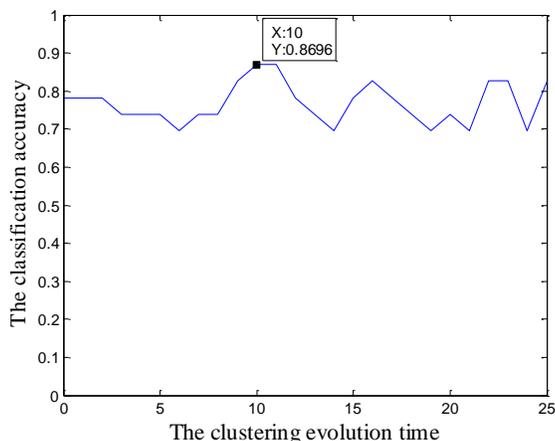


Fig. 3. The changes of the accuracy during the clustering evolutions.

repeating the step for several times, 300 decision trees were constructed and formed the initial random forest. Then, the step size of clustering evolution was set to 10, and the random forest was clustering-evolved for 25 times. We calculated the classification accuracy of random forest on test set after each evolution. The relationship between evolution times and classification accuracy was shown in Fig. 3. When the initial random forest was clustering-evolved for 10 times, the accuracy of the CERF reached the highest. Therefore, when the amount of initial decision trees and the optimal clustering evolution times were 300 and 10 respectively, we obtained a CERF.

In order to obtain the optimal CERF, we further used the optimization strategy mentioned in Section II to optimize the two free parameters of CERF. Firstly, we assumed that the optimal number of initial decision trees and the clustering evolution times for random forest were in the interval of [300,500] and [1,25] respectively. Then, the grid search strategy was used to search for the optimal combination of parameters. Specifically, we gradually increased the number of initial decision trees from 300 to 320, 340, 360, ..., 500. According to the method of clustering evolution mentioned above, random forests with different initial base classifiers numbers were clustering-evolved to peak of classification

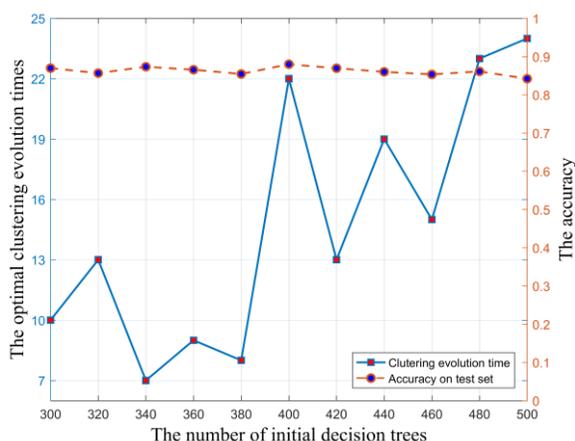


Fig. 4. The relationship curve between the optimal evolution times and the quantity of initial decision trees.

accuracy and we obtained the optimal clustering evolution times corresponding to the different quantities of initial decision trees. In Fig. 4, we showed the highest classification accuracies of random forests with different initial base classifiers numbers and their corresponding optimal clustering evolution times. We found the peak values of random forests with different initial decision trees are close, but the optimal clustering evolution times of random forest with 340 initial decision trees are the least. Therefore, the optimal combination of parameters was (340, 7), because the evolutionary times required to construct the optimal CERF based on this parameter combination are the least under the same classification accuracy peak level.

C. The Extraction of Optimal Fusion Features

The classification accuracy of the final random forest (the initial number of decision trees =340, the clustering evolution times =7) was close to 90%. It is shown that the decision trees with more effective classification features in random forest were preserved through hierarchical clustering evolutions. Therefore, by analyzing the selected features of each decision tree in the final random forest, we could find out the “important brain region-gene pairs” which contributed to classification greatly. The process was as follows.

We extracted all the “brain region-gene pairs” selected by each decision tree in the final random forest, and then counted the frequency of each “brain region-gene pair”. The higher the frequency was, the greater difference between HC and AD patients was. As a result, the top 400 “brain region-gene pairs” with higher frequencies were considered to be the “important brain region-gene pairs”, and the features with frequencies greater than 12 were listed in Table II. The frequency was used as a criterion to select “important brain region-gene pairs”, which may contain some inefficient or redundant features. With the intention of finding the “optimal brain region-gene pairs” with the strongest distinguishing ability, it is necessary to

