Classification of Alzheimer's Disease, Mild Cognitive Impairment and Normal Control Subjects Using Resting-State fMRI Based Network Connectivity Analysis

ZHE WANG[1], YU ZHENG[1], DAVID C. ZHU[2], ANDREA C. BOZOKI[3], AND TONGTONG LI[1]

[1]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48823, USA
[2]Departments of Radiology and Psychology, Michigan State University, East Lansing, MI 48823, USA
[3]Department of Neurology and Ophthalmology, Michigan State University, East Lansing, MI 48823, USA

CORRESPONDING AUTHOR: T. LI (tongli@egr.msu.edu)

ABSTRACT This paper proposes a robust method for the Alzheimer's disease (AD), mild cognitive impairment (MCI), and normal control subject classification under size limited fMRI data samples by exploiting the brain network connectivity pattern analysis. First, we select the regions of interest (ROIs) within the default mode network and calculate the correlation coefficients between all possible ROI pairs to form a feature vector for each subject. Second, we propose a regularized linear discriminant analysis (LDA) approach to reduce the noise effect due to the limited sample size. The feature vectors are then projected onto a one-dimensional axis using the proposed regularized LDA. Finally, an AdaBoost classifier is applied to carry out the classification task. The numerical analysis demonstrates that the purposed approach can increase the classification accuracy significantly. Our analysis confirms the previous findings that the hippocampus and the isthmus of the cingulate cortex are closely involved in the development of AD and MCI.

INDEX TERMS fMRI, Alzheimer's disease, brain connectivity analysis.

I. INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia, and causes problems with memory, thinking and behavior. It is a degenerative brain disorder, characterized by progressive deterioration of nerve cells, eventually leading to cell death. Mild Cognitive Impairment (MCI) is a condition in which people show a slight, but noticeable and measurable decline in cognitive capabilities, beyond what is considered normal for their age. MCI is a transitional stage of dementia between NC and AD [1]. Older people with MCI may or may not progress to AD, though they have a higher risk of doing so. Accurate distinction of AD and MCI from normal control (NC) subjects is critical for early diagnosis and treatment of brain disorders.

Traditional AD and MCI diagnosis methods are generally based on positron emission tomography (PET) and cerebrospinal fluid (CSF) [2]. In recent years, there has also been an increasing interest in noninvasive diagnosis methods based on electroencephalography (EEG) [3], structural magnetic resonance imaging (sMRI) [4], and functional magnetic resonance imaging (fMRI) [5], [6].

In literature, the majority of existing noninvasive classification approaches rely on MRI and EEG [7]–[10]. Pritchard *et al.* [7] analyzed spectral-band measures of EEG data acquired from AD patients and NC subjects. They found that classifiers based on multivariate discriminant analysis [8] and the nearest neighbor approach could typically achieve a two-category (AD and NC) classification accuracy of 80% when applied to EEG data. Magnin *et al.* [9] applied the support vector machine (SVM) classifier to the whole-brain anatomical MRI data acquired from AD patients and NC subjects. They formulated feature vectors for classification using gray matter information extracted from T_1-weighted MR images of AD and NC subjects, and achieved a two-category classification accuracy of 94.5%. Korolev *et al.* [10] combined data from clinical biomarkers, MRI signals, and plasma biomarkers and developed a classifier to predict whether an MCI patient would

develop Alzheimer's disease over a three-year period. Their prediction accuracy was 80%.

More recently, fMRI, which maps brain activities to metabolic changes (such as the blood-oxygen-level dependent (BOLD) contrast) in cerebral blood flow, has also been used to classify AD, MCI and NC subjects [5], [6]. The underlying mechanism of fMRI is that cerebral blood flow and neuronal activation are coupled. That is, when a particular brain region becomes active, blood flow to that region also increases.

Compared with EEG, fMRI data can display active brain areas more directly, and has much better spatial resolution throughout the brain. Unlike structural MRI which mainly reflects the anatomical information of brain tissues and structure, fMRI focuses on functional brain activities, and can provide more direct measurement on how different brain regions are involved in particular brain activities, hence provides more insight on the changes of functional brain connectivity during the evolution of MCI and AD.

Wang *et al.* [5] extracted two intrinsically anti-correlated networks using resting state fMRI data from 14 AD patients and 14 NC subjects, and applied a Pseudo-Fisher Linear Discriminative Analysis (pFLDA) on the high dimensional feature vectors. Their *two-category* classification accuracy was 83%. Chen *et al.* [6] applied the same technique to larger datasets. Similarly, the accuracy of *two-category* classification of AD patients and NC subjects was 82%. In addition, Challis *et al.* developed a logistic regression (LR) based AD and MCI classification method. Their *two-category* classification accuracies for MCI/NC and AD/MCI were 75% and 97%, respectively [11]. Using a set of graph theory based network connectivity measurements as the feature vectors, Jie *et al.* [12] successfully distinguished MCI from NC with an accuracy of 92%. Similar techniques were also used by Khazaee *et al.* [13] to classify AD patients and NC subjects. More recently, classification methods that combine multi-modality data have been proposed for AD, MCI and NC classification [14]–[19]. In all these methods, two or more measurements, including functional MRI (fMRI), structural magnetic resonance imaging (sMRI), fluorodeoxyglucose (FDG), positron emission tomography (PET) and florbetapir-PET, are utilized. A comprehensive comparison of existing classification methods on AD, MCI and NC subjects can be found in [20].

