Biomedical Signal Processing and Control 53 (2019) 101564

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

High-order correlation detecting in features for diagnosis of Alzheimer's disease and mild cognitive impairment

Yi Ding*, Chuanji Luo*, Chang Li, Tian Lan, ZhiGuang Qin

The School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan Province, China

ARTICLE INFO

Article history:
Received 28 December 2018
Received in revised form 15 April 2019
Accepted 10 May 2019
Available online 22 May 2019

Keywords:
Alzheimer's disease diagnosis
Mild cognitive impairment diagnosis
Hypergraph
High-order correlation
Feature selection

ABSTRACT

As shown in the literature, the identification of discriminative features of Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI) can improve the diagnostic accuracy. Besides, many researches have proven that the feature selection method, which also considers the relationship among features, shows improvements in performance for AD diagnosis. However, most existing feature selection methods only consider the pairwise correlation between features. Instead, when adopting these features to diagnose the AD, the high-order correlated relationships among features need to be taken more into consideration. In this paper, a novel classification framework for diagnosing the AD and MCI has been proposed to address these problems. This framework mainly consists of three processes: feature extraction, feature selection and classification. The feature extraction is mainly used to extract texture features and morphometric features from brain MR images. The feature selection process is to select discriminative features. This process firstly measures the high-order correlation among features by borrowing the idea of hypergraph theory and then to generate the optimal feature subset based on the high-order correlation. The classification process will adopt the feature subset to finish the classification task. The main contribution of this framework is to consider the high-order correlation among features instead of pairwise correlation when classifying the AD. Experiments on the Alzheimer's Disease Neuroimaging Initiative ADNI database shows that the proposed method obtains better performance than other state-of-the-art counterpart methods. Overall, the experiment result demonstrates the effectiveness of the proposed framework.

1. Introduction

Alzheimer's Disease(AD) is characterized by the progressive impairment of neurons and their connections, which result in loss of cognitive functions. The estimated number of affected people will double in the next two decades, so that one out of 85 persons will have the AD by 2050 [10]. Mild cognitive impairment (MCI), an intermediate stage between normal cognition and dementia, has a high risk of progressing to AD. The conversion rate from MCI to AD is reported to be approximately 10%–15% per year [15]. Even though there is no cure for AD, some medications can delay the onset of some symptoms, such as memory loss [18]. Therefore, accurate diagnosis is critical for timely treatment and possible delay of AD and MCI.

Neuroimaging technology has been widely used in clinical practice because it can non-invasively detects the structural and functional changes associated with the development of diseases, such as Magnetic Resonance Imaging (MRI) [21] and functional MRI (fMRI) [23]. Besides, we can obtain different clinical phases of AD with the help of structural MRI, which is the most standard modality in clinical practice [25]. So, we choose structural MRI images to implement classification experiments.

Several types of features can be extracted from structure MRI, such as gray matter densities [28], texture measures [30], and morphometry [31]. Combining different features can provide more complementary information and eventually improve the diagnostic accuracy for AD [32]. In recent years, the texture analysis technique which can sensitively reflect the subtle changes of the body has been widely used for AD diagnosis [33]. On the other hands, the voxel-based morphometry (VBM) [34], which is a common technique for measuring morphometric changes in the brain, has been widely applied to the study of gray-matter alterations in AD [31,35]. Our previous work has innovatively integrated the VBM and texture features for diagnosing the AD and it has achieved a better performance [36]. It should be noted that, except for these features referred in this paper, there are also some other imaging features or clinical features have been adopted. For example, some Low-Rank

* Corresponding authors.
E-mail addresses: yi.ding@uestc.edu.cn (Y. Ding), luochuanji@std.uestc.edu.cn (C. Luo).

https://doi.org/10.1016/j.bspc.2019.101564
1746-8094/

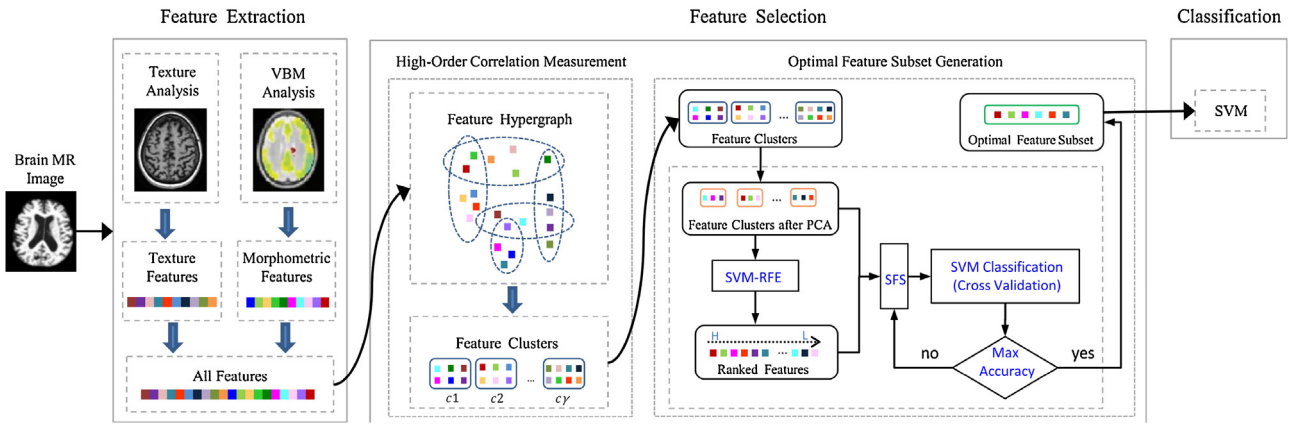

Fig. 1. The framework of our High-Order Correlation Detecting in Features (HOCDF) method, which consists of three main processes: 1) feature extraction, 2) feature selection, and 3) classification.

methods which combines imaging features and cognitive scores have shown good performance [37,38]. However, how to extract effective original features is not the work of this paper. The aim of this paper is to find an effective feature selection method so as to improve the performance for diagnosing the AD/MCI.

Feature selection, which has the capability of overcoming dimensional curse issues after removing the indiscriminative features, is one of the most important steps for building the robust learning model. It is broadly categorized into two main methods, which is based on whether the relationship among features is considered. If the relationship among features is not considered, the feature selection method will just consider the discriminability of the feature itself, such as Principal Components Analysis(PCA) [39], and the Support vector machine recursive feature elimination(SVM-RFE) [40]. These methods have been successfully implemented in various neuroscience applications [41,42]. Another type of methods will take the relations between features or samples into account while selecting features. One of the typical method SVM-RFE with Covariance has proved that the feature selection can be improved by considering the relationship among features so as to raise the accuracy of AD diagnosis [36] In addition, some researches, such as the Low-Rank method [43] and Regression method [44], shows that the classification performance of AD/MCI can be improved by considering the relations among the features, modalities [45], or samples. Except for the pairwise correlation among features which have been taken into account by these mentioned methods, there are more complex relations among features that are needed to be considered. The high-order relations among features have been measured and adopted into a multi-kernel method by Zhang et al, and this method has achieved outstanding performance in image classification tasks [46]. And the high-order correlation of brain regions have been employed to improve the diagnosing method for AD/MCI by Zhang et al [47,48]. Liu et al [24] have also proposed a view-aligned hypergraph learning (VAHL) method for AD/MCI diagnosis which has obtained good results by exploring the underlying coherence among views of multi-modality data. However, to the best of our knowledge, there is no research which has adopted the idea of high-order correlation among features to improve feature selection in the area of AD classification.

In this paper, a novel classification framework to precisely diagnose AD/MCI has been proposed. The main innovation of the proposed method is that the high-order correlation among features was considered for capturing more discriminative features to improve the classification accuracy. To be more specific, borrowing the idea of the hypergraph [49], which is a natural descriptor for high-order correlation among multiple objects, it will be adopted here to discover high-order correlation among features and then combine with the SVM-RFE method to implement feature selection step. This proposed method has been evaluated on the Alzheimer's Disease Neuroimaging Initiative(ADNI) database and shows more excellent performance than other state-of-art counterpart methods. To the best of our knowledge, there is no researches referring to consider the high-order correlation among features when diagnosing the AD/MCI as used in this paper. The rest of this paper is organized as follows. In the 'Method' section, the proposed method is described in detailed. In the 'Experiment and Result' section, the details of experiment are introduced, including material, experiment setting, and result. In the 'Discussion' section, the influence of several parameters is discussed. Finally, In the 'Conclusion' section, we briefly conclude this study.

