Combining Multiple Network Features for Mild Cognitive Impairment Classification

Lipeng Wang
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
China, 210016, Email: wanglipeng@nuaa.edu.cn

Fei Fei
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
China, 210016, Email: feifei@nuaa.edu.cn

Biao Jie
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
China, 210016, Email: jbiao@nuaa.edu.cn

Daoqiang Zhang*
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
China, 210016, Email: dqzhang@nuaa.edu.cn

***Abstract*—Connectivity-network-based techniques have been recently developed for the diagnosis of Alzheimers disease (AD) as well as its prodromal stage, i.e., mild cognitive impairment (MCI). However, most existing methods focus on using only a single property of connectivity networks (e.g., the correlation between paired brain regions), which can not fully reflect the topological information among multiple brain regions. To address that problem, in this paper we propose a novel connectivity-network-based framework to fuse multiple properties of network features for MCI classification. Specifically, two different types of network features (i.e., brain region and subgraph) are respectively used to quantify two different properties of networks, where two kinds of feature selection methods are further performed to remove the irrelevant and redundant features. Then, multi-kernel learning technique is adopted on those corresponding selected features to obtain the final classification results. We evaluate our proposed method on a real MCI dataset containing 12 MCI patients and 25 healthy controls. The experimental results show that by using multiple properties of network features our method achieves better performance than traditional methods using only single property of network features.**

***Index Terms*—Mild cognitive impairment(MCI), Brain connectivity network, Subgraph, Multiple features.**

I. Introduction

Alzheimer's disease (AD), characterized by a progressive impairment of cognitive and memory functions, is one of the most prevalent neurodegenerative brain disease in elderly people. Its prodromal stage called mild cognitive impairment (MCI), an intermediate state of cognitive function between normal aging and dementia, has also gained a large amount of attentions recently due to its high risk of progressing to AD within years.

Recently, machine learning and pattern recognition techniques have been widely used to identify AD and MCI at an individual level, rather than at a group level, i.e., comparison between different clinical groups. Most of these works focus on using regions-of-interest (ROIs, i.e., brain regions) or voxel-wise features extracted from single modality. Among them, brain regions features are commonly used. However, these approaches cannot reflect the topological information among multiple brain regions. It can enable us to obtain a better understanding of the brain pathology through exploring structural or functional interactions among brain regions. Therefore, studies of human connectome has attracted a huge of attentions [1][2].

Existing studies on brain connectivity can be roughly divided into two categories: 1)group analysis based methods [3][4]; 2)machine learning based methods [5]. In the first category, graph theoretical analysis is often used to demonstrate topological differences of the brain network between patients with disease and healthy controls (HC), and a series of abnormal connectivities have been reported in those studies. While in the second category, features based on vectors are often extracted in those methods and used for classification.

Brain region based features have been widely used in current connectivity-network-based classification of brain diseases. For example, local cluster coefficient [6], characteristic path length [7], betweenness centrality [8] are often used as the features extracted from the connectivity-networks. However, one limitation of these method is missing the topological information of the whole brain network. Recently, subgraph-based feature has been demonstrated effective by many studies [9][10]. Notably, sugraph-based methods are not sensitive to the change of single brain region. To the best of our knowledge, few works ever use both kinds of features (i.e., brain region and subgraph) for identifying individuals with disease from HC.

It is worth noting that, Jie et.al [11] combined node-related property and whole network-topology-related property for classification. However, an obvious disadvantage of their method is that they performed their experiments by using the whole brain connectivity network without subgraph selection, which may involve irrelevant and redundant features. In other words, they didn't consider the disease may be only related to a small network instead of the whole brain connectivity network.

In this paper, we propose a new method which employs both brain region and subgraph as the features to classify MCI from HC. Specifically, we extract the brain region features through

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.98

computing the local cluster coefficient of each brain region. At the same time, we extract the subgraph features through subgraph mining. And then feature selection is conducted on both of the two types of features, After then, we construct a kernel with the brain region feature as most studies did. For subgraph features, we can also obtain a kernel matrix on the reconstructed graph data through graph kernel, which is adopted to measure the similarity between two graphs. In the end, we combine the two kernel matrices and choose kernel-Support Vector Machine (k-SVM) as our classifier. The experimental results show that by using multiple properties of network features our method achieves better performance than traditional methods using only single property of network features.