While structural MRI has been widely applied to clinical diagnosis of brain disorders, fMRI has mainly been used for research purposes. As a result, the size of fMRI data samples is generally quite limited, which has become a major bottleneck in fMRI based AD, MCI and NC classification. This is because that, when the sample size is small, most existing classifiers suffer from severe noise effects, due to both biological variability and measurement noise [21].

Motivated by this observation, in this paper, we develop a reliable method for AD, MCI and NC classification that is robust with respect to *size limited* fMRI data samples, by exploiting *brain network connectivity pattern analysis*.

The underlying argument is that: due to variability in the brain connectivity of each individual, the connectivity between two brain regions alone may not be sufficient to distinguish NC subjects from patients with cognitive impairments; brain network connectivity pattern analysis, which looks for subtle changes in the pattern of connectivity among multiple or all regions in the sub-network, has shown to be able to provide more accurate information in neuroimage classification [5], [6], [11]–[13].

The proposed classification scheme can be described as follows. *First*, we select an ROI sub-network and formulate the feature vectors by calculating the Pearson correlation coefficients between all pairs of ROIs. In this paper, we formulate the ROI sub-network by selecting regions within the default mode network (DMN), which denotes the network of brain regions that are active when the brain is at the resting state [22]. Previous resting-state fMRI studies have demonstrated that the DMN is affected by AD [23]–[26]. More specifically, in this paper, we select the right and left hippocampi and isthmus of the cingulate cortices (ICCs) (4 regions) as our ROI sub-network. This is because that both hippocampus and ICC are part of the DMN, and can be well defined anatomically through the FreeSurfer software [26], even in brains with abnormal anatomy [25]. It has also been demonstrated in [25] that the functional connection between hippocampus and ICC was reduced in AD. *Second*, we propose a regularized linear discriminant analysis (LDA) approach, where we take shrinkage based regularization procedures to reduce the noise effect (including both biological variability and measurement errors) due to limited sample size. The feature vectors are then projected onto a one-dimensional axis using the proposed regularized LDA, where the differences between AD, MCI and NC subjects are maximized. Based on the Central Limit Theorem, we show that when used for fMRI based brain functioning classification, LDA is equivalent to the optimal maximum likelihood based classification method. *Finally,* a decision tree based *multi-class* AdaBoost classifier, which is robust to noise effect, is applied to the projected one-dimensional vectors to carry out the classification task.

The major results of this paper can be summarized as:

1) *We propose a regularized LDA approach, which aims to reduce the noise effect by using two shrinkage methods*. The first shrinkage method moves the estimated mean of each class towards the overall mean, and the second one shifts the estimated covariance matrix for each class towards the identity matrix. Numerical analysis shows that: in comparison with the original LDA approach [27], the regularized LDA can reduce the noise effect and increase the classification accuracy significantly.

2) *We investigate the relationship between LDA-based and Maximum Likelihood (ML) based classification or decision making methods*. Recall that LDA aims to separate two or more classes by projecting them into a sub-space or direction where different classes show most

significant differences [27]. In this paper, we prove that when the original data are normally distributed, LDA is equivalent to maximizing the log-likelihood function of the projected data. Note that there are millions of neurons within one fMRI voxel, according to the Central Limit Theorem, the overall fMRI signal corresponding to each voxel follows the normal distribution approximately. This implies that when used for fMRI based brain functioning classification, LDA is equivalent to the optimal ML based classification method.

3) *We conduct the connectivity pattern classification of AD, MCI and NC subjects by applying the regularized LDA and AdaBoost classifier based approach.* First, we calculate the Pearson correlation coefficients between all possible pairs of the ROIs within the group to formulate the feature vectors. Second, the feature vectors are then projected onto a one-dimensional axis using the proposed regularized LDA, where the differences between AD, MCI and NC subjects are maximized. Finally, we construct the decision tree based on the projected feature vectors and carry out the classification using the multi-class AdaBoost classifier.

In this paper, we choose to utilize the AdaBoost classifier instead of the naive Bayesian classifier, since it has been observed consistently in literature that: the AdaBoost classifier could achieve significantly higher classification accuracy than the naive Bayesian classifier when the sample size is very limited [28]. Our numerical results demonstrate that: (i) LDA-Bayesian classifier can achieve a three-category (AD, MCI and NC) classification accuracy of 42%; (ii) LDA-AdaBoost classifier can increase the accuracy to 69%; (iii) when AdaBoost is combined with the regularized LDA, the accuracy can be further increased to 75%.

As expected, it is also observed that compared with AD and NC subjects, it is more difficult for the classifier to identify MCI subjects. The classification accuracy for AD and NC are as high as 80% and 83%, respectively, while the accuracy for MCI is only 63%. Our analysis also confirms the previous findings that the hippocampus and the isthmus of the cingulate cortex are closely involved in the development of AD and MCI [25].

The rest of this paper is organized as follows. In Section II, we present the proposed regularized LDA approach, and explore the relationship between LDA based and the Maximum Likelihood based classification methods. In Section III, we describe the ROI sub-network formulation, and elaborate how to perform AD, MCI and NC classification through connectivity pattern analysis. In Section IV we present the numerical results, and we conclude in Section V.

II. REGULARIZED LINEAR DISCRIMINANT ANALYSIS

In this section, first, we revisit the Linear Discriminant Analysis method. Second, we integrate two shrinkage methods with the original LDA to formulate the regularized LDA.

Finally, we investigate the relationship between LDA and the ML estimation method.

A. LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis aims to separate two or more classes by projecting them into a subspace or direction where different classes show most significant differences [27]. LDA is a general framework that can be applied to classification problems with two or more classes. Here, we are considering the classification of three different groups: AD, MCI and normal subjects. Without loss of generality, we illustrate the basic idea of LDA using the three-class case. Suppose we have a set of $d-$dimensional vector samples $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, where n_1 of them are from the first class, denoted as C_1, and n_2 of them are from the second class, denoted as C_2, and the remaining $n_3 = n - n_1 - n_2$ of them are from the third class, denoted as C_3. For $i = 1, 2, 3$, the mean and scatter matrix (i.e., the scaled covariance matrix) of each of the three classes are defined as:

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in C_i} \mathbf{x}, \tag{1}$$

$$S_i = \sum_{\mathbf{x} \in C_i} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^t. \tag{2}$$

Consider the projection of vectors in X to a new $d-$dimensional space:

$$\mathbf{y} = W\mathbf{x}, \quad \mathbf{x} \in X, \tag{3}$$

where W is a $d \times d$ matrix to be determined by the LDA algorithm. In this paper, we only utilize the first dimension y of projected vector $\mathbf{y}$ where the differences among three classes are maximized. As a result, Equation (3) can be rewritten as:

$$y = \mathbf{w}^t \mathbf{x}, \tag{4}$$

where $\mathbf{w}^t$ is the first row of the matrix W. For $i = 1, 2, 3$, let

$$\tilde{C}_i = \{y = \mathbf{w}^t \mathbf{x} \mid \mathbf{x} \in C_i\}. \tag{5}$$

Define $\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ as the overall mean, $S_W = \sum_{i=1}^{3} S_i$ as the within-class scatter matrix, and the between-class scatter matrix S_B as:

$$S_B = \sum_{i=1}^{3} n_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^t. \tag{6}$$

LDA seeks a transform vector $\mathbf{w}$ that maximizes the following objective function:

$$J(\mathbf{w}) = \frac{\mathbf{w}^t S_B \mathbf{w}}{\mathbf{w}^t S_W \mathbf{w}}. \tag{7}$$

It can be proved [8] that Equation (7) can be maximized when $\mathbf{w}$ is the eigenvector corresponding to the largest eigenvalue of the matrix $S_W^{-1} S_B$ as $\mathbf{w}$. As will be shown in Section 3, various classifiers, such as the Bayesian classifier and the AdaBoost classifier can then be applied to the projected vectors $\{y_i = \mathbf{w}^t \mathbf{x}_i\}_{i=1}^{n}$ for further classification.

B. REGULARIZED LDA

The original LDA algorithm has been widely applied in supervised learning problems [8]. However, as mentioned earlier, when the total number of subjects is small, the estimated statistics suffer considerably from the noise effect caused by both biological variability and measurement error, leading to low classification accuracy. For our fMRI based AD, MCI and NC classification, due to the very limited sample size, LDA together with the Bayesian classifier could only achieve an accuracy that is under 50%. To reduce the noise effect, we propose to regulate the original LDA using two shrinkage methods.

1) SHRINKAGE OF THE MEAN

The first shrinkage method, originally proposed by Tibshirani *et al.* [29] for gene expression profiling, adjusts the estimated mean vectors. In our case, let $C = \bigcup\limits_{i=1}^{3} C_i$ be the whole sample set. Recall that for any $x \in C$, $x = [x(1), \cdots, x(k), \cdots, x(d)]^t$. For $i = 1, 2, 3$, $k = 1, \cdots, d$, define $\mu_{i,k} \triangleq \frac{1}{n_i} \sum\limits_{x \in C_i} x(k)$, and $\mu_k \triangleq \frac{1}{n} \sum\limits_{x \in C} x(k)$, where $n_i = |C_i|$ and $n = n_1 + n_2 + n_3$. Let $\boldsymbol{\mu}_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,d}]^t$, and $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_d]^t$. The algorithm first calculates the following variances:

$$\xi_k^2 = \frac{1}{n-3} \sum_{i=1}^{3} \sum_{\mathbf{x} \in C_i} [x(k) - \mu_{i,k}]^2, \quad k = 1, \cdots, d. \quad (8)$$

Then for $i = 1, 2, 3$, and $k = 1, \cdots, d$, the scaled distance of $\mu_{i,k}$ to the centroid μ_k is calculated as:

$$d_{i,k} = \frac{\mu_{i,k} - \mu_k}{m_i \xi_k}, \quad (9)$$

where $m_i = \sqrt{1/n_i + 1/n}$. After that, the distance is shrunken as follows:

$$d_{i,k} \leftarrow sign(d_{i,k}) max(0, |d_{i,k}| - \delta), \quad (10)$$

where δ is a positive step size determined through cross-validation [8]. Now based on Equation (9), the shrinkage is achieved as follows:

$$\mu_{i,k} \leftarrow \mu_k + m_i \xi_k d_{i,k}. \quad (11)$$

As can be seen from (10) and (11), each dimension of $\boldsymbol{\mu}_i$ has been shrunken towards the overall mean. This shrinkage method is essentially based on the t test between the mean of each class and the overall mean at every dimension. Recall that the t score of two sets of random variables $\{z_1\}$ and $\{z_2\}$, which have the same standard deviation S_z, is defined as:

$$t = \frac{\mu_{z_1} - \mu_{z_2}}{s_z \sqrt{\frac{1}{n_{z_1}} + \frac{1}{n_{z_2}}}}, \quad (12)$$

where $\mu_{z_1} = E\{z_1\}$, $\mu_{z_2} = E\{z_2\}$, and n_{Z_1} and n_{Z_2} are the sample sizes [30]. In this shrinkage method, for $i = 1, 2, 3$, $k = 1, \cdots, d$, the t score between each $\mu_{i,k}$ and μ_k pair is defined as a distance in (9). If the distance $d_{i,k}$ is small,

i.e., if $d_{i,k} < \delta$, then most likely it is caused by the noise effect. In this case, the shrinkage method forces $d_{i,k}$ to be zero, and therefore reduces the noise effect.