2. Method

In this section, an overview of the proposed method High-Order Correlation Detecting in Features (HOCDF) will be presented. As shown in Fig. 1, there are 3 main processes in HOCDF: 1) feature extraction: in this one, there are totally 292 original features are obtained by using Texture analysis and VBM analysis on brain MRIs, which includes 284 texture features and 8 morphometric features. And then, all these 292 features are integrated together for next process. 2) feature selection: Firstly, a hypergraph is constructed to reveal the high-order correlation among extracted features. By computing this hypergraph and using a clustering method, these features are separated into some clusters. Secondly, before making these clusters to obtain the optimal feature subset, the PCA method are employed to reduce the dimensionality for each clusters. After the PCA, a feature ranking method named SVM-RFE is adopted to rank all left features in these clusters and generate a ranking result of these features. Last but not the least, by adapting the ranking result and the high-order relations in clusters together, these fetures will be iteratively chosen and verified to construct the optimal feature subset. 3) classification: in this process, the SVM method is employed to classify the AD/MCI by adopting the optimal feature subset. In the following, each process will be explicitly described.

2.1. Feature extraction

Our previous work has proved that combining morphometric features and texture features can obtain better results than being used separately. So, the feature extraction part of this paper is as same as our previous work in [36]. According to Fig. 1, in the feature extraction progress, we obtain texture features and morphomet-

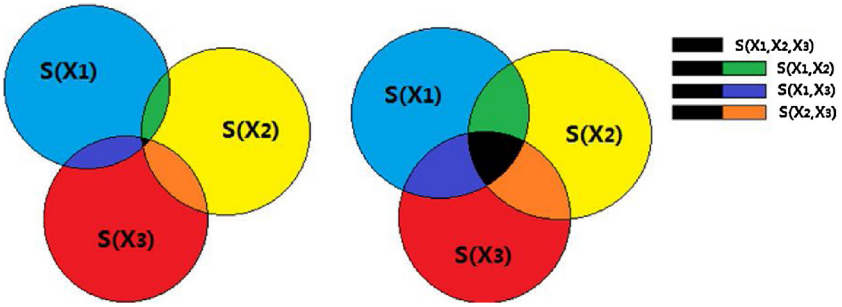

Fig. 2. Graphic explanation of the shared entropy. The three circles represent the shared entropy of three different features, blue for S(X1), yellow for S(X2), red for S(X3). Black + green for S(X1, X2), black + purple for S(X1, X2, X3), black + orange for S(X2, X3). Black for S(X1, X2, X3).

ric features from brain MR images by texture analysis method and VBM analysis method respectively. Besides, the texture features are consist of Gray-level Co-occurrence Matrix (GLCM) [50] and Gabor filter [33]. And the morphometric features are the gray-matter volume of Automated Anatomical Labeling(AAL) regions.

To be more specific, for GLCM, 11 descriptors are employed, which are consisted by the angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum averages, sum variances, sum entropy, entropy, difference variance, and difference entropy. In here, different GLCMs can be generated by changing the distance and direction parameters and a total of 220 features are finally extracted, A total of 220 features are extracted; For Gabor, different orientations and frequencies are applied for the Gabor filter to form a Gabor filter bank. We constructed features from the Gabor response by calculating the mean value and standard deviation of pixels, and totally 64 Gabor features are obtained; For VBM analysis, because the VBM is capable of investigating gray matter abnormalities across the whole brain, features extracted from clusters with significant gray matter are different between AD patients and NC. The main steps of VBM include a)VBM-DARTEL analysis, b)Statistical analysis, c) Extracting gray-matter volume of Automated Anatomical Labeling(AAL) regions. There are totally 8 morphometric features by adopting the VBM analysis. Finally, we simply concatenate all features together as the input of the following procedure.

2.2. Feature selection

In our previous work [36], we innovatively combined covariance with feature ranking method for diagnosing the Alzheimer's Disease, and achieved an excellent classification performance. It can be proven that adopting the correlation of features in the process of feature selection could obtain a better feature subset. But there are still some shortcomings that the covariance can only measure the pairwise correlation between features and don't have the ability to measure the high-order correlation among multiple features. Therefore, in order to make better use of the high-order correlation among features for feature selection, our proposed method HOCDF divided the feature selection progress into two steps. The first step is to clusters all features for measuring high-order correlation among features, and the second step is to use feature clusters to generate optimal feature subset.

2.2.1. High-order correlation measurement

In order to measure the high-order correlation, our clustering method makes these features with strong correlation appear in the same cluster and with weak correlation appear in the different clusters. In this paper, the clustering method adopted the shared entropy [46], hypergraph theory and community learning by graph approximate (CLGA) [51].

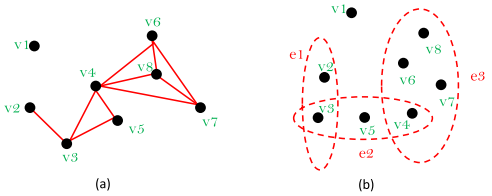

Fig. 3. (a) is a classic graph, which depicts the pairwise correlation among vertices. If the correlation between two points is strong enough, then use an edge to connect them. (b) is a hypergraph. When there is strong high-order correlation among multiple vertices, such as (v4, v6, v7, v8), a hyperedge e3 can be constructed to represent this correlation.

a) Shared Entropy: In order to cluster features based on the high-order correlation among features, first of all, a method is needed to measure this correlation. In this paper, the shared entropy illustrated in Zhang et al [46] is used as the measuring method.

For a set of features $\{X_1, X_2, \ldots, X_m (m \geq 2)\}$, the amount of information contained in these features can be measured by using joint entropy, which is presented in Eq.(1).

$$J(X_1, X_2, \ldots X_m) = -\sum_{x_1, x_2, \ldots x_m}^{m} P(x_1, x_2, \ldots x_m) \log_2 P(x_1, x_2, \ldots x_m) \quad (1)$$

Where $P(x_1, x_2, \ldots x_m)$ is the joint probability of m features. The definition of shared entropy is based on the joint entropy, and it is expressed in Eq.(2):

$$S(X_1, X_2, \ldots X_m) = (-1)^0 \sum_{i=1}^{m} J_i + (-1)^1 \sum_{1 \leq i < j \leq m} J_{ij} + \ldots + (-1)^{m-1} J_{1,2,\ldots,m} \quad (2)$$

Where J_i is an abbreviated of $J(X_i)$, J_{ij} is an abbreviated of $J(X_i, X_j)$, and $J_{1,2,\ldots m}$ is an abbreviated for the joint entropy $J(X_1, X_2, \ldots X_m)$ of m features.

As shown in Fig. 2, if the value of shared entropy is large, the correlation among features becomes higher. It should be noted that only nominal features can be measured with shared entropy, so the continuous features obtained before should be discretized. The discretization method adopted in this paper could be described as follows. All features are firstly normalized into [0,1] by Min-Max normalization, and then [0,1] is divided into 100 equivalent intervals. Finally, these features are discretized to the nearest points respectively.

b) Hypergraph: As shown in Fig. 3, The biggest difference between hypergraph and classic graph lies in the definition of edges. Each edge of classic graph connects two vertices, while each edge of hypergraph, which is called hyperedge can connect any number of vertices. Therefore, in nature, the classic graph can only

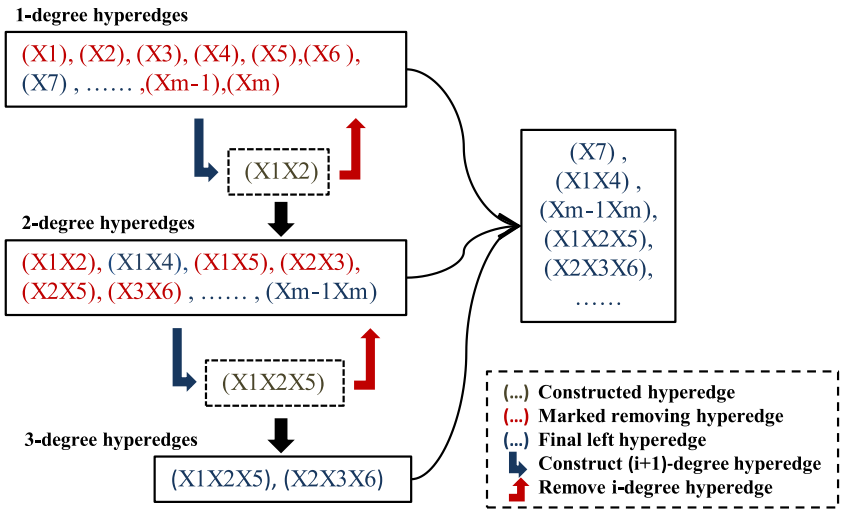

Fig. 4. Graphical illustration of hypergraph construction process [when (X1, X2) becomes a newly constructed hyperedge, (X1) and (X2) are marked as removing hyperedge; similarly, when (X1, X2, X5) becomes a newly constructed hyperedge, (X1, X2), (X2, X5) and (X1, X5) are marked as removing hyperedge].

represent pairwise correlation among vertices, while hypergraph can be used to represent the high-order correlation among multiple vertices.

