DETECTION OF OBSESSIVE COMPULSIVE DISORDER USING RESTING-STATE
FUNCTIONAL CONNECTIVITY DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SONA KHANEH SHENAS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE GRADUATE OF MASTER OF SCIENCE
IN
BIOMEDICAL ENGINEERING

SEPTEMBER 2013

Approval of the thesis:

DETECTION OF OBSESSIVE COMPULSIVE DISORDER USING RESTING-STATE FUNCTIONAL CONNECTIVITY DATA

Submitted by **SONA KHANEH SHENAS** in partial fulfillment of the requirements for the degree of **Master of Science in Biomedical Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen   _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof.Dr. Vasıf Hasırcı   _____-___
Head of Department, **Biomedical Engineering**

Prof.Dr.Uğur Halıcı   _____
Supervisor, **Electrical and Electronics Engineering Dept., METU**

Prof.Dr.Metehan Çiçek   _____
Co-supervisor, **Physiology Dept., Ankara University**

Examining Committee Members

Prof.Dr. Nevzat G. Gençer   _____
Electrical and Electronics Engineering Dept., METU

Prof.Dr.Uğur Halıcı   _____
Electrical and Electronics Engineering Dept., METU

Prof.Dr.Metehan Çiçek   _____
Physiology Dept., Faculty of Medicine, Ankara University

Prof.Dr. Kemal Leblebicioğlu   _____
Electrical and Electronics Engineering Dept., METU

Assist.Prof. Dr. Senih Gürses   _____
Engineering Science Dept., METU

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: **Sona Khaneh Shenas**

Signature:

ABSTRACT

DETECTION OF OBSESSIVE COMPULSIVE DISORDER USING RESTING-STATE
FUNCTIONAL CONNECTIVITY DATA

Khaneh Shenas, Sona
M.Sc., Department of Biomedical Engineering
Supervisor: Prof. Dr. Uğur Halıcı
Co-Supervisor: Prof. Dr. Metehan Çiçek

September 2013, 79 Pages

Obsessive Compulsive Disorder (OCD) is a serious psychiatric disease that might be affiliated with abnormal resting-state functional connectivity (rs-FC) in default mode network (DMN) of brain. The aim of this study is to discriminate patients with OCD from healthy individuals by employing pattern recognition methods on rs-FC data obtained through regions of interest (ROIs) such as Posterior Cingulate Cortex (PCC), Left Inferior Posterior Lobe (LIPL) and Right Inferior Posterior Lobe (RIPL). For this purpose, two different approaches were implemented as feature extraction step of pattern recognition. In the first approach the rs-FC fMRI data were subsampled and then the dimensionality of the subsampled data was reduced using the Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA) and Linear Discriminant Analysis (LDA) alternatives. In the second approach, feature vectors having already low dimensions were obtained by measuring cosine similarity, dot product similarity and correlation similarity to the separate means of the rs-FC data of subjects in OCD and healthy groups. Afterwards the healthy and OCD groups were classified using Support Vector Machine (SVM) and Gaussian Mixture Models (GMMs). In order to obtain more reliable performance results, Double LOO-CV method that we proposed as a version of Leave-One-Out Cross Validation (LOO-CV) was used and the best performance (73%) was obtained by using cosine similarity for feature extraction and GMMs for classification.

Keywords: Functional MRI, Resting-state functional connectivity, Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Gaussian mixture Model (GMM), double leave-one-out cross-validation (D-LOO-CV)

ÖZ

OBSESİF KOMPULSİF BOZUKLUĞUN DİNLENME-DURUMU FONKSİYONEL BAĞLANTI VERİLERİ KULLANILARAK SAPTANMASI

Khaneh Shenas, Sona

Yüksek Lisans, Biyomedikal Mühendisliği Anabilim Dalı

Tez Yöneticisi: Prof. Dr. Uğur Halıcı

Ortak Tez Yöneticisi: Prof. Dr. Metehan Çiçek

Eylül 2013, 79 Sayfa

Obsesif Kompulsif Bozukluk (OKB), beyindeki Olagan Mod Ağındaki (DMN, Default Mode Network) anormal dinlenme-durumu Fonksiyonel Bağlantıları (rs-FC) ile ilişkilendirilen ciddi bir piskiyatrik hastalıktır. Bu çalışmada, OKB olan hastaları sağlıklı kişilerden ayırtedebilmek üzere fMRI uzerinde Posterior Singulat Cortex (PCC Posterior Cingulate Cortex), Sol Inferior Posterior Lob (LIPL, Left Inferior Posterior Lobe) ve Sağ Inferior Posterior Lob (RIPL, Right Inferior Posterior Lobe) olarak adlandırılan ilgi bölgelerinden elde edilen rs-FC verisi üzerinde örüntü tanıma yöntemlerinin kullanılması amaçlanmıştır. Bu amaçla örüntü tanımanın öznitelik çıkarma adımında iki değişik yaklaşım ele alınmıştır. İlk yaklaşımda rs-FC fMRI verisi üzerine altörnekleme yapıldıktan sonra Temel Bileşenler Analizi (PCA, Principal Component Analysis), Kernel Temel Bileşenler Analizi (KPCA) ya da Lineer Ayıraç Analizi (LDA, Linear Discriminant Analysis) yapılarak veri boyutu düşürülmüştür. İkinci yaklaşımda ise, yeni gelen örnek ile hasta ve sağlıklı gruptaki kişilerin rs-FC verilerinin ayrı ayrı ortalamaları üzerinde üzerinde kosinüs benzerliği, iç çarpım benzerliği ya da korelasyon benzerliği ölçütleri kullanılarak doğrudan düşük boyutlu öznitelik vektörleri elde edilmiştir. Daha sonra Destek Vektör Makinası (Support Vector Machine) ya da Gausyan Karışım Modelleri (GMMs, Gaussian Mixture Models) kullanılarak sağlıklı ve OKB gruplar sınıflandırılmıştır. Başarı ölçümünde, daha güvenilir sonuçlar elde etmek üzere, bir örnek dışarıda bırakan çapraz geçerleme (LOO-CV, Leave One Out Cross Validation) yönteminin bir versiyonu olarak önerdiğimiz çiftli LOO-CV kullanılmış ve en iyi başarı (%73), öznitelik çıkarma için kosinüs benzerliği ve sınıflama için GMMs kullanılarak elde edilmiştir.

