

# Pathogeny Detection for Mild Cognitive Impairment via Weighted Evolutionary Random Forest with Brain Imaging and Genetic Data

Xia-an Bi\*, Member, IEEE, Zhaoxu Xing, Wenyan Zhou, Lou Li, Luyun Xu\*

**Abstract**—Medical imaging technology and gene sequencing technology have long been widely used to analyze the pathogenesis and make precise diagnoses of mild cognitive impairment (MCI). However, few studies involve the fusion of radiomics data with genomics data to make full use of the complementarity between different omics to detect pathogenic factors of MCI. This paper performs multimodal fusion analysis based on functional magnetic resonance imaging (fMRI) data and single nucleotide polymorphism (SNP) data of MCI patients. In specific, first, using correlation analysis methods on sequence information of regions of interests (ROIs) and digitalized gene sequences, the fusion features of samples are constructed. Then, introducing weighted evolution strategy into ensemble learning, a novel weighted evolutionary random forest (WERF) model is built to eliminate the inefficient features. Consequently, with the help of WERF, an overall multimodal data analysis framework is established to effectively identify MCI patients and extract pathogenic factors. Based on the data of MCI patients from the ADNI database and compared with some existing popular methods, the superiority in performance of the framework is verified. Our study has great potential to be an effective tool for pathogenic factors detection of MCI.

**Index Terms**—Imaging genetics, Mild cognitive impairment, Weighted evolutionary random forest.

## I. INTRODUCTION

MILD cognitive impairment (MCI) is a kind of cognitive impairment syndrome that has a great risk of turning to dementia [1]. The major symptom of MCI includes the decline of cognitive function, and various functions may also be affected depending on the different brain lesions [2]. With the rapid increment of the worldwide aging population, MCI is

This work was supported by the National Natural Science Foundation of China (Grant No. 62072173), the Natural Science Foundation of Hunan Province, China (Grant No. 2020JJ4432), the Key Scientific Research Projects of Department of Education of Hunan Province (Grant No. 20A296), the Key open project of Key Laboratory of Data Science and Intelligence Education (Hainan Normal University), Ministry of Education (Grant No. DSIE202101), the Hunan Provincial Science and Technology Project Foundation (Grant No. 2018TP1018), and Innovation and Entrepreneurship Training Program of Hunan Xiangjiang Artificial Intelligence Academy.

Xia-an Bi ([bixiaan@hnu.edu.cn](mailto:bixiaan@hnu.edu.cn)), Zhaoxu Xing, Wenyan Zhou, and Lou Li are with Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, Hunan 410081, P.R. China, College of Information Science and Engineering, Hunan Normal University, Changsha, Hunan 410081, P.R. China, and also with Hunan Xiangjiang Artificial Intelligence Academy, Changsha, Hunan 410081, P.R. China.

Luyun Xu ([xuluyun@hunnu.edu.cn](mailto:xuluyun@hunnu.edu.cn)) is with the College of Business, Hunan Normal University, Changsha, Hunan 410081, P.R. China.

\* indicates the corresponding author

increasingly prevalent year by year [3]. Scholars have found that about 10% - 20% of MCI patients convert into Alzheimer's disease (AD) every year, and the rate is ten times higher than that of the normal elderly [4]. Based on the degree of course development, it can be divided into early MCI (EMCI) and late MCI (LMCI), where the latter are at more risk of deterioration. As a progressive disease, a clear understanding of the MCI pathogenesis is of great significance to timely diagnosis and intervention to prevent dementia.

As artificial intelligence technology flourishes, researchers tried to make full use of multi-omics data to deeply mine the hidden pattern that implies the cause of the brain diseases like MCI. On the one hand, radiomics data like functional magnetic resonance imaging (fMRI) can qualitatively and quantitatively detect the regions of interest (ROI) for specific therapy of brain diseases [5-7]. On the other hand, plenty of single nucleotide polymorphisms (SNP) in human genes have been found related to MCI in numerous genome-wide association studies (GWAS) [8, 9], indicating the complex pathogenetic background of MCI.

Based on imaging data and gene data, a new field named imaging genetics emerged to comprehensively study pathogeny with fusion data [10]. However, in the context of big data, the biomedical data are becoming larger in quantity and more complicated in structure, leading to the high dimension of the fusion feature and poor generalization ability of the models [11]. Generally, linear dimensionality reduction methods were commonly applied, including independent component analysis (ICA) [12] and principal component analysis (PCA) [13]. Though these methods could effectively mitigate the impact of high dimensions, useful information may be lost to a certain degree in the procedures.

Furthermore, data fusion [14], feature selection [15], and sample categorization [16] are three core problems in the study of brain diseases. Extensive studies have acquired promising achievements [17]. However, most studies only solve a single problem, failing to construct an integrated framework for comprehensive disease research. Therefore, though certain progress has been made, the results are still insufficient to meet the clinical needs. Hence, an overall framework combining the fusion features construction, features selection, and sample classification is significant to the MCI diagnosis and treatment.

The main objective of this study is to find the discriminative features between EMCI and LMCI patients. We utilize fMRI data and SNP data to conduct an imaging genetics study and propose a novel weighted evolutionary random forest (WERF)

model to reduce the dimension of fusion features. Our main contributions are summarized as follows.

- 1) Through Pearson correlation analysis, fusion features are constructed by combining the time series of brain regions with gene sequences.
- 2) A novel WERF model is proposed to reduce the data dimension through the random selection of sample features and the evolution of base learners.
- 3) Some discriminant genes and ROIs are discovered based on the discriminative sample features.

The above three parts integrated framework to assist the diagnosis practice, which is shown in Fig. 1.

The rest part of this article is organized as follows. Related works are listed in Section II. In Section III, we introduce the methodology and proposed a novel weighted evolutionary random forest framework. In Section IV, we present the experiment results in detail. The discussion and conclusions are described separately in Section V and Section VI.

## II. RELATED WORKS

### A. Pathogeny Mining Methods with Multi-omics Data

Machine learning (ML) methods have received great attention in multi-omics study of brain diseases, some of which have been well-performed in recognizing brain diseases pathogenesis [18]. Based on the connectivity analysis of resting-state fMRI, Khazaei et al. [19] utilized graph measures of functional brain networks to train support vector machine (SVM), achieving high accuracy in the identification of MCI and AD patients from healthy controls (HCs). Lei et al. [20] applied a subspace learning method on fMRI and diffusion tensor imaging (DTI) data and then established a multi-task learning model to effectively extract characteristic features of MCI patients.

Base on gene data, on the one hand, multiple genes are found relevant to the development mechanism of MCI. Bottero et al. [21] analyzed the gene expression data in the blood to identify potential biomarkers that may help distinguish MCI. Zou et al.

[22] found that dihydrolipoamide S-succinyltransferase (DLST) promoter methylation interacted with APOEε4 for the first time, thus affecting the pathogenesis of MCI. Varatharajah et al. [23] used an optimized machine learning approach and found that the expression of CR1 has a great effect on the development of MCI by affecting the immune pathways. On the other hand, extensive research on brain and gene networks shows intensive interactions between each mode [24]. Accordingly, fusion analysis among neuroimaging and genomics data is promising to offer new insight for further exploration of the MCI pathology and promote the diagnosis performance.

### B. Dimensionality Reduction Methods

In the existing multimodal data-based studies, fusion features are usually constructed by calculating the correlation coefficients among data, which may cause excessively high dimensionality that is hard to analyze. Recently, a series of novel dimension reduction methods are presented. For example, Cao et al. [25] proposed a multi-core dimension reduction and oversampling method to enhance the sparsity of brain regions, through which a subset of related brain regions was selected out. Wang et al. [26] proposed a filter-based method to obtain feature subsets with positive information. However, these methods are still prone to fall into the problems of the local optimal solution and slow convergence speed.