II. Materials and method

A. Data acquisition

Subjects used in the current study were recruited by the Duke-UNC Brain Imaging and Analysis Center (BIAC), Durham, North Carolina, USA. There are 37 participants, including 12 MCI patients and 25 healthy controls. Demographic information of the participants is shown in Table 1. Informed consent was obtained from all participants, and the experimental protocols were approved by the institutional ethics board. All recruited subjects were diagnosed by expert consensus panels. A 3.0-Tesla GE Signa EXCITE scanner was used to acquire resting-state functional MRI (fMRI) volumes. The fMRI volumes of each participant were acquired with the following parameters: TR/TE=2000/32ms, flip angle=77°, acquisition matrix=64×64, FOV=256×256 mm^2, voxel resolution=4×4×4 mm^3, 34 slices, 150 volumes, and voxel thickness=4mm. During scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen, which lasted for 5 minutes

TABLE I
Characteristics of the participants in this study

Group	MCI	HC
NO. of subjects (male/female)	6/6	9/16
Age(mean±SD)	75.0±8.0	72.9±7.9
Years of education (mean±SD)	18.0±4.1	15.8±2.4
MMSE(mean±SD)	28.5±1.5	29.3±1.1

Note:MMSE=Mini-Mental State Examination

B. Method

It has been reported that the connectivity of brains with AD/MCI differ from those of normal brains [3][12]. Thus, these abnormal connectivity may serve as potential biomarkers for diagnosis of AD/MCI. In this paper, we propose a new method which utilizes multiple features of brain network to improve the disease diagnosis performance. Fig 1 has shown the framework of our method. Specifically, we construct connectivity network from the respective fMRI data through an appropriate thresholded preprocessing for each subject. Then, in this paper, different from existing studies which employ multiple features from multiple modalities [13][14], which cannot be satisfied in most cases, we extract multiple features (i.e., brain region features and subgraph features) from single modality. After then, feature selection is performed on both of two types of features respectively, and corresponding kernel matrices are computed to quantify two different types of network features. Finally, the multiple kernel support vector machine (SVM) is adopted to fuse these two heterogeneous kernels for distinguishing the individuals with MCI from the healthy controls. The core of our proposed method is summarized below and will be discussed comprehensively in the subsequent sections.

* Extraction of multiple features of connectivity network using appropriate threshold;
* Feature selection of the brain region features, and compute the corresponding kernel matrix;
* Feature selection of the subgraph features, and compute the corresponding kernel matrix;
* Integration of different yet complementary network features using a multi-kernel SVM;

1) Image preprocessing and network construction: The preprocessing step of the fMRI images, which includes slice timing correction and head-motion correction, are performed using the Atatistical Parametric Mapping software package (SPM8, available at http://www.fil.ion.ucl.ac.uk.spm). Since the regions of ventricles and white matter(WM) contain a relatively high proportion of noise caused by the cardiac and respiratory cycles, we utilized only the blood oxygenation level dependent (BOLD) signals extracted from gray matter (GM) tissue to construct the functional connectivity network. Then the fMRI scans are further parcellated into 116 ROIs by warping the AUtomated Anatomical Labeling (AAL) [15] template. After that, for each subject, the mean timeseries of each ROI is computed by averaging the fMRI time series over all voxels in that particular ROI. Finally, by using pairwise Pearson correlation coefficient, a functional connectivity network is constructed with the nodes of network corresponding to the ROIs and the weights of edges corresponding to the Person correlation coefficients between a pair of ROIs.

2) Brain regions features based method: In this section, we regard brain regions as the feature of connectivity network. However, in connectivity network, there exists a large number of low level features related to brain regions. Therefore, it is crucial to extract meaningful features from the connectivity networks. In our study, we choose the local cluster coefficient as our measurement of feature, which has been demonstrated its effectiveness in [16][17].

Given the connectivity network (matrix) $G = [\omega(i,j)] \in R^{n\times n}$ and the threshold T, the connectivity network was thresholded by using the following formulation:

$$\omega_m(i,j) = \begin{cases} 0 & if \ \ \omega(i,j) < T \\ \omega(i,j) & otherwise \end{cases} \quad (1)$$