Anahtar Kelimeler: Fonksiyonel MRI, Dinlenme-durumu fonksiyonel bağlantısı, Principal Component Analizi (PCA), Kernel Principal Component Analizi (KPCA), Linear Diskriminant Analizi (LDA), Support Vector Makinesi (SVM), Gaussian Mixture Modeli (GMM)

To My Family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

FMRI: Functional Magnetic Resonance Imaging

BOLD: Blood Oxygen-Level Dependent Signal

rsFC: Resting-state functional connectivity

PCA: Principal Component Analysis

KPCA: Kernel Principal Component Analysis

LDA: Linear Discriminant Analysis

SVM: Support Vector Machine

GMM: Gaussian Mixture Model

EM: Expectation Maximization

COS: Cosine similarity measure

DOT: Dot product similarity measure

CORR: Correlation measure

DMN: Default Mode Network

ROI: Region of interest

LOO-CV: Leave-One-Out Cross Validation

D-LOO-CV: Double Leave-One-Out Cross Validation

PDF :   Probability Density Function

LIST OF SYMBOLS

H: Set of healthy subjects $\{h_1, h_2, \ldots, h_{12}\}$

n_H: Number of healthy subject, $n_H = 12$

h_i: ith healthy subject, $h_i \in H$ $i = 1,2,\ldots,12$

P: Set of patient (OCD) subjects $\{p_1, p_2, \ldots, p_{12}\}$

n_P: Number of patient subject, $n_P = 12$

p_j: jth patient subject, $p_j \in P$ $j = 1,2,\ldots,12$

S: Set of all subjects, $S = H \cup P$

n_S: Size of S, $n_S = n_H + n_P$, $n_S = 24$

s_k: Index for all subjects, whether healthy or patient; $s_k \in S$ $k = 1,2,\ldots,24$

$H - k$:   Set of healthy subjects without subject $k \in S$.

n_{H-k}:   Size of the set $H - k$

$$n_{H-k} = \begin{cases} n_H - 1 & k \in H \\ n_H & k \in P \end{cases}$$

$H - k - s$: Set of healthy subjects without subjects $k, s \in S$.

n_{H-k-s}:   Size of the set $H - k - s$

$$n_{H-k-s} = \begin{cases} n_{H-k} - 1 & s \in H , s \neq k \\ n_{H-k} & s \in P \ or \ s = k \end{cases}$$

$P - k$:   Set of patient subjects without subject $k \in S$.

$n_{P-k:}$ Size of the set $P - k$

$$n_{P-k} = \begin{cases} n_P - 1 & k \in P \\ n_P & k \in H \end{cases}$$

$P - k - s$:   Set of patient subjects without subjects $k, s \in S$.

n_{P-k-s}:   Size of the set $P - k - s$

$$n_{P-k-s} = \begin{cases} n_{P-k} - 1 & s \in P, s \neq k \\ n_{P-k} & s \in H \text{ or } s = k \end{cases}$$

$c(k)$:   The class that the subject belongs to

$$c(k) = \begin{cases} +1 & k \in H \\ -1 & k \in P \end{cases}$$

$\mathbf{I}(k)$:   The data vector for kth subject obtained from preprocessed 3D functional connectivity fMRI data by containing elements layer, row, column order. Also it can be considered as connectivity image of the kth subject.

$\mathbf{F}(k, h, p)$:   Leave one out feature vector for subject k when $h \in H$ and $p \in P$ is reserved for testing. Note that the feature vector varies according to the method used for feature extraction. (See methods given in chapters 2, 3)

$Train(s)$:   The training set prepared by leave-one-out subject, s:

$$Train(s) = \{(\mathbf{F}(k, h, p), C(k) | k \neq s; (s = h) or (s = p); s \in S\}$$

$Test(s)$:   The test set prepared by leave-one-out-subject, s:

$$Test(s) = \{(\mathbf{F}(k, h, p) | k = s; (s = h) or (s = p); s \in S\}$$

$mean\ \mathbf{I}$:   mean of all data

$sum\mathbf{I}_H$:   Sum of data of healthy subjects

$$sum\mathbf{I}_H = \sum_{i \in H} \mathbf{I}(i)$$

$mean\ \mathbf{I}_H$:   Mean of data of healthy subjects (mean of all healthy data).

$$mean\ \mathbf{I}_H = \frac{sum\mathbf{I}_H}{n_H} = \frac{1}{n_H} \sum_{i \in H} \mathbf{I}(i)$$

$sumI_{H-s}$:   Sum of data of healthy subjects without subject $s \in H$.

$$sumI_{H-s} = \sum_{\substack{i \in H \\ i \neq s}} I(i)$$

$mean\ I_{H-s}$:   Mean of data of healthy subjects without subject $s \in H$.

$$mean\ I_{H-s} = \frac{sumI_{H-s}}{n_{H-s}} = \frac{\sum_{\substack{i \in H \\ i \neq s}} I(i)}{n_{H-s}}$$

