Multi-way Feature Extraction and Selection for Alzheimers Disease Early Detection

AA AA, AA AA

***Abstract*—Recently machine learning methods had gain lots of publicity among researchers in order to analyze the brain images such as functional Magnetic Resonance Imaging(fMRI) to obtain better understanding of the brain and brain related disease such as Alzheimer's disease. Classification methods has been deployed in order to discriminate Alzheimer's disease from normal controls. The majority of deployed techniques rely on constructing the functional connectivity (FC) and use the vectorized FC as the input for the classifiers which has two main drawbacks 1) The need for constructing the FC 2) The loss of possible valuable structural information in the vectorization step. Considering these problems and based on the 4D nature of the data, we have came up with a novel framework which omits the FC construction part and preserve the structural integrity of data for the classification. The proposed framework uses the High Order Singular Value Decomposition (HOSVD) in order to prune the classes and select the proper basis for each of them. This framework also allows us to obtain an FC matrix based on a class but not a single sample which helps us to shed more lights on the brain abnormalities in the Alzheimers disease at its early stages. Extensive experiments using the ADNI dataset demonstrate that our proposed framework effectively boosts the fMRI classification performance in Alzheimer's disease.**

***Index Terms*—IEEE, IEEEtran, journal, LATEX, paper, template.**

I. Introduction

ALZHEIMERS disease (AD) is a progressive neurodegenerative disorder with a long pre-morbid asymptomatic period [1] which affects millions of elderly individuals worldwide. It is predicted that the number of affected people will double in the next 20 years, and 1 in 85 people will be affected by 2050 [2]. The predominant clinical symptoms of AD include a decline in some important brain cognitive and intellectual abilities, such as memory, thinking, and reasoning. Precise diagnosis of AD, especially at its early warning stage: early Mild Cognitive Impairment (eMCI), enables treatments to delay or even avoid such disorders [3].

In recent years, brain imaging techniques like Positron Emission Tomography(PET) [21], Electroencephalography (EEG)[22] and functional Magnetic Resonance Imaging (fMRI)[23] have been used in analysis of AD. Due to the high spatial resolution and relatively lower costs, fMRI is vastly used among researchers in order to monitor brain activities especially in AD and all it's stages in which detecting abnormalities within small brain regions is essential. An fMRI sample is naturally a 4D tensor consisting of 3D voxels moving in time, and each voxel contains an intensity value that is proportional to the strength of the Blood Oxygenation Level Dependent(BOLD) signal , which is a measure of the changes in blood flow, to estimate the active regions in the brain[7]. Resting State fMRI(rs-fMRI) is an fMRI technique in which the patient is asked to rest during the whole scan, focuses on the low-frequency $(< 0.1Hz)$ oscillations of BOLD signal, which presents the underlying neuronal activation patterns of brain regions[8][10]. rs-fMRI is usually used in order to analyze the brain diseases like AD or autism[33], [34]

Since each fMRI volume consist of hundreds of thousands of voxels which are often highly correlated with the surrounding voxels in the brain volume, parcellation of the brain for further analysis has moved toward the use of anatomical atlases. These atlases are strictly defined using anatomical features of the brain, like locations of common gyri and do not rely on any functional information. To generate data using an atlas based approach, the BOLD signal from all voxels is averaged within each brain region called Region of Interest(ROI)[9]. By putting together the average time-series for all the ROIs, the ith volume would become $X_i \in \mathbb{R}^{T \times R}, i = \{1, 2, \cdots, S\}$ in which R, T and S are the number of ROIs, time points and samples respectively. The process of obtaining such matrix is shown in Figure (1).

There are two major studies associated with rs-fMRI data: finding common brain disorders caused by diseases like Alzheimer's or autism, and more recently detecting patients with brain disorders using classification techniques. Due to the high dimensionality of data and the nature of diseases like eMCI which does not show any reliable clinical symptoms, researchers moved towards advanced machine learning techniques in order to achieve more reliable analysis.

A powerful tool that is commonly used in order to achieve aforementioned goals is Functional Connectivity(FC) network. FC is a $region \times region$ matrix $\bar{X}$ in which $\bar{x}_{ij}$ represents the functional connectivity between the ith and jth ROI. Functional connectivity is an observable phenomenon quantifiable with measures of statistical dependencies, such as correlations, coherence, or transfer entropy([][]). Recent studies have shown that some brain disorders like AD could alter the way that some brain regions interact with each other. For example, compared with the healthy, AD patients have been found decreased functional connectivity between hippocampus and other brain regions, and MCI patients have been observed increased functional connectivity between the frontal lobe and other brain regions[4]. Although FCs show promising results, they bring their own challenges. Variety of methods such as Pairwise Pearsons correlation coefficient [10], [11], sparse representation [10], [12], [13] and Sparse Inverse

M. Shell was with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: (see http://www.michaelshell.org/contact.html).

J. Doe and J. Doe are with Anonymous University.

Manuscript received April 19, 2005; revised August 26, 2015.

Covariance Estimation (SICE) exists to obtain an FC, While the first two are easy to understand and can capture pairwise functional relationship based on a pair of ROIs, the latter can account for more complex interactions among multiple ROIs, but the estimation of partial correlation involves an inversion of a covariance matrix, which may be ill-posed due to the singularity of the covariance matrix. These methods results in vastly different networks. On the other hand, computing the correlation based on the entire time series of fMRI data simply measures the FC between ROIs with a scalar value, which is fixed across time. This actually implicitly hypothesizes the **Stationary** interaction patterns among ROIs. As a result, this method may overlook the complex and dynamic interaction patterns among ROIs, which are essentially time-varying [16]–[19]. In order to overcome this issue, **Non-stationary** methods have been proposed which results in more complex networks.

Finding an FC that highlights the patterns caused by a disease has been a common goal in rs-fMRI study for a long time. Several approaches exists to find common patterns among different brain scans. Data driven methods such as kernel-PCA or clustering techniques have been proposed for this task. But ultimately most of them rely on calculating a network for each volume. This may overlook the role of noises or outliers within the data. As we will show later, our proposed method does not rely on any individual FC calculations and treats each class as a whole in order to calculate the FC.

FCs are also used as features for classification. So instead of using X_is, their corresponding FCs i.e. $\bar{X}_i$s are classified in order to detect the patients with eMCI. A straightforward way is to vectorize the FCs into high-dimensional vectors as features as in [33], which are then used to train a SVM or k-NN classifier. this method produces poor performance due to the phenomena known as Curse of Dimensionality. Variety of techniques have been proposed in order to reduce the dimensionality of these FCs but since they all use conventional classifiers like SVM or k-NN, one way or an other they have to vectorize the data in order to feed it to these classifiers. Alongside the curse of dimensionality, vectorization also destroys potential information that are embedded in the structure of data. This problem have been studied specially in image processing in which vectorization destroys the spatial relations within an image.

In this paper, based on high order tensor decomposition, we have created a framework in which the mentioned goals i.e. finding an FC for a whole class and detecting a disorder via classification could be achieved via a single High Order Singular Value Decomposition (HOSVD). The proposed framework reveals both functional connectivities related to eMCI that are confirmed by empirical methods along with novel connectivities. The proposed classifier also outperforms state of the art eMCI classification methods.