TABLE II
THE “IMPORTANT BRAIN REGION-GENE PAIRS” WITH FREQUENCY GREATER THAN 12

Frequency	Brain region-gene pairs
17	HIP.R-KAZN
16	PCUN.L-NRXN1, THA.L-PTPRD, MTG.L-CTNNA3
15	INS.L-DAB1, LING.R-DLGAP2
14	ORBinf.R-ASTN2, ORBsupmed.R-GRM7, INS.R-CTNNA3, ACG.R-ROBO2, PHG.R-FRMD4A, PCUN.L-MAGI2, PCUN.L-RBFOX1, TPOmid.L-CTNNA2
13	SFGdor.R-ROBO2, MFG.L-OPCML, MFG.L-RYR2, IFGoperc.R-CTNNA2, IFGtriang.L-PCDH15, OLF.R-DAB1, TPOmid.L-DAB1, OLF.R-DAB1, OLF.L-MACROD2, DCG.R-CTNNA2, AMYG.L-RF00019, CAL.L-CTNNA3

screen the “important brain region-gene pairs”. We divided the “important brain region-gene pairs” into several subsets according to the frequency. For example, the first 70 “important brain region-gene pairs” with higher frequencies were the subset 1 and the first 75 “important brain region-gene pairs” were the subset 2. Then the classification performances of these subsets were tested by the traditional random forest composed of 340 decision trees, and the result was shown in Fig. 5. When we used the first 290 “important brain regions-gene pairs” to construct a subset, the classification accuracy of the random forest reached the highest value of 91.3%. Therefore, the first 290 “brain region-gene pairs” were the “optimal brain region-gene pairs”. The top 40 “optimal brain region-gene pairs” were displayed in Fig. 6.

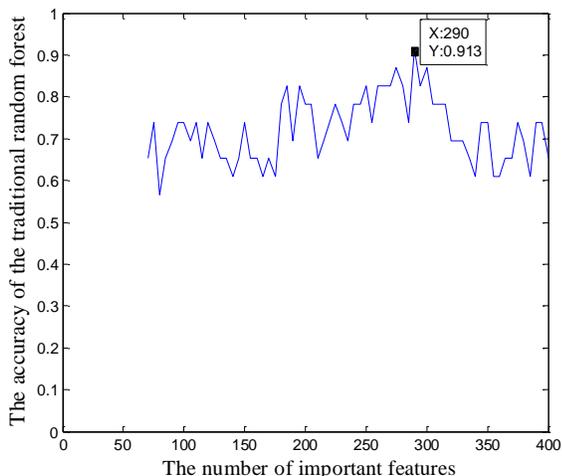


Fig. 5. The accuracies of the traditional random forests based on different subsets.

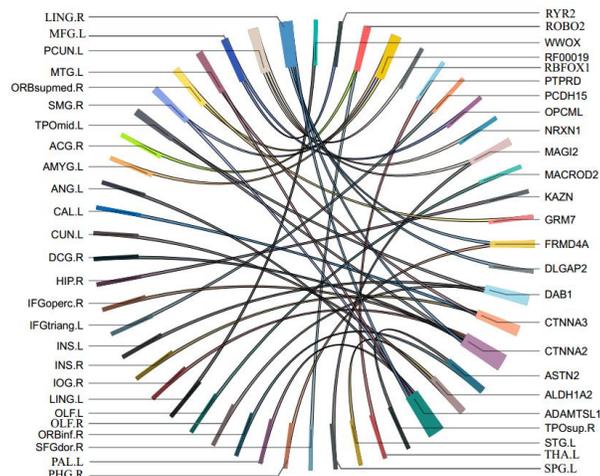


Fig. 6. The top 40 “optimal brain region-genes pairs”. Nodes denote lesion brain regions or pathogenic genes, while edges denote the associations between brain regions and genes.

D. Comparison with Other Methods in Feature Fusion Extraction and Classification

In part B and part C, we constructed fusion features based on Pearson correlation analysis and extracted the optimal fusion features using CERF. Besides the methods mentioned above, we also applied the CCA and DCA to construct fusion features,

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	Discoveries	Classification accuracy of SVM	Overlap with our method
Pearson + CERF	290	0.862	—
Pearson + RF	580	0.827	205 (p = 1.32605e-46)
Pearson + RSVMC	630	0.724	168 (p = 7.29436e-90)
Pearson + t-test	351	0.793	188 (p = 0.000386718)
CCA + t-test	313	0.689	141 (p = 1.00512e-10)
DCA + t-test	329	0.758	182 (p = 0.005151893)

and used other feature extraction methods to extract the optimal fusion features. These feature extraction methods include random forest (RF), random SVM cluster (RSVMC) and two-sample t-test. Finally, we used SVM to evaluate the classification ability of the optimal features extracted by different methods, and the result was shown in Table III.