$mean\ I_{H-k-s}$:   Mean of data of healthy subjects without subjects $k \in S, s \in H$.

$$mean\ I_{H-k-s} = \frac{sumI_{H-k-s}}{n_{H-k-s}} = \frac{\sum_{\substack{i,s \in H \\ i \neq s,k}} I(i)}{n_{H-k-s}}$$

$sumI_P$:   Sum of data of patient subjects

$$sumI_P = \sum_{j \in P} I(j)$$

$mean\ I_P$:   Mean of data of patient subjects

$$mean\ I_P = \frac{sumI_P}{n_P} = \frac{1}{n_P} \sum_{j \in P} I(j)$$

$sum\ I_{P-s}$:   Sum of data of patient subjects without subject $s \in P$.

$$sum\ I_{P-s} = \sum_{\substack{j \in P \\ j \neq s}} I(j)$$

$mean\ I_{P-s}$:   Mean of data of patient subjects without subject $s \in P$.

$$mean\ I_{P-s} = \frac{sumI_{P-s}}{n_{P-s}} = \frac{\sum_{\substack{i \in P \\ i \neq s}} I(i)}{n_{P-s}}$$

$mean\ I_{P-k-s}$:   Mean of data of patient subjects without subjects $k \in S, s \in P$.

$$mean\ I_{P-k-s} = \frac{sumI_{P-k-s}}{n_{P-k-s}} = \frac{\sum_{\substack{i,s \in P \\ i \neq s,k}} I(i)}{n_{P-k-s}}$$

CHAPTER1

INTRODUCTION

1.1 Motivation of the Study

In medical science, it was always a question of whether patients can be diagnosed by means of biomedical engineering methods. Over the years, neuroscientists attempted to find methods to distinguish between patients with mental problems (Alzheimer, Schizophrenia, attention deficit disorder, obsessive compulsive disorder) and healthy people. The functional magnetic resonance imaging (FMRI) of the brain has received much attention in recent years and widely used in experiments to investigate the regions of the brain that have functional abnormalities of neural activity [1]. Studies have shown that the connectivity analysis between regions of the brain in which there are functional abnormalities may provide functional and structural relationship between remote regions of the brain. Up to now, neuroscientists have introduced three models for brain connectivity analysis:

1) Structural (anatomical) connectivity: refers to axonal connections of neurons through synapses.
2) Functional connectivity: refers to correlations between active regions of the brain and evaluate the effect of the certain brain regions on each other and other brain voxels.
3) Effective connectivity: considers the connectivity between brain regions in detail by providing causal models of influence that brain regions apply on each other.

On the other hand, brain researchers have concluded that connectivity analysis on the resting – state of the brain is much successful than task-related state since the brain is in its optimal state while resting [2]. So, the resting-state connectivity analysis of the brain can be introduced as suitable pattern to diagnose patient brain. Some machine learning techniques have been used to classify diseases and health states using connectivity patterns obtained.

In this study we focused on discriminating between a special case of mental disorder called Obsessive Compulsive Disorder (OCD) and healthy cases using their resting-state functional connectivity (rsFC) data. OCD is a psychiatric disorder which is characterized by obsessions and triggered by compulsions and like the other mental disorders suggested to affect the functional correlations between different regions of the brain [3]. The differentiation between OCD and other mental disorders is difficult using MRI or other conventional imaging methods since they can only distinguish the structural disorders of the brain and cannot detect the functional abnormalities in early stages of the disease [4]. Therefore the functional MRI technique is used to analysis the functional associations or dissociations between brain regions in OCD.

The pattern recognition approach was considered to extract features from rsFC data using feature extraction methods such as dimensionality reduction (PCA, KPCA and LDA) and similarity measures (correlation, dot product and cosine similarity). Afterwards the known classifiers SVM and GMM are implemented to differentiate two categories. The SVM method defines a separating boundary called hyper surface to classify two groups. Whereas the GMM method, recently used in brain studies, models the data as Gaussian mixtures using methods such as EM to find the mixture parameters and then classifies the groups according to their probability density distributions.

1.2 Contribution of the Study

The functional MRI data of the 12 healthy and 12 OCD individuals while the brain was in resting-state is used in this study and a set of alternative pattern recognition methods for preprocessing, feature extraction and classification steps were considered in order to discriminate healthy and OCD data. The Block Scheme for the proposed approach is shown in Figure 1.1 representing preprocessing, feature extraction and classification phases and their alternatives. Firstly the preprocessing steps were applied to the data to remove the noise and correct the faults happen during the data acquisition and match the coordinate of all brain maps. In the next step ROIs where functional abnormalities have been established in OCD brains are determined. After finding ROIs, the functional connectivity between these regions or between ROIs and all voxels of the brain is computed using functional connectivity toolbox (CONN). The FC data derived from resting-state activity of the brain is used to identify healthy from OCD groups. The pattern recognition approach is applied to discriminate between OCD and healthy individuals. The pattern recognition steps are as follows. For the first step two different approaches as feature extraction step of pattern recognition were proposed, in the first approach the dimensionality of the data was reduced using the dimensionality reduction methods as PCA, KPCA and LDA. In the second approach new features were extracted for classification using similarity measures as cosine, correlation and dot product similarity measures already having low dimensions. The second step of the pattern recognition relates to applying classification methods as SVM and GMM to the data. Before classification, the training and test sets were provided using special case of leave-one-out cross-validation (LOO-CV) called double LOO-CV (D-LOO-CV) approach. The SVM classifier was then learnt with training set and an ideal linear or nonlinear separating hyper surface was defined. Performance of the classifier was tested using test data and finally the accuracy rate of classification was calculated. The GMM first models the training data with mixture of Gaussians and guess initial parameters for mixture model then using iterative algorithm called Expectation-Maximization (EM) optimize the parameters the model was tested using the test data with optimized parameters and again the performance accuracy of the model was evaluated.

