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 a better understanding of the brain and its related disease such as Alzheimer's disease. Different methods have been deployed in order to discriminate Alzheimer's disease from normal ones which is a hard task, especially in early stages (eMCI) case. The majority of deployed techniques rely on constructing the functional connectivity (FC) for each person 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 multidimensional nature 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 a general FC pattern for normal and eMCI classes 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 and reveals novel connectivity patterns in Alzheimer's disease at its early stages.

Keywords:

1. Introduction

Alzheimers disease (AD) is a progressive neurodegenerative disorder with a long pre-morbid asymptomatic period which affects millions of elderly individuals worldwide[1]. 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)[16], Electroencephalography (EEG)[17] and functional Magnetic Resonance Imaging (fMRI)[18] 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 its stages in which detecting abnormalities within small brain regions is essential [5]. 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 activity of different brain regions[6]. 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 brain diseases like AD or Autism[27, 28].

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)[7]. 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 a 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, Autism, schizophrenia and etc. and more recently detecting patients with brain disorders using classification techniques [29, 30]. 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 [31].

A powerful tool that is commonly used in order to achieve aforementioned

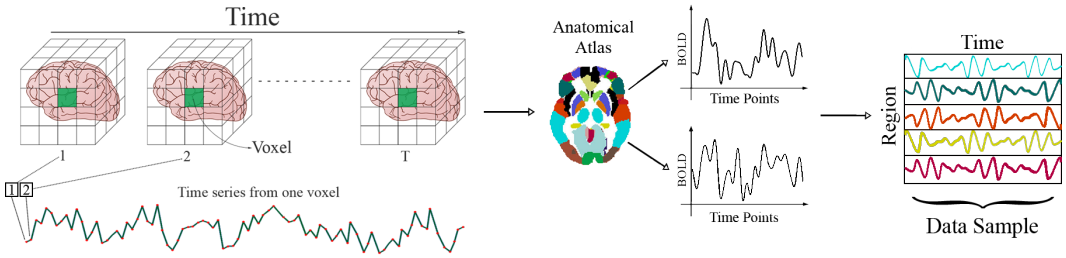


Figure 1: The process of extracting ROI time-series from the original 4D volume.

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 [32]. 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 the hippocampus and other brain regions, and MCI patients have been observed increased functional connectivity between the frontal lobe and other brain regions[5]. So, Finding an FC that highlights the patterns caused by a disease, i.e. a **General** functional connectivity, has been a common goal in the rs-fMRI study for a long time. Several approaches exist to find common patterns among different brain scans. Data-driven methods such as kernel-PCA or clustering techniques have been proposed for this task REF. But ultimately most of them rely on calculating a network for each volume which may overlook the role of noises or outliers within the data[47, 48].

In recent years FCs are also used as features in classification. So, instead of using X_i as the i^{th} sample, its corresponding FC i.e. $\bar{X}_i$ is used as a feature. Although FCs show promising results, they bring their own challenges. The computational cost of FC is usually high and also its quality massively affects the performance of the learning process. Also, Since the conventional classifiers like **S**upport **V**ector **M**achine(SVM) and or k-NN works on data in vector format, these matrix features should be vectorized in order be fed to these classifiers. This vectorization leads to high-dimensional vectors which produce poor performance due to the phenomena known as the Curse of Dimensionality. Alongside the curse of dimensionality, vectorization also destroys potential information that are embedded in the structure of data. This

problem has been studied especially in image data in which vectorization destroys the spatial relations within an image.Rezghi and S.Ahmadi paper

In this paper, based on high order tensor decomposition, we have created a framework in which the aforementioned goals i.e. finding a general FC and detecting a disorder via classification could be achieved via a single **H**igh **O**rder **S**ingular **V**alue **D**ecomposition (HOSVD) of each class. Here based on latent variables obtained by HOSVD a general representative pattern of FC for eMCI and Normal controls are obtained. The majority of connectivity patterns detected by this method have been observed and studied in several seperated researches which shows the reliability and power of the proposed method. Along with these connections, we have also detected novel connectivities specially regarding the Cerebellum which is usually discarded in analysis of AD. 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 on latent space of samples that uses the test data to enhance the quality of the training set without forcing any a prior 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 works with the X_is as features. Having the FC calculation step omitted in classification not only heavily affects the computational performance of the method, but it also saves us from the troubles of FCs which will be discussed in the next section.