Viewing each class as a tensor allows us to work with *time* and *region* features separately but simultaneously. This multilinear view ables us to design a proper dimension reduction relative to the nature of each feature along with a discriminant function based on linear regression that uses the test data to enhance the quality of the training set without forcing any apriory knowledge to the classifier, a task which is not possible through well known classifiers like SVM, logistic regression or k-NN. It is also notable that the proposed discriminant function directly classifies the X_is. Having the FC calculation step omitted not only heavily affects on the computational performance of the method, it also saves us from the troubles of FCs mentioned before.

To verify our approach, we conduct an extensive experimental study on rs-fMRI data from the benchmark dataset ADNI[1] As will be seen, the results well demonstrate the effectiveness and advantages of our method. Specifically, the proposed TBNA system, not only grants us superior classification accuracy to that from other methods, it is also much faster. We have also confirmed our achieved FC matrix using empirical data on the eMCI functional connectivity patterns.

II. eMCI CLASSIFICATION TECHNIQUES

As it was discussed before, classification techniques have become a favorable method for Alzheimer disease early detection. early classification methods used X_is as the representative for each subject and used them directly in the classification process. As the functional connectivity matrix gain popularity among researchers, the majority of these classification methods shifted towards classifying the functional connectivity matrices. There are two main paradigms towards obtaining the FCs. Stationary and non-stationary methods, Stationary methods use a single scaler value in order to determine the functional connectivity between two ROIs. Non-stationary methods consider a more complex relation between ROIs that can not be best captured with a single scaler value.

In order to demonstrate the power of our proposed framework, we have implemented an state of the art method from each of these paradigms. The first method which is an stationary one uses the Sparse Inverse Covariance matrix in order to establish a functional connectivity matrix and then using the SPD property of SICE, propose a dimension reduction technique in order to find a set of low dimensional features for classification. The second method which is non-stationary, uses sliding windows to construct a more complex connectivity matrix and uses clustering in order to obtain low dimensial features for classification.

- **Compact SICE:** In this method, first the SICE matrix is extracted from the data sample S using the following optimization:

$$S^* = \arg\max_{S\succ 0} \quad \log\left(\det(S)\right) - \mathrm{tr}(CS) - \lambda \left\|S\right\|_1 \quad (1)$$

where C is the sample-based covariance matrix; $\det()$, $\mathrm{tr}()$, and $\|.\|_1$ denote the determinant, trace, and the sum of the absolute values of the entries of a matrix. Since the dimensions of SICE matrix is R^2, with R representing the number of ROIs, the obtained matrix is still relatively large. Principal Component Analysis(PCA) has proved itself to be one of the most powerful methods of dimension reduction and this method uses Kernel-PCA with a Gaussian kernel in order to extract the key features

[1] http://adni.loni.usc.edu/

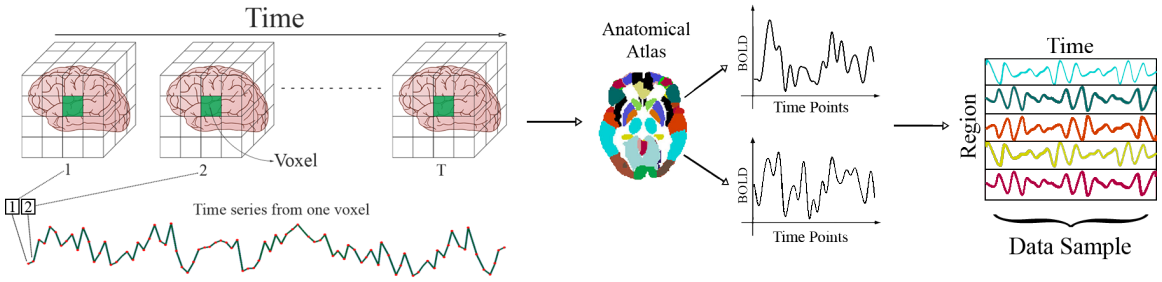


Fig. 1. Simulation results for the network.

of SICE. Since SICE is an SPD matrix, specific distance functions such as Log-Euclidean distance or Root Stein divergence can be used as the distance function in the Gaussian kernel. After extracting key features, conventional methods like SVM or Knn could be deployed for classification.

- **High Order Networks(HON)** This method uses so called High Order Networks as features for classification purposes. it uses the sliding window technique in order to split the time-series into smaller pieces and then find the relation between them. Let $x_i^{(l)}(k) \in \mathbb{R}^N$ denote the k-th segment of subseries extracted from $x_i^{(l)}$, which comprises N image volumes. by taking each $x_i^{(l)}(k)$ as a node A network can be constructed with edges defined using:

$$C_{ij}^{(l)}(k) = \text{corr}\left(x_i^{(l)}(k), x_j^{(l)}(k)\right)$$

which represents the pairwise Pearsons correlation coefficients between the i-th and the j-th ROIs of the l-th subject using the k-th segment of subseries. by taking

$$y_{ij}^{(l)} = \left[C_{ij}^{(l)}(1), C_{ij}^{(l)}(2), \cdots, C_{ij}^{(K)}(1)\right] \in \mathbb{R}^K$$

as new nodes, an other network can be calculated as follow:

$$H_{ij,pq}^{(l)} = \text{corr}\left(y_{ij}^{(l)}, y_{pq}^{(l)}\right)$$

for each pair of correlation time series y_{ij} and y_{pq} thus, $H_{ij,pq}^{(l)}$ indicates how the correlation between the i-th and the j-th ROIs influence the correlation between the p-th and the q-th ROIs. The total number of the high-order correlation coefficients $\{H_{ij,pq}^{(l)}\}$ is proportional to R^4 which will lead to a large-scale high-order FC network, containing at least thousands of vertices and millions of edges. In order to overcome this issue, the correlation time series within each subject is grouped into different clusters. Then, the correlation calculation between the original correlation time series can be converted into that between the respective mean correlation time series in clusters. After reducing the network size, the weighted-graph local clustering coefficients was used to select the key features for each network and then SVM is used in order to classify the obtained features.

It is noteworthy that non of these techniques consider the multi-linearity nature of the data, and since both methods use traditional classifiers like SVM or KNN, they follow a rather complex path to find vector features as the representative of each FC matrix.

III. Proposed fMRI analysis Framework Based On Tensor framework

All main mentioned eMCI classification methods use FC of each sample as input feature. So, the computational complexity of these method due to construction of FC matrices are high and also the quality of these FC matrices have a direct effect in the quality of classification.

Also know of them did not used the multi-linear propriety of such data. In this section by tensor viewpoint we show that only by one tensor decomposition we could doe the following analysis on fMRI data

- Classification: Based on HOSVd decomposition we define a discriminant function to classify matrix data X_i without folding and without construction of FC matrix per- sample.
  - By the obtained base vector from HOVD By using High order singular value eecomposition (HOSVD), all region-time matrices are projected to smaller matrices (without folding the data to vectors)
  -
  - As the first time we propose a novel method that alow test data have a contribution in constructions of discriminant function of each class. This idea hav a big effect in the quality of the propped method

  Since this method directly works on region-time matrices as input features, its computatonal complexity is more smaller than state of the aret methods that use FC matrices as input feature. Also the experimental results confirm the quality of this method in recognition of normal and obnormal data.
- By the previous HOSVD decomposition we proposed a novel method that could find a general representative FC for eMCI and Normal data. In this method computing the FC of all smaple is not necessary to find FC of all samples. In the experiments the obtained FC contains a relations that recently founded expermentaly(cilinically).