We found some interesting conclusions from Table III. Firstly, it is observed that the number of the optimal fusion features extracted by CERF was the least, but the classification performance of these features was the highest. Secondly, the optimal features extracted by other methods overlapped with those extracted by CERF. At the same time, through hypergeometric test, we found that these overlaps were not random. Finally, we also noted that the more overlaps these methods had with CERF, the higher the classification accuracy they had. Furthermore, by analyzing the overlaps of features, we found that some significant associations between genes and brain regions including ACG.R-LRP1B, ANG.L-ASIC2, STG.L-CDH13, IPL.R-CNTN5, IOG.R-CTNNA2 and PCUN.L-MAGI2. These associations could be detected by different correlation analysis methods and feature extraction methods, which suggests that the associations between brain regions and genes may be strong, the difference between AD and HC may be more obvious. It may need to be focused in future research. Based on the above analysis, we could find that the method of Pearson + CERF extracted the least number of optimal features and was the most reliable among all the above methods, because it could avoid the false positive situation effectively. Therefore, the optimal fusion features extracted by Pearson + CERF were the most reasonable.

With the purpose of further testing the performance of the Pearson + CERF method, we used 340 as the optimal amount of initial decision trees, and set the clustering evolution times to 2, 4, 6, 7, 8 respectively. Under different circumstances, we conducted 50 independent experiments to test the classification performance of this method. In addition, the method was also compared with unimodal and multimodal t-test to comprehensively evaluate its performance. In unimodal experiments, t-test was used as feature selection tool and SVM was used as sample classifier. In fMRI-based t-test, functional connectivity is the sample feature, and in SNP-based t-test, single SNP is the sample feature. The results were show in Fig.

7.

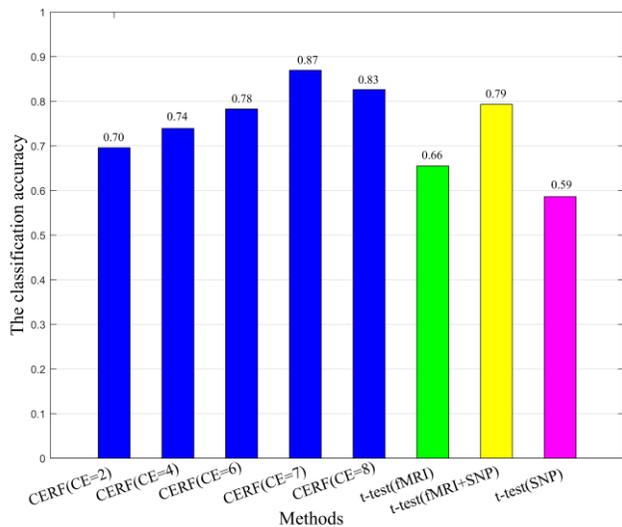


Fig. 7. Classification accuracies of CERFs in different clustering evolution times, and comparison with unimodal and multimodal t-test. CE is the cluster evolution times, and the classification accuracy of all CERFs is obtained by multimodal test. T-test means extracting features by t-test and using SVM as classifier.

At the initial stage of clustering evolution, there was a positive correlation between the classification accuracy of this method and the clustering evolution times. When the peak performance was reached, the classification performance of this method might decline if the clustering evolution was continued. This is because the quantity of decision trees is insufficient and the performance of ensemble learner is fluctuant. Therefore, we set the number of initial base classifiers to 340 and the number of clustering evolution times to 7, which is the best balance between performance and resource consumption. Then, we compared the performance of t-test in multimodal fusion features with that in the unimodal features. We found that the multimodal fusion features effectively improve the performance of the t-test, indicating that the gene and fMRI data fusion have complementary advantages. Furthermore, compared with t-test in multimodal fusion features, we found that the CERF has obvious advantages.

Finally, in order to verify the robustness and generalization of the model under the small sample size, the Pearson + CERF model was extended to the multimodal data fusion study of Parkinson's disease (PD) and early mild cognitive impairment (EMCI). We constructed fusion features and optimal CERF based on the two different diseases data. The experimental dataset information and the key parameters of CERF were listed in Table IV. We also conducted 20 independent performance tests on CERF based on these two other different datasets, and the results were shown in Fig. 8.

The results in Table IV show that Pearson + CERF model had good generalization, which can achieve satisfactory classification performance in different types of brain disease multimodal data analysis tasks through simple parameter adjustment. At the same time, the performance curve in Figure 8 shows that the Pearson + CERF method is relatively stable in different data classification tasks, which also proves the

stability of the method. The above analysis proved that the Pearson + CERF method has satisfactory ability in feature extraction and sample classification.