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 framework, not only grants us superior classification accuracy to that from other methods, it is also much faster and more stable against different data selection schemes. We have also confirmed our achieved General FC matrix using empirical data on the eMCI and Normal functional connectivity patterns.

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

2. eMCI classification techniques

As it was mentioned before, obtaining and classifying FC matrices have become the dominant approach towards eMCI classification. Variety of methods such as Pairwise Pearsons correlation coefficient [8, 9], sparse representation [8, 10, 11] and Sparse Inverse Covariance Estimation (SICE)[13] 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 result in vastly different networks[29]. On the other hand, computing the correlations, 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. In order to overcome this issue, **Non-stationary** methods have been proposed which results in more complex networks[14]–[15].

In the following we briefely discuss two state of the art eMCI classification techniques belonging to these two paradigms:

- **KernelCompactSICE(K − SIEC)** :

This method which belongs to the stationary paradigm was proposed in [12]. In this method, the SICE matrix is chosen as the FC for each sample, then using the SPD property of this matrix, kernel-PCA is deployed in order to reduce its dimensions.

The SICE matrix is extracted from the data sample S using the following optimization:

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

where C is the sample-based covariance matrix; det(), tr(), and $\|.\|_1$ denote the determinant, trace, and the sum of the absolute values of the entries of a matrix. In order to reduce the dimensionality of the obtained SICE matrix, Kernel-PCA with a Gaussian kernel is used to extract the key features. Since SICE is an SPD matrix, specific distance functions such as Log-Euclidean distance[42] or Root Stein divergence[43] can be used as the distance function

in the Gaussian kernel. Analogous to linear PCA, and by defining $\Phi(\cdot) = Sym_d^+ \mapsto \mathcal{F}$ as the kernel map, for a given SICE matrix S,$\Phi(S)$ can then be projected onto the top m eigenvectors to obtain an m-dimensional principal component vector. These vectors are used to train an SVM classifier. The performance of this method heavily relies on the choice of sparsity parameter λ and the number of top eigenvectors m.

- **HighOrderNetworks**(**HON**) :

This method which belongs to non-stationary paradigm 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[45]–[46]. 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 an SVM classifier is trained 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.

3. Proposed fMRI analysis Framework Based On Tensor framework

All main mentioned eMCI classification methods use FC of each sample as an input feature. So, the computational complexity of these methods are high due to the construction of FC matrices. In addition, the quality of the obtained FC heeavily affects the classification performance. Moreover, none of these methods attend the multi-linear propriety of such data. In this section by tensor viewpoint we show that only by one tensor decomposition we could do the following analysis on fMRI data:

• **Classification** : Viewing each class as a 3D tensor ables us to separately project each mode into a smaller one, without the necessity of unfolding it into matrices or vectors, and thus preserving the structural integrity of data along with reducing its dimensions. Transferring each sample matrix $X_i \in \mathbb{R}^{T\times R}$ into the new feature space granted by tensor viewpoint, obsoletes the necessity of constructing the FC for each sample by producing a high quality and low-dimensional feature $\bar{X}_i \in \mathbb{R}^{\bar{T}\times\bar{R}}$ in which $\bar{T}$ and $\bar{R}$ are drastically smaller that T and R. This viewpoint also ables us to design a novel discriminant function in which the test data is used in the training process without forcing any apriory knowledge about its label to the classifier, a task which is impossible via the conventional classifiers such as SVM, K-NN or logistic regression.

• **General FC** : Using the components of HOSVD extracted in the previous step, a general FC could be constructed for each class that reveals patterns common shared among all samples within each class. This technique allows us to discard the role of outliers or noisy samples.

3.1. A Classification of Region-Time data of eMCI and normal data based on Tensor framework

For reducing the computational complexity and reducing the occurrence of over fitting, specially for data with large features (Like fMRI data), using dimensionality reduction is inevitable. Although each sample $X \in \mathbb{R}^{T\times R}$ has two different time and region features, the classical methods like PCA, SVD

,... only could work on vectorized version $x = vec(X)$ of such data. Although this approach is easy to deploy, it has several drawbacks like occurrence of Curse of dimensionality and mixing different features(time and region). Recently multilinear dimension reduction methods like MPCA, GLRAM have been proposed that could works with multidimensional data, without folding them into vectors. In these methods there is a freedom to select specific reduction for each kind of feature. In this section we will used well known tensor decomposition named HOSVD for dimension reduction and also classification of fMRI data.