Previous studies have analyzed OCD using rs-FC obtained from fMRI data but no pattern recognition methods have been used to classify OCD and healthy individuals. The novelty of this study is doing classification by using rs-FC data as features for pattern recognition methods. Also similarity measure method is a novel approach that extracted new feature vectors from rs-

FC data which were useful to separate the healthy and OCD individuals from each other. Therefore, using the rs-FC data might help to develop the diagnostic analysis of OCD and other mental disease in future. Another uniqueness of this research was applying GMM as a classification method to rs-FC data of OCD which has not been tried on fMRI data of any mental disorder in order to detect it.
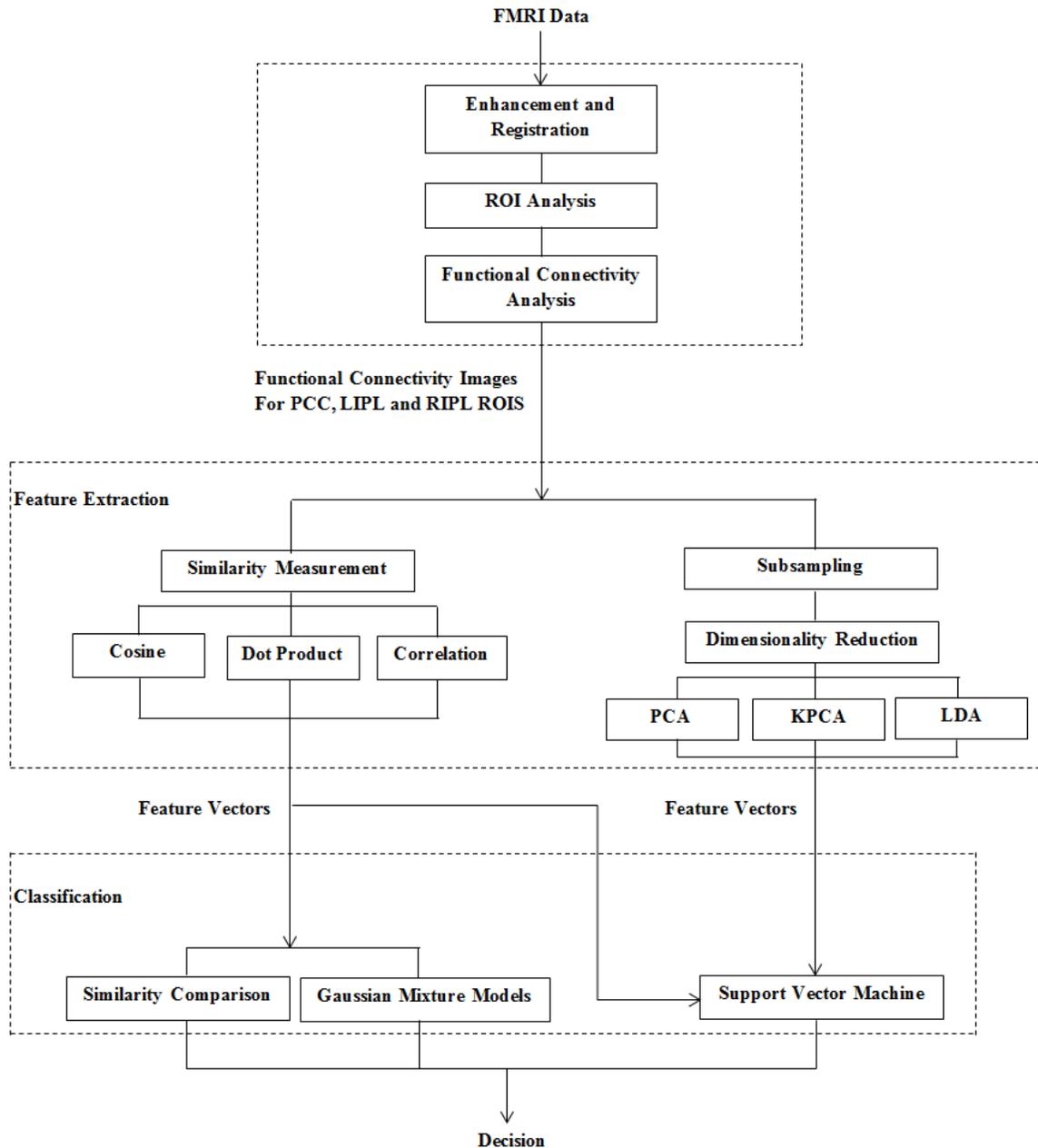
Figure 1.1: Alternative pattern recognition steps considered in this study