As mentioned above, the shared entropy can be used to measure the relations among multiple features; it is naturally an analogy to the hyperedge which may contain more than two vertices. For m features $\{X_1, X_2, \ldots, X_m\,(m \geq 2)\}$, each feature can be represented by a vertex in Fig. 3, and it can be called as feature vertex. Therefore, a hypergraph can be constructed to show the relations among feature vertices. Before the specific construction, there are several parameters needed to be determined. Δd is the maximum degree of hyperedges (Given a hyperedge e, its degree $d(e)$ is the number of features within this hyperedge), Δs is the minimum shared entropy in hypergraph, and these two parameters can be used to control the scale of hypergraph. Besides, the parent-child relation between two hyperedges can be defined as $PC(e_i, e_j)$, and the definition is shown as Eq.(3). If e_i and e_j satisfy $PC(e_i, e_j) = 1$, then e_i is called as the parent hyperedge of e_j and conversely, e_j is called as the child hyperedge of e_i.

$$PC\left(e_i, e_j\right) = \begin{cases} 1 if (d(e_j) - d(e_i) == 1 and V_{e_i} \subset V_{e_j}) \\ 0 otherwise \end{cases} \tag{3}$$
$$V_{e_i} = \{v | v \in e_i\}, V_{e_j} = \{v | v \in e_j\}$$

Where V_{e_i} represents the feature vertex set contained in e_i, and V_{e_j} represents the feature vertex set contained in e_j.

The specific hypergraph construction process is shown in Fig. 4. Suppose that there are m features, at the first iteration, these hyperedges are created with 1-degree ($d(e) = 1$) and the shared entropy S_i for each feature vertex, $S_i = S(X_i)\{0 < i \leq m\}$ will be calculated. One hyperedge will be created if its shared entropy $S_i > \Delta s$ and then, the set of all hyperedges with 1-degree will be represented by $E_1 = \left\{e_1, e_2, \ldots e_k \; k \leq m\right\}$. In the second iteration, all hyperedges with 2-degree ($d(e) = 2$) will be created. Firstly, the shared entropy S_{ij} of each pair for features will be calculated, $S_{ij} = S(X_i, X_j)\{0 < i < j \leq m\}$. if $S_{ij} > \Delta s$, it means that there is strong enough correlation between feature vertex X_i and X_j. Then, a hyperedge can be constructed to represent this correlation, and the shared entropy S_{ij} will be the weight W_{ij}. When all second level hyperedges are completed, they will be presented by a hyperedge set E_2. The final step of second iteration is to remove all parent hyperedges of these hyperedges in E_2 from E_1. For example, as shown in Fig. 4, the parent hyperedges of hyperedge (X_1X_2) are the hyperedge (X_1) and the hyperedge (X_2). In the same way, when constructing the t-degree hyperedges, the shared entropy of each t–degree feature set is needed to be calculated. If the shared entropy is larger than the Δs, a hyperedge will be constructed to represent this relationship. Finally, before completing the t-degree hyperedge construction, all t-degree hyperedge will be integrated into E_t, and all parent hyperedges from E_{t-1} will be removed.

After constructing the hypergraph, it can be expressed as Eq. (4):

$$G = \{V, E, W\} \tag{4}$$

Where $V = \{v_1, v_2, \ldots, v_m\}$ is the set of all feature vertices, $E = \{e_1, e_2, \ldots, e_r\}$ is the set of all hyperedges, $W = \{w_1, w_2, \ldots, w_r\}$ is the set of all hyperedge weight. The weight is used to indicate the strength of the correlation. In this paper, the weight is expressed by the shared entropy. In addition, an incidence matrix H (size $|V| \times |E|$) is employed to represent the relationship of point-edge. The definition of H is shown in (5).

$$H(v, e) = \begin{cases} 1 if v \in e \\ 0 otherwise \end{cases} \tag{5}$$

The W(size($|E| \times |E|$)is a diagonal matrix containing all weights of hyperedges, then the adjacent matrix A (size $|V| \times |V|$) can be defined as in (6).

$$A = HWH^T \tag{6}$$

Where H^T (size $|E| \times |V|$) is the transpose of H, and A represents the relationship between each pair of features.

c)CLGA: After using hypergraph to represent the correlation between features, we need to cluster them, which means to divide the high-dimensional feature set into some low-dimensional feature clusters. To be more specific, firstly, the parameter γ will be set, which represents the number of clusters. And then, the matrix A calculated in (6) was taken as the input of CLGA method to generate a feature cluster set $C = \{c_1, c_2 \ldots, c_\gamma\}$ and a partition membership matrix B (size $|V| \times |C|$), which is used to indicate whether a feature is in a cluster.

$$B(v, c) = \begin{cases} 1 if v \in c \\ 0 otherwise \end{cases} \tag{7}$$

2.2.2. Optimal feature subset generation

After high-order correlation measurement process, all features have been distributed to different feature clusters based on the high-order correlation. In order to make full use of this correla-

tion, a feature selection method is needed to generate the optimal feature subset from these feature clusters.

*a) SVM-RFE:*SVM-RFE was initially proposed for ranking genes from gene expression data for cancer classification [52]. This method takes all features as the input and removes one feature out based on backward sequential selection, which is least significant for classification at each time. Although the SVM-RFE still has the ability to obtain the feature subset according to the feature order, the subset is hard to achieve a better classification result. The reason behind this situation is that the SVM-RFE method only considers the discrimination of feature itself and ignore the correlation among features. In this paper, the SVM-RFE is adopted as a comparison method.

b) SVM-RFE with Feature Clusters: In order to combine the discrimination of feature itself with the high-order correlation among features in the process of feature selection, a novel feature selection approach is proposed to combine SVM-RFE with feature clusters.

The second part of the feature selection in Fig. 1 outlines the proposed feature selection approach. The first step is to adopt the feature clusters as the input of the method. All features in the final optimal subset are generated from these clusters. In the second step, the feature dimensionality reduction process will be implemented for each feature cluster. Due to the strong correlation among features in the same cluster, these features can be regarded as reflecting the characteristics of the target object from similar angles. Therefore, stacking too many similar features together will not only cause redundancy of features, but also waste computing resources. That's why the feature dimensionality reduction process is very important and we finally empirically keep 3–5 features in each clusters after executing the PCA method. In the third step, the optimal feature subset can be generated by adopting the Sequential Forward Selection (SFS) method. This specific process is listed as follows: By starting with an empty set, the SFS iteratively selects one feature at a time and adds it to the current feature set. According to the ordered features set and the feature clusters, the chosen feature is the highest ranking among unselected features or is in the same clusters with the highest ranked feature. Which means, after selecting the highest ranked feature, K features that are in the same clusters with the selected feature will be chosen, and vice versa. It can be noted that the classification performance of the feature subset should be measured again when changes have happened. The measuring method is to implement SVM cross-validation experiment on a sample set, and record its average accuracy. So, this method will be terminated when there is no feature left in all clusters which means no more different feature subset can be generated. The final output of the proposed method will be the feature subset with the highest accuracy. Finally, the feature subset with the highest accuracy is the output of the method and the optimal feature subset.

2.3. Classification

After getting the optimal feature subset, we also need to combine it with the classifier to detect AD and MCI. The classification experiments in this paper are carried out on SVM. A grid-search method has been used to tune all parameters to get the best performance classifier.

3. Experiments and results

3.1. Material

The dataset adopted in this paper was obtained from the ADNI database (http://adni.loni.usc.edu/). In this paper, The selected T1-weighted MR images were obtained using 1.5 T scanner and there are a total of 170 subjects which includes 54 AD patients(AD), 58 normal controls(NC) and 58 mild cognitive impairment patients(MCI). These subjects were chosen as samples in the experiment. The basic information of subjects was introduced in Table 1.

Table 1
Basic information of the subjects.

	AD(N = 54)	MCI(N = 58)	NC(N = 58)
Gender(Male/Female)	22/32	27/31	30/28
Age(Mean ± SD2)	75.7 ± 7.1	74.8 ± 6.2	75.2 ± 5.6
MMSE3(Mean ± SD)	22.8 ± 2.3	26.7 ± 3.6	29.1 ± 5.0

[1] N, Number of subjects.
[2] SD, Standard Deviation.
[3] MMSE, Mini Mental State Examination.

3.2. Experiment setting

In order to evaluate the effectiveness of our proposed approach HOCDF, we designed relevant experiments to measure the diagnosis performance of AD/MCI on ADNI. Our experiment includes four classification tasks: 1) NC vs. AD classification, 2) NC vs. MCI classification, 3) MCI vs. AD classification, 4) 3-way classification(NC-MCI-AD).