A. Dimension reduction and classification based on Tensor framework

In region based fMRI data each data is a region-Time matrix $X \in \mathbb{R}^{T \times R}$, where T, R denote the length of time and regions

of the data. For reducing the computational complexity and reducing the occurrence of over fitting, specially for data with large features (Like fMRI data), using dimesionality reduction is inevitable. Although each sampe X have two different king features (time and region), the classical methds like PCA, SVD ,... only could works on vectorizaed version $x = vec(X)$ of such data. In these methods by having projection matrix U(Due to method), the projected dat will be $y = U^{\mathsf{T}}x$. Although this approach is easy to deploy, it has several drawbacks, one main drawback would be the Curse of dimensionality that appears when the proportion of the number of features to the number of samples is relatively high (which results in over-fitting). Also in this view different kind of features(like *Time* and *region* in our case) would be mixed together which may discard some important information within these features. The second approach is to deploy multilinear methods.

This vectorizatin forthermore inceasing the computatina coplexity, mixed the metioed two type of vectors. Recently some worls like MPCA, GLRAM methods have been proposec that are able to reduce sample with matrix data types witout folding them to vectores. In these methods there is a freedom to selectio spesific reduction for each sample. In this mutilinear methods for each feare its corresponding projection metrix is obtained. In this section we will used wellknown tensor decompostion named HOSVD for dimension reduction and also classification of fMRI data.

Let $X_1, \ldots, X_{s_1}$ and $X_1, \ldots, X_{s_2}$ be the train data of eMCI and Normal classes, respectively. Now we substitute these data at tensors $\mathcal{X} \in \mathbb{R}^{T\times R\times S_i}$, where each mode-2 slice show sample data. Since the same process will be done on these tesors, So for simplicity we explain our proposed process on tensor $\mathcal{X}$ without up-script.

Let $\mathcal{X} \in \mathbb{R}^{T\times R\times S}$, where each slice $X(:,:,i)$ denotes the time-region feature of the i^{th} sample. If

$$\mathcal{X} = \left(U^{(i)}, V^{(i)}, W^{(i)}\right) \cdot \mathcal{S}, \tag{2}$$

where otrhogonal matrices $U \in \mathbb{R}^{T\times T}, V \in \mathbb{R}^{R\times R}$ and $W \in \mathbb{R}^{S\times S}$ are modes-1,2,3 singular matrices of $\mathcal{X}$, and S is the corresponding core tensor. As it was mentioned in (**??**) U is a base of all mode-1 fibers $\mathcal{X}(:,i,j)$ which indicates the behavior of ith region of the jth sample in all times. Also V is a base of all mode-2 fibers $\mathcal{X}(i,:,j)$ which indicates the behavior of all regions of ith jth sample in the ith time. Also due to the properties of HOSVD inherited from svd, the first columns of mode-k (k=1,2,3) has more ability in construction of main parts of kth fibers and on the other hand, the last columns of this singular matrix, are corresponding to noise parts of the fibers of the corresponding mode. Therefore for appropriate values of k_1 and k_2 projection mode-1 and mode-2 fibers into space spanned by the first k_1 and k_2 singular vecotes of modes-1,2, i.e., matrices $U_{k_1} = [u_1, \ldots, u_{k_1}]$ and $V_{k_2} = [v_1, \ldots, v_{k_2}]$ is a suitable dimension reduction. This dimensionality reduction could be done as

$$\mathbb{R}^{k_1\times k_2\times S} \ni \bar{\mathcal{X}} = \left(U_{k_1}^{\mathsf{T}}, V_{k_2}^{\mathsf{T}}\right)_{1,2} \cdot \mathcal{X} \tag{3}$$

So, by having mode-2 and mode-2 singular matrices of tensor $\mathcal{X}$, its reduced versin $\overline{\mathcal{X}}$ without folding the data into vecores could be otained. Here its clear that the reduced version of the i^{th} sample, i.e.,$\overline{\mathcal{X}}(:,:,i) \in \mathbb{R}^{k_1\times k_2}$ could be reduced separatlty in each mode. This means thnat in this case the structure of data is preserved in the reducing process. ?? So, if

$$\mathcal{X}^{(i)} = \left(U^{(i)}, V^{(i)}, W^{(i)}\right) \cdot \mathcal{S}^{(i)}, \tag{4}$$

be the HOSVD decomposition of the data of i^th class, for apploerite k_1^i, k_2^i, indices, its reduced data will be

$$\mathbb{R}^{k_1^i\times k_2^i\times S_i} \ni \bar{\mathcal{X}}^{(i)} = \left(U_{k_1^i}^{(i)\mathsf{T}}, V_{k_2^i}^{(i)\mathsf{T}}\right)_{1,2} \cdot \mathcal{X}^{(i)}. \tag{5}$$

The error analysis and the property of HOSVd, it could b shown that by small number of k_1^1, k_2^i, the reduced data have good reconstruction error.

In the following in spired by the structure of this reduction, we present a tensor based discrimination function. By substituting $\mathcal{X}^{(i)}$ in ?? by its HOSVD decomposition (4), the projected data $\overline{\mathcal{X}}^{(i)}$ becomes

$$\begin{aligned}\bar{\mathcal{X}}^{(i)} &= \left(\begin{bmatrix} I_{k_1^i} & 0 \end{bmatrix}, \begin{bmatrix} I_{k_2^i} & 0 \end{bmatrix}, W\right) \cdot \mathcal{S}^{(i)} \\ &= (W)_3 \cdot \mathcal{S}^{(i)}(1:k_1, 1:k_2, :)\end{aligned}$$

By this equation, ne can see that each sample of the i^{th} class in the the reduced space have the following form

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,:,k) &= \left(W^{(i)}(k,:)\right)_3 \cdot \mathcal{S}^{(i)}(1:k_1^i, 1:k_2^i, :) \\ &= \sum_{k'=1}^{S_i} W^{(i)}(k,k') \cdot \mathcal{S}^{(i)}(1:k_1^i, 1:k_2^i, k').\end{aligned}$$

For simplicity, if we define $\overline{\mathcal{S}}^{(i)} = \mathcal{S}^{(i)}(1:k_1^i, 1:k_2^i, :)$, it is cear that each sample in the i^{th} calss could be represented as linear combination of slices $\overline{\mathcal{S}}^{(i)}$. So if a test data like $X \in \mathbb{R}^{T\times R}$, be in the i^{th} class its natural to expect that it's projected version into principle region and times spaces spanned by $U_{k_1^i}, V_{k_2^i}$, i.e,

$$Z^{(i)} = ??$$

could be approximated well as a linear combination of slices of $\overline{\mathcal{S}}^{(i)}$ as follows

$$Z^{(i)} \approx \sum_{k=1}^{S_i} \lambda_k^i \overline{\mathcal{S}}^{(i)}(:,:,k).$$