1.3 Thesis Outline

The second chapter of this study gives brief background information about functional MRI, resting-state functional connectivity, methods used for discriminative analysis of fMRI data and previous studies conducted for these methods are described. In chapter three, the process of data acquisition, ROI analysis, the toolbox used to calculate the functional connectivity and preprocessing steps such as enhancement and registration are explained then the feature extraction, dimensionality reduction and classification methods implemented on the fMRI data are described in detail. The fourth chapter relates to the results and discussions. Here methods applied are compared with each other and the one that results in the best performance is determined. Finally, the study is concluded in the fifth chapter.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is a magnetic resonance imaging tool to assess neural activity of the brain and analyze its functions by measuring the signals caused by metabolic changes related to functionally active parts of brain. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled. When an area of the brain is in use, blood flow to that region also increases [1], [5] which appears as lighter areas in the Figure 2.1. fMRI is a safe technique and has no radiation hazard for patient in contrast with other imaging techniques like CT,PET [1]. Neuroscientists use this diagnostic method to consider the performance of healthy, patient or injured brain, following progression of disease (such as Alzheimer, Attention defecate, Schizophrenia), studying subjects before and after treatment, drug therapy and the other clinical applications of fMRI. Further, using fMRI helps to monitor the brain tumors and their growth and the activity [5].
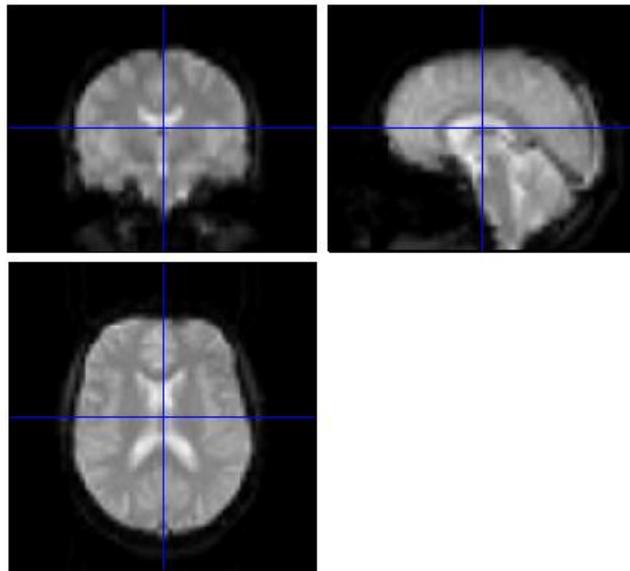
Figure 2.1: Functional MRI

2.1.1 Blood Flow and Neural Activity

Studies suggest that when a region of the brain becomes active the blood flow to that region increase to provide the oxygen and glucose needed for neural activity of cells of the active area and conversely, when the brain performs no task the blood flow and accordingly the oxygen level decreases [6]. This dynamically regulation of the blood flow is called Hemodynamic response. The fMRI measures the signals produced from oxygen level change which is known as blood-oxygen-level-dependent (BOLD) signal.
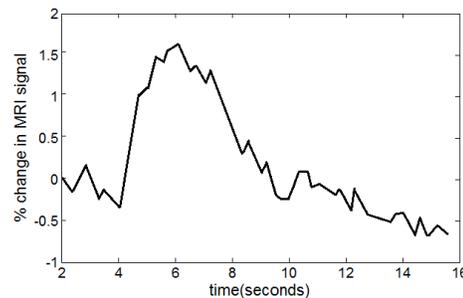
Figure 2.2: Hemodynamic responses [3]

2.2 Functional Connectivity

Temporal and spatial correlations between "spontaneous low-frequency BOLD signal fluctuations "of remote regions of the brain are defined as functional connectivity [5]. There are many definitions for functional connectivity in literature but the most known one is the definition by Friston and Büchel [7], "functional connectivity as the correlations between spatially remote neurophysiological events", and by Aertsen and Preissl [8] "functional connectivity as groups of neurons that act together in a coherent fashion".

Functional connectivity gives information about correlations between brain regions but cannot interpret how they are correlated. Functional connectivity analysis can be performed on regions of interest which are regions in the brain consists of one or several voxels, or on the level of the individual voxels. As ROIs give results about brain regions directly, they are popular than other connectivity analysis methods [9].

2.2.1 Resting-State Functional Connectivity

Recently many fMRI studies have been concentrated on measuring the signals extracted from resting state of the brain. Subjects lie in the scanner under resting conditions (eyes closed with no-task perform) and the correlation analysis is done on brain regions. The correlation results can give the view about the functional activity of the neural systems.

The reasons why resting-state fMRI is preferred to task-related state are as follows [1, 4]:

1) Cerebral Energetics: The brain in resting condition consumes more metabolic energy of the body (about 20%) compered to task-related state (about 5%) so exhibit high activity during resting state.

2) Signal to Noise: Resting state of the brain has large signal to noise ratio than the task-based state.

3) Ease of use for patient populations who are not able to exactly perform tasks asked them in the fMRI scanner.

4) Resting-state fMRI measures functional connectivity more precisely in brain regions in contrast with task-based studies.

The two most common method for resting state functional connectivity (rsFC) data analysis is seed-driven rsFC in which the Pearson's correlation coefficients between the BOLD signal time course of the seed region and the BOLD signal time course of all other voxels are computed [10]. Let the bold signal of a region of interest is x_{ROI} and the bold signal of a voxel of the brain is x_v. Then the Pearson's correlation coefficient is calculated as:

$$r_v = \frac{\text{cov}(x_{ROI}, x_v)}{\sigma_{ROI}\sigma_v} = \frac{\sum (x_{ROI} - \overline{x_{ROI}})(x_v - \overline{x_v})}{\sqrt{\sum (x_{ROI} - \overline{x_{ROI}})^2}\sqrt{\sum (x_v - \overline{x_v})^2}} \tag{2.1}$$

where cov represents the covariance of the x_{ROI} and x_v vectors. The σ_{ROI} and σ_v, are the standard deviations of x_{ROI}, x_v vectors, respectively. $\overline{x_{ROI}}$, is the mean of all voxel values inside the ROI and $\overline{x_v}$ is the mean of all voxels of the brain. Notice that $\sqrt{\sum(x_v - \overline{x_v})^2}$ = $|x_v - \overline{x_v}|$ since only a single voxel value is considered.