TABLE IV
MODEL VALIDATION EXPERIMENTS ON DIFFERENT DATASETS

Dataset	Base classifier number	CE times	Optimal features number	Average accuracy
37AD + 35HC	340	7	290	81.0%
37EMCI + 36HC	400	8	305	80.0%
55PD + 49HC	400	5	260	83.0%

Notes: CE is the cluster evolution times.

37EMCI+36HC: The dataset was obtained from ADNI.

55PD + 49HC: The dataset was obtained from Parkinson's progression markers initiative (PPMI).

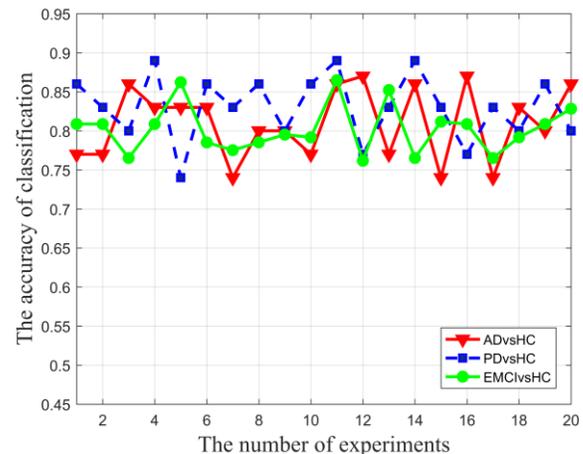


Fig. 8. Classification accuracy curve of Pearson + CERF models based on different datasets.

E. The Extraction of Abnormal Brain Regions and Genes

From the above experimental results, we could learn that the “optimal brain region-gene pairs” extracted by the method of Pearson + CERF had the best recognition ability for samples, and could be used to extract abnormal brain regions and pathogenic genes of AD. These features differed greatly between AD patients and HC, and were associated with brain regions and genes. Thus, we counted the frequencies of the brain regions and genes in “optimal brain region-gene pairs” as weights respectively. The brain regions and genes with the largest weights were abnormal brain regions and pathogenic genes of AD, which included precuneus, lingual gyrus, angular gyrus, insula, thalamic, DAB1 gene and LRP1B gene. The locations of the abnormal brain regions were displayed in Fig.9, and the main information of pathogenic genes was shown in Fig. 10.

IV. DISCUSSION

A. Comparison with Existing Studies

Finding clear disease mechanism of AD across multiple data sets is a common challenge in brain science. In recent years, the related studies of AD have been carried out by some researchers. Our study is similar but different from the existing researches.

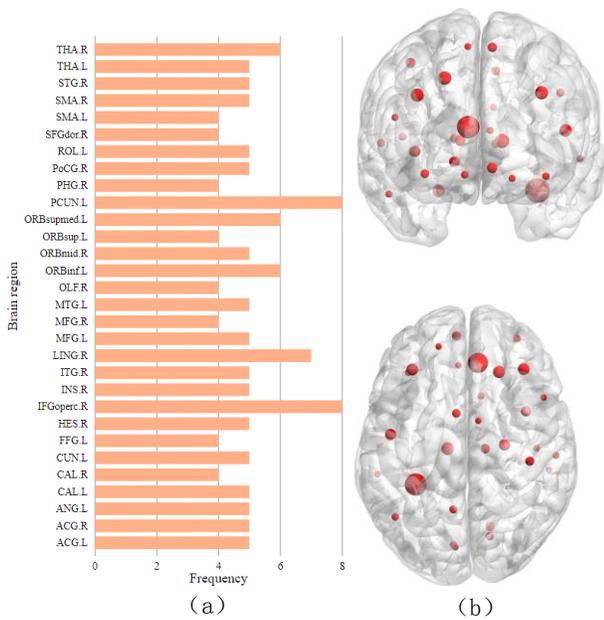


Fig. 9. The main abnormal brain regions. The weights of different brain regions are displayed in Fig. 9 (a). The locations of different brain regions in the brain are showed in Fig. 9 (b). The larger the node size of the brain region is, the more obvious correlation between AD and the brain region is.