## III. METHODOLOGY

### A. Independence Test Methods

The independence test can help to describe data relationships. Two independence test methods this study involved with are introduced in this part. The chi-square test is one common method for independence testing, whose statistic is defined as

$$\chi^2 = \sum_{i=1}^N \frac{r_i - e_i}{e_i}, \quad (1)$$

where  $r_i$  represents the real value in the sample,  $e_i$  represents the mathematical expectations of each value under the

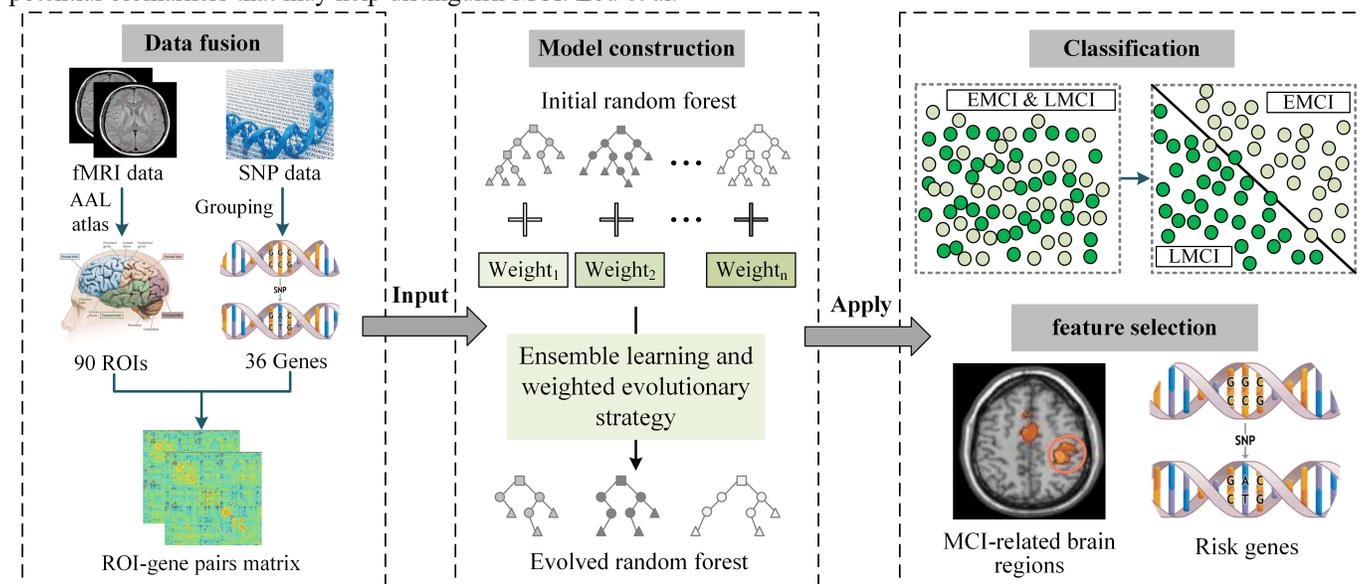


Fig. 1. The overall framework for MCI data analysis. The framework includes three parts: data fusion, model construction and classification and feature selection.

hypothesis that the two factors of the data are independent, and  $N$  represents the number of statistical values. In this paper, we use the chi-square test to assess that the gender factor in our data makes no statistical difference.

The two-sample t-test is another commonly-used method to test the independence between two variables. Let  $A$  and  $B$  be two samples to be tested, the corresponding statistic can be described as

$$t = \frac{\bar{A} - \bar{B}}{\sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}, \quad (2)$$

where  $n_A$  and  $n_B$  represent the sample sizes,  $S_A^2$  and  $S_B^2$  represent the sample variances, and  $\bar{A}$  and  $\bar{B}$  represent mean values. In this study, the two-sample t-test is firstly utilized to verify the independence of the source data. Additionally, in the comparison experiment, we apply a two-sample t-test as a baseline method for feature selection.

### B. Correlation Analysis Methods

In this study, fMRI and SNP data are converted into time series and digital sequences, respectively, and fusion features are therefore constructed. Subsequently, based on correlation analysis, the matrix recording the correlation coefficient between each pair of time series and digital sequences is calculated. By this means, the fusion among different omics of data is realized. In this section, several correlation analysis methods are reviewed. Let  $b$  and  $g$  be column vectors representing the time series and SNP sequences, respectively, and  $b_i$  and  $g_i$  represent each component of these vectors. Thus, the Pearson correlation coefficient can be defined as

$$Pear_{b,g} = \frac{\ell \sum b_i g_i - \sum b_i \sum g_i}{\sqrt{\ell \sum (b_i^2) - (\sum b_i)^2} \sqrt{\ell \sum (g_i^2) - (\sum g_i)^2}}, \quad (3)$$

where  $\ell$  denotes the length of each time series or gene sequence.

Canonical correlation analysis (CCA) is another common correlation analysis method, which can be expressed as

$$Canon_{b,g} = \frac{u^T \Sigma_{12} v}{\sqrt{u^T \Sigma_{11} u} \sqrt{v^T \Sigma_{22} v}}, \quad (4)$$

where  $u$  and  $v$  denote two weighting parameters maximizing the correlation value of paired genes and brain regions,  $\Sigma_{11}$  and  $\Sigma_{22}$  denote the autocorrelation matrices of  $b$  and  $g$ , and  $\Sigma_{12}$  denotes the cross-correlation matrix derived from  $b$  and  $g$ .

The distance correlation coefficient is described as

$$Dis_{b,g} = \frac{dcov(b,g)}{\sqrt{dcov(b,b)} \sqrt{dcov(g,g)}}, \quad (5)$$

where  $dcov(*,*)$  represents the distance covariance, which is a function of two vectors. The distance covariance is defined as follows.

$$dcov(u,v) = S_1 + S_2 - 2S_3, \quad (6)$$

$$S_1 = \frac{1}{n^2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} (dist(u_i, u_j) \cdot dist(v_i, v_j)), \quad (7)$$

$$S_2 = \frac{1}{n^4} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} dist(u_i, u_j) \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} dist(v_i, v_j), \quad (8)$$

$$S_3 = \frac{1}{n^3} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \sum_{k=1}^{\ell} ((dist(u_i, u_k) \cdot dist(v_j, v_k))), \quad (9)$$

where  $dist(*,*)$  represents the function computing the distance.

At last, the Kendall correlation measures the association strength of the cross tabulations, which is defined as

$$\tau = \frac{4P}{\ell(\ell-1)} - 1, \quad (10)$$

where  $P$  represents the number of concordant pairs between two series.

In this study, the Pearson correlation analysis is utilized to construct fusion features, and the others are applied in the comparative experiments.

### C. Traditional Decision Tree and Random Forest

The direct fusion of the data causes the overwhelmingly high dimension and the inefficiency of ML methods. Therefore, it is of great importance to recognize and retain the discriminative features. The decision tree, as a common machine learning method, has shown superiority in dimensionality reduction [27]. It uses a tree structure for sample classification, where the branch nodes include discriminant conditions, which are composed of a series of value comparisons among certain features. It is worth mentioning that the features are not selected randomly but through a factor named information gain. Suppose that the classification task has  $K$  target classes,  $H$  represents the original dataset, and  $\{H_1, \dots, H_T\}$  are subsets of  $H$  divided through all possible values of a certain feature  $f$ . The information gain brought by this data division is described as

$$Gain_{H,f} = \sum_{k=1}^K Ent_{k,H} - \sum_{t=1}^T \sum_{k=1}^K \frac{|H_T|}{|H|} Ent_{k,T}, \quad (11)$$

$$Ent_{k,H} = -P_{k,H} \log_2 P_{k,H}, \quad (12)$$

where  $P_{k,H}$  represents the proportion of samples that are labeled as the  $k$ -th class in the dataset  $H$ , and  $|*|$  measures the size of a set. When constructing a decision tree, the features with low classification ability are excluded, by which the dimensionality is reduced.

Random forest (RF) is an ensemble learning method using decision trees as the base learners, which is increasingly popular in disease classification [28]. When building each decision tree in RF, some features will be deleted randomly. Then, features with better classification ability will be selected out through ensemble learning to improve the overall performance. In this study, we improve the RF method with the weighted evolution process and realize the self-optimization of RF according to the overall performance.

### D. Weighted Evolutionary Random Forest Design Idea and Algorithm

Albeit the robust classification ability in bioinformatics, RF still has some defects [29, 30]. One significant flaw is that the image noise and high-dimension may make traditional RF difficult to obtain stable and robust generalization ability, which may result in the low classification accuracy of the method [31]. Based on the deficiencies mentioned above, this paper combines the idea of ensemble learning with weighted

evolutionary strategy and proposes a novel WERF model constructed in the following 3 steps.