Let tensors $\mathcal{X}^{(i)} \in \mathbb{R}^{T\times R\times S_i}$, consists normal and eMCI data, for i=1,2, respectively. Here S_1, S_2 are the number of normal and eMCI data. For tensor $\mathcal{X}^{(i)}$, the decomposition

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

is known as higher order singular decomposition(HOSVD), where orthogonal matrices $U^{(i)} \in \mathbb{R}^{T\times T}, V^{(i)} \in \mathbb{R}^{R\times R}$ and $W^{(i)} \in \mathbb{R}^{S_i\times S_i}$ are known as modes-1,2,3 singular matrices of and $\mathcal{S}^{(i)}$ is the corresponding core tensor Ref?. Here $U^{(i)}$ is a base of all mode-1 fibers $\mathcal{X}(:, l, k)$. Here $\mathcal{X}^{(i)}(:, l, k)$ indicates the behavior of lth region of the kth sample of the ith class in all times. Also $V^{(i)}$ is a base of all mode-2 fibers $\mathcal{X}(l, :, k)$ which indicates the behavior of all regions of lth sample of the ith class in the kth 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 mode ref??. Therefore, with appropriate values of k_1^i and k_2^i projection of mode-1 and mode-2 fibers into space spanned by the first k_1 and k_2 singular vectors of modes-1,2, which are denoted by $U^{(i)}_{k_1^i}$ and $V^{(i)}_{k_2^i}$, respectively, is a suitable dimension reduction. This dimensionality reduction could be done as

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

Here it's clear that the reduced version of the k^{th} sample, i.e.,$\overline{\mathcal{X}}(:, :, k) \in \mathbb{R}^{k_1\times k_2}$ could be reduced separately in each mode without folding. This means that the structure of data is preserved in the reducing process. Base on the properties of the 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 HOSVD decomposition of $\mathcal{X}^{(i)}$ the projected data $\overline{\mathcal{X}}^{(i)}$ in equation (3) becomes

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

So, each sample of the i^{th} class in the reduced space has the following form

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

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

$$Z^{(i)} = \left(U_{k_1^i}^{(i)\mathsf{T}}, V_{k_2^i}^{(i)\mathsf{T}}\right)_{1,2} \cdot X$$

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

$$Z^{(i)} \approx \sum_{k=1}^{S_i} \lambda_k^i \overline{\mathcal{S}}^{(i)}(:,:,k). \tag{4}$$

Based on this viewpoint, each test data X could be assigned to a class that its prjected version have the best approximation in the form??. Due to importance of core tensor elements with small indices in reconstruction of the signal part of data in comparison with the last parts, the small number $k_3^i < S_i$ of slices $\overline{\mathcal{S}}^{(i)}(:,:,k)$ could be used in (4).

In this viewpoint each test data X assignment to the l^{th}class, if

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

where

$$r_i = \min_{\lambda^i} \|Z^{(i)} - \sum_{k=1}^{k_3^i} \lambda_k^i \overline{\mathcal{S}}^{(i)}(:,:,k)\|, \quad \lambda^i = \begin{pmatrix} \lambda_1^i \\ \vdots \\ \lambda_{s_i} \end{pmatrix} \tag{5}$$

shows the reconstruction error of the projected version of X in the i^{th} class. The minimization (5) 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 principle properties of the i^{th} class could be reflected in the slices of $\overline{\mathcal{S}}^{(i)}$. 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 last ones do not this role and sometimes represents the wast parts. Now consider the test 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

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

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 it will changes the behavior of last slices of core tensor of new tensor. This meas that before 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 belongs to this class or will changed the last Slices when dose not belong. 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 belongs to it, the first slices in (5), will be adapted to have good reconstruction for this data. But for a class that dose not belong to, this addition have small changes in the slices in (5). It should be mentioned that since X added to both data sets, so we did not use its label and the method in true. After computing the HOSDVD of this new data sets $\widetilde{\mathcal{X}}^{(i)}$, we apply the method in (5) for classification. Algorithm 1 summarized the proposed classification method.

In this section by HOSVD decomposition on both normal and eMCI data, we proposed a novel tensor based classification. The benifits of the proposed tensor based classification method could be summarized as follow

- This method uses multi linear property of region-Time matrix data and could works with time and region features separately based on their specific bases(corresponding singular matrices). This means that the classification works on data without folding them into vectors.