The 3D functional connectivity image for kth subject derived from Pearson's correlation coefficient between each ROI and all brain voxels will be represented as $I(k)$ notation through this study.
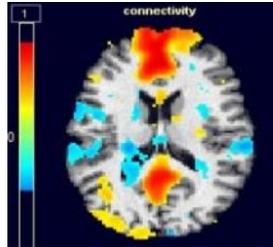
Figure 2.3: Funcional connectivity using Pearson's correlation coeffitient

2.3 Obsessive Compulsive Disorder

The Obsessive Compulsive Disorder (OCD) is a psychiatric brain abnormality that is characterized by the presence of repetitive invasive impulses or thoughts called obsessions that cause recurrent behaviors which are called compulsions [3, 11]. The OCD usually occurs in early adulthood or late adolescence [11] and rare in children such that 1- 4% of children and adolescents catch OCD [12], also it may become worse and chronic if it left untreated.

The first neuroimaging studies on OCD were done by Baxter and associates [13]. They used positron emission tomography (PET) to investigate resting cerebral glucose metabolism in patients with OCD [14]. Recently, biological markers such as fMRI signals that have diagnostic information for OCD have become popular [3, 4]. Since functional MRI produce better spatial resolution than radio imaging techniques (PET) therefore smaller brain regions are identified [15]. Also fMRI unlike other medical imaging techniques does not have any ionizing radiation hazard thereby it is a safe imaging tool which can be used widely in clinical applications.

Reviewing literature several functional connectivity studies has been done in OCD. Jang et al. [16] investigated the functional connectivity between fronto-subcortical regions of OCD brain during resting state. Koçak et al. [2] considered the resting state functional connectivity differences between healthy and OCD individuals. Harrison et al. [4] investigated the abnormal functional connectivity in regions of OCD brain using resting state fMRI and also suggested that rs fMRI gives more precise measurement of FC of brain regions.

2.4 Default Mode Network and Region of Interest

The regions of the brain that are functionally active during resting-condition (no mental task is performed) are called default mode network (DMN) [17]. There are many reports based on that the DMN region of the brain in patients with obsessive compulsive disorder may have suppression problems. Koçak et al. [18] hypothesized in his paper that there is some abnormalities in right posterior partial region of OCD brain. His findings represented that there is more brain functional activity in post cingulate corpus (PCC) and superior frontal gyrus (SFG) in OCD group in contrast with healthy group during the resting-state condition and interestingly where both of these regions belong to DMN region. Also a decreased brain functional activity is recorded in right inferior posterior lobe (RIPL) of patients with OCD which belongs to the DMN again. According to this paper there may be significant functional activity changes in RIPL, SFG and PCC brain regions of OCD group. Jang and Kim [16] suggested that an abnormal rsFC in DMN may cause the obsessions in patients with OCD and fronto-subcortical regions belong to DMN such as PCC exhibit lower functional connectivity in OCD. According to Koçak left inferior posterior lobe which is part of DMN shows different functional activity in patients with OCD compared to healthy group in control condition. As reported by Hon, Epstein [19] and Suchan, Botko [20] ,the IPL and PCC ROIs have key roles in imagination control and imaginative control of visual information. For this reason, the LIPL, RIPL and PCC are the regions that existence of some differences in OCD is proved so these

regions of interest (ROIs) is thought to assist us to discriminate between the OCD and healthy brains [2].
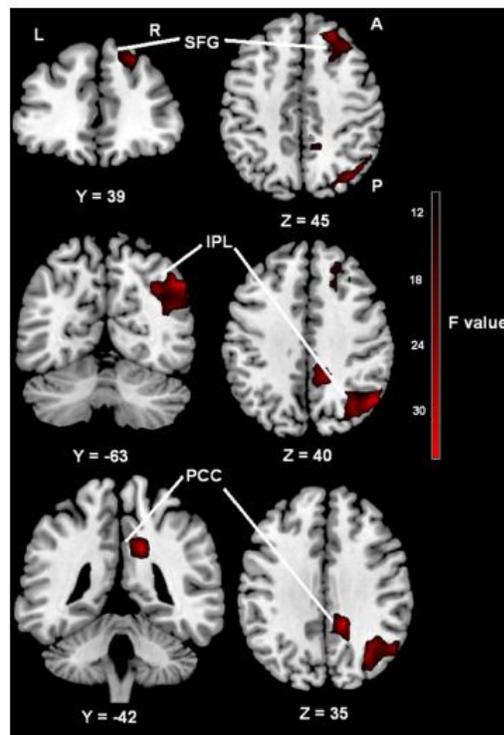
Figure 3.1: Default Mode Network consisting PCC, IPL and SFG [64]

2.5 Pattern Recognition

In machine learning, pattern recognition is defined as automatic detection of patterns of the input data and uses these patterns to classify or describe the data [21]. In literature many pattern recognition methods has been developed and applied to various fields such as computer vision, signal and image processing, medicine, psychology and finance. Recently pattern recognition has received much attention in brain connectivity researches where the correlations between different regions of brain have been identified as patterns. Craddock et al., [22] used pattern recognition to predict the disease state from rsFC. Shen et al., [23] used rsFC data to detect schizophrenia disease by pattern recognition methods.

Most of pattern recognition methods consist of following parts:
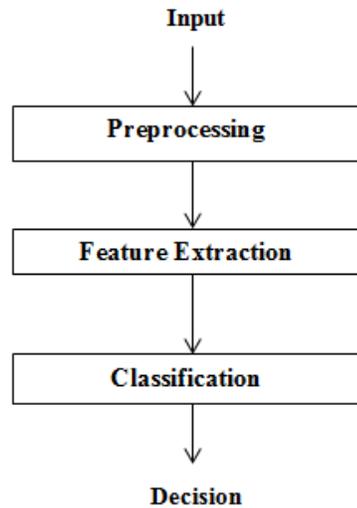
Figure 2.4: Pattern recognition components

2.5.1 Preprocessing

Preprocessing is a required step in pattern recognition for improvement of data quality. The preprocessing steps applied on fMRI data are explained in chapter 3 in detail.