- 1) Constructing base classifiers. Using the conventional method in the construction of RF, multiple decision trees are built as base classifiers by randomly selecting samples and features;
- 2) Weighting. Each base classifier is given a weight. In this paper, we use respective classification accuracies as weights of base classifiers;
- 3) Evolving. Weighted evolutions are carried out to get the final WERF model, during which redundant features are constantly eliminated, and the dimension is therefore reduced.

Fig. 2 demonstrates the design idea of the WERF model. The detailed implementation process is as follows.

Let  $B = \{\{x_1, y_1\}, \dots, \{x_n, y_n\}\}$  be the initial sample set,  $F$  be the set of all fusion features,  $x_n$  be the  $n$ -th sample,  $y_n \in \{-1, +1\}$  be the corresponding labels, and the label "+1" and "-1" denote LMCI and EMCI patients, respectively. At first,  $B$  is divided into the training set  $B_{\text{train}}$ , validation set  $B_{\text{validate}}$ , and test set  $B_{\text{test}}$ . In specific,  $B_{\text{train}}$  is used to train the primitive decision trees to obtain a traditional RF model,  $B_{\text{validate}}$  evaluates the performances of the base classifiers to obtain the weights of each base classifier, and  $B_{\text{test}}$  tests the overall classification performance of WERF. Then, random sampling without replacement is adopted to obtain a certain number of samples and features from  $B_{\text{train}}$  for constructing the base classifiers. Therefore, some samples and features are arbitrarily extracted from  $B_{\text{train}}$  and  $F$  to construct and train each base classifier. For a certain base classifier, the corresponding features of each sample are the same. According to the common

practice, we set the feature number to a value close to the square root of the total feature number [32]. Formally, assuming that the initial dimensionality is  $d$ , the dimensionality corresponds to each base classifier is determined as

$$s = \text{fix}(\sqrt{d}), \quad (13)$$

where  $\text{fix}(\ast)$  denotes an integral function, which returns the largest integer less than the input value.

Using the extracted samples and features and repeating the procedure for  $N$  times,  $N$  base classifiers are accordingly obtained to construct a conventional RF model. We introduce  $D_i$  to denote the  $i$ -th decision tree and  $D = \{D_1, \dots, D_N\}$  to denote the set of all classifiers. Subsequently, we employ validation set  $B_{\text{validate}}$  to obtain the classification accuracy of each base classifier, and the accuracy is taken as the weight of the corresponding base classifier. The equation calculating the weight is shown as

$$W_i = \frac{T'_i}{T}, \quad (14)$$

where  $T'_i$  denotes the quantity of samples in  $B_{\text{validate}}$  which are properly categorized by the  $i$ -th base classifier, and  $T$  denotes the size of  $B_{\text{validate}}$ .

In order to optimize the classification effect, we need to selectively delete the features that will adversely affect the overall classification performance. Firstly, we select the base classifiers whose classification accuracy is less than 50%. Such classification performance is inferior to that of random prediction. Then, the features frequently appearing in these base classifiers are redundant features to be reduced. Supposing that  $D_{\text{weak}} \subseteq D$  denote the set includes all base classifiers with weak classifying ability, the evolved feature set is described as

$$F_{\text{evolve}} = F - \{h \in F \mid \sum_{i=1}^{|D_{\text{weak}}|} W_i t_{h,i} < r\}, \quad (15)$$

where  $W_i$  denotes the corresponding weight of the  $i$ -th base classifier in  $D_{\text{weak}}$ ,  $r$  is a threshold, and  $t_{h,i}$  indicates whether the feature  $h$  is selected by the  $i$ -th classifier (If so, then  $t_{h,i}=1$ , otherwise  $t_{h,i}=0$ ). In each time of weighted evolution, the base classifiers will be reconstructed, and the features are randomly extracted from the evolved feature set. For each base classifier, the features corresponding to each sample are also the same.

Through the above steps, some redundant or invalid features are removed, therefore obtaining an updated feature set  $F_{\text{evolve}}$ . We repeat such evolution for several times to obtain the optimal feature set. In each iteration, the  $F_{\text{evolve}}$  obtained from the last iteration is used as  $F$ , and  $N$  decision trees as base classifiers are thus rebuilt to construct a new RF model. As the number of evolutions increases, when the overall classification accuracy of the model tends to be stable, the evolution is stopped, and the optimal number of evolutions is therefore determined. The number of the remaining features after the  $i$ -th evolution is defined as

$$S_i = |F| - \sum_{k=1}^i N_k, \quad (16)$$

where  $N_k$  indicates how many features are deleted in  $i$ -th iteration. This equation is applied to evaluating the effect of

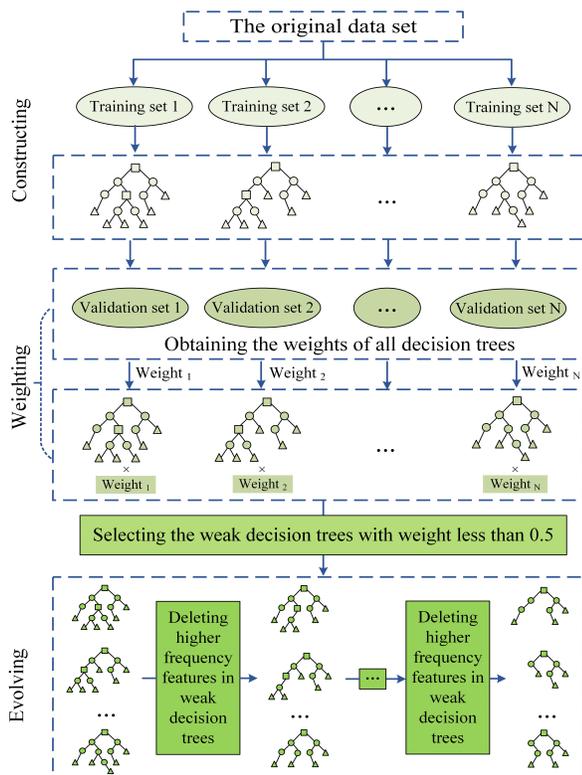


Fig. 2. The design idea of WERF.

every iteration of evolution.

Algorithm 1 illustrates the process of construction and evolution of the WERF model, of which the input is the original sample set  $B$  and feature set  $F$ , and the output is the eventual WERF model.

### E. Classification with WERF

The WERF model is applied to classifying the samples in the test set, and the results of WERF are obtained via the weighted voting mechanism. Specifically, each base classifiers predicts the category of samples, and the weighted sum of the results is obtained. The weights of results are exactly that of base classifiers. Let  $a \in \{-1, +1\}$  be one of the two possible resulting labels, "+1" represents LMCI, and "-1" represents EMCI, the voting value of each label is defined as

$$S_a = \sum_{k=1}^m I_a(f_k(x))W_k, \quad (17)$$

$$I_a(f_k(x)) = \begin{cases} 1, & f_k(x) = a \\ 0, & f_k(x) \neq a \end{cases} \quad (18)$$

where  $x$  denotes each sample from the test set and  $f_k(x) \in \{-1, +1\}$  represents the classification result generated by the  $k$ -th base classifier for the  $x$ . For the samples in  $B_{\text{test}}$ , the label having the most votes will be selected as the final classification label, which can be written as

$$Rst = \arg \max_a (S_a). \quad (19)$$

---

**Algorithm 1** The constructing process of weighted evolutionary random forest

---

**Input:** Experimental sample set  $B$  and feature set  $F$ .

**Output:** The evolved model  $WERF$ .

- 1: Initialize  $B, F, N, tmax$
  - 2:  $B, F$  are the set of samples and features, respectively;
  - 3:  $N$  represents the number of decision trees in WERF;
  - 4:  $tmax$  is the maximum number of iterations.
  - 5: Partition the  $B$  into  $B_{\text{train}}, B_{\text{validation}},$  and  $B_{\text{test}}$
  - 6: **do**
  - 7:   **for**  $i = 1:N$  **do**
  - 8:     Extract some samples from  $B_{\text{train}}$  without replacement;
  - 9:     Randomly select a subset of  $F$ ;
  - 10:     Selected samples and features  $\rightarrow$  Decision tree  $D_i$ ;
  - 11:      $B_{\text{validation}} \rightarrow$  the classification accuracy of  $D_i$  as its weight  $W_i$ ;
  - 12:   **end**
  - 13:    $W = \{W_1, \dots, W_N\}$ ;
  - 14:    $WERF =$  Ensemble of decision trees  $D_1, D_2, \dots, D_N$ ;
  - 15:   Apply  $WERF$  to classifying the samples in  $B_{\text{test}}$ , get assembled classification results and Calculate the overall classification accuracy  $Acc$ ;
  - 16:   According to  $W$ , features with low weights in  $F$  are deleted, and the remains compose  $F_{\text{evolved}}$ ;
  - 17:    $F = F_{\text{evolved}}$ ;
  - 18: **until**  $Acc$  reaches the peak **or** iteration times come to  $tmax$ .
- 

In other words,  $Rst$  is the label maximizing the voting value statistic  $S_a$ .