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.
3) Compute $U_{k_1^i}, V_{k_2^i}$ and $\mathcal{S}(1 : k_1^i, 1 : k_2^i, k_3^i)$ of $\widetilde{\mathcal{X}}^{(i)}$.
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 (5)
6) Assign X to class l, if $l = \arg\min_i\{r_i\}$

- In this classification, the test data has a role in construction of bases for both classes, without using its label. This increase the quality of the discrimination. This could not applied on important classification methods like SVM, logistic regression classifier and neuronal networks.

- 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 based on FC features. Also in this method we do not wary about the quality of FC methods. It should be mentioned that our method also could be applied on FC matrices as input features. But in experiments the proposed method gives better results on original time-region features.

In the experiments we compare the quality of the method with some of the sate of the art methods for recognition of eMCI like mentioned k-SICE and HON . Experimental results confirm the quality of the proposed method. In the next section we show that based the obtained HOSVD decomposition for both classes one could obtain appropriate General patters for normal and eMCI data.

3.2. General Functional Connectivity for Normal and eMCI by tensor framework

As it was mentioned before, one of the main studies associated with rs-fMRI analysis is finding common functional disorders cause by a disease like AD. This can be done via constructing proper FC matrices. The majority of techniques calculate the FC matrix for each individual subject(like SICE). This may overlook tiny but common connectivities shared within a class.

It is also noteworthy that the majority of non-stationary methods are not capable of constructing a relatable FC since the concept of a physical ROI is not well defined in them.

In this section we show that based on HOSVD decomposition one could obtain one representative functional connectivity for each normal or eMCI classes without obtaining the per-sample FC of all data separately. Here the classification is not our goal, instead finding a general and representative relation of regions for eMCI and normal person is our demand. This general pattern could be used clinically by expert cognitive scientists to found functional relation disorders caused by 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. In the laset section by property of these matrices we designed our classification method. know 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 should be mentioned that here we did not use the FC of per-samples to construct this general FC matrix.

In tensor $\mathcal{X}^{(i)}$ of the l^{th} class, the slice $\mathcal{X}^{(i)}(:, l, :)$ denotes the behavior of l^{th} region of all samples in all times in the i^{th} class. So this could be consider as a feature 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 singular matrices in modes-1,3, for appropriate values k_1^i, k_3^i, each region $\mathcal{X}(:, l, :)$ could be reduced in both time and sample features, separately based on mode-1 and mode-3 truncated singular matrices $U_{k_1^i}^{(i)}$ and $W_{k_3}^{(i)}$ as follows

$$\mathcal{Y}^{(i)}(:, l, :) = \left(U_{k_1^i}^{(i)^\mathsf{T}}, W_{k_3}^{(i)^\mathsf{T}}\right)_{1,3} \cdot \mathcal{X}^{(i)}(:, l, :). \tag{6}$$

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,

$$\mathbb{R}^{k_1^i \times R \times k_3^i} \ni \mathcal{Y}^{(i)} = \left(U_{k_1^i}^{(i)^\mathsf{T}}, W_{k_3^i}^{(i)^\mathsf{T}}\right)_{1,3} \cdot \mathcal{X}^{(i)} \tag{7}$$

denote all reduced regions of the i^{th} class. By this structure and substituting

the HOSVD decomposition of $\mathcal{X}^{(i)}$ in (7), 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

$$\begin{aligned}\mathcal{Y}^{(i)}(:,k,:) &= \sum_{k'}^{R} V^{(i)}(k,k')\bar{\mathcal{C}}^{(i)}(:,k',:) \\ &= \left(V^{(i)}(k,:)\right)_2 \cdot \mathcal{C}^{(i)}\end{aligned} \tag{8}$$

in which

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

The equation (8), show that the reduced version of each region in the i^{th} class could be written as the linear combinations of mode-2 slices of $\mathcal{C}^{(i)}$. So the coefficients 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 mentioned the first slices are better than the last ones to reflect the principle properties of the data. So for appropriate k_3^i we could select only the fits coefficient in (8) as new features for the l^{th} region. Mathematically this means each region in the i^{th} class could be represented by new feature $V(l,1;k_3^i) \in \mathbb{R}^{k_3^i}$.

