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, achieving both above goals requires advanced data analysis tools such as machine learning techniques.

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 $\bar{X}_i$ 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, there are two main criticisms towards using FCs. There are variety of methods such as Pairwise Pearsons correlation coefficient [10], [11], sparse representation [10],

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.

[12], [13] and Sparse Inverse Covariance Estimation (SICE) to obtain an FC which their resulting network differs drastically. 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 use FC networks for finding common brain alterations caused by a disease, an FC is usually extracted for each sample, and then a representative is obtained from the combination of these FCs. There are two main issues associated with this approach:

- There are variety of methods for obtaining the FC matrix which heavily affect the quality of the network. Pairwise Pearsons correlation coefficient [10], [11], sparse representation [10], [12], [13] and Sparse Inverse Covariance Estimation (SICE) are of the popular ones [14], [15]. 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.

Since the common patterns are not considered in the first place, additional data such as noise or outliers could heavily affect the results.

Since FCs have been proved to highlight alteration within brain connectivity patterns, they became the favorite candidate for classification purposes. 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. Two issues are generally involved with this strategy:

- **Obtaining the proper FC:**
- **Extracting key features from FCs:** Since the majority of classification techniques such as SVM or KNN uses vectors as their input, the obtained FC should be vectorized which leads to serious problems such as curse of dimensionality or over-fitting. Moreover, vectorization may destroy some useful property embedded within data. For example SICE matrix is naturally SPD, but this property gets destroyed in the vectorization. Several dimension reduction methods have been proposed in order to overcome this issue. As an example [] uses the SPD property of SICE in order to reduce it's dimensionality using kernel PCA and obtaining a much smaller set of features for each sample.

in this paper we show that tensor representation is the natural form to present fMRI images. by this representation, we would be able to work with time and region features seperately in the same time. also with this representation we an easy and affordable multilinear dimension reduction that works with time and region features seperately this will be done by high order singular value decomposition of of the normal and eMCI subjects.

also we show that by this representation and using the obtained HOSVD, dimension reduction step we can obtain tensor based discriminat features for classification. this discriminant method furthermore simplicity is very effective. here we show that by this framework the test data also could have effect in construction of discriminant function which is impossible in well known classification methods like SVM, logistic regression and etc.

by this approach the construction of FC feature that was time consuming is ignored and we represent an other method that at the same time with simple structure and low computation cost gives very high quality results with the other state of the art methods.

in fMRI data, one other feature is the fnctional connectivity of normal and abnormal subjects. as the third innovation of this paper we showed that based on hosvd of normal and eMCI patients and effective and high quality functional connectivity network can be extracted for both classes. we will show that some connectivities that are discovered via these FCs are well studied in recent clinical documents and thus . . .

Distinguishing two rs-fMRI volumes in the machine learning viewpoint can be considered as a two class classification problem. Some methods use the vectorized version of samples ie, $x_i = vec(X_i) \in \mathbb{R}^T$ as inputs of the learning process. However as we will discuss later, this vectorization may cause the loss of some important properties within the structure of data, and because of that the dominant approaches tend to find different set of features for each sample. These methods construct a $region \times region$ matrix $\bar{X}_i \in \mathbb{R}^{R\times R}$ which is often called functional connectivity (FC) matrix and indicates the between correlation of ROIs. Since in most fMRI samples the number of timepoints is larger that the number of ROIs i.e. $T > R$, substituting X_i with $\bar{X}_i$ reduce the number of features and the effects of the *curse of dimensionality*, also it has been shown that using FC can enhance small between region connections hence turns FC into a powerful tool for rs-fMRI data analysis. For correlation computation, different methods have been explored, among which the pairwise Pearsons correlation coefficient [10], [11], sparse representation [10], [12], [13] and Sparse Inverse Covariance Estimation (SICE) are of the popular ones [14], [15]. 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. On the other hand, computing the correlation based on the entire time series of RS-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, some methods have been proposed that consider the time as a non-stationery property. As for an example [] uses sliding-window technique in order to divide the time series into several smaller parts and then calculate the between correlation of these parts.[][] although

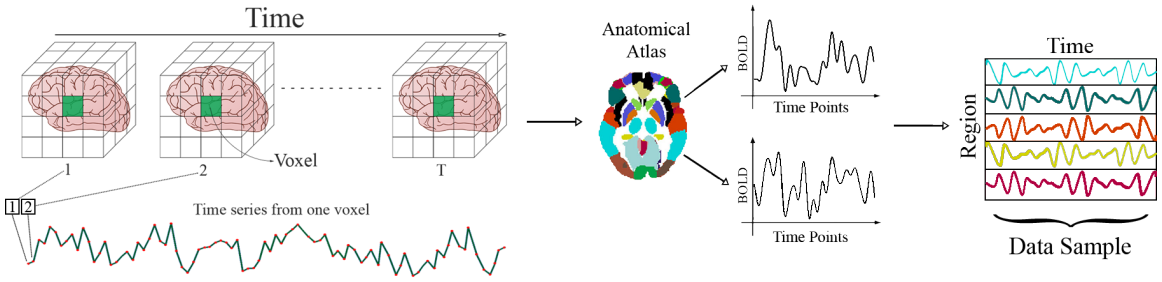


Fig. 1. Simulation results for the network.

these methods show promissing classification results, they are not computational favorable and since the aspect of region is not preserved in the final model, the relation between the high-order FC and the real world FC is somewhat ambigues. After calculating the FC, conventional classifiers like SVM or KNN is deployed in order to classify the vectorized FC for each sample [20].