In our proposed method HOCDF, there are some parameters need to be set. In the feature extraction section, we completely adopt the feature extraction method in our previous work, and the related parameters are exactly the same [36]. The second stage is feature selection which is divided into two parts. In the first part high-order correlation measurement, there are two parameters need to be determined. The first one is Δd which is the maximum degree of hyperedges. Due to the exponential growth of the calculation complexity caused by this parameter, it is set empirically to 4 (Δd =4) so as to balance the efficiency and accuracy of measuring the correlation in features. It means that the number of feature vertices in a hyperedge can't exceed 4. The second parameter that needs to be explicitly determined in feature clustering process is Δs which represents the minimum shared entropy for hyperedge construction. It means a constructed hyperedge will be preserved if the shared entropy of these feature vertices in it exceeds the Δs, otherwise, this hyperedge will be removed. In the experiment, Δs is empirically set to 0.4(Δs = 0.4). In another part optimal feature subset generation of feature selection, when selecting K related features from a certain feature cluster, the selecting strategy is to first choose the top-ranked and unselected features in this cluster. If all features in this cluster are selected, then stop the process of selecting related features. Besides, a 10-fold cross-validation method is adapted to obtain the optimal feature subset and reduce the impact of sample selection randomness. The training sample set is divided into 10 small batches. Each time, it selects one sample batch from them as the test set. The other nine sample batches are regarded as a training set, then the classification experiment will be evaluated. This process will be repeated ten times, and make sure that the sample batch, which is chosen as the test set, is different each time. The final result is the average accuracy of ten times. Note that, the training samples used in feature selection process will not appear in the final test sample set.

The proposed method HOCDF has been evaluated on four different classification tasks, and has been judged by the accuracy(ACC), sensitivity(SEN), specificity(SPEC), positive predictive value(PPV), negative predictive value(NPV), and the area under the receiver operating characteristic (ROC) curve (AUC) respectively. The ACC is the most common chosen metric as the comparison standard among methods. The SEN, SPEC, PPV and NPV describe how well the experiment captures the true presence or absence of the dis-

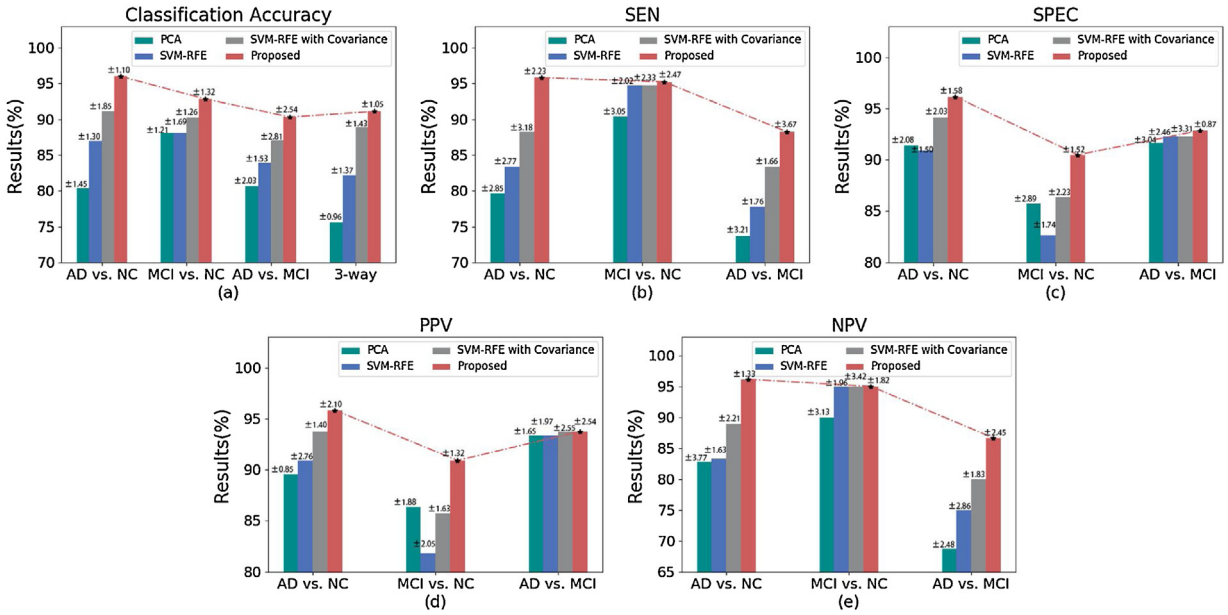

Fig. 5. Comparison between our proposed method and 3 methods without high-order correlation: (a)Accuracy(ACC), (b)Sensitivity(SEN), (c)Specificity(SPEC), (d)Positive Predictive Value(PPV), (e)Negative Predictive Value(NPV).

ease. Regarding of the SEN and SPEC, the higher value these metrics are, the lower chance the misdiagnosing will happen. The PPV and NPV are the prevalence of the disease in the target population.

In addition, the proposed method HOCDF is also compared with some feature selection methods and some AD/MCI classification methods. These counterpart methods are divided into two groups. The first group includes: 1)PCA, 2)SVM-RFE, 3)SVM-RFE with Covariance [36]. This group is adopted to prove the significance of high-order correlation of features in the process of feature selection. Another group of methods mainly includes some well-known competing methods. It is used to prove the effectiveness of the method HOCDF on AD/MCI classification.

3.3. Result

3.3.1. Comparison with methods without high-order correlation

Making full use of high-order correlation among features to feature selection is the biggest innovation in our proposed method. In order to test whether the high-order correlation is effective, the proposed method is compared with some other methods without high-order correlation. These methods are PCA, SVM-RFE, SVM-RFE with Covariance. PCA method is a classic feature dimensionality reduction method. It is clear that it does not consider the correlation among features, especially the high-order correlation. SVM-RFE is a simple and commonly used feature ranking method, which can select features according to the ranking results. However, this method also ignores the correlation among features during feature selection. The SVM-RFE with Covariance is a method proposed in our previous work which considers the pairwise correlation among features in feature selection, but ignoring the high-order correlation. This is the biggest difference between these two methods. The proposed method has been compared from five different perspectives and the comparison results is presented in Fig.5.

From the experiment result, it can be found that it can achieve better performance for AD/MCI classification when adopting the high-order correlation among features to the process of feature selection. For example, according to the comparison result in Fig.5(a), the SVM-RFE with Covariance and the proposed method which adopts the idea of the correlation among features can achieve a great improvement in all four classification tasks. When analyzing the classification result of these four tasks separately, the advantage of these two methods is more obvious. All five evaluation metrics are improved by least 3% in these two methods. Furthermore, the proposed method even increased the classification result by at least 5%. In addition, based on the experiment result, it can be found that the classification performance which adopts the high-order correlation is better than the one which only considers the pairwise correlation. After analyzing the comparison result between our proposed method HOCDF and SVM with Covariance in Fig.5(a), it can be proven that the proposed method HOCDF obtained a higher accuracy in all four tasks. The best improvement is up 4.82% on the AD vs.NC task. And on the other three tasks (MCI vs. NC, AD vs. MCI, 3-ways), it increases to 2.62%, 3.22% and 2.22% respectively. Except for the ACC, the proposed method also greatly improves on other four evaluation metrics. Overall, the HOCDF method obtains the best result on AD/MCI diagnosis, especially in ACC which is one of the most important metrics.

In addition, the confidence intervals of these methods are also presented in Fig. 5. It can be noticed that the proposed method is more stable in most situation. In other words, it can evaluate the power of the proposed method from another perspective.

3.3.2. Comparison with well-known classification methods based on ADNI

In order to further verify the effectiveness of the proposed method, it is also compared with some latest state-of-the-art classification methods. These methods include DL(Deep Learning method), SVM (classic SVM-based method), MKL-SVM (Multiple Kernel Learning Support Vector Machine method), HG (Hypergraph method). And we ensure that these methods are all based on ADNI database. The comparison results about AD vs. NC, MCI vs. NC, AD vs. MCI, 3-way are presented in Table 2, Table 3, Table 4, Table 5, respectively. The evaluation metrics, which are most commonly used to evaluate AD/MCI diagnostic methods, include ACC, SEN, SPEC, AUC.

Comparison results about AD vs. NC are listed in Table 2. The comparison method includes the multi-kernel learning, deep learning, and classic SVM-based methods. It can be seen in Table 2 that the proposed method achieves the highest classification accuracy. It reaches 96%, which even exceeds these deep learning methods.

Table 2
Comparison with existing studies using MRI data of ADNI for MCI vs. NC classification.

Method	Classifier	ACC (%)	SEN (%)	SPEC (%)	AUC (%)
Suk et al., 2014 [2]	DP	85.67	**95.37**	65.87	–
Suk et al., 2016 [5]	SVM	80.11	93.89	53.67	–
Payan et al., 2015 [8]	DP	92.11	–	–	–
Shi et al., 2015 [11]	SVM	80.91	79.07	82.7	–
An et al., 2016 [14]	SVM	79.90	85	67.5	86.9
Lei et al., 2017 [17]	SVM	80.32	64.35	86.67	86.2
Liu et al., 2017 [20]	MKL-SVM	86.35	89.49	85.68	91.07
Liu et al., 2017 [24]	HG	80.00	86.19	68.78	80.49
Suk et al., 2017 [22]	DP	72.91	77.60	68.22	73.61
Yao et al., 2018 [27]	SVM	86.70	–	–	–
HOCDF(Proposed)	SVM	**92.86** ± **1.32**	95.24 ± 2.47	**90.48** ± **1.52**	**97.05** ± **1.89**

Table 3
Comparison with existing studies using MRI data of ADNI for AD vs. MCI classification.