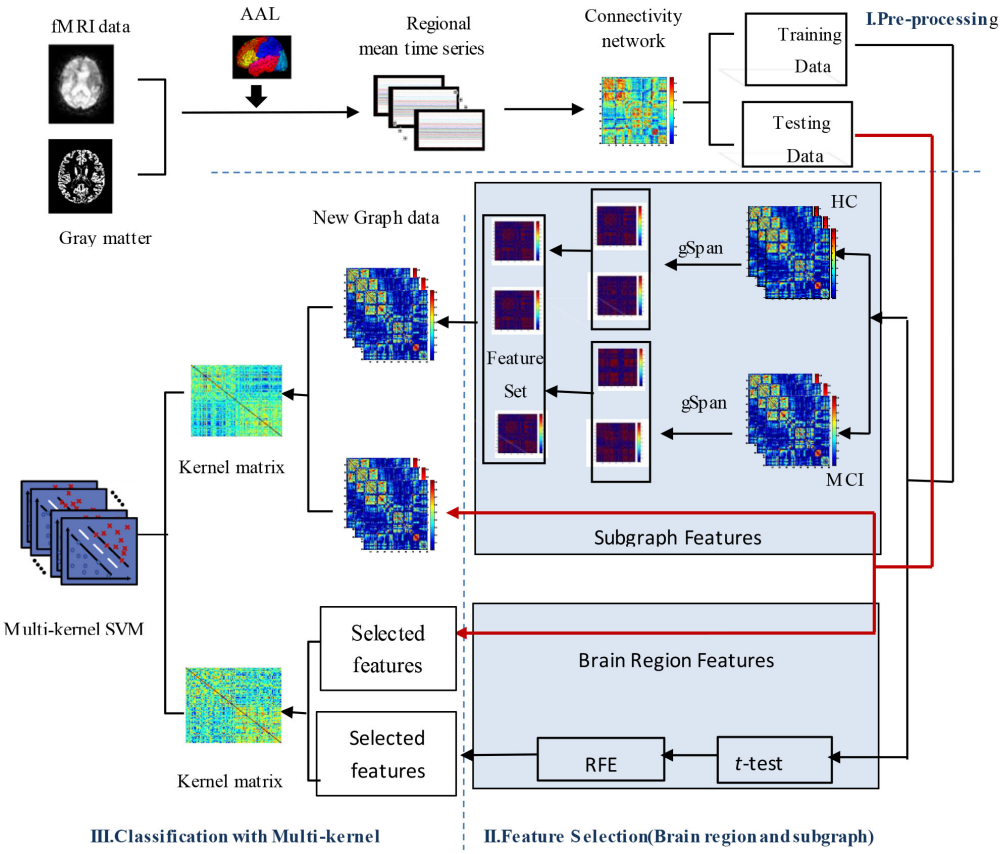


Fig. 1. The framework of proposed method

where n is the number of ROIs, and T is the threshold, thus, we can obtain a new connectivity network $G_m = [\omega_m(i,j)]$, which will be used as feature. The local cluster coefficient is defined as follows:

Definition 1 (Local Cluster Coefficient)

Given a connectivity network G, the local cluster coefficient of node i is defined as [17]:

$$f_i = \frac{2}{n_i(n_i-1)} \sum_{j,k} (\omega_m(i,j)\omega_m(j,k)\omega_m(k,i))^{1/3} \quad (2)$$

where n_i is the number of neighboring nodes around i, and j,k are other nodes in the connectivity network.

It is worth noting that: each brain region corresponds to one cluster coefficient, hence there are n features (n is the number of ROIs).

In our study, we adopted a two-stage feature selection strategy to select the discriminative brain regions as Jie et.al has done in [18]. Specifically, a standard t-test, which has been widely used in the neuroimaging analysis, was first performed to screen out those features that are not significant for discrimination between MCI patients and HC. For instance, given training samples, the p-value of each feature was first computed via t-test, as is known to us, the smaller the p-value is, the more important the feature is. Here, the feature refers to the brain region. Consequently, the features with p-value larger than a given threshold will be excluded. Further more, to obtain a good classification performance, we use recursive feature elimination (RFE) strategy for further selection of discriminative features, i.e., brain regions. Accordingly, we can obtain a kernel with the features selected above.

3) Subgraph features based method: In this section, we regard subgraph as the feature of connectivity network, corresponding we utilize the well-known gSpan algorithm (i.e., graph-based substructure pattern mining) to mine frequent subgraphs from connectivity networks because of its effectiveness. Before giving the details of our method, we start with some preliminaries:

Definition 2 (Labeled Undirected Graph)

Let $G = (V, E, L, l)$ be a labeled undirected graph, where V is a set of nodes and $E \subseteq V \times V$ is a set of edges, $e = u, v$ indicates a edge between the nodes u and v. L is a set of labels ,and l is a mapping function that assigns labels to vertices V

and edges E.