Although FC grants us certain useful structural properties such as being SPD and having relatively low dimensionality, using it in the classification process has its own drawbacks such as calculating the proper FC and extracting key features from it. Considering these issues we have come up with a framework that eliminate the need for finding the FC in the classification process. Also the proposed framework works directly with the sample matrix and preserve its structural integrity by avoiding the vectorization using novel multilinear techniques. (more about vectorization?)

In this paper we claim that by considering $X_i \in \mathbb{R}^{T\times R}$ as the sample, the dataset could be considered as an order three tensor. Using this representation we have shown that by extremely lower computational cost compared with other methods, and without the need to construct FC, a framework can be designed that only by one High Order Singular Value Decomposition (HOSVD) for each class could give us the desired results. This framework which we call it **TBNA** (High Order Feature Selection and Extraction) does three main tasks:

- **Dimension Reduction:** The proposed frameworks allows us to select key features of each mode (i.e. *Time*, *Region* and *sample*) of the constructed tensor seperately and differently. This feature extraction would cause a massive reduction of the dimensions.
- **Classification:** Based on the properties of the HOSVD, we have eliminated the vectorization at any step and hence assure that no information is lost in the vectorization process. Moreover we have formed a novel classification enhancement technique that allows the test samples to cooperate in the training phase. This technique increases the tendency of the test sample to it's true class without forcing any apriory knowledge about it's label to the trainer.
- **Functional Connectivity Network Construction:** Being able to see the class as a whole ables us to derive a Functional Connectivity network for each class. Such FCs allows us to investigate on shared properties within each class such as functional abnormalities of some ROIs or the connection between them.

change

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

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

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 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. Preliminaries

In this section, we briefly introduce some preliminary knowledge from tensor algebra. For a deeper introduction to the concepts and terminology, we refer to [17]. Before proceeding, we introduce some basic notations that will be used throughout this paper. Tensors (*i.e.*, multidimensional arrays) are denoted by calligraphic letters ($\mathcal{A}, \mathcal{B}, \mathcal{C}, \cdots$), matrices by boldface capital letters ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \cdots$), vectors by boldface lowercase letters ($\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \cdots$), and scalars by lowercase letters ($a, b, c, \cdots$). The columns of a matrix are denoted by boldface lower letters with a subscript, *e.g.*, $\boldsymbol{a}_i$ is the ith column of matrix $\boldsymbol{A}$. The elements of a matrix or a tensor are denoted by lowercase letters with subscripts, *i.e.*, the $(i_1, \cdots, i_n)$ element of an n-th order tensor $\mathcal{A}$ is denoted by $a_{i_1,\cdots,i_n}$. In formulas we will sometimes use MATLAB-style notations. For instance, we define the mode-1 fibers of a 3-tensor $\mathcal{A}$ to be the column vectors $\mathcal{A}(:, j, k)$.

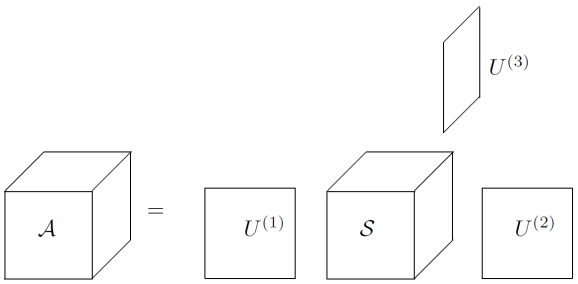


Fig. 2. Simulation results for the network.

A. Basic Tensor Operations

Let $\mathbb{R}^{I_1 \times \cdots \times I_M}$ be endowed with the usual Euclidean geometry and $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$. where I_k's are positive integers; the vector space $\mathbb{R}^{I_1 \times \cdots \times I_M}$ has dimension $I_1 \times \cdots \times I_M$, and $\mathcal{A}$ is a tensor in this space. If $M = 2$, $\mathcal{A}$ would be a matrix in $\mathbb{R}^2$.

A fiber is a subtensor, where all indices but one are fixed. For example mode-2 fibers of $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, have following form
$$\mathcal{T}(i_1, :, i_3) \in \mathbb{R}^{I_2}$$

Similarly, we define *slices* of a tensor to be the sub-tensors obtained by fixing the index in one mode, *e.g.*
$$\mathcal{T}(:, i_2, :) \in \mathbb{R}^{I_1 \times I_3}$$
is an slice of $\mathcal{T}$ in mode-2.