2.5.2 Feature Extraction

In pattern recognition, the selection of proper feature extraction method is the key factor of obtaining high classification performance [24]. Here the question is which feature extraction method is appropriate for resting-state functional connectivity (rsFC) discriminative analysis?
In fMRI studies, some feature extraction methods such as dimensionality reduction methods (PCA and LDA) are frequently used to reduce the dimensionality of the data. Since the dimensionality of rsFC data is much more than the number of samples, these methods is applied to data to reduce its dimensionality. For instance, Tang et al. [25] used PCA dimensionality reduction method to identify schizophrenia patients using rsFC data. Another feature extraction methods that used for fMRI data analysis are similarity measurement methods that Shinkareva et al., [26] used cosine similarity measurement to compute the similarity scores between connectivity matrices to classify them.

In this study instead of using fMRI images directly, the images computed by functional connectivity are used. Since the size of the fMRI images and so the size of the functional connectivity images are very high, they need to be subsampled before using as feature vectors. Furthermore, due to difficulties arising from high dimensionality, dimension reduction methods

such as PCA, LDA are used. In addition to obtaining of feature vectors in this way, also the similarities of connectivity image of individual subjects to the means of connectivity images of healthy and patient subjects are used as features.

In the following, first the distance (similarity) measures used are explained and then the dimensionality reduction algorithms are presented.

2.5.2.1 Similarity Measures

A. Dot Product Measurement

The simplest similarity measure of two vectors is dot product which finds the Euclidean distance between these vectors.

Assume two d-dimensional vectors $\mathbf{x}, \mathbf{y}$, the dot product of them is defined as [27]:

$$dot(\mathbf{x}, \mathbf{y}) = \mathbf{x}.\mathbf{y} = \sum_{k=1}^{d} x_k y_k \tag{2.2}$$

In literature dot product is a method to find the measure of similarity between vectors. For instance, Yang et al. [28] used dot product to compute the similarity between multivariate time series. Garret and Kovacevic [29] used inner product measurement method to measure the similarity between different brain patterns .

Conventionally, the classification is done directly by choosing the highest similarity, however in this thesis it is used in feature extraction of feature vectors as explained in chapter 4.

B. Cosine Similarity Measurement

There are several ways to measure the similarity between vectors, The common method frequently used for similarity measurement in data mining and information retrieval [30] is the cosine similarity. The cosine similarity measure is the cosine of the angle between the vectors as shown in Figure 2.5 and computed using the Euclidean dot product of vectors and their magnitude. Experimental results demonstrated that the feature selection algorithms that select the significant features needed for classification are complex and expensive to implement. Instead cosine similarity method occupy small storage space, lower the computation cost, make process faster and give improved performance with negligible error rate thanks to produce small features [31]. Assume two d-dimensional vectors $\mathbf{x}, \mathbf{y}$, the cosine similarity of them is defined as:

$$\cos(x, y) = \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\sum_{i=1}^{d} x_i y_i}{\sqrt{\sum_{i=1}^{d} x_i^2} \sqrt{\sum_{i=1}^{d} y_i^2}} \tag{2.3}$$
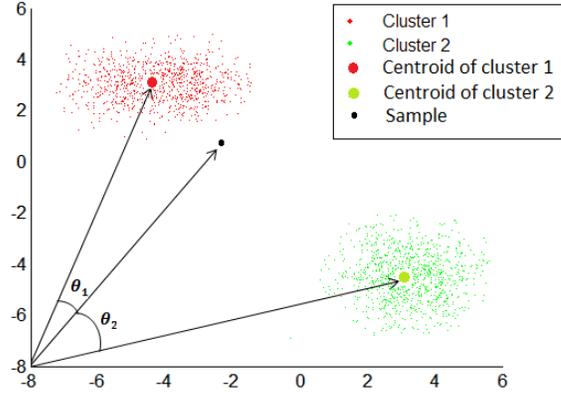
Figure 2.5: The cosine measure between a sample and mean of two classes of clusters is represented by $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ values

Baharudin and Muflikhah [32] used Cosine similarity in document clustering to measure the similarity between two different document vectors and resulted in dimension reduction of the document matrix. Movellan evaluated the performance of the face recognition problem calculating cosine similarity between coefficient vectors of test and train sets in k-nn algorithm. Borge-Holthoefer [33] in 2011 used this approach in a medical study for patients suffer Alzheimer's disease. Shinkareva and Gudkov [26] also used cosine similarity to classify brain functional connectivity matrices. In their paper the cosine similarity between vectors of train and test connectivity matrices is computed and two pairs with best similarity score are selected. Mitchell and Shinkareva [34] applied this technique to consider human brain activity when predicted and actual images are introduced. In this paper the cosine similarity between vectors of voxel values of two predicted and observed images is calculated. According to ranked similarity scores, two pairs of predicted and actual images that brain can match them are selected.

In this thesis the cosine similarity is used as the feature extraction method which will be explained in chapters 3 and 4.