Definition 3 (Subgraph)

For two labeled undirected graphs $G = (V, E, L, l)$ and $G_s = (V_s, E_s, L_s, l_s)$, we say G_s is a subgraph of G if $V_s \subseteq V$,$E_s \subseteq E, L_s \subseteq L$ and $l_s \subseteq l$

Definition 4 (Subgraph Frequency)

Given a set of graph G, the frequency of a subgraph g_s is defined as:

$$f_q(g_s \mid G) = \frac{\mid g_s \; is \; a \; subgraph \; of \; g \; and \; g \in G \mid}{\mid G \mid} \tag{3}$$

Definition 5 (Frequent Subgraph Mining)

Given a set of labeled undirected graphs G,and a support parameter s where $0 \leq s \leq 1$, find all undirected graphs that are subgraphs in at least $s \cdot \mid G \mid$ of the input graphs.

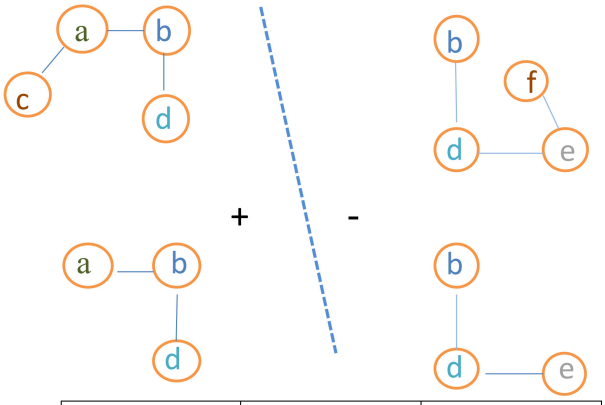
In a variety of frequent subgraph mining algorithm, gSpan is a widely used and effective mining algorithm. gSpan construct a new lexicographic order among graphs, and maps each graph into a unique minimum DFS(depth-first search) code as its canonical label. Based on this lexicographic order, gSpan utilizes the DFS(depth-first search) strategy to traverse the pattern search space. On the basis of a frequent subgraph p, Extended to produce its child $p^{'}$, and compute the support of $p^{'}$, and examine whether the support of the each child of p satisfied the minimum support, If not, all of the subtree would be pruned, otherwise, we continue the search strategy on all the children of $p^{'}$. More details about gSpan algorithm in [19].

After graph mining i.e. subgraph feature extraction of brain connectivity network, what we should do is feature selection. As is known to us, the main problem of feature selection is how to select the set of discriminative features. For graph, it comes to how to mine a set of subgraph patterns in order to effectively perform graph classification. Formally, let us introduce the following notations:

* D: $D = \{D_n, D_p\}$ where D_n denotes the negative samples and D_p denotes the positive samples.
* G: $G = \{G_n, G_p\}, G_p = \{g_{p1}, g_{p2}, ...g_{pm}\}$ refers to the set of all the subgraph features of positive samples mined by gSpan, and $G_n = \{g_{n1}, g_{n2}, ...g_{nk}\}$ refers to the set of all the subgraph features of negative samples, which are used together to predict class membership of new graph instances.
* T^*: the optimal set of subgraph features, $T^* = T_1^* \sqcup T_2^*$ and $T_1^* \subseteq G_p, T_2^* \subseteq G_n$, obviously, $T^* \subseteq G$.
* $J(T)$: The evaluation criterion to estimate the usefulness of subgraph feature subset T.
* $S(g_s)$: We denote the discriminative score of g_s as:

$$S(g_s) = \mid f_q(g_s \mid D_p) - f_q(g_s \mid D_n) \mid \tag{4}$$

the discriminative score of a subgraph pattern g_s is simply defined as the difference between its positive and negative frequency. We can see that the larger the score, the larger the difference of the pattern between two groups. $S(g_s) = 1$ means the subgraph g_s exists in all graphs of the group of healthy subjects and there is no such pattern in any graphs of the group of MCI.



subgraph	positive	negative
ⓐ-ⓑ	√	×
ⓑ-ⓓ	√	√
ⓓ-ⓔ	×	√

Fig. 2. The discrimination of different subgraph, the subgraph feature 'a-b' and 'd-e' is more useful than 'b-d' and other subgraphs to positive or negative classes

Intuitively, Fig 2 has shown the discrimination of different subgraph. We propose the following general optimization framework to select the optimal subgraph feature set:

$$T^* = \underset{T_1 \subseteq G_p, T_2 \subseteq G_n}{argmax} J(T) \qquad s.t \mid T_1 \mid \leq t_1, \mid T_2 \mid \leq t_2 \tag{5}$$

where $\mid \bullet \mid$ denote the size of the feature set and t_1, t_2 are the maximum numbers of the feature selected from each group, so we can obtain that:

$$J(T) = \sum_{i \leq t_1} S(g_{pi}) + \sum_{j \leq t_2} S(g_{nj}) \tag{6}$$

For each subgraph we we can compute its discriminative score through equation. Suppose the score of all subgraphs are denoted as:

$$S(g_p^1) \geq S(g_p^2) ... \geq S(g_p^m), \;\; S(g_n^1 \geq S(g_n^2) ... \geq S(g_n^k) \tag{7}$$