### F. Extraction of MCI-associated Genes and ROIs

The proposed WERF is mainly used for feature selection. After continuous weighted evolutions, the remaining features are regarded as the significant features to distinguish EMCI from LMCI. Then, the discriminative genes and ROIs are obtained by calculating the occurrence frequency of each ROI and gene in the optimal multimodal features. The frequency indicates the importance of the corresponding pathogenetic factor in classification.

### G. Parameter Optimization

Evolution time is a parameter to be optimized in the WERF model. At each time of evolution, some redundant or useless features (i.e., features with poor classification ability) will be eliminated, enhancing the model performance. However, interminable evolution is useless in a long term. In this study, we set different evolution times to construct the WERF model until the overall classification performance becomes stable, the time of evolution when the performance starts to be stable is defined as the optimal evolution time.

## IV. EXPERIMENT AND RESULTS

### A. Data Acquiring and Preprocessing

The data applied in this paper are from the data of EMCI and LMCI patients in the ADNI database (<http://adni.loni.usc.edu/>), an open platform for sharing data including clinical, genetic, and MRI images. The subjects are screened to ensure the data homogeneity and obtain the clinical characteristics of EMCI and LMCI. As the result, data of 63 participants are selected in this study, including 37 EMCI patients (11 males, 26 females, mean age:  $72.97 \pm 7.38$ ) and 26 LMCI patients (14 males, 12 females, mean age:  $72.45 \pm 7.47$ ). All data collection is approved and supported by relevant agencies. All participants have signed the informed consent. In addition, the multimodal data used in this paper has been approved and authorized by ADNI, and the data use conforms to the standards.

We utilize the chi-square test to assess the gender difference between EMCI and LMCI patients, and the age difference between the two groups is tested by the two-sample t-test. The information of participants and p-values are shown in Table I. It can be concluded that there exist no statistical differences in the sex and age of the subjects.

We conduct our experiment using MATLAB platform. Concretely, we use the DPARSF toolbox within MATLAB to preprocess fMRI data. The processing steps are as follows:

TABLE I  
BASIC INFORMATION OF THE DATA

Variables	EMCI (n = 37)	LMCI (n = 26)	p-value
Gender (M/F)	11/26	14/12	0.054
Age (years) <sup>a</sup> (Mean $\pm$ SD)	$72.97 \pm 7.38$	$72.45 \pm 7.47$	0.783

<sup>a</sup> The p-value of Kolmogorov-Smirnov test is 0.8928 ( $>0.05$ ), indicating that the ages are conform to a normal distribution. Moreover, the F value in variance equivalence test is 0.022 ( $<0.05$ ), indicating that the variances of the two classes are equivalent.

- 1) Transform the format of image data to NIFTI;
- 2) Delete the first 10 time points to decrease the adverse impact produced by the magnetic field;
- 3) Correct the time difference between layers;
- 4) Correct head motion, and eliminate data exceeding the range of 2.5mm;
- 5) Normalize images with echo-planar imaging template to compensate the differences among the anatomical structures of subjects during data acquisition;
- 6) Gaussian smooth the normalized image [33];
- 7) Retain signals with pathological significance using the delinearized drift;
- 8) Filter signal in the range of 0.1 Hz-0.8 Hz to reduce the effect of physiological noise;
- 9) Remove the interference from whole-brain signals and white matter signals with linear regression.

Similarly, it is necessary to perform preprocessing on gene data to ensure its quality. We acquired the SNP data on the Illumina Omni 2.5M BeadChip, and preprocessed them by PLINK software [34]. The details are as follows:

- 1) Set sample's call rate threshold at 95% in order to assess the total quality of the genetic data;
- 2) Set the minimum allele frequency threshold, genotyping threshold, and the Hardy-Weinberg equilibrium test to 0.03, 0.99, and  $1e-5$  respectively for eliminating the SNP with inferior quality.

### B. Construction of Fusion Features and WERF

The first contribution of this study is the construction of fusion features using imaging genetic data. At first, fMRI and gene data are serialized and encoded into the numeric format. For fMRI data, 90 ROIs are separated using the automatic anatomical labeling (AAL) atlas [35] and the first 60 time points of each ROI are selected as the representative time series. The length of 60 is finalized through multiple attempts. For SNP data, according to the reference quantity of the preprocessed SNPs, we group the SNPs to represent the genes they belong to, and 36 gene groups with more than 30 SNPs are extracted. Afterward, the 4 bases (i.e., A, T, C, and G) of the gene are encoded into 4 discrete numbers (1, 2, 3, and 4) to shape a digital sequence based on the PLINK (1.07) documentation. The numbers 1, 2, 3, and 4 are just marks. It has been verified through experiments that the replacements and reordering of these marks do not affect the eventual results. Noting 30 SNPs are directly matched with 60 bases, the gene sequences can further match with the time series of ROIs at the equivalent length of 60. Subsequently, the correlations between each pair of sequences are calculated by Pearson correlation analysis (Eq. (3)), bringing about 3240 ( $90 \times 36$ ) fusion features, which, in this paper, are named as ROI-gene pairs.

We built WERF model and conducted weighted evolution to reduce the discriminative fusion features. Firstly, according to the certain ratio of 10:5:6, 63 samples are arbitrarily divided into three groups. Concretely, 30 samples formed the training set, 15 samples composed the validating set, and the test set included the rest 18 samples. Subsequently, according to Eq. (13), 57 features were randomly extracted from the whole 3240

features as input features to construct a base classifier. The rationality of number 57 is further verified by multiple experiments. When the number of input features is too small, a multitude of basic classifiers need to be constructed to obtain the satisfactory performance, which will increase the modeling complexity. However, if there are too many input features, the diversity of base classifiers will decrease, which will increase the time cost of the weighted evolution.

To determine the number of base classifiers, we repetitively construct WERF with different number of base classifiers. As shown in Fig. 3, when the number of base classifiers reaches 300, the performance growth of ensemble learning tends to be flat, indicating that the random forest containing more than 300 decision trees will optimize less to performance but bring extra burden to the overall calculation. Therefore, the base classifier number is initialized as 300. Then, continuous evolutions were conducted to sift out the optimal fusion features.

### C. Parameter Optimization and Fusion Features Extraction

To optimize the overall performance of the whole model, a proper time of evolution was needed. As shown in Fig. 4, the performance of the WERF model was estimated through classification accuracy, and when the evolution time was 19, the overall classification accuracy of the model reached the highest value of 88.9% and had tended to be stable, indicating that evolving more than 19 times makes the little effect. Thus, the optimal evolution time was determined as 19. It is worth noting that the accuracy may sometimes decrease when the evolution time is over 19. This can attribute to the instability caused by the continuous selection of base classifiers. When the base classifiers are too few, the performance of the model will be unstable.