So, for each test data X, the best approximation its projection versions for spaces of two normal and obnormal is computed and then X assighen to one class with best approximation. mathematically this could be written as follows: For given test data X, its projection data $Z^{(i)}$ for two calss is obtained. Then the best approximation of these projected data in coreesponding spaces are obtained as follows

$$r_i = \min_{\lambda^i} \|Z^{(i)} - \sum_{k=1}^{S_i} \lambda_k^i \overline{\mathcal{S}}^{(i)}(:,:,k)\|, \quad \lambda^i = \begin{pmatrix} \lambda_1^i \\ \vdots \\ \lambda_{s_i} \end{pmatrix} \tag{6}$$

Here $r^{(i)}$ show the reconstruction error of the projected version of X in the i^{th} class. So we assigen X to l^{th} class if

$$r_l = \min_{i=1,2} r_i.$$

The minimization problem (6) is a simple least square problem that could be solved easily. The proposed method have an interesting property that bring us to propose a novel method to improve its quality. By the properties of HOSVD, the priciple propertires of the i^{th} class could be reflected in the slices of $\overline{\mathcal{S}}^{(i)}(:,:,i)$. Also for small indices these Slices have a better role in construction of main properties of this class. So the first slices of $\overline{\mathcal{S}}$ could represent signal parts, while laste slices donse not this role and sometimes represents the wast parts. Now consider a data X added to data set $\mathcal{X}^{(i)}$ of the i^{th} class. So new data set will be $\widetilde{\mathcal{X}} \in \mathbb{R}^{T\times R\times(S_i+1)}$. For example

$$\begin{aligned}\widetilde{\mathcal{X}}^{(i)}(:,:,1:S_i) &= \widetilde{\mathcal{X}}^{(i)},\\ \widetilde{\mathcal{X}}^{(i)}(:,:,S_i+1) &= X.\end{aligned}$$

If X belongs to this class then in HOSVD of this tensor X reinforce the slices of core tensor with small indices. But when it dose not belong to this class then it will changes the behavior of last slices of core tensor of new tensor. This meas that befor classification of a test data X, We could added it to both classes. This addition, will changed the first slices of core tensor when it beloges to this class or will cahged the last Slices when dose not belog to this class. But in our classification method only the first Slices of core tensor are exits. Also the elements of S in the first and second modes are filtered and we work with $S(1:k_1^i, 1:k_2^i, 1:k_3^i)$. This means that by addition of X in both classes for class that it beloge to it, the first slices in ??, will be adapted to have good reconstruction for this data. But for a class that dose not beloge to, this addition have small chages in the slices in ??.

It shoul be mentioed that since X added to both data sets, so we did not its labe and the method in true. After computing the HOSDVD of this new data sets $\widetilde{\mathcal{X}};(i)$, we apply the approach in ??. In summery this algorith coul be summerized as follows: In this section by HOSVD decomposition on both normal and

Algorithm 1 TNBeMCI: Tensor based Classification method

1) **Input**: Normal train data $\mathcal{X}^{(1)}$, eMCI train data $\mathcal{X}^{(2)}$ $k_i^j, i=1,2,3, \quad j=1,2.$
   Test data X
2) Construct $\widetilde{\mathcal{X}}^{(i)}$ for $i=1,2$ by adding X as ??
3) Compute $U_{k_1^i}, V_{k_2^i}$ and $\mathcal{S}(1:k_1^i, 1:k_2^i, k_3^i)$ for $\widetilde{\mathcal{X}}^{(i)}$.
4) Compute $Z^{(i)} = \left(U_{k_1^i}^{(i)\mathsf{T}}, V_{k_2^i}^{(i)\mathsf{T}}\right)_{1,2} \cdot X$, $i=1,2$.
5) Comput r_1, r_2 from ??
6) Assign X to class l, if $l = \arg\min_i\{r_i\}$

eMCI data, we proposed a novel tensor based classification. The benifits of the proposed tensor based classification method could be summerized as follows

- This method use multilinear propert of these data and works could works with time and region features separatly based on their specific onbatned bases(corresponding singular matrices).
- The classification works on data without folding them ti vectors.
- As we mentioed the test data has a role in construction of bases for both classes, without knowing the class of test data. This incearse the quality of the discrimination. This could not applied on important classification ,methods like SVM, logistic regression classifier and neroual networks. This is also could be applied on classifications based on matrix factorization.
- This method works with time-region features, and dose not need to have FC matrices for each sample. So its computational complexity is more less than other state of the art methods like methods mentied in this paper. It should be mentioed that our method also could be applied on FC matrices as input featur. But in experments the propoed method gives better results on original time-region faetres. So, in this feature we do not warry about the qaulity of FC methods.

In the expermental results we compaire the quality of the method with some of the sate of the art methods for recognition of eMCI based on accuray nad elapsed times.

B. General Functional Connectivity for normal and eMCI by tensor framework

Functional connectivity simply means the relation between different ROIs. As we mentioed the FC could be used for two different pouroses. Some classification methods for each samle by time-region matrix compute a region-region matrix wich shows the FC per sample. These method use this new fearure insead of time-region feature. But, in other view point we want to obtaine only one representative functional connectivity for each normal or eMCI classes. Here the classification is not our goal, instaed finding a gereral and represetative realtion of regions for eMCI and normal preson is our demand. This general pattern could be used clinically by expert congnitive scintists to found functional relation Disorders accrued by eMCI.

??Although several approaches have been proposed in order to find the functional connectivity, the majority of them focus on the individual samples rather that the whole class. This may overlook tiny but common connectivities shared within a class like the class of people in their early Alzheimer's disease.

As we know in HOSVD the singular matrices U, V, W, are the base of time, region features and also the samples, receptively. By property of these matrices we desigen our classification method. In this section we show that , the same idea could helps us to define a general FC matrix based on HOSVd decomposition for normal and eMCI data. It show be mentioed that here we did not use the FC of per-samples to construct this gerenral FC matrix.

Due to structure of data for $\mathcal{X}^{(i)}$, the slice $\mathcal{X}^{(i)}(:,l,:)$, denotes the behavior of l^{th} region of all samples in all times. So this could be consider as features of the l^{th} region of the i^{th} class and each region is shown with an Times-sample feature matrix. Viewing each region as a slice would allow us to consider it's behavior in all time points and across all samples, which itself allows us to shed more light on common properties and ignore individual differences that is highly possible due to the presence of noise and outliers. So, by the properties of singula matrices in modes-1,3, for approprite values k_1^i, k_3^i, each regin $\mathcal{X}(:,l,:)$ could be reduced

in bith time and sample dfeatures based on mode-1 and mode-3 truncated singular matrices $U_{k_1^i}^{(i)}$ and $W_{k_3}^{(i)}$ as follows

$$\mathcal{Y}^{(i)}(:,l,:) = \left(U_{k_1^i}^{(i)^\mathsf{T}}, W_{k_3}^{(i)^\mathsf{T}}\right)_{1,3} \cdot \mathcal{X}^{(i)}(:,l,:). \tag{7}$$