Method	Classifier	ACC (%)	SEN (%)	SPEC (%)	AUC (%)
Gray et al., 2012 [3]	MKL-SVM	83.50	79.90	86.40	–
Suk et al., 2015 [6]	MKL-SVM	83.70	–	–	–
Wee et al., 2013 [9]	MKL-SVM	79.24	78.03	80.46	88.82
Apostolova et al.,2014 [13]	SVM	70.00	–	–	61.00
Jin et al., 2016 [16]	MKL-SVM	88.63	91.55	86.25	90.70
Cheng et al.,2016 [19]	SVM	82.70	89.20	69.6	90.60
Payan et al., 2015 [8]	DP	86.84	–	–	–
Liu et al., 2017 [20]	MKL-SVM	**90.85**	**91.77**	89.56	93.55
HOCDF(Proposed)	SVM	90.32 ± 2.54	88.24 ± 3.67	**92.85** ± **0.87**	**97.50** ± **2.09**

Table 4
Comparison with existing studies using MRI data of ADNI for AD vs. NC classification.

Method	Classifier	ACC (%)	SEN (%)	SPEC (%)	AUC (%)
Suk et al., 2014 [2]	DP	93.35	94.65	95.22	–
Suk et al., 2016 [5]	SVM	95.09	92.00	**98.00**	–
Payan et al., 2015 [8]	DP	95.39	–	–	–
Shi et al., 2015 [11]	SVM	88.73	84.86	91.69	–
An et al., 2016 [14]	SVM	92.10	85.70	95.90	97.30
Lei et al., 2017 [17]	SVM	94.68	**97.90**	91.38	97.92
Liu et al., 2017 [20]	MKL-SVM	95.24	94.26	95.74	97.54
Suk et al., 2017 [22]	DP	91.33	92.72	89.94	92.72
Liu et al., 2017 [24]	HG	93.10	90.00	95.65	94.83
Liu et al., 2018 [26]	DP	91.09	88.05	93.50	95.86
Lu et al., 2018 [29]	DP	93.58	91.54	95.06	–
HOCDF(Proposed)	SVM	**96.00** ± **1.10**	95.83 ± 2.23	96.15 ± 1.58	**98.40** ± **0.95**

Note: DP means Deep Learning method; SVM means classic SVM-based method; MKL-SVM means Multiple Kernel Learning Support Vector Machine method; HG: Hypergraph method.

Table 5
Comparison with existing studies using MRI data of ADNI for 3-way classification.

Method	Classifier	ACC (%)	Computational Time
Hosseini-Asl et al., 2016 [1]	DP	89.10	–
Gupta et al., 2014 [4]	DP	85.00	–
Zhu et al., 2016 [7]	SVM	63.90	–
Zhu et al., 2016 [12]	SVM	73.35	–
Payan et al., 2015 [8]	DP	89.47	–
HOCDF(Proposed)	SVM	**91.11** ± **1.05**	< 1min

In the AUC metric, the proposed method also shows the best performance and reaches 98.40%. Although the HOCDF method is not the best one on the other two metrics SEN and SPEC, the result of SEN is lower than the result of Lei et al., 2017 [17] and the result of SPEC is slightly lower than the result of Suk et al., 2016 [5], our method is still in the second place and the value is more than 95%. In summary, all results are good enough to demonstrate the effectiveness of the proposed method on AD vs. NC classification task. In other words, the HOCDF method can precisely identify the AD from NC.

By analyzing the comparison result on MCI vs. NC shown in Table 3, the proposed method achieves the highest value on three metrics (ACC,SPEC,AUC). The value of ACC reaches 92.86% and is far more than other classic SVM-based methods, even higher than the other three deep learning methods. In SPEC, the proposed method reaches 90.48%, which is higher than the second-place method nearly 4%, and shows a distinct advantage. About the AUC, the proposed method is 6% higher than the multi-kernel method which is in the second-place, and the value reaches 97.05%. It is 10% higher than the SVM-based method. Besides, the proposed method reaches 95.24% on the SEN. Although it is slightly lower than the result 95.37% of deep learning method [2], our experiment result is much better than other methods. Therefore, the proposed method shows more excellent classification performance in MCI vs. NC classification tasks.

From Table 4 which presents the comparison result of AD vs. MCI, it can be noted that although the proposed method is slightly lower on the ACC and SEN metrics than the multi-kernel method [20], it is still in the second place and is just 0.53% lower than the first one. It shows great advantage comparing with the classic SVM-based method and is almost 8% higher than the highest one among them. Besides, our result is 3.5% higher than the deep learning method in Payan et al., 2015 [8]. In addition, the proposed method achieves the best result on both SPEC and AUC. On SPEC, the proposed method reaches 92.85%, which is at least 3% higher than the second-place method. And On AUC, the proposed method

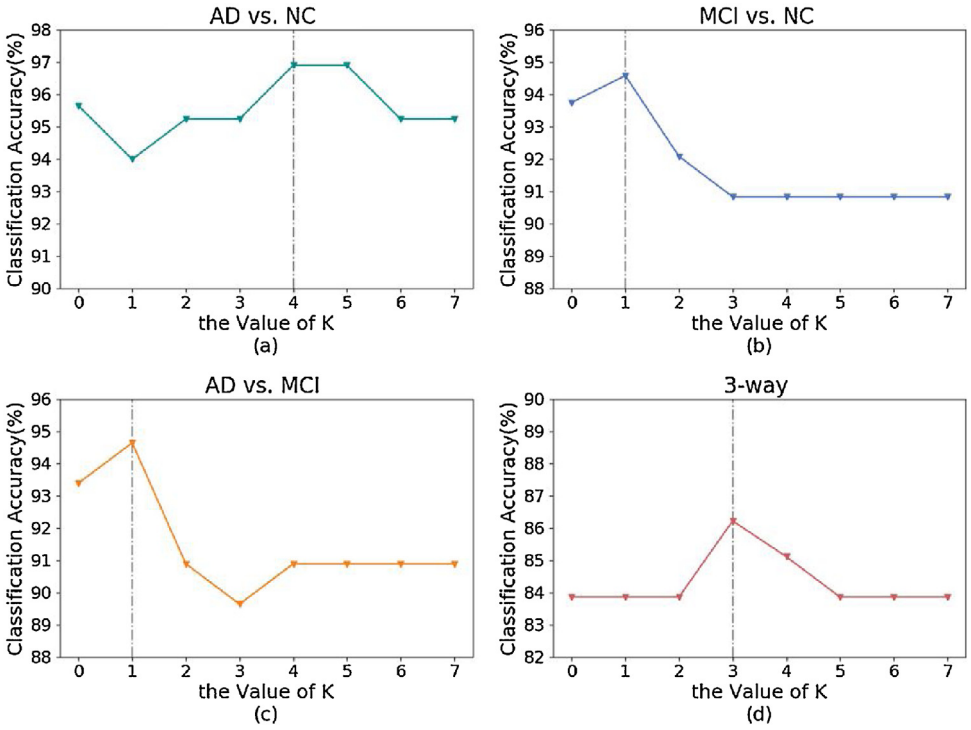

Fig. 6. The classification accuracy versus value of K on 4 tasks. (a)NC vs. AD, (b)NC vs. MCI, (c)MCI vs. AD, (d)3-way.

reaches 97.50%, which is almost 4% higher than the second-place method. Overall, the proposed method still has a good performance in AD vs. MCI classification task.

In the 3-way classification task, the ACC of the proposed method is the highest which its value reaches 91.11%. As shown in Table 5, it can be observed that most deep learning methods are significantly better than the SVM-based method, the result of ACC is at least 10%. But the proposed method shows an excellent performance in multi-class classification task and is better than all comparison methods, which also includes these deep learning methods. Comparing with the SVM-based method, the ACC of the proposed method exceeds almost 18% (17.76%). Such performance proves the effectiveness of our method on multi-class classification task.

All experiment results with the confidence interval of the proposed method is listed as follows: the ACC is 96%(±1.10%) on AD vs. NC, 92.86%(±1.32%) on MCI vs. NC, 90.32%(±2.54%) on AD vs. MCI, and 91.11%(±1.05%) on 3-way. For the other four metrics, the result of the proposed method is almost more than 90%. In detailed, the SEN is 95.83%(±2.23%) on AD vs. NC, 95.24%(±2.47%) on MCI vs. NC and 88.24%(±3.67%) on AD vs. MCI; The SPEC is 96.15%(±1.58%) on AD vs. NC, 90.48%(±1.52%) on MCI vs. NC and 92.85%(±0.87%) on AD vs. MCI; The PPV is 95.80%(±2.10%) on AD vs. NC, 90.91%(±1.32%) on MCI vs. NC and 93.75%(±2.54%) on AD vs. MCI; The NPV is 96.10%(±1.33%) on AD vs. NC, 95.00%(±1.82%) on MCI vs. NC and 86.67%(±2.45%) on AD vs. MCI.