During the above experiment, we simultaneously recorded the number of features that are deleted in each evolution and the number of remaining features after each feature selection. Fig. 5 delineates the results in the first 20 evolutions. After the 19-th evolution, the overall classification accuracy of the WERF

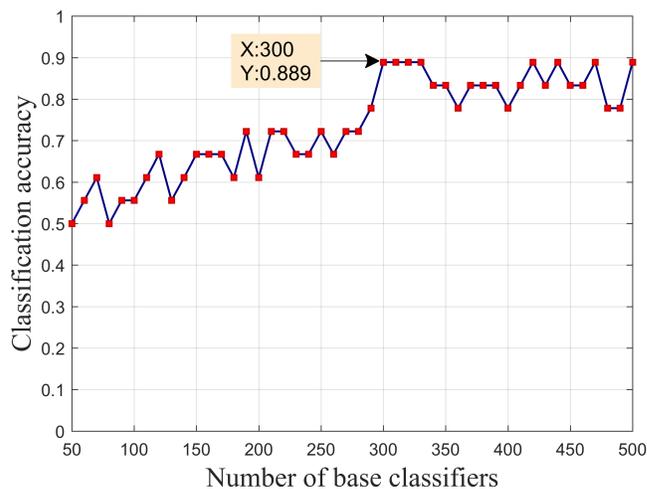


Fig. 3. The classification accuracies of WERF in different number of base classifiers. We adjust the base classifier number in a certain interval of [50,500] within a step of 10 and find that when WERF contains 300 base classifiers, the overall performance tends to be stable. Thus, the number 300 is finalized as the optimal number of base classifiers.

model reached the highest point and began to be stable. At the same time, the number of retained features was 342, which meant that most of the inefficient features were filtered out, and the corresponding 342 features in the ensemble learner were therefore taken as the optimal fusion features.

#### D. Performance of the Proposed Framework

As mentioned before, the Pearson correlation analysis was utilized to construct fusion features, and the WERF model was proposed for sample classification and feature extraction. To verify the superiority of the "Pearson + WERF" framework, we combined different methods for feature construction and extraction to build different analysis frameworks and compared their performance. Specifically, we utilized Pearson correlation analysis, distance correlation (DC) [36], CCA [37], and Kendall [38] to build fusion features using the same data samples, and used decision tree, random forest, two-sample t-test, and WERF model to extract the optimal fusion features. We recorded and counted each batch of optimal features generated by these frameworks and respectively measured the classification accuracy of these features leveraging SVM [39]. SVM is a common machine learning method that defines a hyperplane to classify two classes of samples. In this paper, the hyperplanes are determined by the values of the extracted features. Therefore, it can be a fair method to evaluate the classification ability of the feature set. Furthermore, we compared the overlaps of the feature sets extracted by other frameworks over that by WERF to verify the reliability of WERF. All comparative results were shown in Table II.

From Table II, it could be observed that the proposed "Pearson + WERF" framework extracted a relatively small quantity of optimal fusion features in all frameworks (third to the last), whereas it achieves the highest classification accuracy of 88.9%. Meanwhile, we found that there are significant overlaps among the optimal fusion features extracted by the "Pearson + WERF" framework and other frameworks. Through hypergeometric tests, it had been proved that these overlaps were not occurred accidentally. It is also worth noting that the

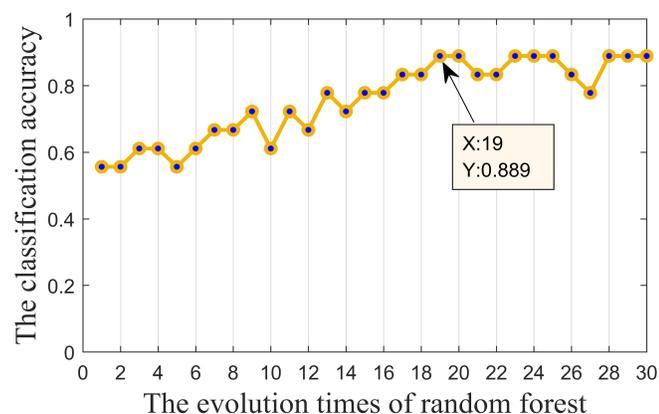


Fig. 4. The classification accuracies of WERF under different evolution times. When the classification accuracy reaches its peak and remains stable, the number of evolutions is the optimal. The figure shows that the optimal number of evolutions is 19. Noting that the optimal number of the evolution time is determined in a wider search range. In order to show the parameter changes in this interval more carefully, we only depict the search results in the interval [0,30].

more the quantity of overlapping optimal fusion features was, the higher the categorization accuracies of these corresponding frameworks showed, indicating that the features extracted by the proposed method were those that can improve the performance. According to the comparative results, we further drew receiver operating characteristic (ROC) curves, and the results were delineated in Fig. 6. Comparing to others, the area under curve (AUC) of the "Pearson + WERF" framework had reached the highest value of 0.889, which shows its superiority from another aspect.

After the preprocessing of fMRI data and genetic data, we acquired the time series of 90 ROIs and the digital series of 36 genes. Considering the difference in lengths among these sequences and series, and in order to facilitate the correlation analysis between two different omics of data, we need to intercept them to a unique length, which, in this study, was eventually set to 60. In other words, the value of  $\ell$  in Eq. (3) is set to 60. This value was determined through repeated experiments. In concrete, we firstly utilized different values of  $\ell$  to construct fusion features. Then, different models were built

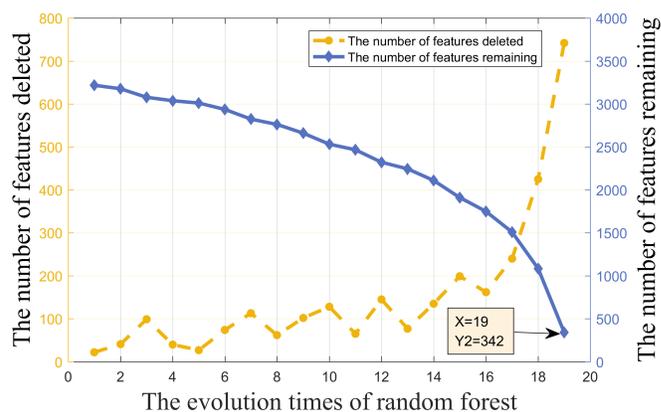


Fig. 5. The changes of the feature numbers during the evolution. The abscissa in the figure indicates the evolution times. The ordinate on the left side indicated the number of deleted features, while the right side indicated the number of remained features.

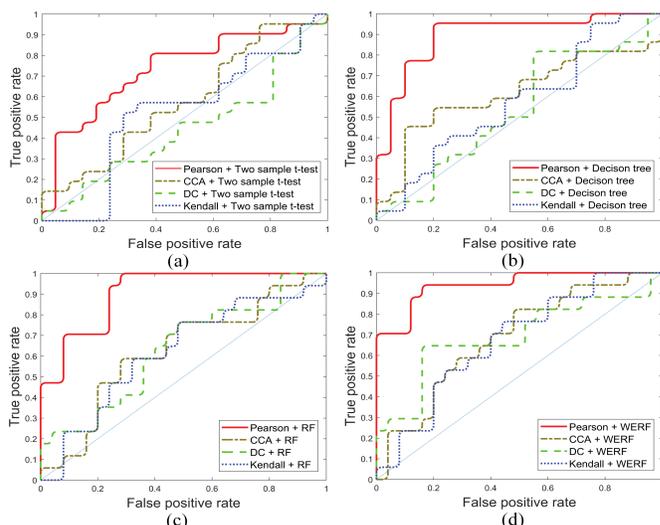


Fig. 6. The performance comparison of different frameworks. (a) denotes the ROC curves of two-sample t-test with 4 different fusion feature construction methods. (b), (c), and (d) denote the ROC curves of decision tree, RF and the proposed method (WERF), respectively.

in the same way for fair comparison. Finally, we utilized these models to do classification, and the ROC curves were shown in Fig. 7 (a), which showed that when  $\ell$  was 60, the performance of the model was the best. One plausible explanation is that if  $\ell$  is too small, the fusion features are not able to differentiate the samples. Furthermore, we extended the framework to the study of AD & HC and Parkinson's disease (PD) & HC datasets to demonstrate the generalization ability and robustness of "Pearson + WERF". The data of AD and PD patients are obtained from ADNI and Parkinson's Progression Markers Initiative (PPMI) database, respectively. As shown in Fig. 7 (b) and Fig. 7 (c), the "Pearson + WERF" framework performed well, achieving the AUC values of 0.862 and 0.850. Also, we noted that in these two datasets, the optimal intercept lengths were 60 and 80, respectively. The origin of such difference was that we did not randomly choose but determined the optimal threshold through repeating trial experiments. Therefore, the optimal intercept length might be affected by the datasets we use.