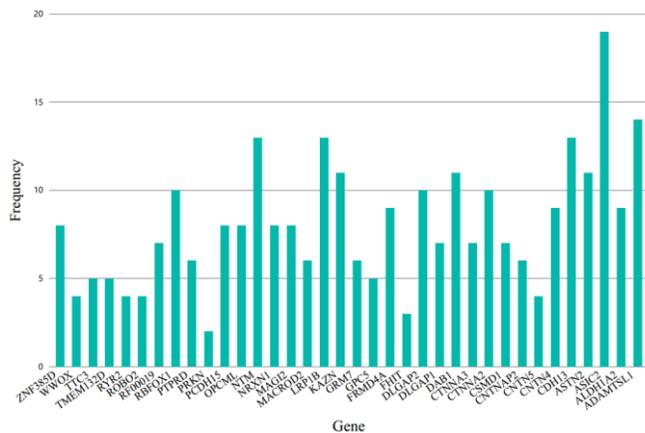


Fig.10. The frequency of main pathogenic genes. The higher the frequency of the gene is, the more significant correlation between the gene and AD is.

Firstly, in the selection of multimodal data, the existing researches mainly focus on the fusion of different modal neuroimaging data, or the fusion of neuroimaging data and cognitive performance scale data. For example, de Vos et al. [18] combined cortical thickness, cortical curvature and subcortical volume to predict AD using elastic network logistic regression. Tong et al. [19] also fused the regional MRI volumes and FDG-PET signal intensities to explore the multiple biomarkers of AD and mild cognitive impairment (MCI). Altaf et al. [20] combined image information with clinical features to classify AD and normal people by gray level co-occurrence matrix. These studies were less likely to reveal etiologies of AD at other levels besides neuroimaging. To fulfill this gap, fMRI and gene data were fused in our study, which expanded the search scope for AD etiologies. We could not only detect abnormal brain regions of AD, but also further identify pathogenic genes.

Secondly, in the classification of samples, the CERF is

proposed as a novel ensemble learning method. Compared with the existing classification methods, the classification accuracy of CERF model is close to 90%. In comparison with the classification methods of single base learner [21], [22], the decision trees in CERF are screened by clustering evolutions. In our method, only the decision trees with higher accuracies are retained, which overcomes the performance fluctuation of single base learner. By comparing with traditional ensemble learning classification methods [23], [24], the proposed model introduces the idea of hierarchical clustering to carry out clustering evolutions of random forest, which enhances the diversities among base learners in ensemble learning model. Furthermore, compared with the current popular deep learning methods, such as CNN [25], [26], the advantages of our method in the case of small samples mainly come from the following aspects. Firstly, at the feature learning aspect, CERF learns partly specific feature subsets each time by randomly selecting samples and features, and has good applicability to different types of data features, which can play the multimodal data complementary advantages of gene and fMRI, while CNN emphasizes the learn of fMRI image features. On the other hand, our method optimizes the ensemble learner by clustering evolutions, which reduces the sensitivity of the method to the data scale. However, most of CNN's parameter optimization is based on back propagation, so it is less likely to get a satisfactory result in small sample size.

Ultimately, in the extraction of features, there is less possibility to explain the extracted features intuitively in these existing methods [27], [28], especially the improved methods based on PCA or ICA. Compared with these previous studies, we used CERF to select the discriminative features between AD and HC. These features indicate specific brain regions and genes, and thereby the abnormal brain regions and genes could be directly detected.

B. Abnormal Brain Regions and Pathogenic Genes

In this paper, we demonstrated that there were associations between brain regions and genes, and the abnormality of this association mechanism may be a potential factor of neurological diseases. For example, the DAB1-SFGdor.R pair was included in the “optimal brain region-gene pairs” implies that the interaction between DAB1 and SFGdor.R is abnormal. In fact, DAB1 gene is abnormally expressed in SFGdor.R in most patients with neurodegenerative diseases [29]. Additionally, the abnormal brain regions and pathogenic genes of AD were observed based on the “optimal brain region-gene pairs”. We detected some typical brain regions and genes that had been proved to be associated with AD in previous studies such as thalamus, lingual gyrus, angular gyrus and DAB1 gene.

The thalamus is a typical brain region of AD detected by CERF. There is a dependence between thalamus and posterior cerebral cortex, which works on spatial learning together [30]. The thalamus is viewed as the major regulator of numerous fields in AD, and leads to several alterations of AD in cognitive and behavioral [31]. Additionally, the radioactivity abnormalities in thalamus may lead to indifference and irritability in AD patients [32]. The disabled-1 (DAB1) belongs

to protein-coding gene. DAB1 gene plays a role in brain development, guiding the migration of cortical neurons through the formed neurons to reach their appropriate layer [33]. The result of GWAS found that DAB1 gene was one of the susceptible genes for AD [34]. The abnormal expression of DAB1 gene in AD patients results in the deregulation of the cellular proteome [35]. These conclusions are consistent with our experimental results.