According to equations above, the optimal subgraph feature set is:

$$T^* = \{g_p^i, g_n^j \mid i \leq t_1, j \leq t_2\} \tag{8}$$

According to the features we selected using the method mentioned above, we obtain new graph data, which only containing the select subgraph and make two types of samples more discriminative. Kernel can be defined not only on vector data, but also on complex data types, for example, graph data

of the connectivity network, and the kernel is called graph kernel. Graph kernel maps the graph data from the original graph space to the feature space and further measures the similarity between two graphs by comparing their topological structures. In this study, Weisfeiler-Lehman (WL) subtree kernel, which is proposed in [20], is used to measure the topological similarity between paired connectivity networks. It has been shown that this type of graph kernel can effectively capture the topological information from graphs and achieve better performance than the state-of-the-art graph kernels [20]

WL kernel using an iterative approach to relabel the original label, at each subsequent iteration, the label of each node is simultaneously updated based on its previous label and the labels of its neighbors. That is to say, for each node, we augment its label by the sorted set of node labels of its neighboring nodes, and compress these augmented labels into a new shorter label. This process is iterated until the node label sets of G and $G^{'}$ differ from each other, or the number of iteration reaches a predefined maximum value h.

Given two graphs G and H, let L_0 denote the original label set, and L_1 denote the label set after the first iteration, after h-th iterations, we denote the final label set as:

$$L = \{L_0, L_1, ...L_h\} \tag{9}$$

where $L_i = \{l_{i1}, l_{i2}...l_{i|l_i|}\}$ denote the label set after i-th iteration, $\mid l_i \mid$ denote the number of node after i-th iteration, so the map function can be defined as:

$$\varphi(G) = (\rho_0(G, l_{01})...\rho_0(G, l_{0|L_0|})...\rho_h(G, l_{h1})...\rho_h(G, l_{h|L_h|})) \tag{10}$$

and corresponding

$$\varphi(H) = (\rho_0(H, l_{01})...\rho_0(H, l_{0|L_0|})...\rho_h(H, l_{h1})...\rho_h(H, l_{h|L_h|})) \tag{11}$$

where $\rho_i(G, l_{ij})$ and $\rho_i(H, l_{ij})$ are the numbers of occurrences of the letter l_{ij} in G and H, and then the WL-kernel on two graphs G and H with h iterations is defined as:

$$k(G, H) =< \varphi(G), \varphi(H) > \tag{12}$$

and we can obtain the graph kernel matrix.

4) Multi-kernel support vector machine: Recent studies on multi-kernel learning have shown that the integration of multiple kernels could increases the classification and at the same time it enhances the interpretability of the result [21].

Generally, kernel integration is achieved through a linear combination of multiple kernels:

$$k(x, y) = \sum_{i=1}^{M} \alpha_i k_i(x, y) \qquad s.t \sum_{i=1}^{M} \alpha_i = 1 \tag{13}$$

where $k_i(x, y)$ is a basic kernel built for subjects x and y, M is the number of kernel matrix we need to combine, and α_i is a nonnegative weighting parameter.

In this study, the kernels we employed (vector based kernel and graph kernel) are two different types of kernels, so a normalization step must be performed individually as in equation (14) before combining them

$$k^*(x, y) = k(x, y)/\sqrt{k(x, x)\ k(y, y)} \tag{14}$$

It would be specially mentioned that, different from the existing multi-kernel learning method that jointly optimize the weighting parameters α_i together with other classifier parameters, in our study, the optimal weighting parameters α_i is determined via grid search on the training data. Once the optimal weighting parameter α_i are obtained, the multi-kernel learning based classifier can be naturally embedded into the conventional single-kernel classifier framework, and in this paper we select Support Vector Machine (SVM) as our classifier.