### E. Extraction of Discriminant Genes and Brain Regions

Based on the comparison experiments, we observed that the 342-dimensional fusion features extracted by the proposed method had considerable resolution performance. In other words, the differences in these features are more obvious between EMCI and LMCI patients. Therefore, the brain regions contained in the optimized features are more likely to have functional lesions and the genes contained in these features are more likely to have abnormal expressions. Some fusion features, i.e., ROI-gene pairs, are shown in Fig. 8. It is worth noting that each ROI-gene pair, though not explicitly shown in the figure, had an appearance frequency as its corresponding weight. The greater the weight of an ROI-gene pair was, the more significant the ROI-gene pair was in distinguishing EMCI and LMCI.

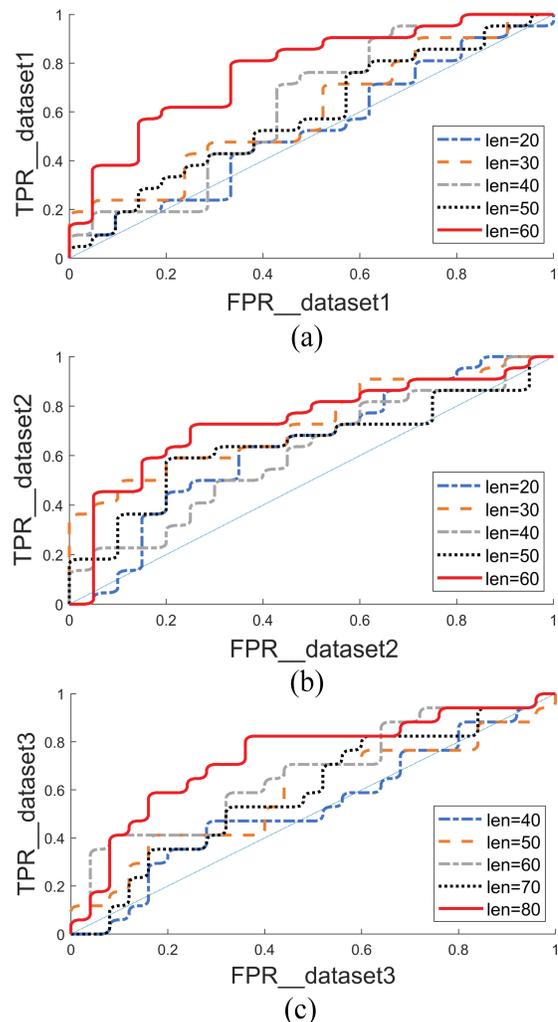


Fig. 7. Comparative results of different intercept lengths of time series and gene sequences. (a) ROC curves under dataset1 (36 EMCI and 27 LMCI, from ADNI database). (b) ROC curves under dataset2 (37 AD and 36 HC, from ADNI database). (c) ROC curves under dataset3 (55 PD and 50 HC, from PPMI database).

TABLE II  
THE COMPARATIVE RESULTS OF THE FRAMEWORKS

Method	Number of extracted fusion features	Classification accuracy of extracted features with SVM		Overlap size of the extracted features with the proposed method
		Number of extracted fusion features	Classification accuracy of extracted features with SVM	
Pearson + WERF	342		88.9%	/
Pearson + Two sample t-test	360		66.7%	132 (p = 1.720884e-08)
CCA + Two sample t-test	294		61.1%	101 (p = 3.954116e-09)
DC + Two sample t-test	308		55.6%	114 (p = 4.095232e-07)
Kendall + Two sample t-test	352		61.1%	105 (p = 1.151482e-16)
Pearson + Decision Tree	800		72.2%	160 (p = 1.004105e-87)
CCA + Decision Tree	560		66.7%	110 (p = 2.072108e-59)
DC + Decision Tree	660		61.1%	125 (p = 5.002476e-75)
Kendall + Decision Tree	600		66.7%	138 (p = 9.139333e-51)
Pearson + Random Forest	560		77.8%	175 (p = 1.606163e-23)
CCA + Random Forest	675		66.7%	130 (p = 2.156843e-75)
DC + Random Forest	365		66.7%	135 (p = 2.817358e-08)
Kendall + Random Forest	405		66.7%	128 (p = 2.933994e-16)
CCA + WERF	590		72.2%	165 (p = 1.280715e-33)
DC + WERF	445		78.8%	185 (p = 3.153446e-05)
Kendall + WERF	394		72.2%	158 (p = 6.760018e-06)

Moreover, we computed the frequencies of each gene and ROI through these ROI-genes pairs. The genes and ROIs with the highest occurrence frequencies were expected to be more capable of identifying EMCI and LMCI. On the one hand, Fig. 9 (a) shows the ROIs with the highest frequencies. We employed frequencies as the corresponding weights of brain regions, as shown in Fig. 9 (b), where the magnitude of the weights was graphically represented as the size of the locating points in the figure. On the other hand, according to the calculated frequency values, we found the genes with the highest frequencies (Fig. 10). The most discriminative ROIs with the highest frequencies included Heschl's gyrus (HES.L and HES.R), Temporal pole: middle temporal gyrus (TPOmid.R), and Median cingulate and paracingulate gyri (DCG.L), while the risk genes included CSMD1, DAB1, CNTN5, and CTNNA2. In some respects, the results can provide important evidences to the pathological studies of MCI development because the results embodied that the EMCI and LMCI are most likely discrepant in these genes and brain regions.

We investigated existing medical researches to verify the reliability and clinical significance of the extracted ROIs and genes. For example, as shown in the results, the Temporal pole had a relatively high occurrence frequency, and it had been reported to be related to visual memory, emotional association, language understanding, and performing function according to other studies. For example, Binder et al. [40] reported that the removal of the anterior temporal lobe might present a risk of decreased language ability and speech memory deficits, indicating that the abnormal temporal lobe may be related to the developmental process of MCI. Cui et al. [41] utilized fractional amplitude of low-frequency fluctuations to identify the characteristic local functional activities specific to amnesic MCI patients, detecting the significant activity enhancements nearby DCG.L. Some highly-ranked ROIs, such as HES.L and TPOmid.R, had not yet been reported to be associated with MCI, but have great prospects for further studies. For instance,

TPOmid.R was thought to be associated with sleep deprivation-related behavior [42], and such dysfunctions might also be the cause of MCI. In conclusion, these ROIs can be employed as new biological markers to detect EMCI and LMCI and provide convenience for the clinical diagnosis of EMCI and LMCI.

We also investigated the risk genes we found. For example, It had been confirmed that abnormal CSMD1 would lead to expression loss and cognitive decline problems among patients with MCI [43]. CNTN5 was expressed in hypothalamus glutamatergic neurons, whose reduction may lead to long-term synaptic enhancement, and the mutation of CNTN5 might be one of the potential mechanisms of post-traumatic stress disorder [44]. The mutation of the CTNNA2 gene could cause abnormal neuronal migration and lead to giant gyrus malformation, which will lead to a series of cognitive problems [45]. Compared with EMCI patients, the mutation of the CTNNA2 gene was more obvious in LMCI patients. Therefore, the genes extracted by the WERF method can be further applied as biomarkers for distinguishing EMCI and LMCI, offering reference to the clinical diagnosis and treatment.

## V. DISCUSSION

It has long been a challenge for researchers to distinguish

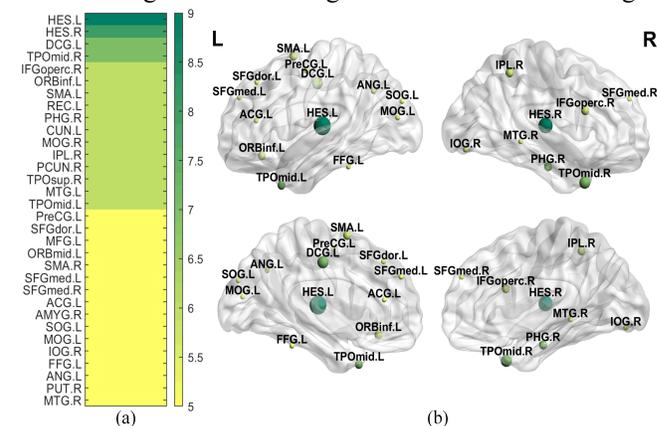


Fig. 9. Locations, sizes, and frequencies of extracted ROIs. (a) The ROIs with high frequencies. (b) The locations and sizes of abnormal brain regions.