2) SHRINKAGE OF THE COVARIANCE MATRIX

The second shrinkage method, proposed by Friedman *et al.* [31], regulates the estimation of covariance matrix S_i for each class as:

$$S_i \leftarrow (1 - \epsilon) S_i + \epsilon I, \quad i = 1, 2, 3, \quad (13)$$

where I is the identity matrix, and ϵ a positive number determined through cross-validation. The basic idea of this shrinkage method is that: when the sample size is small, the estimated covariance matrix S_i, $i = 1, 2, 3$, generally becomes non-invertible. By adding a small perturbation to the slightly scaled covariance matrix, the adjusted or shrunken S_i will generally become invertible as expected.

After the regularized LDA transform, the feature vectors are projected into a set of real valued numbers. After that, a selected classifier can be applied to the transformed data for further classification.

C. LDA AND THE ML ESTIMATION

In this subsection, we demonstrate that when the original data from all classes are normally distributed, then LDA is equivalent to the ML method. For $i = 1, 2, 3$, assume each vector $\mathbf{x}$ in class C_i has the same probability density function (pdf):

$$f_X(\mathbf{x}; \boldsymbol{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{2\pi^d |\Sigma_i|}} e^{\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \Sigma_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)\}}. \quad (14)$$

Consider a general linear transform defined by:

$$\mathbf{y} = W\mathbf{x}. \quad (15)$$

where W is a $d \times d$ matrix. For the transformed data, the probability density function becomes:

$$f_Y(\mathbf{y}; \tilde{\boldsymbol{\mu}}_i, \tilde{\Sigma}_i) = \frac{1}{\sqrt{2\pi^d |\tilde{\Sigma}_i|}} e^{\{-\frac{1}{2}(\mathbf{y}-\tilde{\boldsymbol{\mu}}_i)^t \tilde{\Sigma}_i^{-1} (\mathbf{y}-\tilde{\boldsymbol{\mu}}_i)\}}, \quad (16)$$

where $i = 1, 2, 3$, $\tilde{\boldsymbol{\mu}}_i = W\boldsymbol{\mu}_i$, and $\tilde{\Sigma}_i = W\Sigma_i W^t$.

Recall that in LDA, we try to find W such that the difference among different classes is maximized in the transformed space. Without loss of generality, we assume that the major difference lies in the first dimension of the transformed vector $\mathbf{y}$ only, and the remaining $d - 1$ dimensions make little contributions. Under this assumption, $\tilde{\boldsymbol{\mu}}_i$ and $\tilde{\Sigma}_i$ can be decomposed as:

$$\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix} \tilde{\mu}_i^1 \\ \tilde{\boldsymbol{\mu}}^{d-1} \end{bmatrix}, \quad \tilde{\Sigma}_i = \begin{bmatrix} \tilde{\Sigma}_i^1 & 0 \\ 0 & \tilde{\Sigma}^{d-1} \end{bmatrix}, \quad (17)$$

since for each i, $\tilde{\mu}_i^{d-1} \approx \tilde{\boldsymbol{\mu}}^{d-1}$, $\tilde{\Sigma}_i^{d-1} \approx \tilde{\Sigma}^{d-1}$. Accordingly, the matrix W can also be decomposed as

$$W = \begin{bmatrix} W^1 \\ W^{d-1} \end{bmatrix}. \quad (18)$$

In this case, we have $\tilde{\mu}_i^1 = W^1\mu_i$, $\tilde{\mu}_i^{d-1} = W^{d-1}\tilde{\mu}_i$, $\tilde{\Sigma}_i^1 = W^1\Sigma_i W^{1^t}$ and $\tilde{\Sigma}_i^{d-1} = W^{d-1}\Sigma_i W^{d-1^t}$.

For fairness, in LDA based classification, the sample size of the three classes is assumed to be the same, i.e., $n_1 = n_2 = n_3 = n/3$. With the probability density function given in (16), the log-likelihood function of the original data can be written as:

$$
\begin{aligned}
L(W) &= \sum_{i=1}^{3} \sum_{\mathbf{y}\in\tilde{C}_i} log\,|W|f_Y(\mathbf{y};\tilde{\boldsymbol{\mu}}_i,\tilde{\Sigma}_i) \\
&= nlog|W| - \frac{n}{2}log(2\pi)^d - \sum_{i=1}^{3}\frac{n_i}{2}log|\tilde{\Sigma}_i^1| \\
&\quad - \frac{1}{2}\sum_{i=1}^{3}\sum_{\mathbf{x}\in C_i}(W^1\mathbf{x}-\tilde{\mu}_i^1)^t(\tilde{\Sigma}_i^1)^{-1}(W^1\mathbf{x}-\tilde{\mu}_i^1) \\
&\quad - \frac{n}{2}log|\tilde{\Sigma}^{d-1}| \\
&\quad - \frac{1}{2}\sum_{x\in C}[(W^{d-1}\mathbf{x}-\tilde{\boldsymbol{\mu}}^{d-1})^t(\tilde{\Sigma}^{d-1})^{-1} \\
&\qquad\qquad \times (W^{d-1}\mathbf{x}-\tilde{\boldsymbol{\mu}}^{d-1})].
\end{aligned} \tag{19}
$$