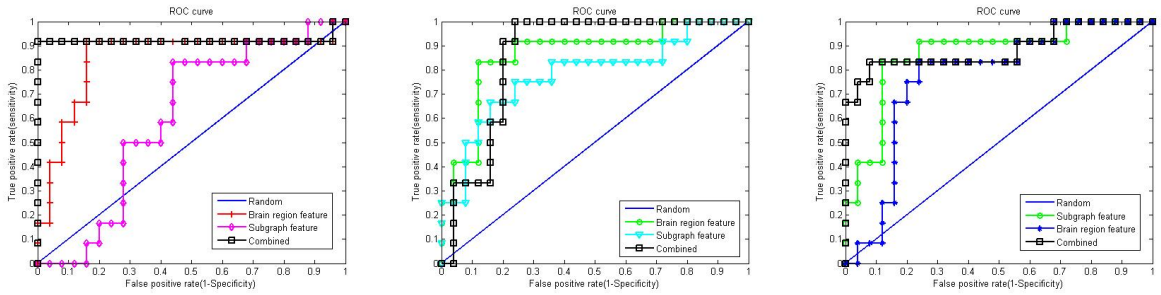
5) Multiple features fusion for classification: In this study, we utilize multi-kernel learning method to perform our classification, in fact, different types of kernels represent different properties of the connectivity network, we fuse multiple features through multi-kernel learning. Specially, vector based kernel describes the correlation between pairwise brain region through the local cluster coefficient, and graph based kernel describes the topological information of the whole network.

III. Experiments and results

To evaluate the performance of our method, Leave-One-Out (LOO) strategy is adopted to enhance the generalization power of the classifier and to avoid the over-fitting on small sample dataset. Specially, for all subjects, one is left out for testing, and the remaining are used for training. This entire process was repeated for each subject thereby yielding an unbiased estimate of classification error rate. The classifier training of standard SVM is implemented using LIBSVM library [22]. Because we use multi-kernel method in our study, in the training step, another LOO cross-validation is adopted through a grid search to obtain the optimal weighting parameter α_i.

A. Classification performance

In our study, we perform our experiments with different thresholds, i.e.,T=[T_1, T_2, T_3]=[0.2,0.3,0.38], which have been demonstrated effective by many researchers [18]. And the classification performance is evaluated based on classification accuracy and the area under receiver operating characteristic (ROC) curve (AUC). Table 2 shows the classification accuracy and AUC(Area Under Curve) of different thresholds. As can be seen from Table 2, our proposed method achieves the best performance by combining the brain region feature and subgraph feature. Specifically, our method yield a classification accuracy of 97.27% with threshold of 0.2, and the AUC is 0.92. for other thresholds, the classification accuracy and the AUC are also improved to varying degrees (except for the AUC of threshold of 0.3 is equal to employing the brain region feature only). Experiment results has shown that brain region features and subgraph features can be complementary for better classification performance. And the ROC curves of different threshold with our proposed method is shown in Fig 3.



(a) (b) (c)

Fig. 3. The ROC curve of proposed method, where the left figure corresponds to the threshold of T_1, the middle figure corresponds to the threshold of T_2, and the right corresponds to the threshold of T_3