In addition, although the feature extraction and feature selection stage will cost a long time, the computational time for classifying each MRI image in test dataset is less than one minute. Unfortunately, the computational time for other comparison methods can't be found in their articles, and also didn't provide any discussion on the computational efficiency. But, according to the experience, the computational time of deep learning method for classification or segmentation is usually less than one minute (as shown in Table 5). From this perspective, the computational efficiency of the proposed method can be accepted when classifying the MRI image.

In general, based on the excellent evaluation performance of the proposed method on all 4 tasks, it can be proven that it is an effective way for AD/MCI diagnosis and is able to help doctors to diagnose the AD/MCI in clinical setting with the high value of the ACC. In addition, the proposed method belongs to the traditional machine learning and its focus is on feature analysis and feature selection. The experiment result has proven two factors. The first one is that the research on traditional machine learning methods for diagnosing AD/MCI is still be significant. And, the second one is that the AD/MCI diagnostic method can be improved by using the high-order correlation among features. On the other hand, by comparing the proposed method with the deep learning method, it is not appropriate to simply draw conclusion that traditional machine learning methods are better than deep learning methods in AD/MCI diagnosis just through the final result metrics. Because the final results are related to specific testing sets which cann't be guaranteed to be same between comparison methods. we just make sure these methods are all based on ADNI database. Our results have proved the traditional machine learning method is still a good choice for AD/MCI diagnosis.

4. Discussion

The previous analysis of the experimental result can prove that the proposed method is an effective AD/MCI diagnosis method. The main innovation is to emphasize the high-order correlation among features. Therefore, how the high-order correlation affects the final classification result will be investigated by analyzing several related parameters and the feature selection process.

4.1. Effects of the parameter K

The K indicates the number of related features to be chosen after a top-ranked feature is selected. Therefore, analyzing these parameters can reflect how the high-order correlation affects the process of feature selection. When evaluating the parameter K, the ten fold cross-validation strategy was adopted, and the final result is the average one. As can be seen from Fig. 6, the accuracy varies with the changing of K. When K = 0, it means the proposed method will

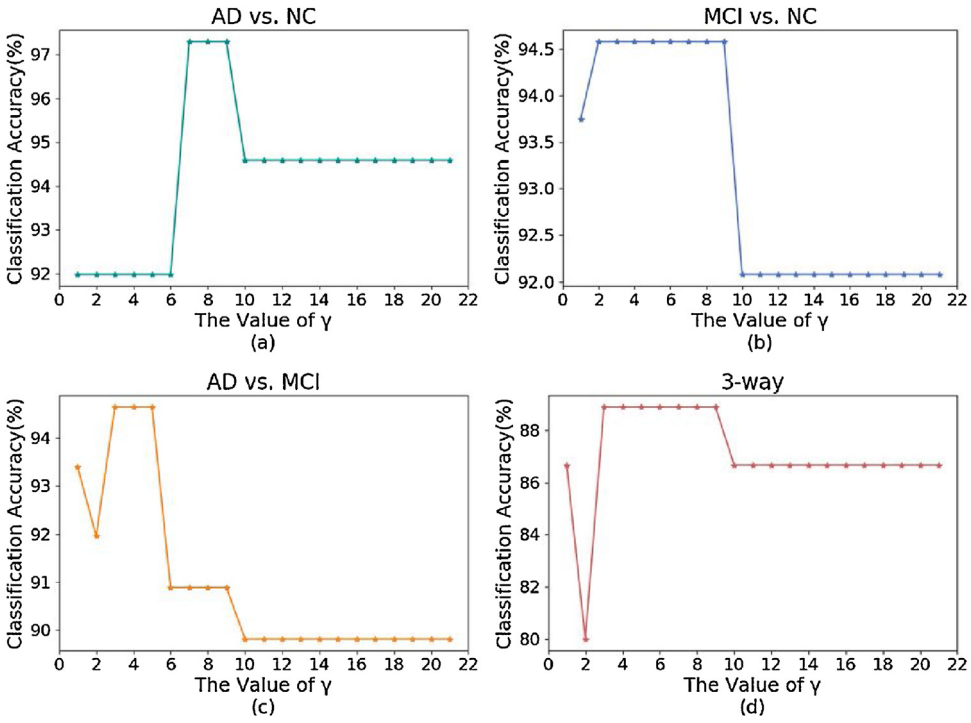

Fig. 7. The classification accuracy versus value of γ on 4 tasks: (a)NC vs. AD, (b)NC vs. MCI, (c)MCI vs. AD, (d)3-way.

select features just according to the ranking result of features and ignore the correlation among features. In other words, the proposed method will be the same as the SVM-RFE one.

What can be found from Fig. 6 is listed as follows: First, the curve of the accuracy rate will appear as a clear peak and eventually tend to be stable with the increase of K. The growth of accuracy before reaching the highest point indicates that appropriately employing the high-order correlation among features can improve the final classification accuracy. However, when the classification accuracy reaches the highest point, the accuracy will start to decrease with the continuously increasing of the K. This process indicates that overemphasizing the high-order correlation and neglecting the strength of feature itself may cause opposite effect. Note that especially in Fig. 6(b) and Fig. 6(c), the final stable accuracy is even lower than the accuracy obtained where K = 0. Second, for different classification tasks, the optimal value of K may be different. From Fig. 6, it can be found that the K value corresponding to the highest accuracy in the four figures are not the same. In Fig. 6(a), the optimal value of K is 4. In Fig. 6(b) and Fig. 6(c), the optimal value of K is 1, and in Fig. 6(d), the optimal value of K is 3. Therefore, for different classification tasks, the parameter K needed to be individually determined for a specific sample set. Third, even with 10-fold cross-validation method, the average accuracy obtained by this method is still considerable. In Fig. 6(a), the highest accuracy is close to 97%. In Fig. 6(b) and Fig. 6(c), both the highest accuracy are close to 95%. And in Fig. 6(d), the accuracy of 3-way is over 86%. On the other hand, these results can prove the robustness of the proposed method.

4.2. Effects of the parameter γ

The parameter γ indicates the number of feature clusters. Before the feature selection, all features will be clustered according to the high-order correlation among features. Furthermore, features contained in the same cluster will be ensured to have a strong high-order correlation and to have a weak or no high-order correlation with these features in other clusters. The number of feature clusters γ is a parameter that needs to be set in advance and it will have an impact on the final classification result. With the value of γ increasing, the number of features grouped to each feature cluster will decrease and the measurement of high-order correlation will be more precise. However, if γ is set to be too large, the number of features in some clusters will be too small (even less than 2) to seek related features. Therefore, how to set a proper number of feature clusters is very important for the proposed method. Some experiments have been evaluated to analyze the impact of γ on the classification accuracy, and all results are presented in Fig. 7. Note that, γ = 1 indicates all features are put into one cluster. It means the relationship among features is ignored and features are chosen just by the ranking score. In other words, this method is changed into the original SVM-RFE method.

From Fig. 7, It can be found that as the number of feature clusters increases, the accuracy rate will increase at the beginning, then decrease and finally be stable on all four classification tasks. For example, in Fig. 7(a), when the number of feature clusters changes from 6 to 7, the accuracy increases by 5%. However, when the value changes from 9 to 10, the accuracy decreases by 3%. Eventually, with the value of γ increasing continuously, the accuracy will stabilize to 94.5%. So, setting a proper number of feature clusters will significantly improve the classification accuracy. In Fig. 7 (b), when the γ is set between 2–9, the accuracy is 2.5% higher than when set between from 10 to 20. In Fig. 7(a) and Fig. 7(c), the accuracy is going to vary about 5% depending on whether γ is set properly or not. In Fig. 7(d), when the γ is set between 3–9, the accuracy is almost 10% higher than the γ is set to 2.

In order to clearly show how the classification performance is affected by the hyperparameters of K&γ, a 3-way experimental result is presented in a color map (as shown in Fig. 8). The brightness of the color indicates the accuracy level, where the bright color presents the higher accuracy rate and the dark color indicates lower accuracy rate. As shown in Fig. 8, on the one hand, the point representing the highest accuracy rate is concentrated in the area where K is in 3–5, and γ is in 3–5. In other words, the accuracy of 3-way task exceeds 90% in this area, which explores that the classification

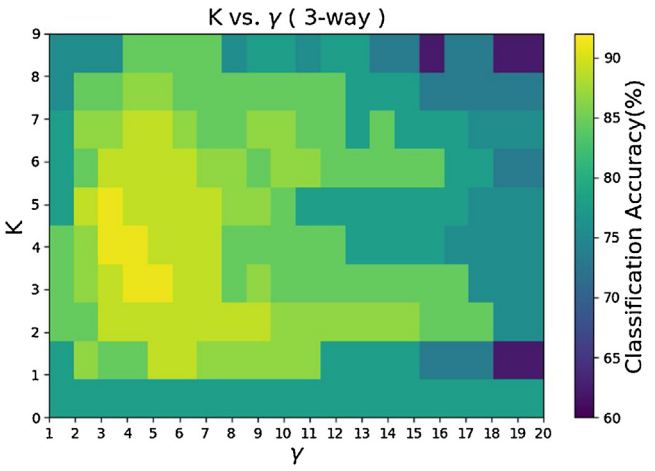

Fig. 8. The classification accuracy versus values of K& γ on 3-way task.

performance can be significantly improved by setting the proper value of K&γ. On the other hand, when the point is far away from the high-accuracy concentration area, the final classification accuracy will decrease a lot by setting the value of K&γ too large or too small. Especially when γ is set to be greater than 15, and K is set to 1 or 9, the accuracy is even lower than 70%. Overall, this experiment result explores that the final classification accuracy can be improved by appropriately adopting the high-order correlation among features. However, it should be noticed that overemphasizing the usefulness of the high-order correlation and neglecting the strength of feature itself may cause opposite effect.