The main benefit of this approach is that each region could be represented only be a vector with size k_3^i, instead large feature matrix time-sample. Now different methods could be used to find the correlation of reigns based on this new type of feature and clean matrix $V(:,1:k_3^i)$ which its rows corresponding to different regions and columns are new features. Due to quality of Sparse Inverse Covariance , this method is applied the obtained feature-region matrix $V(:,1:k_3^i)^\mathsf{T}$ in order to find the functional connectivity of the i^{th} class.

4. EXPERIMENTAL STUDY

4.1. 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

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

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].

4.2. 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 ref. 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

Table 1: 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
Proposed	**80.43**	**82.20**	**84.60**	**75.59**	**60.20**	**80.09**

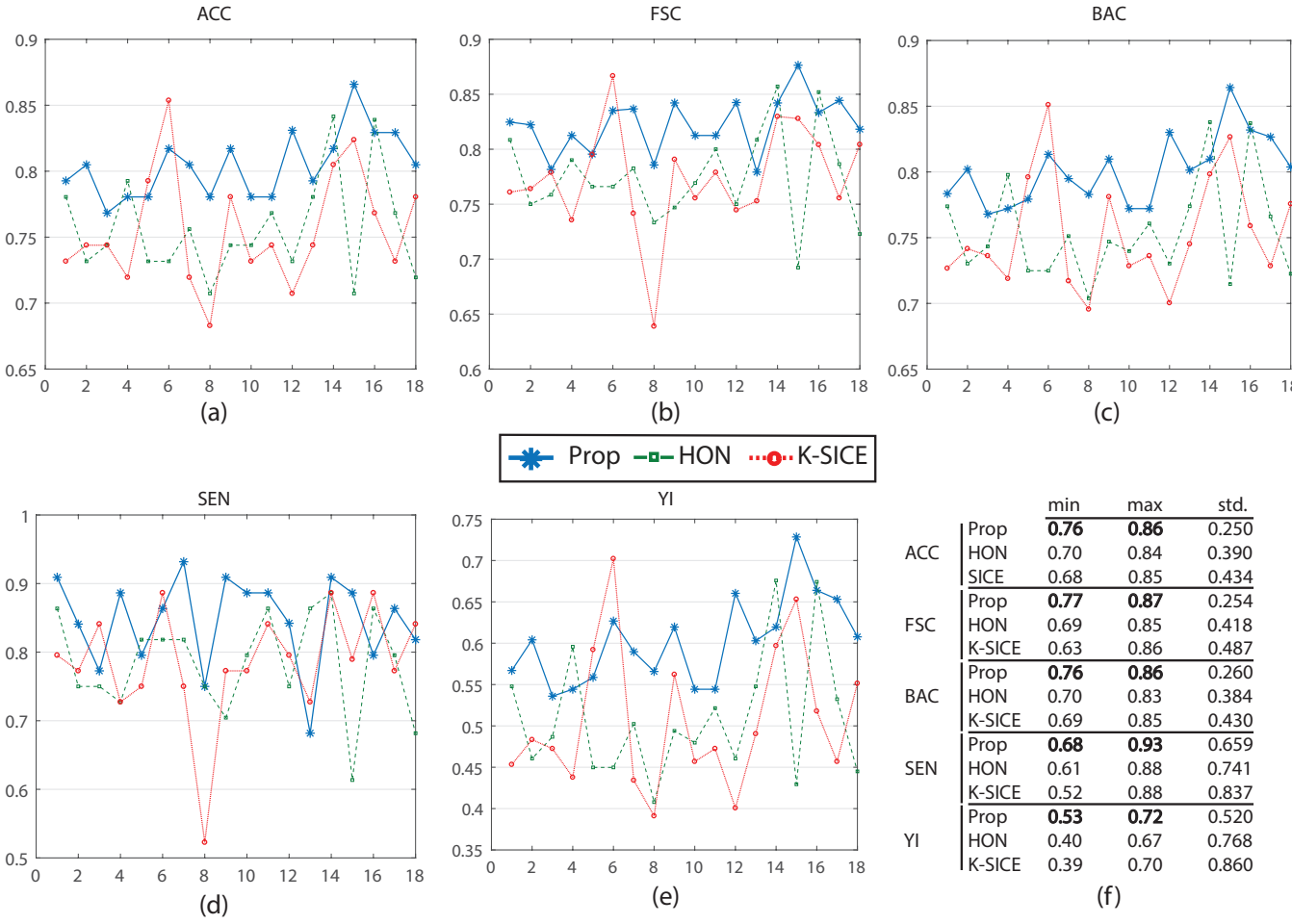
use five evaluation measures: accuracy (ACC), sensitivity (SEN), Youdens index(YI), F-score, and balanced accuracy (BAC). Tan Book

In this article, we treat the eMCI samples as positive class and the NC samples as negative class.