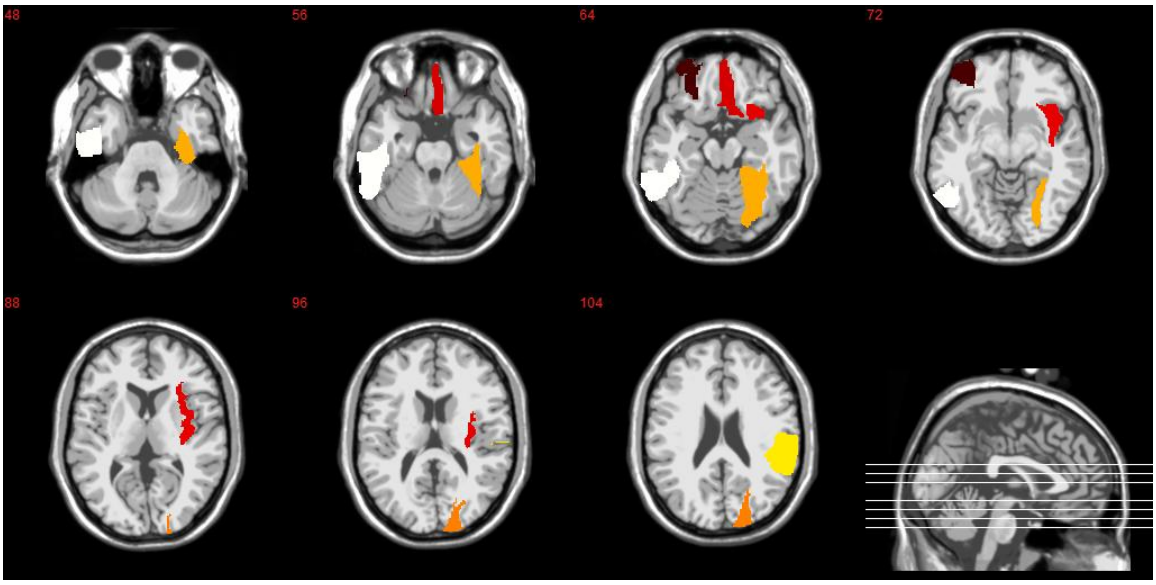


Fig. 4. The top 8 discriminative brain region selected

TABLE II
ACCURACY OF DIFFERENT THRESHOLD AND DIFFERENT TYPES OF FEATURE

	Accuracy(AUC)		
	T_1(%)	T_2(%)	T_3(%)
brain region feature	86.49(0.85)	83.78(0.86)	75.67(0.77)
subgraph feature	64.86(0.60)	72.97(0.65)	81.08(0.84)
combined	**97.27(0.92)**	**91.89(0.86)**	**86.47(0.88)**

B. The most discriminative brain regions

Besides reporting classification of our proposed method, another important issue is to mining the features (both brain region and subgraph) those are sensitive to MCI. Table 3 lists the top 8 brain regions are selected by feature selection method we adopt in our study, which coincide with other researchers. and Fig 4 shows those brain regions for visual inspection.

C. The most discriminative connectivity patterns

As mentioned in the section of the most discriminant brain regions, we also shows the most discriminative connectivity patterns in Fig 5, which suggest possible disruptions in connectivity between these regions and is consistent with previous research [12][23]. Comparing with the discriminative brain regions we find above, we can find that most of the ROIs are selected in both feature types.

TABLE III
THE MOST DISCRIMINATIVE 8 BRAIN REGION FOUND IN OUR STUDY

ID	Cortical Regions	Abbreviation
16	Right orbital part of inferior frontal gyrus	ORBinf.R
35	Left posterior cingulate gyrus	PCG.L
37	Left hippocampus	HIP.L
38	Right hippocampus	HIP.R
47	Left lingual gyrus	LIG.L
84	Right superior temporal pole	TPOsup.R
87	Left middle temporal pole	TPOmid.L
88	Right middle temporal pole	TPOmid.R

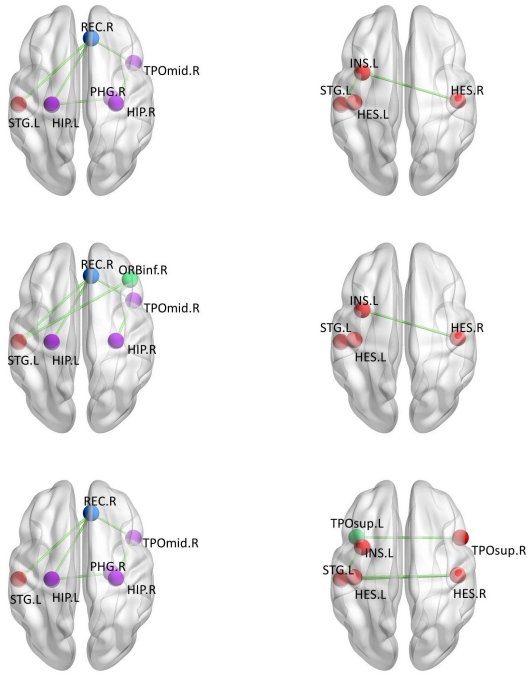


Fig. 5. The most discriminative connectivity pattern, where the left is the discriminative subgraph of MCI, and the right is the discriminative subgraph of HC, and ORBinf.R= Right orbital part of inferior frontal gyrusREC.R= Right gyrus rectusINS.L= Left insulaHIP.L= Left hippocampusHIP.R= Right hippocampus, PHG.R= Right parahippocampal gyrus, STG.L= Left superior temporal gyrus, TPOsup.L= Left superior temporal pole, TPOsup.R= Right superior temporal pole, TPOmid.R= Right middle temporal pole

IV. DISCUSSION

This paper presents a new classification framework that fused multiple features (brain region features and subgraph features) through multi-kernel learning approach. We evaluated our method on a real MCI dataset, the experimental result shows our proposed method can significantly improve the classification performance over the single feature. Many studies have suggested that brains of MCI are different from the HC in both of brain regions and the connectivity between the brain regions [3][12]. In our study, we employ both of the two types of features for automatic brain disease classification. What's more, we also detect the discriminative brain regions and the discriminative connectivity patterns which are sensitive to MCI. However, the current study is limited by the following factors. First, the network construction of the brain network is a very important step, and different network construction may exhibit different properties. In our study, we use the thresholds that many former researchers used, to eliminate the influence of different threshold, we perform our experiment with different threshold. Second, as is shown in the paper, the accuracy of subgraph features didn't achieve a good performance, and in the subgraph selection step, the number of feature(i.e.,t_1,t_2) is artificial specified, which may lose information. Finally, our study is limited by the small number of the MCI dataset, which may reduce its generalization ability on MCI classification.