4.3. Process of features selection

In this section, how the number of sequence features (sequence features mean the feature subset that generated in each iteration) affects the classification accuracy will be evaluated and the result is presented in Fig. 9. From the Fig. 9, it can be found that the accuracy is improved sharply at the beginning, and the optimal subset is identified in a very quick way. That is because top-ranked features and its highly relevant features are added into the feature subset at the beginning. For example, in Fig. 9(a), the accuracy is significantly increased when the number of features in subset changes from 0 to 11. When the number of features equals to 11, the best accuracy is achieved, more than 95%. In the other three figures of Fig. 9, the proposed method can obtain the optimal feature subset within the first ten choices. To be more specific, Fig. 9(b) can achieve the best accuracy at the 5^{th} choosing iteration, Fig. 9(c) is in the 9^{th} iteration, and Fig. 9(d) is 7^{th} iteration. We also found that even though only eight morphometric features have been extracted totally in feature extraction part, they almost make up half of the final optimal subset. In addition, after going through the best point, the accuracy rate begins to decrease significantly when more features are chosen into the feature subset. It means that selecting too many features not only can not improve the feature subset, but also may interfere with other features so as to cause the opposite result. Another finding is that the accuracy rate tends to be stable eventually. This is because a large number of features have been added to the feature subset in the early stage, and the rest of features selected in the final stage are regarded as less important ones in the proposed method. These features have less impact on the accuracy. Overall, it can be proven that the proposed method is an efficient feature selection method because it has the ability to obtain the optimal feature subset quickly.

5. Conclusion

In this paper, an effective classification method for diagnosing the AD/MCI, which is based on detecting high-order correlation among features, has been proposed. This method has the ability to make full use of more complicated relations among features to generate better feature subset. Comparing with our previous work which only considers the pairwise relationship between features, the HOCDF can obtain better feature subset in the process

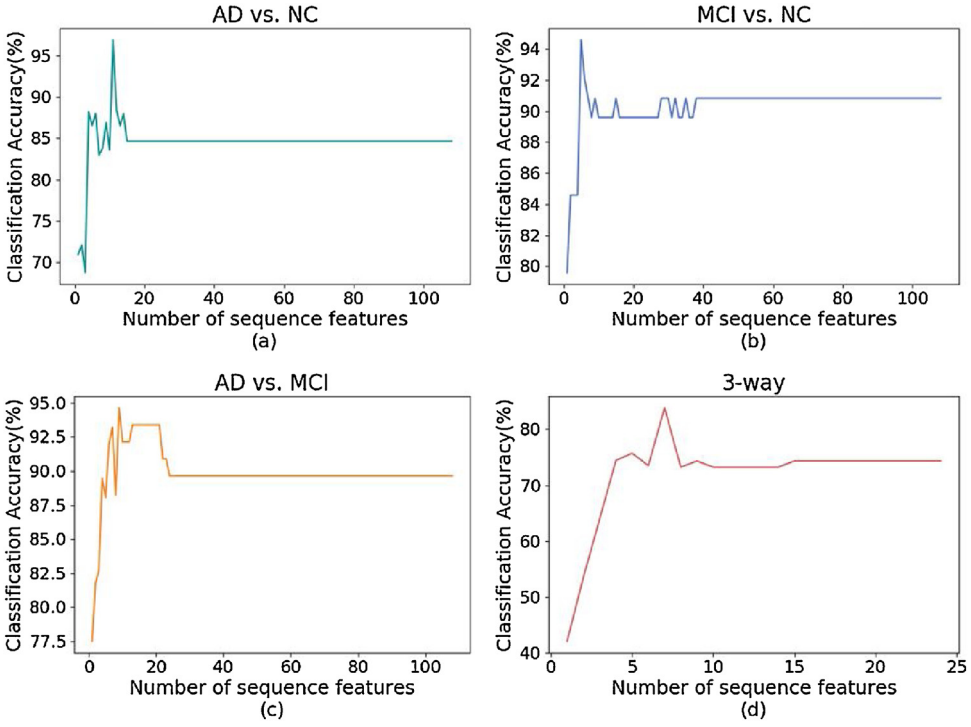

Fig. 9. The classification accuracy versus the number of sequence features on 4 tasks: (a)NC vs. AD, (b)NC vs. MCI, (c)MCI vs. AD, (d)3-way.

of feature selection so as to facilitate the process of diagnosing the AD/MCI. This proposed method has also been evaluated on the ADNI database. According to the experiment result, it can be seen that the proposed method is an effective method for improving the AD/MCI classification performance and achieves a significant improvement by comparing with other state-of-the-art counterpart methods.

Beyond that, there are some future works needed to be done to improve the proposed method. Firstly, more datasets like pMCI, sMCI and multi-dimensional medical images such as FDG-PET should be evaluated on the proposed framework so as to show the generality of the proposed framework. Secondly, except the texture and morphometric features, more types and more numbers of features should be extracted from original images, especially the feature which can be interpreted from the neurobiological perspective. So it can be seen that how the optimal feature subset related to the pathology of AD and MCI. Besides, more diverse and more robust features can be extracted by using the method mentioned in [53]. Thirdly, the computational time is an important metric to show the efficiency of the proposed method. In this paper, the current implementation efficiency of our method is relatively low, especially the feature extraction process and feature selection process take a relatively long time. In future, we plan to borrow the idea from the [54,55] to further improve the computational efficiency and classification performance of the proposed method. Last but not the least, the visualization of intermediate results, such as features selected with a high-order correlation and corresponding hypergraphs, would be useful to clearly show the effectiveness of the proposed method. But it is still an important and challenge work, which will be considered as a future research work.

Acknowledgments

This work was supported by the Natural Science Foundation of Guangdong Province [grant number 2018A030313354], the Neijiang Intelligent Showmanship Service Platform Project [grant number 180589], the Sichuan Science-Technology Support Plan Program [grant number 2019YJ0636, 2018GZ0236, 18ZDYF2558], and the National Science Foundation of China - Guangdong Joint Foundation [grant number U1401257].