The mode-n product of an order-M tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ by a matrix $X \in \mathbb{R}^{K \times I_n}$ is defined as

$$\mathcal{B} = (X)_n \cdot A \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times K \times I_{n+1} \times \cdots \times I_M} \tag{2}$$

where,
$$b_{i_1,\cdots,i_n} = \sum_{l=1}^{I_n} x_{i_n,l} a_{i_1,\cdots,i_{n-1},l,I_{n+1},\cdots I_M}.$$

This means that all mode-n fibers of $\mathcal{A}$ are multiplied by the matrix X. The notation (2) was suggested by Lim [23]. An alternative notation was earlier given in [24]. $(X)_n \cdot A$ is the same as $\mathcal{A} \times_n X$ in that system.

The Frobenius norm of the order-M tensor $\mathcal{A}$ can also be defined as

$$\|\mathcal{A}\| = \left(\sum_{i_1,\cdots,i_m} a^2_{i_1,\cdots i_m} \right)^{1/2}.$$

B. Matricization

Sometimes it is useful to rearrange the tensor into a matrix. A tensor can be *matricized*[2] in many different ways[26], [27]. In special case rearranging the mode-n fibers of a tensor $\mathcal{A}$ is an operation where the mode-n fibers of $\mathcal{A}$ are aligned as the columns of a matrix, denoted $K^{(n)}$[26][27]. The mode-n matricization $A^{(n;1,\cdots n-1,n+1,\cdots,M)}$ is a map between the element $a_{i_1,\cdots i_n}$ of $\mathcal{A}$ to (i_n, j) of the matrix $A^{(n)}$ where

$$j = 1 + \sum_{k=1,k\neq n}^{M} (i_k - 1)J_k, \quad J_k = \prod_{m=1,m\neq n}^{k-1} I_m$$

C. Higher order singular value decomposition

Higher order singular value decomposition (HOSVD) is one common extension of singular value decomposition to the tensors [24]. Using HOSVD, every order-M tensor $\mathcal{A} \in \mathbb{R}^{I_1\times\cdots I_M}$ can be decomposed as

$$\mathcal{A} = \left(U_{(1)},\cdots,U_{(M)}\right)\cdot\mathcal{S}, \tag{3}$$

where orthogonal matrices $U^{(i)}$ are singular matrices of tensor $\mathcal{A}$. Here, $U^{(i)}$ is the left singular matrix of $\mathcal{A}^{(i)}$. The core tensor $\mathcal{S}$ is a real tensor of the same dimensions as $\mathcal{A}$ and

$$\mathcal{S} = \left(U^T_{(1)},\cdots,U^T_{(M)}\right)\cdot\mathcal{A}. \tag{4}$$

Although this core tensor is not diagonal as in the case of SVD of matrices, but satisfies the following conditions:

- Any two different slices along the same mode are orthogonal. This property of core tensor S is named as all orthogonality.
- The values $s^k_j = \|\mathcal{S}(:,\cdots,:,j,:,\cdots,:)\|$, where j is in the kth mode of $\mathcal{S}$, are named mode-k singular values of $\mathcal{A}$. It can be shown that for every k

$$s^k_1 \geq s^k_2 \geq \cdots \geq s^k_n \geq 0, \quad k = 0,\cdots,M. \tag{5}$$

  are equal to the singular values of the matrix $A^{(k)}$. This means that the norms of the slices along every mode are ordered.

The ordering property (3) demonstrates that, in the same way as matrices, singular values measure the 'energy' of the tensor. So, it is easy to see that the energy of core tensor $\mathcal{S}$ focused on the elements of $\mathcal{S}$ with small indices, especially in $\mathcal{S}(1,1,\cdots,1)$. HOSVD for a 3-tensor is illustrated in figure (2).

The computation of the HOSVD of a tensor $\mathcal{A} \in \mathbb{R}^{I_1\times\cdots\times I_M}$ can be done by separately computing the orthogonal matrices $U_{(1)}, U_{(2)},\cdots,U_{(M)}$, as the left singular matrices of $A^{(m)}, m = 1, 2,\cdots M$. *i.e.*:

1) Compute the SVDs:
$$A^{(i)} = U^{(i)}S^{(i)}(V^{(i)})^T,$$
without forming the $V^{(i)}$s explicitly
2) Compute the core tensor: $\mathcal{S} = \left(U^T_{(1)},\cdots,U^T_{(M)}\right)\cdot\mathcal{A}$

IV. Proposed fMRI analysis Framework

By using the properties of HOSVD, the proposed framework allows us to reduce the dimensionality of data, create a discriminant function and enhance it and also calculate an FC network. These four features are discussed in full detail in the following sections:

A. Dimension Reduction and Feature Extraction

Let $x \in \mathbb{R}^n$ be a sample data, in this case, $y = U^T_k x \in \mathbb{R}^k$ is a linear feature extraction and $\bar{x} = U_k U^T_k x$ denotes an approximation of x in space spanned by base columns $\{u_1,\cdots,u_k\}$. In fact y is in a *latent* space spanned by $\{u_1,\cdots,u_k\}$, and $\bar{x}$ is the filtered version of x that in it only the effects of basis vectors $\{u_1,\cdots,u_k\}$ are preserved. So by choosing appropriate values for $\{u_1,\cdots,u_k\}$ we could remove the unnecessary parts of x in the original space or in the feature space. These unnecessary parts can be accounted for noise or outliers in the data.