V. CONCLUSION

In this paper, we proposed a new multiple features fusion based approach to identify whether the individual is MCI or not. In the proposed framework, brain region feature and subgraph feature were employed to reflect multiple properties. Then, two different types of kernels were used to quantify two different yet complementary network properties, and a multi-kernel learning technique was further adopted to fuse these heterogeneous kernels. Thus, multiple features were used to describe the connectivity network. Experiment result has shown that this method not only can significantly improve the performance of classification, but also can potentially detect the brain regions and the discriminative connectivity patterns that are sensitive to disease. In the future work, we will investigate how to select the appropriate number of subgraphs and apply our work on larger MCI dataset.

VI. ACKNOWLEDGEMENT

The authors would like to thank Duke-UNC Brain Imaging and Analysis Center (BIAC), Durham, North Carolina, USA and Dr. Dinggang Shen for providing the data we used in our paper. And this paper is supported by National Natural Science Foundation of China under grant numbers 61422204 and 61473149, by the Jiangsu Natural Science Foundation for Distinguished Young Scholar under grant number BK20130034, and by the NUAA Fundamental Research Funds under grant number NE2013105.

REFERENCES

[1] O. Sporns, "From simple graphs to the connectome: networks in neuroimaging," *Neuroimage*, vol. 62, no. 2, pp. 881–886, 2012.

[2] T. X. andYong He, "Mapping the alzheimers brain with connectomics," *Magnetic resonance imaging of disturbed brain connectivity in psychiatric illness*, p. 21, 2011.

[3] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in alzheimer's disease," *PLoS computational biology*, vol. 4, no. 6, p. e1000100, 2008.

[4] C. Stam, B. Jones, G. Nolte, M. Breakspear, and P. Scheltens, "Small-world networks and functional connectivity in alzheimer's disease," *Cerebral Cortex*, vol. 17, no. 1, pp. 92–99, 2007.

[5] C.-Y. Wee, P.-T. Yap, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Resting-state multi-spectrum functional connectivity networks for identification of mci patients," *PloS one*, vol. 7, no. 5, p. e37828, 2012.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-worldnetworks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[7] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz, "Generalizations of the clustering coefficient to weighted complex networks," *Physical Review E*, vol. 75, no. 2, p. 027105, 2007.

[8] M. Barthelemy, "Betweenness centrality in large complex networks," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 163–168, 2004.

[9] X. Kong, A. Ragin, X. Wang, P. Yu, X. Kong, S. Y. Philip, X. Wang, and A. B. Ragin, "Discriminative feature selection for uncertain graph classification," *Proceedings of Computing Research Repository (CoRR)*, 2013.

[10] X. Kong and P. S. Yu, "Semi-supervised feature selection for graph classification," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 793–802.

[11] B. Jie, D. Zhang, W. Gao, Q. Wang, C. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Transactions on Biomedical Engineering (TBME)*, 2013.

[12] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao, J. Jia, Y. Han, and Y. He, "Disrupted functional brain connectome in individuals at risk for alzheimer's disease," *Biological psychiatry*, vol. 73, no. 5, pp. 472–481, 2013.

[13] C. Hinrichs, V. Singh, G. Xu, and S. C. Johnson, "Predictive markers for ad in a multi-modality framework: an analysis of mci progression in the adni population," *Neuroimage*, vol. 55, no. 2, pp. 574–589, 2011.

[14] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[15] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[16] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[17] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, "Intensity and coherence of motifs in weighted complex networks," *Physical Review E*, vol. 71, no. 6, p. 065103, 2005.

[18] B. Jie, D. Zhang, C.-Y. Wee, and D. Shen, "Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification," *Human brain mapping*, 2013.

[19] X. Yan and J. Han, "gspan: Graph-based substructure pattern mining," in *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*. IEEE, 2002, pp. 721–724.

[20] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, "Weisfeiler-lehman graph kernels," *The Journal of Machine Learning Research*, vol. 12, pp. 2539–2561, 2011.

[21] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *The Journal of Machine Learning Research*, vol. 5, pp. 27–72, 2004.

[22] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[23] J. Shao, Q. Yang, A. Wohlschlaeger, and C. Sorg, "Discovering aberrant patterns of human connectome in alzheimer's disease via subgraph mining," in *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 86–93.