Interestingly, besides these typical brain regions and genes, we also found some brain regions and genes subtly associated with AD containing the precuneus, insula, and LRP1B gene, which play crucial roles in the process of AD and are easy to be overlooked.

The precuneus with a higher frequency plays an essential role in classification, which has a significant impact on human memory [36]. Quantitative cerebral blood flow (CBF) values and the utilization rate of glucose in the precuneus were decreased, which can reduce oxygenation of the precuneus in early AD patients [37]. These metabolic abnormalities in the precuneus are significant causes of cognitive impairment in AD patients and can be used as biomarkers. The insula also contributes greatly to the sample classification, and is considered to be associated with working memory [38]. According to the existing studies, amyloid protein tau is the key biomarker of AD, and the insula cortex of AD patients is affected by tau immunoreactive neurofibrillary tangle pathology [39]. The changes of insula also are a main cause of cognitive abnormalities in AD patients [40]. The Haplotypes in lipoprotein receptor-related protein 1B (LRP1B) can protect the aged from cognitive decline [41]. The abnormal expression of LRP1B as a ligand is a key factor in the pathogenesis of AD, which can inhibit C1q-mediated neuroprotection [42]. The genetic association between LRP1B and AD is further demonstrated by subsequent studies. For instance, increasing evidences suggest that the interaction between β -amyloid precursor protein and LRP1B is also the cause of AD [43]. The discovery of above-mentioned abnormal brain regions and genes is helpful to understand the changes of emotion and memory in AD patients.

C. Limitation and Future Directions

Multimodal data fusion remains an open challenge in brain researches. Although our research has made progresses in the classification and feature extraction of AD, there are still some limitations in our work. On the one hand, we use classical correlation analysis methods to detect associations between brain regions and genes, which may overlook some significant associations. We will design more appropriate indicators to capture the correlation between fMRI and gene data in the follow-up work. On the other hand, our research mainly considers the fusion of fMRI and gene data, and the data scale is still small. However, AD has many biomarkers including proteins and metabolites. Therefore, in the future work, we will expand the scale of data, and use transfer learning and other methods to integrate MRI data from OASIS and other data sets into our research. At the same time, we will also consider fusing more multimodal data in the future work.

V. CONCLUSIONS

In this paper, we carry out the multimodal data fusion research of AD. The contributions of the work are highlighted as follows. Firstly, we apply the correlation analysis to detecting the associations between brain regions and genes. Our method can efficiently fuse the data from different modalities, and facilitate the follow-up analysis. Secondly, the CERF is proposed to analyze “brain region-gene pairs”, and extract the discriminative fusion features between AD and HC. Finally, the CERF is integrated into the comprehensive diagnostic framework of AD, which included fusion feature construction, feature selection and sample classification. Based on this framework, we find out the abnormal brain regions and pathogenic genes of AD such as thalamic, lingual gyrus, angular gyrus, precuneus, insula DAB1 gene and LRP1B gene. But it needs to be stated that validation of results with larger data sets as the need to validate these conclusions and “brain region-gene pairs” in future work. Our study is of great significance for the diagnosis of AD and the development of computational medicine.

REFERENCES

- 1 S. Mancini, C. Balducci, E. Micotti, D. Tolomeo, G. Forloni, M. Masserini, *et al.*, "Multifunctional liposomes delay phenotype progression and prevent memory impairment in a presymptomatic stage mouse model of Alzheimer disease," *Journal of Controlled Release*, 258(1):121-129, 2017.
- 2 "2019 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, 15(3):321-387, 2019.
- 3 A. Tramutola, C. Lanzillotta, M. Perluigi, and D. A. Butterfield, "Oxidative stress, protein modification and Alzheimer disease," *Brain Research Bulletin*, 133:88-96, 2017.
- 4 R. Marchitelli, M. Aiello, A. Cachia, M. Quarantelli, C. Cavaliere, A. Postiglione, *et al.*, "Simultaneous resting-state FDG-PET/fMRI in Alzheimer Disease: Relationship between glucose metabolism and intrinsic activity," *NeuroImage*, 176:246-258, 2018.
- 5 R. Sims, S. J. van der Lee, A. C. Naj, C. Bellenguez, N. Badarinarayan, J. Jakobsdottir, *et al.*, "Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease," *Nature Genetics*, 49(1):1373-1384, 2017.
- 6 T. Zhou, M. Liu, K. Thung, and D. Shen, "Latent Representation Learning for Alzheimer's Disease Diagnosis with Incomplete Multi-modality Neuroimaging and Genetic Data," *IEEE Transactions on Medical Imaging*, 38(10):2411-2422, 2019.
- 7 C. He, Q. Liu, H. Li, and H. Wang, "Multimodal medical image fusion based on IHS and PCA," *Procedia Engineering*, 7:280-285, 2010.
- 8 W. Hu, D. Lin, S. Cao, J. Liu, J. Chen, V. D. Calhoun, *et al.*, "Adaptive Sparse Multiple Canonical Correlation Analysis With Application to Imaging (Epi)Genomics Study of Schizophrenia," *IEEE Transactions on Biomedical Engineering*, 65(2):390-399, 2018.
- 9 S. Liang, Y. Li, Z. Zhang, X. Kong, Q. Wang, W. Deng, *et al.*, "Classification of First-Episode Schizophrenia Using Multimodal Brain Features: A Combined Structural and Diffusion Imaging Study," *Schizophrenia Bulletin*, 45(3):591-599, 2018.
- 10 X. Zhu, H.-I. Suk, and D. Shen, "Low-rank dimensionality reduction for multi-modality neurodegenerative disease identification," *World Wide Web*, 22(2):907-925, 2019.
- 11 J. Fang, C. Xu, P. Zille, D. Lin, H. Deng, V. D. Calhoun, *et al.*, "Fast and Accurate Detection of Complex Imaging Genetics Associations Based on Greedy Projected Distance Correlation," *IEEE Transactions on Medical Imaging*, 37(4):860-870, 2018.
- 12 N. Jie, M. Zhu, X. Ma, E. A. Osuch, M. Wammes, J. Thøerge, *et al.*, "Discriminating Bipolar Disorder From Major Depression Based on