When samples are not vectors, two approaches can be followed, the first one is to simply vactorize the data and use linear feature extractors as described above. 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.

Due to the multilinear nature of our data, it seems natural to take advantage of multilinear techniques in order to reach better performances. As we described before, each sample is considered to be $X \in \mathbb{R}^{T\times R}$ where x_{ij} is the ith observation of the jth node and thus the ith class can be considered as $\mathcal{X}^{(i)} \in \mathbb{R}^{T\times R\times S_i}$ where $\mathcal{X}(:,:,s)$ is the s'th sample in the ith class. figure * shows this structure. Observing the dataset as a multilinear entity would allow us to treat each mode differently. Since each mode represents a different feature(e.g. *time* or *region*) appropriate dimension reduction methods can be deployed separately on each mode.

As an example of a multilinear feature extraction approach let: $\mathcal{X} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ then

$$\mathbb{R}^{I_1\times\cdots I_{i-1}\times k_i\times I_{i+1}\times\cdots\times I_N} \ni \mathcal{Y} = (U^T_{k_i})\cdot\mathcal{X} \tag{6}$$

is the dimension reduced version of $\mathcal{X}$ in the ith mode. We have also

$$\mathbb{R}^{I_1\times\cdots\times I_N} \ni \bar{\mathcal{X}} = \left(U_{k_i}(U_{k_i})^T\right)\cdot\mathcal{X}$$

that is the filtered version of $\mathcal{X}$ in the ith mode according to the basis matrices $\{U_{k_1},\cdots,U_{k_i}\}$. So based on the properties

[2]Alternative terms are *unfolding* [9] or *flattening* [13].

of each feature, both Y and $\bar{X}$ can be used together in order to extract the features.

A key aspect in dimension reduction and feature extraction is finding the appropriate projection matrices $\{U_1, \cdots, U_{k_i}\}$. When data samples are vectors, methods like Non-negative Matrix Factorization(NMF), Principal Component Analysis(PCA) and Singular Value Decomposition(SVD) are traditionally used to obtain such basis vectors. As the number of dimensions grows and therefore the dataset becomes a tensor, multilinear methods (such as MPCA or HOSVD) should be deployed in order to extract the appropriate basis matrices. As we mentioned before, HOSVD is a tensor decomposition technique that provides us with basis matrices for each mode of a tensor. If we denote the ith class with $\mathcal{X}^{(i)} \in \mathbb{R}^{T\times R\times S_i}$ in which T, R and S_i denotes the size of the *Time*, *Region* and *Sample* modes respectively, then it's HOSVD would be:

$$\mathcal{X}^{(i)} = \left(U_1^{(i)}, U_2^{(i)}, U_3^{(i)}\right) \cdot \mathcal{S}^{(i)}$$

where for $k = 1, 2, 3$, U_k is the mode-k singular matrix of $\mathcal{X}^{(i)}$, and $S^{(i)}$ is the core tensor. As it was mentioned in (III) U_1 is a base of all mode-1 fibers $\mathcal{X}^{(i)}(:, i, j)$ which indicates the behavior of ith region of the jth sample in all times. Also due to the properties of HOSVD inherited from svd, the first columns of U_i has more ability in construction of main parts of ith fiber and on the other hand, the last columns of U_i in each mode, are corresponding to noise parts of these fibers and hence projecting the data in mode-i into space of $U_k = [u_1, \cdots, u_{k_i}]$ is a suitable dimension reduction. For example,

$$\mathbb{R}^{k\times R\times S} \ni \mathcal{X}_f = (U_{k_1})_1^{\mathsf{T}} \cdot \mathcal{X}$$

is a dimension reduction in the first mode of $\mathcal{X}$. Note that $\mathcal{X}_f$ is in the feature space induced by $(U_{k_1})^{\mathsf{T}}$.

let $X^{(i)} \in \mathbb{R}^{T\times R\times S_i}$ denotes the ith class. using HOSVD on each class, using the mensioned DR method, the ith class becomes:

$$\bar{\mathcal{X}}^{(i)} = \left(U_1^{(i)}(1:k_1^i,:)^{\mathsf{T}}, U_2^{(i)}(1:k_2^i,:)\right) \cdot \mathcal{X}^{(i)}$$

B. Discriminant Function

In previous section we proposed a novel framework for feature extraction and dimension reduction of fMRI data. We have also introduced a classification technique to distinguish eMCI and Normal subjects. In our case, classifiers like SVM, knn and other convenient classifiers or their multilinear versions such as STM could be applied. Since even in atlas-based approach the number of features are relatively high, the computation cost of the mentioned classifiers are also high and would probably cause the over-fitting issue.