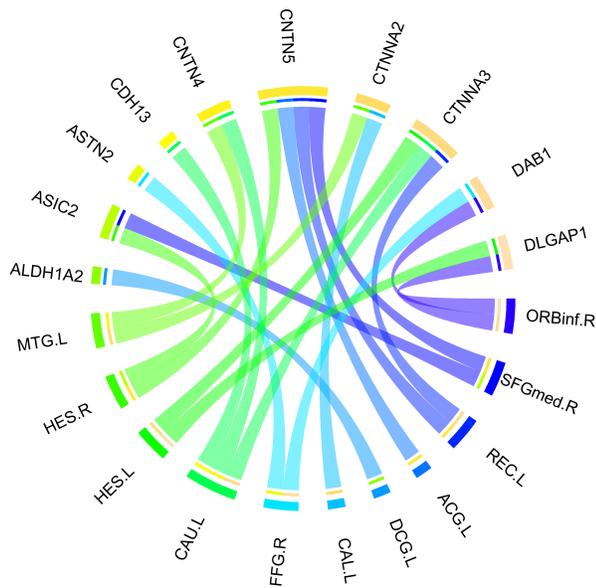


Fig. 8. Some optimal ROI-gene pairs with the highest weight value.

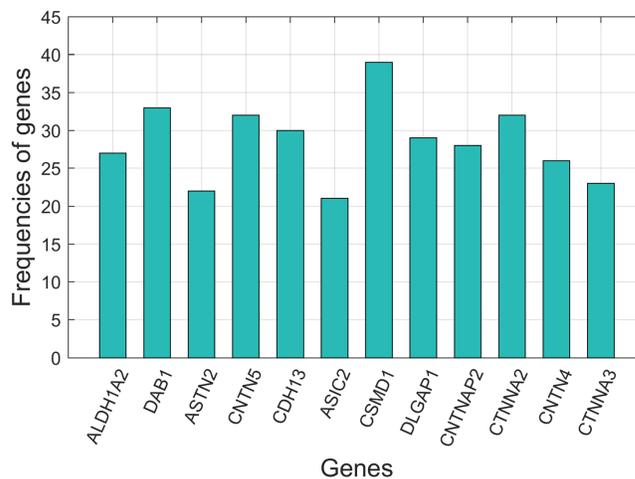


Fig. 10. The frequencies of main MCI-associated gene. The frequency reflects the degree of correlation between gene and MCI. Other genes are with relatively low frequency and are not illustrated in this figure.

MCI patients accurately and find the exact pathogenic factors. In biology, gene expressions can affect the function and structure of the brain. In the traditional research paradigm, this association can be verified by clinical experiments. At the aspect of computer and data science, some researchers have tried to detect such correlation by appropriate algorithms. However, present studies are usually based on single modalities, causing the inadequate detection of factors associated with disease and the detecting method may be weaker in terms of universality. In this paper, we carry out an in-depth exploration of multimodal features specific to the EMCI and LMCI patients. We put forward a novel multimodal data fusion method and further presented a machine learning-based framework of data fusion, sample classification, and pathogenetic factor extraction. We found out the brain regions and genes according to the optimal fusion features, namely, the ROI-gene pairs, which revealed the multifactorial pathogenesis in different stages of MCI to a certain extent. Compared with the existing dimension reduction approaches such as PCA and ICA, our method can effectively detect the important features without reconstruction, maintaining the biological significance of each fusion feature.

To further illustrate the progressiveness of our works, we also made extensive comparisons between our method and other existing methods. In the typical single classifier method, Nanni et al. [46] trained SVM with different data clusters extracted from the whole training set, and the accuracy was more than 75%. On the other hand, the idea of integration is also used in the categorization of MCI. Son et al. [47] adopted the regional volume shrinkage and functional connectivity of the recognition area as the characteristics of random forest, and the classification accuracy reached 53.33%. Lebedeva et al. [48] established a random forest model based on MRI and predicted that the accuracy of MCI and dementia in patients with late-life depression (LLD) was 76%. Using the deep learning method, Basaia et al. [49] utilized a single MRI and convolutional neural network to automatically distinguish the MCI patients that were likely to have AD from those whose conditions were stable. The accuracy was 75%. In contrast, the proposed WERF had better performance and unique advantages of lower cost and greater capability to abstract the discriminant factors between EMCI and LMCI. The advantage of our model intrinsically lies in the following points.

- 1) The combination of the weighted evolution and random forest. The weighting can differentiate features with different categorizing ability, and the evolution removes the redundant or invalid features.
- 2) The optimal number of evolutions was lowered, making the model occupy fewer calculation resources but retain the model efficiency at the same time.
- 3) The effective integration of the multi-omics data fusion method and WERF model. Three core works in disease studies (i.e., data fusion, feature reduction, and sample classification) were assembled by a unified framework in this paper, applying the complementarity information among imaging genetics data to boost the performance.

Although the WERF model achieves satisfying results, there also exist some limitations. Firstly, considering the complexity

of the pathogenic factors, more omics of data are expected to be involved in further researches. Secondly, we only applied the AAL template for brain segmentation. In the future, we can use other templates such as Broadman to match as well. Furthermore, some atypical pathogenic factors are unverified yet. These factors provide a reference for further study of MCI. We will collect more data and design new algorithms for in-depth analysis. In addition, we plan to cooperate with clinicians to explain the role and rationality of this factors.

## VI. CONCLUSION

This paper proposes a novel WERF model combining genetic data and neuroimaging data and constructs an integrated framework integrating data fusion, feature reduction, and sample classification. The proposed method makes full use of complementary information in multi-omics data, effectively explores the pathological mechanism of MCI development, and provides the relevant diagnosis basis. Compared to some existing methods, WERF shows better performance. According to the superior performance in classifying other brain-related diseases, such as PD and AD, the proposed framework shows the great potential of being a general framework for brain disease diagnosis.

## REFERENCES

- [1] D. Müller-Gerards *et al.*, "Subjective cognitive decline, APOE  $\epsilon 4$ , and incident mild cognitive impairment in men and women," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 11, pp. 221-230, 2019.
- [2] H. Zhang *et al.*, "Playing mahjong for 12 weeks improved executive function in elderly people with mild cognitive impairment: A study of implications for TBI-induced cognitive deficits," *Frontiers in neurology*, vol. 11, p. 178, 2020.
- [3] L. M. Lavrencic *et al.*, "Cognitive test norms and comparison between healthy ageing, mild cognitive impairment, and dementia: A population - based study of older Aboriginal Australians," *Australian Journal of Psychology*, vol. 71, no. 3, pp. 249-260, 2019.
- [4] L. Hernández-Domínguez, S. Ratté, G. Sierra-Martínez, and A. Roche-Bergua, "Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 260-268, 2018.
- [5] L. Du *et al.*, "A novel SCCA approach via truncated  $\ell 1$ -norm and truncated group lasso for brain imaging genetics," *Bioinformatics*, vol. 34, no. 2, pp. 278-285, 2018.
- [6] X. Hao *et al.*, "Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease," *Medical image analysis*, vol. 60, p. 101625, 2020.
- [7] H. Li *et al.*, "High-resolution chest x-ray bone suppression using unpaired CT structural priors," *IEEE transactions on medical imaging*, vol. 39, no. 10, pp. 3053-3063, 2020.
- [8] E. Lee *et al.*, "Single-nucleotide polymorphisms are associated with cognitive decline at Alzheimer's disease conversion within mild cognitive impairment patients," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 8, pp. 86-95, 2017.
- [9] Y. Chen *et al.*, "The relationship between four GWAS-identified loci in Alzheimer's disease and the risk of Parkinson's disease, amyotrophic lateral sclerosis, and multiple system atrophy," *Neuroscience letters*, vol. 686, pp. 205-210, 2018.
- [10] R. Bogdan *et al.*, "Imaging genetics and genomics in psychiatry: a critical review of progress and potential," *Biological psychiatry*, vol. 82, no. 3, pp. 165-175, 2017.
- [11] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, and A. S. D. N. Initiative, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern recognition*, vol. 63, pp. 171-181, 2017.