To find the optimal W that maximizes $L(W)$, set $\frac{\partial L(W)}{\partial \tilde{\Sigma}_i^1}=0$ and $\frac{\partial L(W)}{\partial \tilde{\Sigma}^{d-1}}=0$, we get:

$$\tilde{\Sigma}_i^1 = W^1 S_W W^{1^t}, \tag{20}$$

$$\tilde{\Sigma}^{d-1} = W^{d-1} S_B W^{d-1^t}. \tag{21}$$

Substitute (20) and (21) into (19) and remove the constant items, the optimization of $L(W)$ is equivalent to optimizing the following function:

$$
\begin{aligned}
L_{eq}(W) &= nlog|W| - \frac{n}{2}log|W^1 S_W W^{1^t}| \\
&\quad - \frac{n}{2}log|W^{d-1} S_B W^{d-1^t}|.
\end{aligned} \tag{22}
$$

The optimal choice of W will satisfy the differential equation:

$$\frac{dL_{eq}(W)}{dW} = 0. \tag{23}$$

It was shown in [8] that the partial differential equations are satisfied when W is composed of eigenvectors of the matrix $S_W^{-1}S_B$.

If we only keep the eigenvector corresponding to the largest eigenvalue of $S_W^{-1}S_B$, then we obtain the LDA algorithm presented in Section II-A. As can be seen, LDA is equivalent to the ML method.

III. CLASSIFICATION OF AD, MCI AND NC SUBJECTS BASED ON CONNECTIVITY PATTERN ANALYSIS

In this section, we formulate the ROI sub-network, and perform AD, MCI and NC Subjects classification through connectivity pattern analysis, by exploiting the proposed Regularized LDA.

A. ROI SUB-NETWORK FORMULATION AND CONNECTIVITY PATTERN ANALYSIS

The default mode network (DMN) is one of the well studied networks at the resting state [22]. Previous resting-state fMRI studies have demonstrated that the DMN is affected by AD [23]–[26]. Both hippocampus and ICC are part of the DMN, and can be well defined anatomically through the FreeSurfer software [26], even in brains with abnormal anatomy [25]. The paper by Zhu *et al.* [25] specifically demonstrated that the functional connection between hippocampus and ICC was reduced in AD.

Motivated by the observations above, in this paper, we select the right and left hippocampi and ICCs (4 regions) as our ROI sub-network. Our connectivity pattern analysis is carried out following the procedure below.

First, we calculate the Pearson correlation coefficients between all possible pairs of the ROIs within the group to formulate the feature vectors. As we now have 4 regions in the ROI sub-network, for each subject j, we can obtain a $d-$dimensional ($d = 6$) vector $\mathbf{x}_j$, consisting of the Pearson correlation coefficients for each pair of ROIs. When we have n subjects, we get the feature vector set $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$.

Second, using the proposed regularized LDA, we map $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ to a one-dimensional subspace or axis, where the differences between AD, MCI and NC subjects are maximized, and denote the projected vectors as $\{y_1,\cdots,y_n\}$.

Finally, we construct the decision trees based on $\{y_1,\cdots,y_n\}$ and carry out the classification using the multi-class AdaBoost classifier.

In the following subsections, we will provide more details on decision tree construction, and multi-class classification using AdaBoost.

B. BASIC DECISION TREE CONSTRUCTION

In the classification procedure, we will construct $T = 50$ basic decision trees. Each decision tree divides the LDA projected data set $y_j, j = 1, \cdots, n$, into K regions, and each region is called a leaf node. Here, the number of regions, K, and the boundaries for all the regions are chosen by the decision tree algorithm to minimize the Gini impurity coefficient I_G [32]. More specifically, assume $y_j \in c_j$, where $c_j \in \{\tilde{C}_1, \tilde{C}_2, \tilde{C}_3\}$. Here $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3$ denote the projected data set corresponding to AD, MCI and NC subjects, respectively. For $k = 1, 2, \ldots, K$, without loss of generality, suppose m_k data samples $\{(y_{k,1}, c_{k,1}), (y_{k,2}, c_{k,2}), \ldots, (y_{k,l}, c_{k,l}), \ldots, (y_{k,m_k}, c_{k,m_k})\}$ are assigned to node k, where $y_{k,l} \in \{y_j\}, c_{k,l} \in \{\tilde{C}_1, \tilde{C}_2, \tilde{C}_3\}, l = 1, \cdots, m_k$. The Gini impurity coefficient of node k is calculated as:

$$I_G(k) = \sum_{i=1}^{3} f_{k,i}(1 - f_{k,i}), \tag{24}$$

where $f_{k,i} = \frac{number\ of\ occurrence\ of\ \tilde{C}_i\ within\ node\ k}{m_k}$ and $i = 1, 2, 3$.

For any given input y to be classified, if y falls within the boundaries of node k, then it will be assigned to node k,

and paired with the majority class inside this node. Note that in our case, $y_j, j = 1, \cdots, n$, are all real-valued numbers. That is, $\{y_j\} \in \mathbf{R}$. In this case, the boundary between two neighboring regions is reduced to a point, and hence each region corresponds to an interval on the R axis.

The decision tree is a weak classifier. In most applications, it needs to be incorporated with an ensemble classifier to achieve higher accuracy. Some representative ensemble algorithms include Bagging and Boosting [32]. In the following, we will apply the AdaBoost algorithm to construct the ensemble classifier, due to its robustness under noise effect [8], [28].