4.2.1. 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 performs better than k-SICE, the same also holds for 15 datasets comparing to HON. i.e. in 88.8% of datasets, proposed method works better than k-SICE and in 83.3% of datasets, it works better than FON. The highest classification accuracy(86.59%) is achieved with the proposed method 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 proposed method is 0.64 times less than HON and 1.73 times less than k-SICE method. Similar results also holds for other classification measures.

Figure (2) 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 (1) provides the average of several classification measurements scores for all permutations. As it can be seen in this table, the average accuracy of Proposed method 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



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

Figure 2: Comparison of proposed method(Prop) with K-SICE and HON in five different classification evaluation measures along with the standard deviation of each one. As it can be seen, the both the minimum and maximum of the proposed methods in these measurements are higher that the other two. Also the lower std. of our method indicates its stability.

Table 2: Elapsed time of the test and train phase in seconds

Method	HON	k-SICE	proposed method
Elapsed Time	6950	230	11

results in average.

4.2.2. Runtime Comparison

One other key features of proposed method is that it works significantly faster that the other two methods. Table (2) 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, proposed method 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.

4.3. Functional connectivity Network

The functional connectivity networks of the Normal and eMCI classes was obtained via proposed method as it is described in (3.2). 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(3). 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

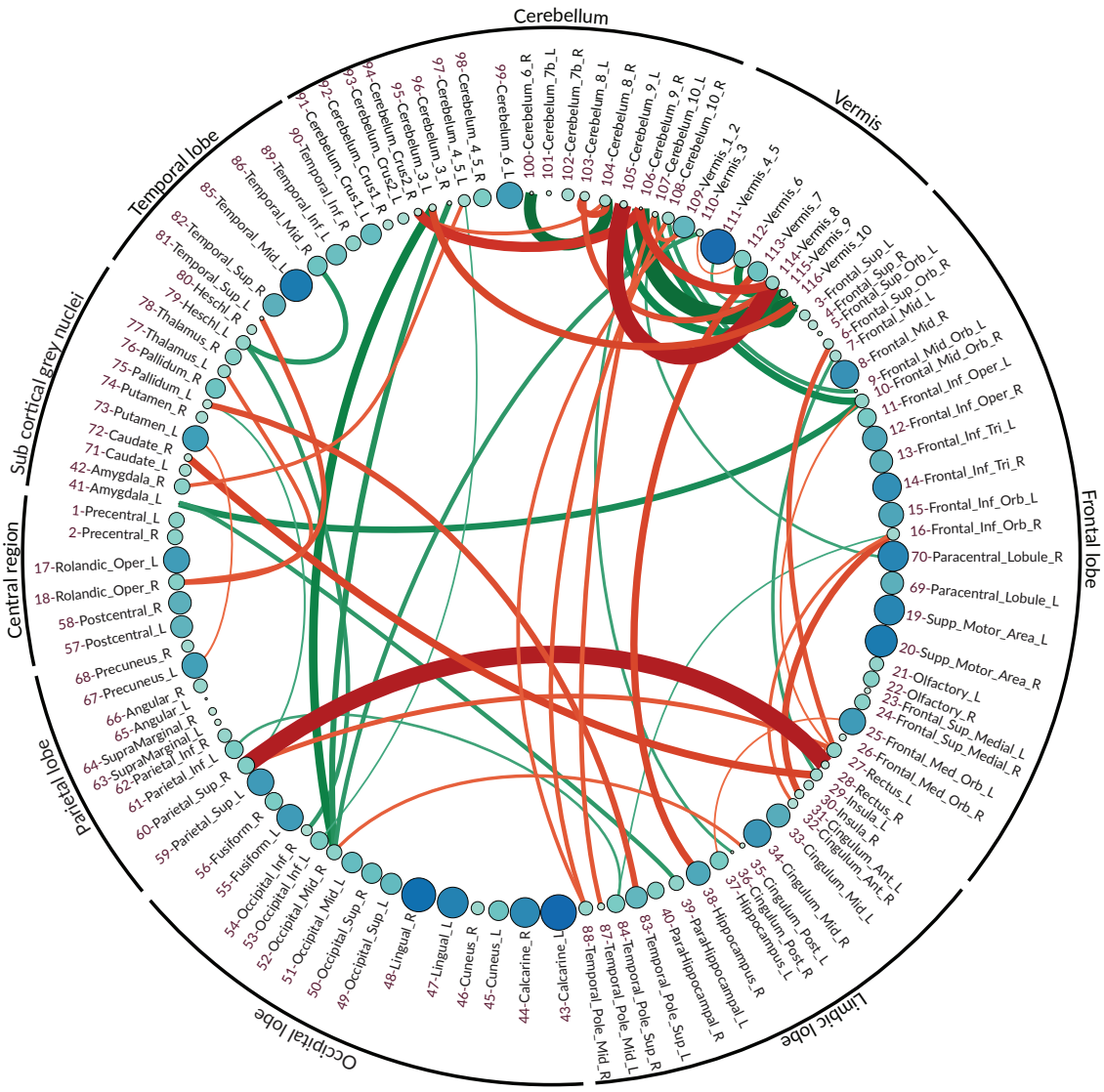


Figure 3: The difference graph. This graph is obtained via subtracting the functional connectivity of eMCI subjects from normal subjects. Each circle represents a ROI in AAL atlas and the color and size of each circle is proportional to the graph clustering coefficient of the difference graph. red = more activity in EMCI, green: less.

index: 47,48)[19, 20], Calcarine sulcus(ROI index: 43, 44)[21, 22], Supplementary motor area(ROI index: 19,20)[22, 23] and Temporal_mid_L(ROI index: 85)[24] are easily detectable. The majority of ROIs located in frontal lobe also shows rather high activities comparing to normal subjects[25, 5].