References

[1] E. Hosseini-Asl, R. Keynton, A. El-Baz, Alzheimer's disease diagnostics by adaptation of 3D convolutional network„ in: IEEE International Conference on Image Processing, 2016.
[2] H.I. Suk, S.W. Lee, D. Shen, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, Neuroimage 101 (2014) 569.
[3] K.R. Gray, R. Wolz, R.A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease, Neuroimage 60 (2012) 221–229.
[4] A. Gupta, A.S. Maida, M. Ayhan, Ñatural image bases to represent neuroimaging data„ in: International Conference on Machine Learning, 2014, pp. 987–994.
[5] H.I. Suk, S.W. Lee, D. Shen, Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis, Brain Struct. Funct. 221 (2016) 2569–2587.
[6] H.I. Suk, S.W. Lee, D. Shen, Latent feature representation with stacked auto-encoder for AD/MCI diagnosis, Brain Struct. Funct. 220 (2015) 841–859.
[7] X. Zhu, H.I. Suk, K.H. Thung, Y. Zhu, G. Wu, D. Shen, Joint discriminative and representative feature selection for alzheimer's disease diagnosis, Mach. Learn. Med. Imaging (2016) 77–85.
[8] A. Payan, G. Montana, Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, Comput. Sci. (2015).
[9] C.Y. Wee, P.T. Yap, D. Shen, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Hum. Brain Mapp. 34 (2013) 3411–3425.
[10] A. S. Association, 2014 Alzheimer's disease facts and figures, Alzheimers Dement. J. Alzheimers Assoc. 10 (2014) e47.
[11] B. Shi, Y. Chen, P. Zhang, C.D. Smith, J. Liu, Nonlinear feature transformation and deep fusion for alzheimer's disease staging analysis, Pattern Recognit. (2015) 304–312.
[12] X. Zhu, H.I. Suk, S.W. Lee, D. Shen, Ŝubspace regularized sparse multi-task learning for multi-class neurodegenerative disease identification„ IEEE Transactions on Biomedical Engineering 63 (2016) 607.
[13] L.G. Apostolova, K.S. Hwang, O. Kohannim, D. Avila, D. Elashoff, C. R. J Jr, et al., ÄpoE4 effects on automated diagnostic classifiers for mild cognitive impairment and Alzheimer's disease, Neuroimage Clin. 4 (2014) 461–472.
[14] L. An, E. Adeli, M. Liu, J. Zhang, D. Shen, *Semi-supervised Hierarchical Multimodal Feature and Sample Selection for Alzheimer's Disease Diagnosis*, Springer International Publishing, 2016.
[15] M. Grundman, R.C. Petersen, S.H. Ferris, R.G. Thomas, P.S. Aisen, D.A. Bennett, et al., M̈ild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials„ Arch. Neurol. 61 (2004) 59.
[16] L. Jin, L. Min, L. Wei, F.X. Wu, P. Yi, J. Wang, C̈lassification of alzheimer's disease using whole brain hierarchical network„ in: IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2016, vol. PP, pp. 1-1.
[17] B. Lei, Y. Peng, T. Wang, S. Chen, N. Dong, Relational-regularized discriminative sparse learning for alzheimer's disease diagnosis, IEEE Trans. Cybern. (2017) 1–12.
[18] S. Salloway, Current and future treatments for alzheimer's disease, Ther. Adv. Neurol. Disord. 6 (2013) 19–33.
[19] B. Cheng, M. Liu, D. Zhang, B.C. Munsell, D. Shen, Domain transfer learning for MCI conversion prediction, IEEE Trans. Biomed. Eng. 62 (2016) 1805–1817.
[20] J. Liu, J. Wang, Z. Tang, B. Hu, F.X. Wu, Y. Pan, Ïmproving alzheimeres disease classification by combining multiple measures„ in: IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2017, vol. PP, pp. 1-1.
[21] D. Zhang, D. Shen, Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers, PLoS One 7 (2012), e33182.
[22] H.I. Suk, S.W. Lee, D. Shen, Deep ensemble learning of sparse regression models for brain disease diagnosis, Med. Image Anal. 37 (2017) 101–113.
[23] H.I. Suk, C.Y. Wee, D. Shen, *Discriminative Group Sparse Representation for Mild Cognitive Impairment Classification*, Springer International Publishing, 2013.
[24] M. Liu, J. Zhang, P.T. Yap, D. Shen, View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data, Med. Image Anal. 36 (2017) 123–134.
[25] T. Tong, R. Wolz, Q. Gao, J.V. Hajnal, D. Rueckert, Multiple instance learning for classification of dementia in brain MRI, Med. Image Anal. 18 (2014) 808–818.
[26] M. Liu, J. Zhang, E. Adeli, D. Shen, Ľandmark-based deep multi-instance learning for brain disease diagnosis„ Med. Image Anal. 43 (2018) 157.
[27] D. Yao, V.D. Calhoun, Z. Fu, Y. Du, J. Sui, An ensemble learning system for a 4-Way classification of alzheimer's disease and mild cognitive impairment, J. Neurosci. Methods 302 (2018).
[28] E. Janoušová, M. Vounou, R. Wolz, K.R. Gray, D. Rueckert, G. Montana, Biomarker discovery for sparse classification of brain imagesin Alzheimer's disease, Neurology Neurosurgery Neurosciences 2012 (2012).
[29] M. Lu, K. Popuri, G. Ding, R. Balachandar, Multiscale deep neural network based analysis of FDG-PET images for the early diagnosis of Alzheimer's disease, Med. Image Anal. (2018).
[30] L. Sørensen, C. Igel, N.L. Hansen, M.J. Lauritzen, M. Osler, E. Rostrup, et al., Validation of hippocampal texture for early Alzheimer's disease detection: generalization to independent cohorts and extrapolation to very early signs of dementia, Alzheimers Dementia 10 (2014), pp. P133–P133.
[31] W.Y. Wang, J.T. Yu, Y. Liu, R.H. Yin, H.F. Wang, J. Wang, et al., Voxel-based meta-analysis of grey matter changes in Alzheimer's disease, Transl. Neurodegener. 4 (2015) 1–9.
[32] G.W. Jiji, G.E. Suji, M. Rangini, An intelligent technique for detecting Alzheimer"s disease based on brain structural changes and hippocampal shape, Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 2 (2014) 121–128.
[33] P. Ghorbanian, D.M. Devilbiss, A. Verma, A. Bernstein, T. Hess, A.J. Simon, et al., Identification of resting and active state EEG features of alzheimer's disease using discrete wavelet transform, Ann. Biomed. Eng. 41 (2013) 1243–1257.
[34] G.F. Busatto, B.S. Diniz, M.V. Zanetti, Voxel-based morphometry in Alzheimer's disease, Expert Rev. Neurother. 8 (2008) 1691–1702.
[35] Y. Guo, Z. Zhang, B. Zhou, P. Wang, H. Yao, M. Yuan, et al., Grey-matter volume as a potential feature for the classification of Alzheimer's disease and mild cognitive impairment: an exploratory study, Neurosci. Bull. 30 (2014) 477–489.
[36] Y. Ding, C. Zhang, T. Lan, Z. Qin, X. Zhang, W. Wang, C̈lassification of alzheimer's disease based on the combination of morphometric feature and texture feature„ in: IEEE International Conference on Bioinformatics and Biomedicine, 2015, pp. 409–412.
[37] Z. Huo, D. Shen, H. Huang, *New Multi-task Learning Model to Predict Alzheimer's Disease Cognitive Assessment*, Springer International Publishing, 2016.
[38] J. Xu, C. Deng, X. Gao, D. Shen, H. Huang, P̈redicting alzheimer's disease cognitive assessment via robust Low-rank structured sparse model„ in: Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 3880–3886.
[39] A.M. Martínez, A.C. Kak, PCA versus LDA, Pattern Anal. Machine Intell. IEEE Trans. on 23 (2001) 228–233.
[40] S.J. Hanson, Y.O. Halchenko, B̈rain reading using full brain support vector machines for object recognition: there is no f̈aceïdentification area, Neural Comput. 20 (2008) 486.
[41] C. Chu, A.L. Hsu, K.H. Chou, P. Bandettini, C. Lin, Does feature selection improve classification accuracy? Impact of sample size and feature selection

on classification using anatomical magnetic resonance images, Neuroimage 60 (2012) 59–70.
[42] A.R. HidalgoMuñoz, J. Ramírez, J.M. Górriz, P. Padilla, Regions of interest computed by SVM wrapped method for Alzheimer's disease examination from segmented MRI, Front. Aging Neurosci. 6 (2014) 20.
[43] X. Wang, K. Liu, J. Yan, S.L. Risacher, A.J. Saykin, S. Li, et al., Predicting interrelated alzheimer's disease outcomes via new self-learned structured low-rank model, Inf. Process. Med. Imaging (2017) 198–209.
[44] X. Zhu, H.I. Suk, D. Shen, Matrix-similarity based loss function and feature selection for alzheimer's disease diagnosis, in: Computer Vision & Pattern Recognition, 2014, pp. 3089.
[45] Y. Shi, H.I. Suk, G. Yang, D. Shen, Joint coupled-feature representation and coupled boosting for AD diagnosis, in: Computer Vision and Pattern Recognition, 2014, pp. 2721–2728.
[46] L. Zhang, Y. Gao, C. Hong, Y. Feng, J. Zhu, D. Cai, Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition, IEEE Trans. Cybern. 44 (2017) 1408–1419.
[47] Z. Yu, Z. Han, X. Chen, S.W. Lee, D. Shen, Hybrid high-order functional connectivity networks using resting-state functional MRI for mild cognitive impairment diagnosis, Sci. Rep. 7 (2017) 6530.
[48] Z. Yu, Z. Han, X. Chen, M. Liu, X. Zhu, D. Shen, Ïnter-subject Similarity Guided Brain Network Modeling for MCI Diagnosis„ 2017.
[49] D. Zhou, J. Huang, Ïearning with hypergraphs: clustering, classification, and embedding„ in: International Conference on Neural Information Processing Systems, 2006, pp. 1601–1608.
[50] M. Torabi, R.D. Ardekani, E. Fatemizadeh, Ïiscrimination between alzheimer's disease and control group in MR-images based on texture analysis using artificial neural network„ in: International Conference on Biomedical and Pharmaceutical Engineering, 2006, pp. 79–83.
[51] B. Long, X. Xu, Z. Zhang, P.S. Yu, Community learning by graph approximation, in: IEEE International Conference on Data Mining, 2007, pp. 232–241.
[52] I. Guyon, Erratum: gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2001) 389–422.
[53] Y. Zhang, S.N. Chang, G. Zhou, J. Jin, A. Cichocki, Temporally constrained sparse group spatial patterns for motor imagery BCI, IEEE Trans. Cybern. (2018) 1–11, vol. PP.
[54] Z. Jin, G. Zhou, D. Gao, Y. Zhang, EEG classification using sparse Bayesian extreme learning machine for brain–computer interface, Neural Comput. Appl. (2018).
[55] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X.Y. Wang, et al., Sparse group representation model for motor imagery EEG classification, IEEE J. Biomed. Health Inform. (2018), vol. PP, pp. 1-1.