- SVM-FoBa: Efficient Feature Selection With Multimodal Brain Imaging Data," *IEEE Transactions on Autonomous Mental Development*, 7(4):320-331, 2015.
- 13 V. M. Vergara, A. R. Mayer, E. Damaraju, K. A. Kiehl, and V. Calhoun, "Detection of Mild Traumatic Brain Injury by Machine Learning Classification Using Resting State Functional Network Connectivity and Fractional Anisotropy," *Journal of Neurotrauma*, 34(5):1045-1053, 2016.
- 14 J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease," *IEEE Journal of Biomedical and Health Informatics*, 22(1):173-183, 2018.
- 15 X. Zhu, H.-I. Suk, L. Wang, S.-W. Lee, and D. Shen, "A novel relational regularization feature selection method for joint regression and classification in AD diagnosis," *Medical Image Analysis*, 38:205-214, 2017.
- 16 A. Chaddad, C. Desrosiers, and T. Niazi, "Deep Radiomic Analysis of MRI Related to Alzheimer's Disease," *IEEE Access*, 6:58213-58221, 2018.
- 17 H. Choi and K. H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behavioural Brain Research*, 344:103-109, 2018.
- 18 F. de Vos, T. M. Schouten, A. Hafkemeijer, E. G. P. Dopper, J. C. van Swieten, M. de Rooij, *et al.*, "Combining multiple anatomical MRI measures improves Alzheimer's disease classification," *Human Brain Mapping*, 37(5):1920-1929, 2016.
- 19 T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognition*, 63(1):171-181, 2017.
- 20 T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, 43(1):64-74, 2018.
- 21 D. Zafeiris, S. Rutella, and G. R. Ball, "An Artificial Neural Network Integrated Pipeline for Biomarker Discovery Using Alzheimer's Disease as a Case Study," *Computational and Structural Biotechnology Journal*, 16(1):77-87, 2018.
- 22 N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, and Y. Li, "A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease," *Neurocomputing*, 320:195-202, 2018.
- 23 X.-a. Bi, Q. Jiang, Q. Sun, Q. Shu, and Y. Liu, "Analysis of Alzheimer's Disease Based on the Random Neural Network Cluster in fMRI," *Frontiers in Neuroinformatics*, 12(60):1-10, 2018.
- 24 X.-a. Bi, Q. Shu, Q. Sun, and Q. Xu, "Random support vector machine cluster analysis of resting-state fMRI in Alzheimer's disease," *PLOS ONE*, 13(3):e0194479, 2018.
- 25 A. Chaddad, M. Toews, C. Desrosiers, and T. Niazi, "Deep Radiomic Analysis Based on Modeling Information Flow in Convolutional Neural Networks," *IEEE Access*, 7:97242-97252, 2019.
- 26 N. Zeng, Z. Wang, H. Zhang, K. Kim, Y. Li, and X. Liu, "An Improved Particle Filter With a Novel Hybrid Proposal Distribution for Quantitative Analysis of Gold Immunochromatographic Strips," *IEEE Transactions on Nanotechnology*, 18:819-829, 2019.
- 27 D. Lin, J. Chen, S. Ehrlich, J. R. Bustillo, N. Perrone-Bizzozero, E. Walton, *et al.*, "Cross-Tissue Exploration of Genetic and Epigenetic Effects on Brain Gray Matter in Schizophrenia," *Schizophrenia Bulletin*, 44(2):443-452, 2017.
- 28 J. M. Ford, B. J. Roach, V. A. Palzes, and D. H. Mathalon, "Using concurrent EEG and fMRI to probe the state of the brain in schizophrenia," *NeuroImage: Clinical*, 12:429-441, 2016.