Here $\mathcal{Y}^{(i)}(:,l,:)$ denotes a reduced version of $\mathcal{X}^{(i)}(:,l,:)$ into space spanned by $U_{k_1^i}^{(i)}$ and $W_{k_3}^{(i)}$ in modes-1,3. So,

$$\mathbb{R}^{k_1^i \times R \times k_3^i} \ni \mathcal{Y}^{(i)} = \left(U_{k_1^i}^{(i)^\mathsf{T}}, U_{k_3^i}^{(i)^\mathsf{T}}\right)_{1,3} \cdot \mathcal{X}^{(i)} \tag{8}$$

denote all reduced regions of the i^{th} class. By this structure and subsituting the HOSVD decomposition of $\mathcal{X}^{(i)}$ in (10), we obtain

$$\begin{aligned}\mathcal{Y}^{(i)} &= \left(\begin{bmatrix} I_{k_1^i} & 0 \end{bmatrix}, V, \begin{bmatrix} I_{k_3^i} & 0 \end{bmatrix}\right) \cdot \mathcal{S}^{(i)} \\ &= (V)_2 \cdot \mathcal{S}^{(i)}(1:k_1^i,:,1:k_3^i)\end{aligned}$$

and so

$$\begin{aligned}\mathcal{Y}^{(i)}(:,k,:) &= \sum_{k'}^{R} V^{(i)}(k,k')\bar{\mathcal{C}}^{(i)}(:,k',:) \\ &= \left(U_2^{(i)}(k,:)\right)_2 \cdot \mathcal{C}^{(i)}\end{aligned} \tag{9}$$

in which

$$\mathbb{R}^{k_1^i, R, k_3^i} \ni \mathcal{C}^{(i)} = \mathcal{S}(1:k_1^i,:,1:k_3^i).$$

In order to calculate the FC, we first reduced the dimensions of our data similar to what we did in the previous part:

$$\mathbb{R}^{k_1^i \times R \times k_3^i} \ni \bar{\mathcal{X}} = \left(U_{k_1^i}^{(i)^\mathsf{T}}, U_{k_3^i}^{(i)^\mathsf{T}}\right)_{1,3} \cdot \mathcal{X} \tag{10}$$

it can be seen that each region slice can be written as:

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,k,:) &= \sum_{k'}^{N_i} U_2^{(i)}(k,k')\bar{\mathcal{C}}^{(i)}(:,k',:) \\ &= \left(U_2^{(i)}(k,:)\right)_2 \cdot \bar{\mathcal{C}}^{(i)}\end{aligned} \tag{11}$$

in which

$$\bar{\mathcal{C}}^{(i)} = \mathcal{C}(1:k_1^i,:,1:k_3^i).$$

The equation (11), show that the reduced version of each region in the i^{th} class could written as the linear combinations of mode-2 slices of $\mathcal{C}$. So the coeffienvt of slices in this linear combination could be considered as a new feature for the l^{the} region of the i^{th} class. Also as we mentioed the first slices are better than to reflect the pricipp of the data So for approperite k_3^i we could select anly the fits coefficient in ?? as new features for the l^{th} region. Mathematically this means each region in the i^{th} class could represen by new feaure $V(l,1;k_3) \in \mathbb{R}^3$. The benefits of this approach is that each region coulb be represented only be a vecor with size k_3, instead large feature matrix time-sample. Now different methods could be used to find the correlation of regins based on this small and clean matrix $V(:,1:k_3)$ which its rows corresponding to different regions and columns are new features. In this due to quality of Sparse Inverse Covariance , this method is applied the obained feature-region matrix $V(:,1:k_3^i)^\mathsf{T}$ in order to find the functional connectivity of the i^{th} class.

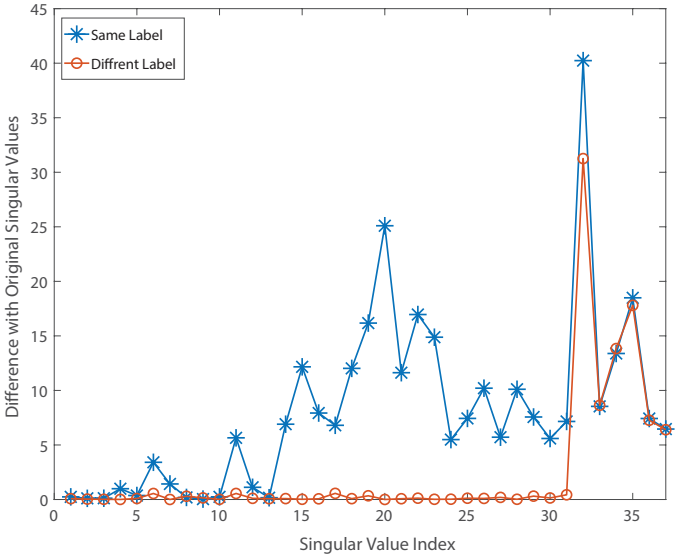


Fig. 2. The blue line with the star indicator, shows the difference between the mode-3 singular values of the original data and the data appended with a *same* label subject. The orange line indicated with circles, shows the difference between the mode-3 singular values of the original data and the data appended with a *different* label subject.

IV. EXPERIMENTAL STUDY

A. Data Preprocessing and Experimental Settings

Rs-fMRI data of 196 subjects were downloaded from the ADNI website[2]. Nine subjects were discarded due to the corruption of data, and the remaining 187 subjects were preprocessed for analysis. After removing subjects that had problems in the preprocessing steps, such as large head motion, 156 subjects were kept, including 26 AD, 44 early MCI, 38 late MCI, 38 NC, and ten significant memory concern labeled by ADNI. We used the 38 NC and the 44 early MCI because our focus in this paper is to identify MCI at very early stage, which is the most challenging and significant task in AD prediction. The IDs of the 82 (38 NC and 44 early MCI) subjects are provided in the supplementary material.

The data are acquired on a 3-T (Philips) scanner with TR/TE set as $3000/30$ ms and flip angle of 80. Each series has 140 volumes, and each volume consists of 48 slices of image matrices with dimensions 64×64 with voxel size of $3.31 \times 3.31 \times 3.31\ mm^3$. The preprocessing is carried out using SPM12 and DPARSFA [40]. The first ten volumes of each series are discarded for signal equilibrium. Slice timing, head motion correction, and MNI space normalization are performed. Participants with too much head motion are excluded. The normalized brain images are warped into automatic anatomical labeling (AAL) [41] atlas to obtain 116 ROIs as nodes. By following common practice [15][17], the ROI mean time series are extracted by averaging the time series from all voxels within each ROI and then bandpass filtered to obtain multiple sub-bands as in [17].

[2]http://adni.loni.usc.edu

B. Classification

Almost every subject in ADNI dataset has several scans. Usually a random scan data is selected and enters the processing step. This random selection may cause several problems. Since the number of train data is very low, a small perturbation in it could drastically change the set of input parameters in order to achieve the highest prediction accuracy and other classification evaluation methods. Also achieving high quality results with a classifier does not guarantee its effectiveness on other datasets even with fine tuning the parameters since the training set may contain outliers and unidentified corrupted data. In order to show that the proposed framework is less sensitive against the choice of different perturbations and is less vulnerable towards the aforementioned issues, we have selected 18 different perturbations and two test state of the art classification methods on them: **HON** and **k-SICE**. To make full use of the limited subjects, a leave-one-out procedure is used for training and test. That is, each sample is reserved for test in turn, while the remaining samples are used for training. We have use five evaluation measures: accuracy (ACC), sensitivity (SEN), Youdens index(YI), F-score, and balanced accuracy (BAC). In this article, we treat the eMCI samples as positive class and the NC samples as negative class.