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) [33, 34], Frontal_Inf_Orb_R(ROI index: 16 in Frontal lobe) and Cingulum_Ant(ROI index: 31,32 in Limbic lobe)[35], 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)[36]. It can also be seen that within activities in frontal lobe also increased in patients with eMCI[37]. 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)[38]. 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[39].

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 since their role was regarded as insignificant[40, 41]. 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[26].

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.

5. 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. Classification
2. General Functional Connectivity network

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.

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] Dennis, Emily L., and Paul M. Thompson. "Functional brain connectivity using fMRI in aging and Alzheimers disease." Neuropsychology review 24.1 (2014): 49-62.

[6] R. Graaf and K. Kevin. Methods and apparatus for compensating eld inhomogeneities in magnetic resonance studies. US Patent No. 8035387, 2011.

[7] Stanley, Matthew Lawrence, et al. "Defining nodes in complex brain networks." Frontiers in computational neuroscience 7 (2013): 169.

[8] Jie, Biao, et al. "Integration of network topological and connectivity properties for neuroimaging classification." IEEE transactions on biomedical engineering 61.2 (2014): 576-589.

[9] Wee, Chong-Yaw, et al. "Resting-state multi-spectrum functional connectivity networks for identification of MCI patients." PloS one 7.5 (2012): e37828.

[10] 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.

[11] Wright, John, et al. "Robust face recognition via sparse representation." IEEE transactions on pattern analysis and machine intelligence 31.2 (2009): 210-227.

[12] Zhang, Jianjia, et al. "Functional brain network classification with compact representation of SICE matrices." IEEE Transactions on Biomedical Engineering 62.6 (2015): 1623-1634.

[13] Huang, Shuai, et al. "Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation." NeuroImage 50.3 (2010): 935-949.

[14] Allen, Elena A., et al. "Tracking whole-brain connectivity dynamics in the resting state." Cerebral cortex 24.3 (2014): 663-676.

[15] Leonardi, Nora, et al. "Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest." NeuroImage 83 (2013): 937-950.

[16] Nordberg, Agneta. "PET imaging of amyloid in Alzheimer's disease." The lancet neurology 3.9 (2004): 519-527.

[17] Jeong, Jaeseung. "EEG dynamics in patients with Alzheimer's disease." Clinical neurophysiology 115.7 (2004): 1490-1505.