In this section, using the computed components of HOSVD from the last section, and without extra computational costs, we have designed a discrimination process. Let $\mathcal{X} = \left(U_{(1)}^i, U_{(2)}^i, U_{(3)}^i\right)$ denotes the HOSVD of the ith class, its correspondent dimension reduced tensor would be

$$\mathbb{R}^{k_1^i\times k_2^i\times S} \ni \bar{\mathcal{X}} = \left(U_{k_1^i}^{(i)^{\mathsf{T}}}, U_{k_2^i}^{(i)^{\mathsf{T}}}\right)_{1,2} \cdot \mathcal{X} \tag{7}$$

where the $U_{k_1^i}^{(i)}$ and $U_{k_2^i}^{(i)}$ are the first k_1^i and k_2^i vectors of the singular matrix $U_1^{(i)}$ and $U_2^{(i)}$ respectively. by (7) the above equation can be written as

$$\begin{aligned}\bar{\mathcal{X}} &= \left(\begin{bmatrix} I_{k_1^i} & 0\end{bmatrix}, \begin{bmatrix} I_{k_2^i} & 0\end{bmatrix}, U_3^{(i)}\right) \cdot \mathcal{S} \\ &= \left(U_3^{(i)}\right)_3 \cdot \mathcal{S}(1:k_1^i, 1:k_2^i, :)\end{aligned}$$

in the next step we develope our multilinear classification method based on HOSVD. from 7 it is easy to show 8.

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,:,k) &= \sum_{k'=1}^{S_i} U_3^{(1)}(k,k') \cdot \mathcal{S}(1:k_1^i, 1:k_2^i, k') \\ &= \left(U_3^{(i)}(k,:)\right)_3 \cdot \mathcal{S}(1:k_1^i, 1:k_2^i, :)\end{aligned}$$

by this equation and by increasing k' from HOSVD we know that $\mathcal{S}(1:k_1^i, 1:k_2^i, k')$ becomes smaller also the $U_3^{(1)}(k,k')$ for large k' show ossilated behaviour.

it is clear that the kth sample from the ith class in the new feature space can be written as $\bar{\mathcal{X}}^{(i)}(:,:,k)$, thus we have:

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,:,k) &= \sum_{k'}^{N_i} U_3^{(i)}(k,k')\bar{\mathcal{S}}^{(i)}(:,:,k') \\ &= \left(U_3^{(i)}(k,:)\right)_3 \cdot \bar{\mathcal{S}}^{(i)}\end{aligned} \tag{8}$$

in which

$$\bar{\mathcal{S}}^{(i)} = \mathcal{S}(1:k_1^i, 1:k_2^i, :)$$

this means that all elements in the ith class in the new domain can be reconstructed by the slices of $\bar{\mathcal{S}}^{(i)}$. Also by the properties of the core tensor we know that the first slices of $\bar{\mathcal{S}}^{(i)}$ Has higher value and since they are multiplied by the first columns of singular matrices, they have higher effects in the reconstruction process so by truncating the $\bar{\mathcal{S}}^{(i)}$ tensor we would have:

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,:,k) &\approx \sum_{k'}^{l} U_3^{(i)}(k,k')\bar{\mathcal{S}}^{(i)}(:,:,k') \\ &= \left(U_3^{(i)}(k,1:l)\right)_3 \cdot \bar{\mathcal{S}}_l^{(i)}\end{aligned} \tag{9}$$

where

$$\bar{\mathcal{S}}_l^{(i)} = \mathcal{S}(1:k_1^i, 1:k_2^i, 1:l)$$

These relations helps us to design a novel discreminant function. Since every sample in the ith class could be reconstructed by $\bar{\mathcal{S}}_l^{(i)}$ to assign a test data $Z \in \mathbb{R}^{T\times R}$ to a class we can do the following process: First we project the Z matrix to the span of singular matrices of the two available classes:

$$\begin{aligned}\bar{Z}^{(i)} &= \left(U_{k_1^i}^{(i)^{\mathsf{T}}}, U_{k_2}^{(i)^{\mathsf{T}}}\right)_{1,2} \cdot Z \\ &= U_{k_1^i}^{(i)^{\mathsf{T}}} Z \;\; U_{k_2^i}^{(i)}\end{aligned} \tag{10}$$

in which $\bar{Z}^{(i)}$ denotes the test data Z after pruning with respect to the ith class. The coordinates of $\bar{Z}^{(i)}$ in the extracted basis of the ith class can be obtained via:

$$x^i = \arg\min_{\mathrm{x}} \left\| \bar{Z}^{(i)} - \sum_j x_j X_j^i \right\|$$

in which X_j^i is the jth frontal slice of $\bar{\mathcal{X}}^{(i)}$. The above minimization is equivalent to:

$$\min_x \left\| B^i x - z^i \right\|$$

in which $b_j^i = \text{vec}(A_j^i)$ and $z^i = \text{vec}(\bar{Z}^{(i)})$. Now we can argue that $\bar{Z}^{(i)}$ belongs to the class $c = \arg\min_i \{r_i\}_{i=1}^I$, in which