1) Classification performance: The classification accuracy measure(ACC), After fine-tuning the input parameter set for each method, shows that for 16 out of 18 different perturbations, our approach works better than k-SICE, the same also holds for 15 datasets comparing to HON. i.e. in 88.8% of datasets, TBNA works better than k-SICE and in 83.3% of datasets, it works better that FON. The highest classification accuracy(86.59%) is achieved with the TBNA in the 15th perturbation. The highest accuracy for the HON (84.15%) is achieved in the 14th, and the highest accuracy for the SICE method (85.37%) is achieved in the 6th perturbation. As it was mentioned before, being stable when the input dataset changes is a very important aspect for a classifier, in order to measure the stability, standard deviation of accuracy along with other measures are calculated. The std. of accuracy for TBNA is 0.64 times less than HON and 1.73 times less than SICE method. Similar results also holds for other classification measures.

Figure (3) shows the performance of these three methods in all five measurements. Some statistical information about these plots are also included in the embedded table. As it can bee seen in this figure, similar to the accuracy, the proposed method in overall works much better than FON and k-SICE. For a better Demonstration, table (I) provides the average of several classification measurements scores for all permutations. As it can be seen in this table, the average accuracy of TBNA which is 80.43% is 4.77% higher than the next method HON, and 4.86% better than k-SICE. It is noteworthy that The other two methods i.e HON and SICE shows similar results in average.

2) Runtime Comparison: One other key features of TBNA is that it works significantly faster that the other two methods. Table (II) shows the average elapsed time (Training plus Testing) of each method for all selected permutations. These methods were executed in matlab R2017b and carried with an intel Core-i7 processor and 16GB of RAM. As it can be seen in this table, TBNA is more that 600 times faster than HON and 20 times faster that SICE. Having a huge execution time specially affects the parameter selection for HON, since it uses cross-validation procedure in order to find the optimal parameters which itself require several runs of the algorithm.

TABLE I
THE AVERAGE OF DIFFERENT CLASSIFICATION MEASUREMENTS IN ALL PERTURBATIONS IN %

Method	ACC	F-Score	SEN	SPE	YI	BAC
k-SICE	75.57	77.36	78.50	72.19	50.69	75.34
FON	75.66	77.44	78.40	72.48	50.89	75.44
TBNA	**80.43**	**82.20**	**84.60**	**75.59**	**60.20**	**80.09**

TABLE II
ELAPSED TIME OF THE TEST AND TRAIN PHASE IN SECONDS

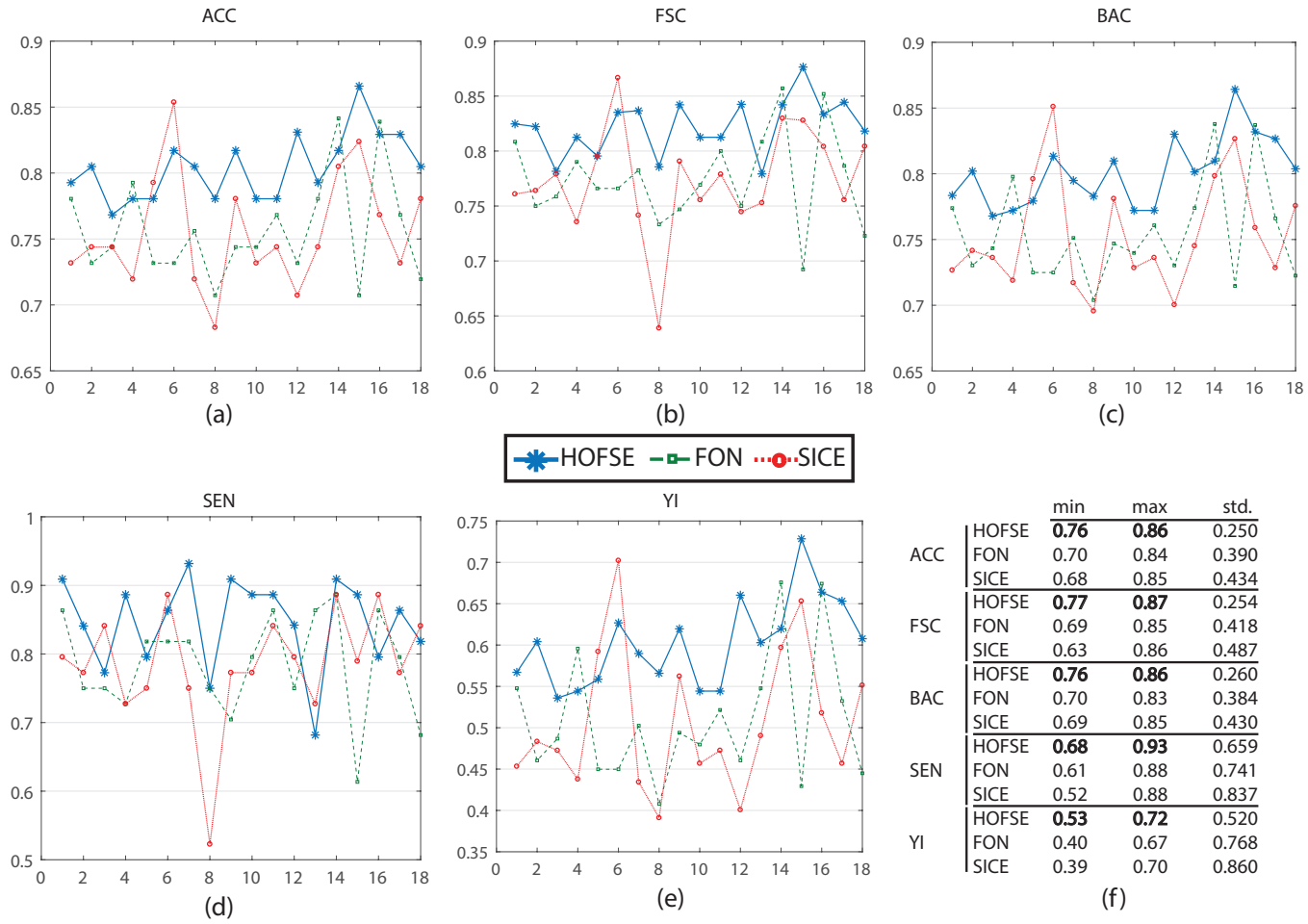
Method	HON	k-SICE	TBNA
Elapsed Time	6950	230	11

C. Functional connectivity Network

The functional connectivity networks of the Normal and eMCI classes was obtained via TBNA as it is described in (III-B). In order to better highlight the differences between Normal and eMCI subjects, a difference graph D is constructed by subtracting the Normal FC from the eMCI FC. This graph could be seen in Figure(4). The nodes of D shows the ROIs according to the AAL atlas. The size of each node is proportional to its graph clustering coefficient, i.e. the bigger node, demonstrates higher activity in eMCI subjects in the corresponding ROI. Similar to nodes, the size of each edge is also proportional to the correlation between two ROI's. In addition, the edges are also color coded in a way that the green edges shows the positive edges in D and the orange edges shows the negative edges in D. In this manner, the green edges demonstrates decreasing in activity between the corresponding nodes in eMCI subjects and vice versa, the orange edges shows increasing activity between corresponding ROIs in the eMCI subjects.