C. THE MULTI-CLASS ADABOOST CLASSIFIER
The multi-class AdaBoost classifier is built upon an ensemble of weak decision tree classifiers [28]. Given a set of labeled data $\{(y_1, c_1), (y_2, c_2), \ldots, (y_n, c_n)\}$, where $c_i \in \{\tilde{C}_1, \tilde{C}_2, \tilde{C}_3\}$, the algorithm first starts with an empty ensemble and T decision trees, as constructed above. Each sample y_j in the data set is given an initial weight $w_j = 1/n$, where $j = 1, 2, \cdots, n$. Then for $t = 1, 2, \cdots, T$, the algorithm will iteratively implement the following procedures:

1) WEIGHTED CLASSIFICATION ERROR CALCULATION
Apply a weak decision tree classifier t to the samples and calculate the weighted classification error. More specifically, let

$$\mathbf{I}(c_j, \tilde{c}_j) = \begin{cases} 1 & if \ c_j \neq \tilde{c}_j, \\ 0 & if \ c_j = \tilde{c}_j, \end{cases} \tag{25}$$

where $\tilde{c}_j$ is the assigned class for sample y_j, and c_j is the true class y_j belongs to. Then the weighted classification error e_t would be

$$e_t = \sum_{j=1}^{n} w_j \mathbf{I}(c_j, \tilde{c}_j). \tag{26}$$

2) VOTING WEIGHT ASSIGNMENT
Based on the weighted classification error e_t, the algorithm will assign a voting weight α_t for the weak decision tree classifier t as follows:

$$\alpha_t = ln \frac{1 - e_t}{e_t} + ln2, \tag{27}$$

and then add classifier t into the ensemble.

3) WEIGHT UPDATE
Before the next iteration, the weight of each data sample y_j is updated as follows:

$$w_j \leftarrow w_j e^{\alpha_t \mathbf{I}(c_j, \tilde{c}_j)}, \tag{28}$$

$$w_j \leftarrow \frac{w_j}{\sum_{j=1}^{n} w_j}. \tag{29}$$

The procedure in (29) ensures that the weights $\{w_j\}, j = 1, 2, \cdots, n$, form a probability distribution with $\sum_{j=1}^{n} w_j = 1$.

As can be seen, after the update, those samples which have been incorrectly classified in current iteration will have higher weights in the next iteration.

4) FINAL CLASSIFICATION
After T iterations, there will be T decision trees in the ensemble. The final classification is a weighted majority votes of each of those classifiers.

As will be shown in the next Section, the combination of regularized LDA and AdaBoost can achieve much higher accuracy in AD, MCI and NC classification than the conventional approach based on the original LDA and the Bayesian classifier.

IV. NUMERICAL ANALYSIS
A. FMRI DATA ACQUISITION AND PRE-PROCESSING
1) SUBJECTS
Ten patients with mild-to-moderate probable AD, 11 MCI patients, and 12 age- and education- matched healthy NC subjects were recruited to participate in this study. The study was approved by the Michigan State University Institutional Review Board. All subjects or their legal representatives provided written informed consent. All subjects were carefully screened to exclude those with a history of stroke, brain tumors, aneurysms, brain surgery, serious head injury, or other significant neurological disease, as well as those with uncontrolled diabetes, hypertension, and hypothyroidism. All subjects were also screened for MR-incompatible metallic implants. NC subjects were community-dwelling older adults recruited from the Greater Lansing area in Michigan. The AD and MCI patients were recruited through the Memory Disorders Clinic in the Department of Neurology at Michigan State University and were diagnosed using standard criteria by a practicing neurologist (NINCDS-ADRDA criteria for clinically probable AD and Petersen criteria for MCI).

2) fMRI DATA ACQUISITION
The MRI experiment was conducted on a GE 3T Signa® HDx MR scanner (GE Healthcare, Waukesha, WI) with an 8-channel head coil. During each session, first and higher-order shimming procedures were carried out to improve magnetic field homogeneity. To study resting-state brain function, echo-planar images, starting from the most inferior regions of the brain, were acquired for 7 minutes with the follow- ing parameters: 38 contiguous 3-mm axial slices in an interleaved order, time of echo (TE) = 27.7ms, time of repetition (TR) = 2500 ms, flip angle = 80°, field of view (FOV) = 22 cm, matrix size = 64 × 64, ramp sampling, and with the first four data points discarded. Each volume of slices was acquired 164 times while a subject was asked to relax and keep his/her eyes open in a dim-light condition. After the functional data acquisition, diffusion-weighted images were acquired with a dual spin-echo echo-planar imaging (EPI) sequence for 12 minutes and 6 seconds with the following parameters: 48 contiguous 2.4-mm axial slices

in an interleaved order, FOV $= 22$ cm$\times 22$ cm, matrix size $= 128\times 128$, number of excitations (NEX) $= 2$, TE $= 77.5$ ms, TR $= 13.7$ s, 25 diffusion-weighted volumes (one per gradient direction) with b $= 1000$ s/mm^2, one volume with b $= 0$ and parallel imaging acceleration factor $= 2$. Finally, 180 T$_1$-weighted 1-mm^3 isotropic volumetric inversion recovery fast spoiled gradient-recalled images (10 minute scan time), with cerebrospinal fluid (CSF) suppressed, were obtained to cover the whole brain with the following parameters: TE $= 3.8$ms, TR of acquisition $= 8.6$ms, time of inversion (TI) $= 831$ms, TR of inversion $= 2332$ ms, flip angle $= 8°$, FOV $= 25.6$ cm$\times 25.6$ cm, matrix size $= 256\times 256$, slice thickness $= 1$ mm, and receiver bandwidth $= \pm 20.8$ kHz.