$$r_i = \left\| z^i - B^i x^i \right\| \tag{11}$$

C. Enhancement of the Discriminant Function

The designed framework allows us to involve the test subject in the training process without inducing any apriory knowledge about its label to the discriminant function. As it was discussed in the previous section, each sample in a class can ideally be reconstructed via the frontal slices of $\bar{\mathcal{S}}_l^{(i)}$ which itself is obtained via HOSVD. Based on this decomposition for the ith class $\mathcal{X}^{(i)}$ we can see that:

$$\mathcal{X}^{(i)} = (U_{k_3}) \cdot \left((U_{(1)}, U_{(2)})_{1,2} \cdot \bar{\mathcal{S}}_l^{(i)} \right) + \left(U_{\bar{k}_3} \right) \cdot \left((U_1, U_{(2)})_{1,2} \cdot \bar{\mathcal{S}}_{\bar{l}}^{(i)} \right). \tag{12}$$

in which

$$\bar{\mathcal{S}}_l^{(i)} = \mathcal{S}(1 : k_1^i, 1 : k_2^i, 1 : l)$$
$$\bar{\mathcal{S}}_{\bar{l}}^{(i)} = \mathcal{S}(k_1^i : end, k_2^i : end, l : end)$$

we have chosen the frontal slices of $\bar{\mathcal{S}}_l^{(i)}$ to be the basis matrices for the ith class since it contains the shared information between samples, and discarded $\bar{\mathcal{S}}_{\bar{l}}^{(i)}$ since it probably contains noise and other non-common information within individual samples.

In order to enhance the obtained basis for each class, the test data Z is appended to each class before the basis extraction process. Doing such has two possible effects: If Z share common patterns with other samples, it will mostly affect the first part of (12), i.e. $\bar{\mathcal{S}}_l^{(i)}$. The second possibility is that Z is not similar to other subjects and hence it will mostly effect the second part of (12), i.e. $\bar{\mathcal{S}}_{\bar{l}}^{(i)}$. This effect is due to the fact that by the properties of HOSVD, the common patterns manifest themselves in the first singular matrices and their corresponding basis matrices. Figure (3) shows a better demonstration of this effect. As it can be seen in this figure, when the test data is similar to the rest of the class the first singular values get heavily affected, on the other hand when the new data does not show similarity to the rest subjects in the class it affects the last singular values which will be discarded along with their corresponding basis matrices.

In order to better demonstrate the classification process it is summarized in algorithm (1).

D. Functional Connectivity

Functional connectivity simply means the relation between different ROIs. 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.

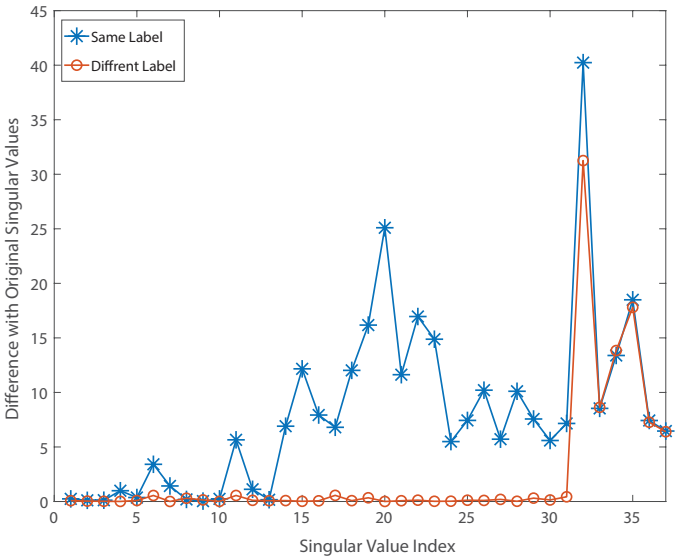


Fig. 3. The blue line with the star indicator, shows the difference between the mode-3 singular values of the original data and the data appended with a *same* label subject. The orange line indicated with circles, shows the difference between the mode-3 singular values of the original data and the data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

Data: $\mathcal{X}^{(1)}, \cdots, \mathcal{X}^{(I)} \in \mathbb{R}^{T \times R \times S}$, $Z \in \mathbb{R}^{T \times R}$
Result: Label of Z
for $i \leftarrow 1$ **to** I **do**
  Append the test data Z to tthe ith class
  Calculatethe *basis* $\bar{\mathcal{S}}_l^{(i)}$ of the ith class
  Calculate $\bar{Z}^{(i)}$ based on (10)
  Calculate and store r_i based on (11)
end
return $c = \arg\min_i \{r_i\}_{i=1}^I$

Algorithm 1: Classification by TBNA

The proposed framework allows us to obtain an FC network for each class. Although we do not use the FCs in the classification process, FC networks would allow further investigations on the brain activities. Let $\mathcal{X} \in \mathbb{R}^{T \times R \times S}$, finding the FC for this class means finding the relation between mode-2 slices of $\mathcal{X}$. 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.