- [12] D. Sacha *et al.*, "Visual interaction with dimensionality reduction: A structured literature analysis," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 241-250, 2016.
- [13] W. Min, J. Liu, and S. Zhang, "Edge-group sparse PCA for network-guided high dimensional data analysis," *Bioinformatics*, vol. 34, no. 20, pp. 3479-3487, 2018.
- [14] C. Lin, H. Mu, R. Xiong, and J. Cao, "Multi-model probabilities based state fusion estimation method of lithium-ion battery for electric vehicles: State-of-energy," *Applied energy*, vol. 194, pp. 560-568, 2017.
- [15] M. Wang *et al.*, "Graph-kernel based structured feature selection for brain disease classification using functional connectivity networks," *IEEE Access*, vol. 7, pp. 35001-35011, 2019.
- [16] L. Salamanca, N. Mechawar, K. K. Murai, R. Balling, D. S. Bouvier, and A. Skupin, "MIC - MAC: An automated pipeline for high - throughput characterization and classification of three - dimensional microglia morphologies in mouse and human postmortem brain samples," *Glia*, vol. 67, no. 8, pp. 1496-1509, 2019.
- [17] B. Lei, S. Chen, D. Ni, and T. Wang, "Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion," *Frontiers in aging neuroscience*, vol. 8, p. 77, 2016.
- [18] M. Tanveer *et al.*, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 1s, pp. 1-35, 2020.
- [19] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, "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*, vol. 10, no. 3, pp. 799-817, 2016.
- [20] B. Lei *et al.*, "Self-calibrated brain network estimation and joint non-convex multi-task learning for identification of early Alzheimer's disease," *Medical image analysis*, vol. 61, p. 101652, 2020.
- [21] V. Bottero and J. A. Potashkin, "Meta-Analysis of Gene Expression Changes in the Blood of Patients with Mild Cognitive Impairment and Alzheimer's Disease Dementia," *International journal of molecular sciences*, vol. 20, no. 21, p. 5403, 2019.
- [22] T. Zou *et al.*, "Association of multiple candidate genes with mild cognitive impairment in an elderly Chinese Uygur population in Xinjiang," *Psychogeriatrics*, vol. 19, no. 6, pp. 574-583, 2019.
- [23] Y. Varatharajah, V. K. Ramanan, R. Iyer, and P. Vemuri, "Predicting short-term MCI-to-AD progression using imaging, CSF, genetic factors, cognitive resilience, and demographics," *Scientific reports*, vol. 9, no. 1, pp. 1-15, 2019.
- [24] J. Fang *et al.*, "Fast and accurate detection of complex imaging genetics associations based on greedy projected distance correlation," *IEEE transactions on medical imaging*, vol. 37, no. 4, pp. 860-870, 2017.
- [25] P. Cao *et al.*, "Nonlinearity-aware based dimensionality reduction and over-sampling for AD/MCI classification from MRI measures," *Computers in biology and medicine*, vol. 91, pp. 21-37, 2017.
- [26] C.-C. Wang, X. Chen, J. Qu, Y.-Z. Sun, and J.-Q. Li, "RFSMMA: a new computational model to identify and prioritize potential small molecule-mirna associations," *Journal of chemical information and modeling*, vol. 59, no. 4, pp. 1668-1679, 2019.
- [27] L. Wang *et al.*, "LMTRDA: Using logistic model tree to predict MiRNA-disease associations by fusing multi-source information of sequences and similarities," *PLoS computational biology*, vol. 15, no. 3, p. e1006865, 2019.
- [28] X. Chen, C.-C. Zhu, and J. Yin, "Ensemble of decision tree reveals potential miRNA-disease associations," *PLoS computational biology*, vol. 15, no. 7, p. e1007209, 2019.
- [29] H. Ruan *et al.*, "Topographic diversity of structural connectivity in schizophrenia," *Schizophrenia research*, vol. 215, pp. 181-189, 2020.
- [30] P. Singh and A. Kumar, "Deciphering the function of unknown Leishmania donovani cytosolic proteins using hyperparameter-tuned random forest," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, no. 1, pp. 1-9, 2020.
- [31] V. Saccà *et al.*, "Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data," *Brain imaging and behavior*, vol. 13, no. 4, pp. 1103-1114, 2019.
- [32] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintia, and S. Kundu, "Improved random forest for classification," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012-4024, 2018.
- [33] H. Yuan, X. Zhu, W. Tang, Y. Cai, S. Shi, and Q. Luo, "Connectivity between the anterior insula and dorsolateral prefrontal cortex links early symptom improvement to treatment response," *Journal of affective disorders*, vol. 260, pp. 490-497, 2020.
- [34] S. Liu *et al.*, "Polygenic effects of schizophrenia on hippocampal grey matter volume and hippocampus-medial prefrontal cortex functional connectivity," *The British Journal of Psychiatry*, vol. 216, no. 5, pp. 267-274, 2020.
- [35] M. Wang, X. Hao, J. Huang, W. Shao, and D. Zhang, "Discovering network phenotype between genetic risk factors and disease status via diagnosis-aligned multi-modality regression method in Alzheimer's disease," *Bioinformatics*, vol. 35, no. 11, pp. 1948-1957, 2019.
- [36] T. Górecki, M. Krzyśko, W. Ratajczak, and W. Wołyński, "An extension of the classical distance correlation coefficient for multivariate functional data with applications," *Statistics in Transition new series*, vol. 17, no. 3, pp. 449-466, 2016.
- [37] L. Du *et al.*, "Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach," *Medical image analysis*, vol. 61, p. 101656, 2020.
- [38] D. Valencia, R. E. Lillo, and J. Romo, "A Kendall correlation coefficient between functional data," *Advances in Data Analysis and Classification*, vol. 13, no. 4, pp. 1083-1103, 2019.
- [39] A. Li *et al.*, "A neuroimaging biomarker for striatal dysfunction in schizophrenia," *Nature medicine*, vol. 26, no. 4, pp. 558-565, 2020.
- [40] J. R. Binder *et al.*, "Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study," *Neuroimage*, vol. 54, no. 2, pp. 1465-1475, 2011.
- [41] L. Cui, Z. Zhang, C.-Y. Zac Lo, and Q. Guo, "Local Functional MR Change Pattern and Its Association With Cognitive Function in Objectively-Defined Subtle Cognitive Decline," *Frontiers in Aging Neuroscience*, vol. 13, p. 289, 2021.
- [42] L. Du *et al.*, "Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort," *Bioinformatics*, vol. 35, no. 14, pp. i474-i483, 2019.
- [43] T. Porter *et al.*, "Cognitive gene risk profile for the prediction of cognitive decline in presymptomatic Alzheimer's disease," *Personalized Medicine in Psychiatry*, vol. 7, pp. 14-20, 2018.
- [44] A. E. Schaffer *et al.*, "Biallelic loss of human CTNNA2, encoding  $\alpha$ N-catenin, leads to ARP2/3 complex overactivity and disordered cortical neuronal migration," *Nature genetics*, vol. 50, no. 8, pp. 1093-1101, 2018.
- [45] Y. Lei, F. Han, and Y. Shi, "Progress in Research on CNTN5, a Member of the Immunoglobulin Superfamily," *Journal of China Medical University*, vol. 46, no. 9, pp. 848-852, 2017.
- [46] L. Nanni, A. Lumini, and N. Zaffonato, "Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment," *Journal of neuroscience methods*, vol. 302, pp. 42-46, 2018.
- [47] S.-J. Son, J. Kim, and H. Park, "Structural and functional connective fingerprints in mild cognitive impairment and Alzheimer's disease patients," *PloS one*, vol. 12, no. 3, p. e0173426, 2017.
- [48] A. K. Lebedeva *et al.*, "MRI-based classification models in prediction of mild cognitive impairment and dementia in late-life depression," *Frontiers in aging neuroscience*, vol. 9, p. 13, 2017.
- [49] S. Basaia *et al.*, "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.