3) RESTING-STATE fMRI INDIVIDUAL-SUBJECT DATA PRE-PROCESSING

For each subject, the acquisition timing difference was first corrected for different slice locations. With the last functional volume as the reference, rigid-body motion correction was done in three translational and three rotational directions. The amount of motion was estimated and then modeled in data analysis. For each subject, spatial blurring with a full width half maximum (FWHM) of 4mm was used to reduce random noise and inter-subject anatomical variation during group analysis. At each voxel, motion-estimation parameters, baseline, linear and quadratic system-induced signal trends were removed from the time courses using the "3dDeconvolve" routine in AFNI. Brain global and CSF mean signals were modeled as nuisance variables and removed from the time courses as well. In order to create the time course from pure CSF regions, the lateral and 3rd ventricles on the high-resolution T$_1$-weighted volumetric images were segmented using FreeSurfer software followed by 1 mm^3 erosion. The cleaned time courses were then band-pass filtered in the range of 0.009 Hz - 0.08 Hz and used for connectivity analyses.

B. GENERATION OF ROIs FOR CONNECTIVITY ANALYSES

We selected relatively small-size, anatomically-defined regions based on the FreeSurfer [26] segmentation as seeds for connectivity analyses in this work, including right and left isthmi of the cingulate cortex (ICCs), and right and left hippocampi. Each of these regions is well-defined by FreeSurfer, with reasonable volumes for connectivity analyses (means of 2089/2381 mm^3 for the right/left ICCs and 3494/3400 mm^3 for the right/left hippocampi for the NC group). The ICCs defined in FreeSurfer overlaps with the boundaries of the PCC/RSC (L1, P50, S26), defined by Buckner *et al.* [33] for their connectivity analyses of the DMN. The right and left hippocampi are also part of the DMN [33] and are known to be affected by AD.

C. PERFORMANCE COMPARISON OF DIFFERENT CLASSIFICATION ALGORITHMS

In this subsection, we present the classification performance of the proposed method and compare it to existing methods.

The performance of the classifier is evaluated using the Leave-One-Out (LOO) cross-validation. As was pointed out in [21], comparing with other method, LOO might have a higher estimate error. However, in our study, as the size of data sample is small, splitting the data into two sets based on a threshold would generate a biased testing set which contains only a few subjects for each category. That is why LOO is chosen to evaluate the performance of the classifiers in this case. As described earlier, the ROIs used are the hippocampus and ICC from both hemispheres of the brain.

TABLE 1. Comparison of 3-category (AD, MCI, NC) classification results.

Method	Accuracy
LDA with Bayesian	42.4%
LDA with AdaBoost	69.7%
LR with L2 penalty	36.4%
SVM with RBF kernel	57.6%
Regularized LDA with AdaBoost	75.8%

TABLE 2. LDA with Bayesian.

		Predicted Class			Recall
		NC	MCI	AD	
Actual Class	NC	6	5	1	50.0%
	MCI	6	4	1	36.4%
	AD	2	4	4	40.0%
Precision		42.9%	30.8%	66.7%	

Table 1 shows the performance of five classifiers. In the first one, a naive Bayesian classifier is employed after the original LDA transform. As can be seen, its final accuracy is only 42.4%. As explained in Section 2, the main reason of such an unsatisfying performance is that: when the number of data samples is small, the estimation of class means and covariance matrices in LDA suffers from severe noise effect, leading to overfitting. In the second one, the original LDA is combined with the AdaBoost classifier. As can be seen, the accuracy is increased to 69.7% by AdaBoost. The accuracies corresponding to Logistic Regression (LR) and Support Vector Machine (SVM) are 36.4% and 57.6%, respectively. The last one is the regularized LDA combined with the AdaBoost classifier that we proposed. The shrinkage operations in the regularized LDA can reduce the noise effect in the estimation, and further improve the accuracy to 75.8%. As can be seen, Clearly, the purposed method is shown to be more robust in distinguishing AD, MCI and NC subjects. Compared to existing fMRI based classification algorithms [5], [6], i.e., LDA with Bayesian classifier, the proposed method has improved the classification accuracy by 33.4% with a p-value of 0.007 using the McNemar test [34].

Table 2 to 6 show the confusion matrices of the five classifiers. It can be seen that the purposed method shows better performance in terms of precision and recall rates. Also, as expected, compared with NC subjects and AD patients,

TABLE 3. LDA with AdaBoost.

		Predicted Class			Recall
		NC	MCI	AD	
Actual Class	NC	9	2	1	75.0%
	MCI	2	7	2	63.6%
	AD	2	1	7	70.0%
Precision		69.2%	70.0%	70.0%	

TABLE 4. LR with L2 penalty.

		Predicted Class			Recall
		NC	MCI	AD	
Actual Class	NC	6	3	3	50.0%
	MCI	7	2	2	18.2%
	AD	2	4	4	40.0%
Precision		40.0%	22.2%	44.4%	

TABLE 5. SVM with RBF kernel.

		Predicted Class			Recall
		NC	MCI	AD	
Actual Class	NC	7	5	0	58.3%
	MCI	2	8	1	72.7%
	AD	1	5	4	40.0%
Precision		70.0%	44.4%	80.0%	

TABLE 6. Regularized LDA with AdaBoost.

		Predicted Class			Recall
		NC	MCI	AD	
Actual Class	NC	10	2	0	83.3%
	MCI	2	7	2	63.6%
	AD	1	1	8	80.0%
Precision		76.9%	70.0%	80.0%	

it is generally more difficult for the classifier to identify MCI patients, and the classification accuracy for MCI patients is much lower than that for AD and NC subjects. MCI is a transitional stage of dementia between NC and AD. Therefore, MCI patients generally share some similar characteristics in functional brain connectivities with AD patients or NC subjects. As a result, the classification algorithm is prone to mis-recognize MCI patients as AD or NC subjects.