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

$$\mathbb{R}^{k_1^i \times R \times k_3^i} \ni \bar{\mathcal{X}} = \left(U_{k_1^i}^{(i)^\mathsf{T}}, U_{k_3^i}^{(i)^\mathsf{T}} \right)_{1,3} \cdot \mathcal{X} \tag{13}$$

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

$$\begin{aligned}\bar{\mathcal{X}}^{(i)}(:,k,:) &= \sum_{k'}^{N_i} U_2^{(i)}(k,k')\bar{\mathcal{C}}^{(i)}(:,k',:) \\ &= \left(U_2^{(i)}(k,:)\right)_2 \cdot \bar{\mathcal{C}}^{(i)} \end{aligned} \tag{14}$$

in which

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

Thus the kth mode-2 slice of $\mathcal{X}$, i.e. the kth region, can be written as the linear combination of

in order to extract low dimensional features for each region, they were reconstructed via the obtained basis matrices and their coefficients were stored as the new representative for each region. In order to do such, each region should be transfered to the space of basis matrices, let the ith region be $B = \mathcal{X}(:,i,:)$ then the transfered version would be

$$\bar{B}^{(i)} = \left(U_{k_1^i}^{(i)\mathsf{T}}, U_{k_3}^{(i)\mathsf{T}}\right)_{1,3} \cdot B$$

After extracting low dimensional features, they were put together as the columns of a matrix and then Sparse Inverse Covariance of this matrix was calculated in order to find the functional connectivity.

V. EXPERIMENTAL STUDY

A. Data Preprocessing and Experimental Settings

Rs-fMRI data of 196 subjects were downloaded from the ADNI website[3]. 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].

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

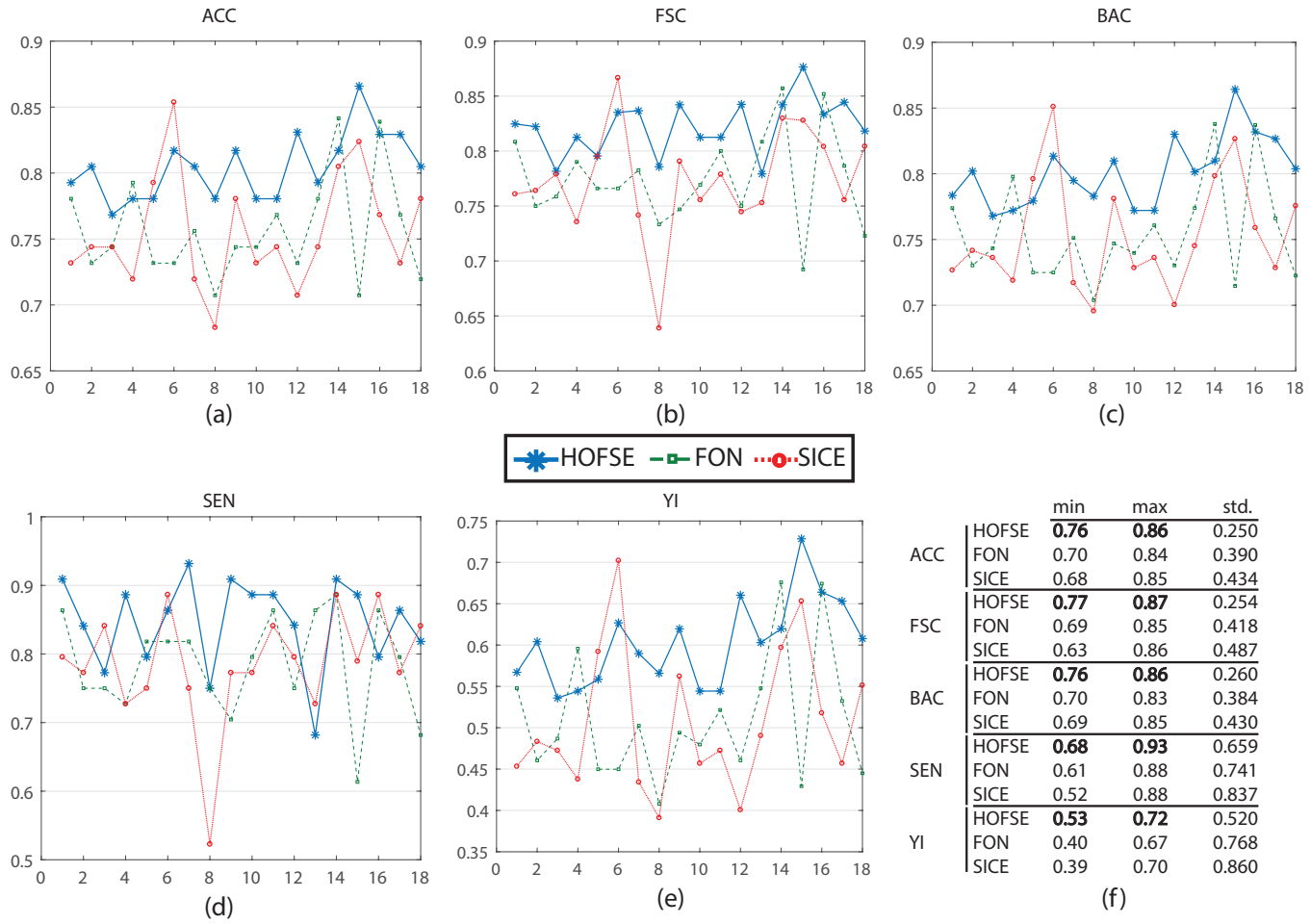
B. Classification

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 (4) 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



		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

Fig. 4. sdsss

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

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.