[18] Jeong, Jaeseung. "EEG dynamics in patients with Alzheimer's disease." Clinical neurophysiology 115.7 (2004): 1490-1505.

[19] Golby, Alexandra, et al. "Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory." Brain 128.4 (2005): 773-787.

[20] 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.

[21] 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.

[22] 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.

[23] Jacobsen, Jrn-Henrik, et al. "Why musical memory can be preserved in advanced Alzheimers disease." Brain 138.8 (2015): 2438-2450.

[24] Kosicek, Marko, and Silva Hecimovic. "Phospholipids and Alzheimers disease: alterations, mechanisms and potential biomarkers." International journal of molecular sciences 14.1 (2013): 1310-1322.

[25] 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.

[26] Jacobs, Heidi IL, et al. "The cerebellum in Alzheimers disease: evaluating its role in cognitive decline." Brain 141.1 (2017): 37-47.

[27] 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.

[28] Cherkassky, Vladimir L., et al. "Functional connectivity in a baseline resting-state network in autism." Neuroreport 17.16 (2006): 1687-1690.

[29] Du, Yuhui, Zening Fu, and Vince D. Calhoun. "Classification and prediction of brain disorders using functional connectivity: promising but challenging." Frontiers in neuroscience 12 (2018).

[30] 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.

[31] Cuingnet, Rmi, et al. "Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database." neuroimage 56.2 (2011): 766-781.

[32] Friston, Karl J. "Functional and effective connectivity: a review." Brain connectivity 1.1 (2011): 13-36.

[33] Brickman, Adam M., et al. "Reconsidering harbingers of dementia: progression of parietal lobe white matter hyperintensities predicts Alzheimer's disease incidence." Neurobiology of aging 36.1 (2015): 27-32.

[34] De Reuck, J., et al. "Topography of cortical microbleeds in Alzheimers disease with and without cerebral amyloid angiopathy: a post-mortem 7.0-tesla MRI Study." Aging and disease 6.6 (2015): 437.

[35] Perani, Daniela, et al. "The impact of bilingualism on brain reserve and metabolic connectivity in Alzheimer's dementia." Proceedings of the National Academy of Sciences 114.7 (2017): 1690-1695.

[36] Subcortical volume changes in dementia with Lewy bodies and Alzheimer's disease. A comparison with healthy aging

[37] Cai, Suping, et al. "Changes in thalamic connectivity in the early and late stages of amnestic mild cognitive impairment: a resting-state functional magnetic resonance study from ADNI." PloS one 10.2 (2015): e0115573.

[38] Ortner, Marion, et al. "Progressively Disrupted intrinsic Functional connectivity of Basolateral amygdala in Very early alzheimers Disease." Frontiers in neurology 7 (2016): 132.

[39] Steketee, Rebecca ME, et al. "Early-stage differentiation between presenile Alzheimers disease and frontotemporal dementia using arterial spin labeling MRI." European radiology 26.1 (2016): 244-253.

[40] Sanz-Arigita, Ernesto J., et al. "Loss of small-worldnetworks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity." PloS one 5.11 (2010): e13788.

[41] Zhang, Daoqiang, et al. "Multimodal classification of Alzheimer's disease and mild cognitive impairment." Neuroimage 55.3 (2011): 856-867.

[42] V.Arsigny et al.,. (2006). Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med.. [Online]. 56(2), pp. 411421. Available: http://dx.doi.org/10.1002/mrm.20965

[43] S. Sra, A new metric on the manifold of kernel matrices with application to matrix geometric mean, in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, K. Q.Weinberger, Eds., New York, NY: Curran Associates, Inc., 2012, pp. 144152.

[44] Allen, Elena A., et al. "Tracking whole-brain connectivity dynamics in the resting state." Cerebral cortex 24.3 (2014): 663-676.

[45] Chang, Catie, and Gary H. Glover. "Timefrequency dynamics of resting-state brain connectivity measured with fMRI." Neuroimage 50.1 (2010): 81-98.

[46] Handwerker, Daniel A., et al. "Periodic changes in fMRI connectivity." Neuroimage 63.3 (2012): 1712-1719.

[47] Supekar, Kaustubh, et al. "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease." PLoS computational biology 4.6 (2008): e1000100.

[48] Contreras, Joey A., et al. "Resting state network modularity along the prodromal late onset Alzheimer's disease continuum." NeuroImage: Clinical 22 (2019): 101687.