- 29 S. H. Fatemi, A. V. Snow, J. M. Strydom, M. Araghi-Niknam, T. J. Reutiman, S. Lee, *et al.*, "Reelin signaling is impaired in autism," *Biological Psychiatry*, 57(7):777-787, 2005.
- 30 J. P. Aggleton, A. Pralus, A. J. D. Nelson, and M. Hornberger, "Thalamic pathology and memory loss in early Alzheimer's disease: moving the focus from the medial temporal lobe to Papez circuit," *Brain*, 139(7):1877-1890, 2016.
- 31 R. Jagirdar and J. Chin, "Corticothalamic network dysfunction and Alzheimer's disease," *Brain Research*, 1702:38-45, 2019.
- M. Torso, L. Serra, G. Giuliatti, B. Spanò, E. Tuzzi, G. Koch, *et al.*, "Strategic Lesions in the Anterior Thalamic Radiation and Apathy in Early Alzheimer's Disease," *PLOS ONE*, 10(5):e0124998, 2015.
- 33 T. Pohlkamp, L. Xiao, R. Sultana, A. Bepari, H. H. Bock, M. Henkemeyer, *et al.*, "Ephrin Bs and canonical Reelin signalling," *Nature*, 539(7630):E4-E6, 2016.
- 34 D. Harold, R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M. L. Hamshere, *et al.*, "Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease," *Nature genetics*, 41(10):1088-1093, 2009.
- 35 T. Muller, C. Loosse, A. Schrotter, A. Schnabel, S. Helling, R. Egensperger, *et al.*, "The AICD Interacting Protein DAB1 is Up-Regulated in Alzheimer Frontal Cortex Brain Samples and Causes Deregulation of Proteins Involved in Gene Expression Changes," *Current Alzheimer Research*, 8(5):573-582, 2011.
- 36 G. Koch, S. Bonnì, M. C. Pellicciari, E. P. Casula, M. Mancini, R. Esposito, *et al.*, "Transcranial magnetic stimulation of the precuneus enhances memory and neural activity in prodromal Alzheimer's disease," *NeuroImage*, 169:302-311, 2018.
- 37 J. S. Miners, J. C. Palmer, and S. Love, "Pathophysiology of Hypoperfusion of the Precuneus in Early Alzheimer's Disease," *Brain Pathology*, 26(4):533-541, 2016.
- 38 Y. Stern, "Cognitive reserve in ageing and Alzheimer's disease," *The Lancet Neurology*, 11(11):1006-1012, 2012.
- 39 K. A. Josephs, M. E. Murray, J. L. Whitwell, N. Tosakulwong, S. D. Weigand, L. Petrucelli, *et al.*, "Updated TDP-43 in Alzheimer's disease staging scheme," *Acta Neuropathologica*, 131(4):571-585, 2016.
- 40 A. L. Foundas, C. M. Leonard, S. M. Mahoney, O. F. Agee, and K. M. Heilman, "Atrophy of the Hippocampus, Parietal Cortex, and Insula in Alzheimer's Disease: A Volumetric Magnetic Resonance Imaging Study," *Cognitive and Behavioral Neurology*, 10(2):81-89, 1997.
- 41 S. E. Poduslo, R. Huang, and A. Spiro Iii, "A genome screen of successful aging without cognitive decline identifies LRP1B by haplotype analysis," *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153B(1):114-119, 2010.
- 42 M. Benoit, M. Hernandez, M. Dinh, F. Benavente, O. Vasquez, and A. Tenner, "C1q-induced LRP1B and GPR6 Proteins Expressed Early in Alzheimer Disease Mouse Models, Are Essential for the C1q-mediated Protection against Amyloid-Neurotoxicity," *Journal of Biological Chemistry*, 288(1):654-665, 2012.
- 43 M.-P. Marzolo and G. Bu, "Lipoprotein receptors and cholesterol in APP trafficking and proteolytic processing, implications for Alzheimer's disease," *Seminars in Cell & Developmental Biology*, 20(2):191-200, 2009.