C. Functional connectivity Network

The functional connectivity networks of the Normal and eMCI classes was obtained via TBNA as it is described in (IV-D). 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(5). 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],

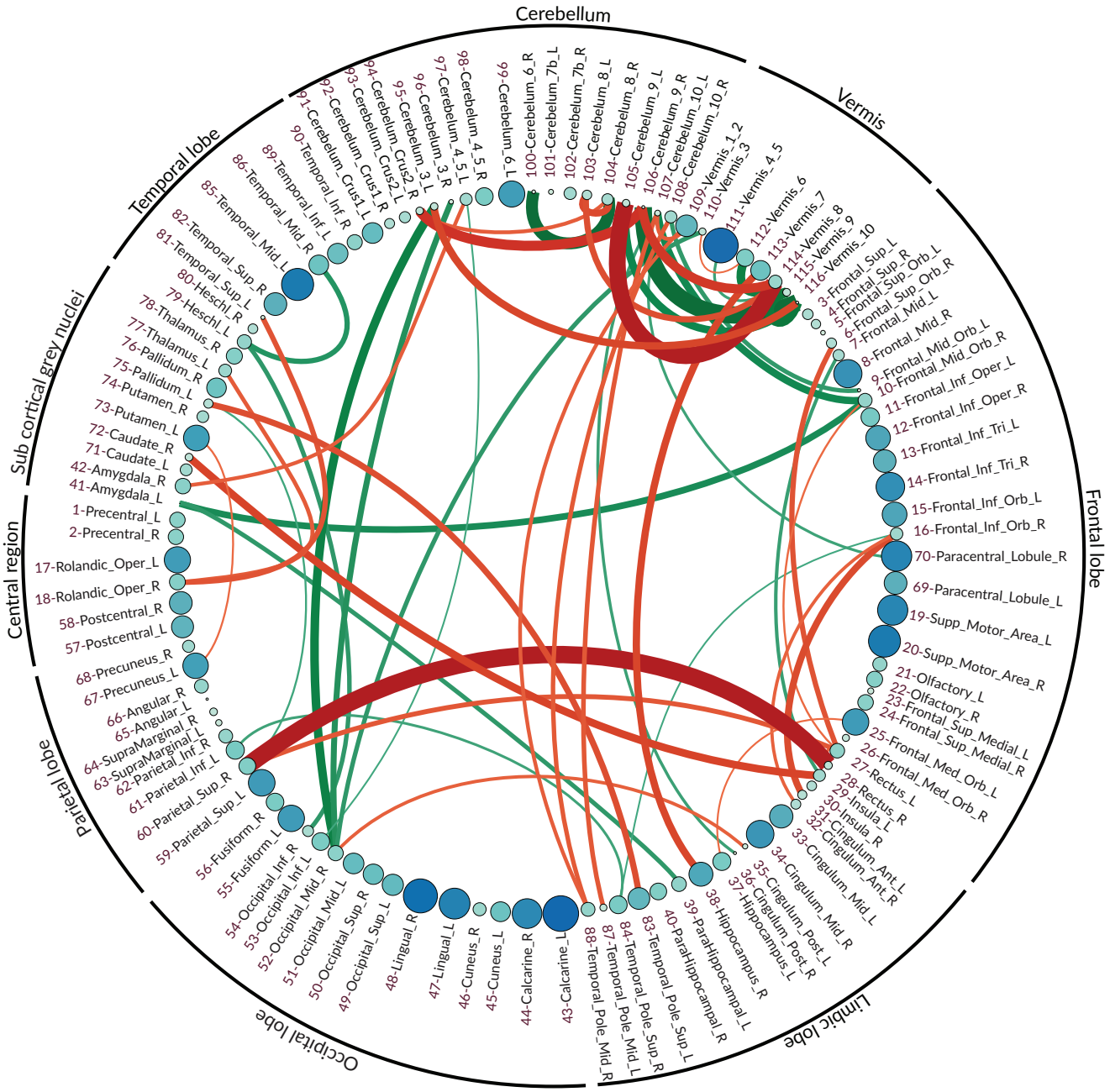


Fig. 5. The difference graph. This graph is obtained via subtracting the functional connectivity of eMCI subjects from normal subjects. Each circle represents a ROI in AAL atlas and the color and size of each circle is proportional to the graph clustering coefficient of the difference graph. red = more activity in EMCI, green: less.

[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 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.

VI. 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
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] de Vos, Frank, et al. "A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease." Neuroimage 167 (2018): 62-72.
[34] Dvornek, Nicha C., Pamela Ventola, and James S. Duncan. "Combining phenotypic and resting-state fMRI data for autism classification with recurrent neural networks." 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018.
[35]

Ali Noroozi

Mansoor Rezghi