As it can be seen in the difference graph, the big nodes i.e. ROIs with higher activities does not necessarily establish strong connections with the other nodes. As an obvious example, higher activities in Lingual gyrus(ROI index: 47,48)[24], [25], Calcarine sulcus(ROI index: 43, 44)[26], [27], Supplementary motor area(ROI index: 19,20)[27], [28] and Temporal_mid_L(ROI index: 85)[29] are easily detectable. The majority of ROIs located in frontal lobe also shows rather high activities comparing to normal subjects[31], [30].

Similar to the nodes, strong edge between two ROIs does not necessarily requires the nodes to be highly active in eMCI. Although a strong edge does indicate high activities and functional connectivity between the two corresponding ROIs. The difference Graph shows significant increase in



		min	max	std.
ACC	HOFSE	**0.76**	**0.86**	0.250
	FON	0.70	0.84	0.390
	SICE	0.68	0.85	0.434
FSC	HOFSE	**0.77**	**0.87**	0.254
	FON	0.69	0.85	0.418
	SICE	0.63	0.86	0.487
BAC	HOFSE	**0.76**	**0.86**	0.260
	FON	0.70	0.83	0.384
	SICE	0.69	0.85	0.430
SEN	HOFSE	**0.68**	**0.93**	0.659
	FON	0.61	0.88	0.741
	SICE	0.52	0.88	0.837
YI	HOFSE	**0.53**	**0.72**	0.520
	FON	0.40	0.67	0.768
	SICE	0.39	0.70	0.860

(f)

Fig. 3. sdsss

connectivity between Rectus(ROI index: 28, 27 in Frontal lobe) and Parietal_Sup_R(ROI index: 60 in Parietal lobe), Frontal_Inf_Orb_R(ROI index: 16 in Frontal lobe) and Cingulum_Ant(ROI index: 31,32 in Limbic lobe), Insula_L, Temporal_Pole_Sup_L(ROI index: 29,83 in Limbic lobe) and Pallidum_R, Caudate_R(ROI index: 29,83 in Sub Cortical Grey Nuclei). It can also be seen that within activities in frontal lobe also increased in patients with eMCI. There is a decrease in connectivity between Amygdala_L(ROI index: 41 in Sub Cortical Grey Nuclei) with Frontal_Mid_Orb_R(ROI index: 10 in Sub Frontal lobe) and ParaHippocampal_L(ROI index: 39 in Sub Limbic lobe). The connectivity between Heschl_L(ROI index: 79 in Temporal lobe) and two ROIs Temporal_Mid_R(ROI index: 86 also in Temporal lobe) and Occipital_Inf_R(ROI index: 54 in Occipital lobe) also decreased in eMCI.

Regarding the Cerebellum and Vermis: In fMRI data analysis and especially in Alzheimer's disease studies, ROIs within the Cerebellum and Vermis are usually excluded[ref] since their role was regarded as insignificant. Recent studies have shown that the traditional assumption that Cerebral area is essential only for the coordination of voluntary motor activity and motor learning is not valid and indicates the significant role of cerebellum in nervous system function, cognition and emotion[32].

As it can be seen in the difference graph that we obtained, ROIs within Cerebellum and Vermis are highly active and both their intra and inter connections are noticeable. There is an increasing functional connectivity between the Limbic lobe especially Hippocampus_R, Temporal_Pole_Mid(ROI index: 38,87,88) and Cerebral areas in eMCI patients. Also, the connectivity between Occipital lobe, especially Occipital_mid_R(ROI index: 52), the Frontal lobe, especially in Frontal_mid_orb(ROI index: 9,10) and Cerebral areas seems to decrease in patients with eMCI.

V. Conclusion

The majority of classification techniques uses the vectorized version of data as the input of the discriminant function. As the number of the dimensions grows, i.e. when data is naturally a high order tensor, this vectorization become problematic and it would highly affect the performance. Taking advantage of the techniques designed for the tensors, we have developed a framework for fMRI data analysis in which the following objectives:

1) Dimension Reduction
2) Classification

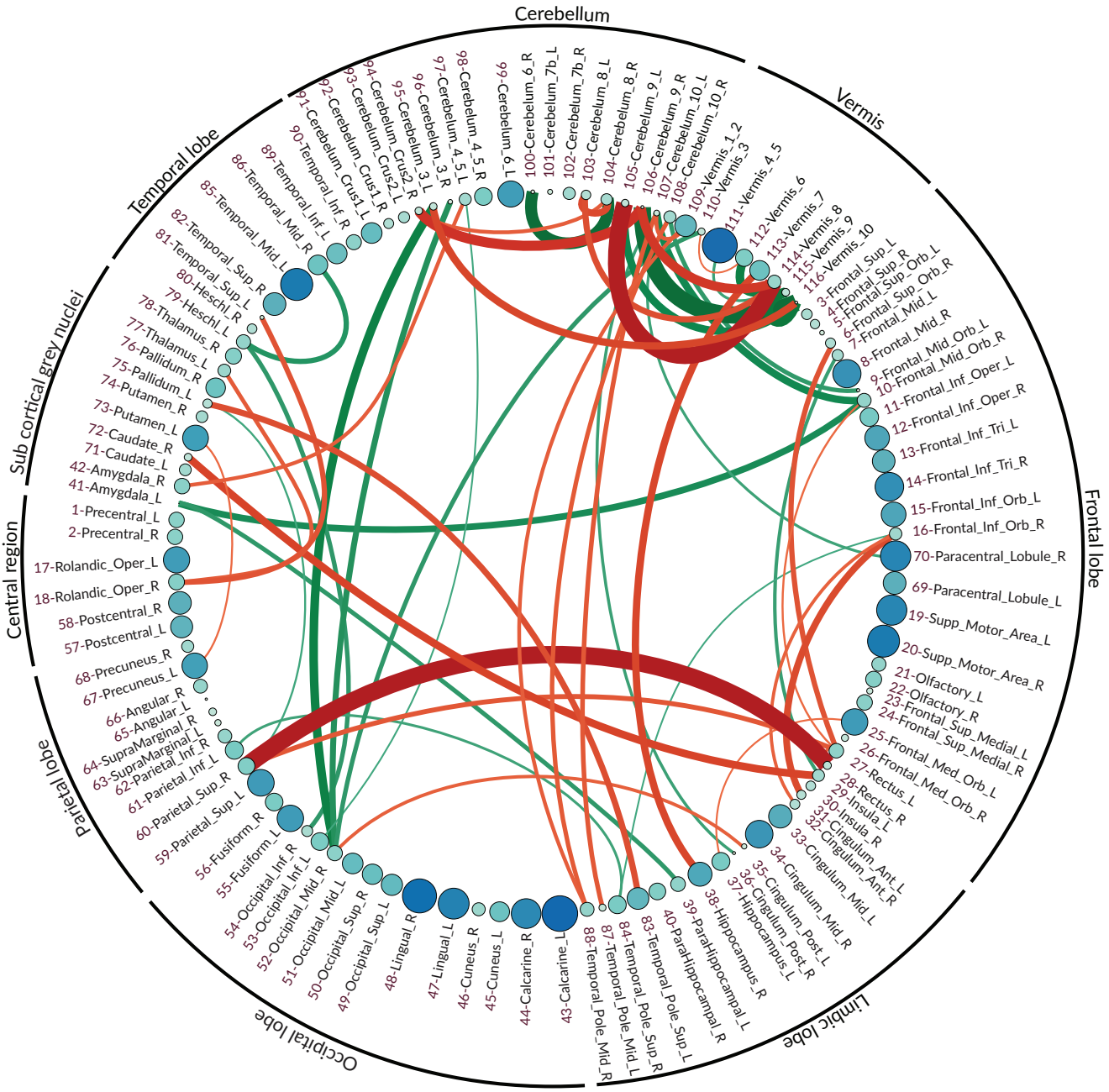


Fig. 4. The difference graph. This graph is obtained via subtracting the functional connectivity of eMCI subjects from normal subjects. Each circle represents a ROI in AAL atlas and the color and size of each circle is proportional to the graph clustering coefficient of the difference graph. red = more activity in EMCI, green: less.

3) Obtaining Functional Connectivity networn

are achieved via a singe **H**igh **O**rder **SVD** of the input tensors.

Extensive studies on the rs-fMRI provided by ADNI shows the superiority of the proposed framework in both classification and functional connectivity. The obtained FC network not only acknowledge the previous discovered connections but also reveals new connectivity patterns previously unknown. The framework proposed in this paper can be easily extended to other studies involved with high order data.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] Caselli, Richard J., et al. "Longitudinal changes in cognition and behavior in asymptomatic carriers of the APOE e4 allele." Neurology 62.11 (2004): 1990-1995.

[2] Brookmeyer, Ron, et al. "Forecasting the global burden of Alzheimers disease." Alzheimer's & dementia: the journal of the Alzheimer's Association 3.3 (2007): 186-191.

[3] Musha, Toshimitsu, et al. "EEG markers for characterizing anomalous activities of cerebral neurons in NAT (neuronal activity topography) method." IEEE Transactions on Biomedical Engineering 60.8 (2013): 2332-2338.

[4] Gould, R. L., et al. "Brain mechanisms of successful compensation during learning in Alzheimer disease." Neurology 67.6 (2006): 1011-1017.

[5] Richiardi, Jonas, et al. "Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity." Neuroimage 62.3 (2012): 2021-2033.
[6] Yang, Xue, et al. "Evaluation of statistical inference on empirical resting state fMRI." IEEE Transactions on Biomedical Engineering 61.4 (2014): 1091-1099.
[7] R. Graaf and K. Kevin. Methods and apparatus for compensating eld inhomogeneities in magnetic resonance studies. US Patent No. 8035387, 2011.
[8] Zhang, Xiaowei, et al. "Resting-state whole-brain functional connectivity networks for mci classification using l2-regularized logistic regression." IEEE transactions on nanobioscience 14.2 (2015): 237-247.
[9] Stanley, Matthew Lawrence, et al. "Defining nodes in complex brain networks." Frontiers in computational neuroscience 7 (2013): 169.
[10] Jie, Biao, et al. "Integration of network topological and connectivity properties for neuroimaging classification." IEEE transactions on biomedical engineering 61.2 (2014): 576-589.
[11] Wee, Chong-Yaw, et al. "Resting-state multi-spectrum functional connectivity networks for identification of MCI patients." PloS one 7.5 (2012): e37828.
[12] Tibshirani, Robert, et al. "Sparsity and smoothness via the fused lasso." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67.1 (2005): 91-108.
[13] Wright, John, et al. "Robust face recognition via sparse representation." IEEE transactions on pattern analysis and machine intelligence 31.2 (2009): 210-227.
[14] Zhang, Jianjia, et al. "Functional brain network classification with compact representation of SICE matrices." IEEE Transactions on Biomedical Engineering 62.6 (2015): 1623-1634.
[15] Huang, Shuai, et al. "Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation." NeuroImage 50.3 (2010): 935-949.
[16] Allen, Elena A., et al. "Tracking whole-brain connectivity dynamics in the resting state." Cerebral cortex 24.3 (2014): 663-676.
[17] Damaraju, Eswar, et al. "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia." NeuroImage: Clinical 5 (2014): 298-308.
[18] Hutchison, R. Matthew, et al. "Dynamic functional connectivity: promise, issues, and interpretations." Neuroimage 80 (2013): 360-378.
[19] Leonardi, Nora, et al. "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest." NeuroImage 83 (2013): 937-950.
[20] Leonardi, Nora, et al. "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest." NeuroImage 83 (2013): 937-950.
[21] Nordberg, Agneta. "PET imaging of amyloid in Alzheimer's disease." The lancet neurology 3.9 (2004): 519-527.
[22] Jeong, Jaeseung. "EEG dynamics in patients with Alzheimer's disease." Clinical neurophysiology 115.7 (2004): 1490-1505.
[23] Jeong, Jaeseung. "EEG dynamics in patients with Alzheimer's disease." Clinical neurophysiology 115.7 (2004): 1490-1505.
[24] Golby, Alexandra, et al. "Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory." Brain 128.4 (2005): 773-787.
[25] He, Yong, et al. "Regional coherence changes in the early stages of Alzheimers disease: a combined structural and resting-state functional MRI study." Neuroimage 35.2 (2007): 488-500.
[26] Bakkour, Akram, et al. "The effects of aging and Alzheimer's disease on cerebral cortical anatomy: specificity and differential relationships with cognition." Neuroimage 76 (2013): 332-344.
[27] Brewer, Alyssa A., and Brian Barton. "Visual cortex in aging and Alzheimer's disease: changes in visual field maps and population receptive fields." Frontiers in psychology 5 (2014): 74.
[28] Jacobsen, Jrn-Henrik, et al. "Why musical memory can be preserved in advanced Alzheimers disease." Brain 138.8 (2015): 2438-2450.
[29] Kosicek, Marko, and Silva Hecimovic. "Phospholipids and Alzheimers disease: alterations, mechanisms and potential biomarkers." International journal of molecular sciences 14.1 (2013): 1310-1322.
[30] Salvatore, Christian, et al. "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach." Frontiers in neuroscience 9 (2015): 307.
[31] Gould, R. L., et al. "Brain mechanisms of successful compensation during learning in Alzheimer disease." Neurology 67.6 (2006): 1011-1017.
[32] Jacobs, Heidi IL, et al. "The cerebellum in Alzheimers disease: evaluating its role in cognitive decline." Brain 141.1 (2017): 37-47.
[33] N. Leonardi et al., Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest, NeuroImage, vol. 83, pp. 937950, 2013.
[34]
[35]

PLACE PHOTO HERE

Mansoor Rezghi

PLACE PHOTO HERE

Ali Noroozi