V. CONCLUSIONS

This paper proposes a reliable method for AD, MCI and NC subject classification that is highly robust under *size limited* fMRI data samples, by exploiting *brain network connectivity pattern analysis*. To do it, *first*, we selected the right and left hippocampi regions and isthmus of the cingulate cortices (ICCs) as our ROI sub-network, and calculated the Pearson correlation coefficients between all possible ROI pairs and used them to form a feature vector for each subject. *Second*, the vectors were projected into a one-dimensional axis using the proposed regularized LDA approach, where the differences between AD, MCI and NC subjects

were maximized. Shrinkage based regularization procedures were taken to reduce the noise effect due to the limited sample size. *Finally*, a decision tree based multi-class AdaBoost classifier, which is robust to noise effect, was applied to the projected one-dimensional vectors to perform the classification.

Both the theoretical and numerical analysis demonstrated that: (i) the regularization methods and the AdaBoost classifier can increase the classification accuracy significantly; (ii) brain network connectivity analysis, which evaluates the changes in the pattern of connectivity among multiple or all regions in the sub-network, can reveal in-depth information about brain connectivity and result in relatively accurate classification of AD, MCI and NC, especially when the sample size is very limited; (iii) our analysis confirms the previous findings that the hippocampus and the isthmus of the cingulate cortex are closely involved in the development of AD and MCI.

Potentially, the proposed framework can be applied to other classification problems as well, especially under limited sample size.

REFERENCES

[1] J. C. Morris *et al.*, "Mild cognitive impairment represents early-stage Alzheimer disease," *Arch. Neurol.*, vol. 58, no. 3, pp. 397–405, 2001.

[2] E. Westman, J.-S. Muehlboeck, and A. Simmons, "Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion," *Neuroimage*, vol. 62, no. 1, pp. 229–238, 2012.

[3] C. Lehmann *et al.*, "Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)," *J. Neurosci. Methods*, vol. 161, no. 2, pp. 342–350, 2007.

[4] R. Cuingnet *et al.*, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, 2011.

[5] K. Wang *et al.*, "Discriminative analysis of early Alzheimer's disease based on two intrinsically anti-correlated networks with resting-state fMRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2006, pp. 340–347.

[6] G. Chen *et al.*, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, 2011.

[7] W. S. Pritchard *et al.*, "EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures," *Electroencephalogr. Clin. Neurophysiol.*, vol. 91, no. 2, pp. 118–130, 1994.

[8] D. G. Stork, P. E. Hart, and R. O. Duda, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2012.

[9] B. Magnin *et al.*, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.

[10] I. O. Korolev, L. L. Symonds, and A. C. Bozoki, "Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification," *PLoS One*, vol. 11, no. 2, p. e0138866, 2016.

[11] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI," *NeuroImage*, vol. 112, pp. 232–243, May 2015.

[12] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 2, pp. 576–589, Feb. 2014.

[13] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with Alzheimer's disease using resting-state fMRI and graph theory," *Clin. Neurophysiol.*, vol. 126, no. 11, pp. 2132–2141, Nov. 2015.

[14] C.-Y. Wee *et al.*, "Identification of MCI individuals using structural and functional connectivity networks," *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, 2012.

[15] 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.

[16] F. Liu, L. Zhou, C. Shen, and J. Yin, "Multiple kernel learning in the primal for multimodal alzheimer's disease classification," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 3, pp. 984–990, May 2014.

[17] L. Xu *et al.*, "Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers," *J. Alzheimer's Disease*, vol. 51, no. 4, pp. 1045–1056, 2016.

[18] Q. Li, X. Wu, L. Xu, K. Chen, L. Yao, and R. Li, "Multi-modal discriminative dictionary learning for Alzheimer's disease and mild cognitive impairment," *Comput. Methods Programs Biomed.*, vol. 150, pp. 1–8, Oct. 2017.

[19] Q. Li, X. Wu, L. Xu, K. Chen, and L. Yao, "Classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning," *Frontiers Comput. Neurosci.*, vol. 11, p. 117, Jan. 2017.

[20] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, "A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages," *NeuroImage*, vol. 155, pp. 530–548, Jul. 2017.

[21] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, pp. 68–77, Oct. 2018.

[22] R. L. Buckner *et al.*, "The organization of the human cerebellum estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 5, pp. 2322–2345, 2011.

[23] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 13, pp. 4637–4642, 2004.

[24] H.-Y. Zhang *et al.*, "Resting brain connectivity: Changes during the progress of Alzheimer disease," *Radiology*, vol. 256, no. 2, pp. 598–606, 2010.

[25] D. C. Zhu, S. Majumdar, I. O. Korolev, K. L. Berger, and A. C. Bozoki, "Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default-mode network: A multi-modal imaging study," *J. Alzheimer's Disease*, vol. 34, no. 4, pp. 969–984, 2013.

[26] B. Fischl *et al.*, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[27] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[28] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class AdaBoost," *Statist. Interface*, vol. 2, no. 3, pp. 349–360, 2009.

[29] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 10, pp. 6567–6572, 2002.

[30] J. A. Rice, *Mathematical Statistics and Data Analysis*. Boston, MA, USA: Cengage, 2006.

[31] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.

[32] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[33] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network-anatomy, function, and relevance to disease," *Year Cogn. Neurosci.*, vol. 1, pp. 1124–1138, 2008.

[34] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Jul. 2004.