C. Correlation Measurement

Correlation similarity measure is another similarity measure method closely similar to the dot product similarity with a difference that in this method the mean value of the data vector is subtracted from itself. Suppose $\mathbf{x}, \mathbf{y} \in R^d$ are input vectors then $\bar{x}, \bar{y}$ are mean values of vectors in $\mathbf{x}, \mathbf{y}$ respectively such that [27]:

$$\bar{x} = \frac{1}{d}\sum_{i=1}^{d} x_i \qquad i = 1, \dots, n \tag{2.4}$$

$$\bar{y} = \frac{1}{d}\sum_{i=1}^{d} y_i \tag{2.5}$$

$$corr(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{d}(x_i - \bar{x})(y_i - \bar{y}) \tag{2.6}$$

12

2.5.2.2 Dimensionality Reduction

A. Principal Component Analysis (PCA)

In classification problems, reducing dimensionality of the data is an essential part before data analysis [35]. PCA is a powerful orthogonal linear transformation technique in which the coordinate system of high dimensional data transforms to new coordinate system called principal components (PCs) such that the largest variance (which has the maximum information) by a chosen projection of the data comes to lie on the first coordinate (called the first principal component), the second largest variance lie on the second coordinate and importantly the second PC is uncorrelated with the first PC and they are orthogonal since the covariance matrix is symmetric [36]. Further principal components exhibit decreasing variance and are uncorrelated with all other principal components [37], [38].

Reversibility property of PCA (the original data can be reconstructed from principal components) makes it a useful method for data reduction, noise rejection, visualization and data comparison among other things. PCA approach has been extensively applied to diverse fields such as image processing, face recognition [39], [40], Computer vision, machine learning [41], [42], [43], data compression, image de-noising [44] and [45] Eigen-image analysis [46]. Lately, neuroscientists have been started to use PCA-based techniques in analysis of functional connectivity, since they are "data-driven methods" which means that the biological aspects of structure such as regions involved in structural connectivity have no significance [47].

PCA Algorithm:
Assume a sample set of $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ where each sample is a d-dimensional vector.

Step1:
The d-dimensional mean of the sample set is obtained as [48]:

$$m = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k \tag{2.7}$$

Step 2:
The $d \times d$ covariance matrix $\mathbf{C}$ is calculated for the sample set such that:

$$\mathbf{C} = \sum_{k=1}^{n}(\mathbf{x}_k - m)(\mathbf{x}_k - m)^t \tag{2.8}$$

Step 3:
The eigenvectors $\mathbf{e}$ and their corresponding eigenvalues λ of covariance matrix $\mathbf{C}$ are computed and arranged in decreasing order of eigenvalues as shown in Figure 2.6 .So that the first eigenvector $\mathbf{e}_1$ that consists of the maximum variance is called first principal component, the second eigenvectors show the second largest variance and so on. The eigenvalue equation is defined as:

$$\lambda\mathbf{e} = \mathbf{C}\mathbf{e} \tag{2.9}$$
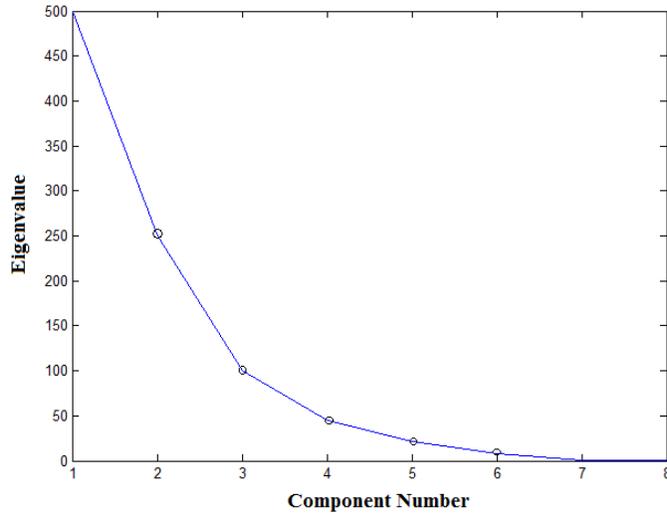
Figure 2.6: Eigenvalues in decreasing order

Step 4:

The M largest eigenvectors $\{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ of covariance matrix correspond to the largest eigenvalues $\{\lambda_1, \dots, \lambda_M\}$ are chosen. So there will be M principal components.

If there is only one significant eigenvector the data is linearly and orthogonally projected onto a line (one-dimensional space) in the direction of the largest eigenvector $\mathbf{e}_1$ known as principal axes as shown in Figure 2.7, but in the case of the M significant eigenvectors, the data is projected onto the M-dimensional space known as principal subspace.
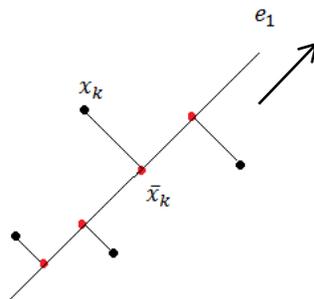
Figure 2.7: Orthogonal projection of the d-dimensional sample $\mathbf{x}_k$ onto a line in direction of the $\mathbf{e}_1$ [7]

Step 5:

The matrix **A** of size $d \times M$ where $M < d$ is formed with M largest eigenvectors in columns.

14

Step 6:

The dimensionality of any sample $\mathbf{x} \in R^d$ is reduced to $M < d$ by formula $\mathbf{x}_r = \mathbf{A}\mathbf{x}$ where $\mathbf{x}_r \in R^M$, is the M-dimensional vector.

In summary, the dimensionality of feature space is reduced by considering only the significant eigenvectors or equivalently directions through which the variance matrix is largest [21, 48].

B. Kernel Principal Component Analysis (KPCA)

Kernel Principal Component Analysis (KPCA) has been introduced recently as a nonlinear expansion of PCA for handling nonlinear problems [35]. The KPCA principle is based on nonlinearly mapping the input space into a feature space by means of nonlinear transformations as kernel functions (e.g. dot product, Gaussian kernels, Radial Basis functions) and then efficiently computes the principal components in high-dimensional feature space [49]. Since KPCA uses alternative kernels in its algorithm, it is able to handle a wide category of nonlinearities.

In fMRI-based studies, this approach has been used as a dimensionality reduction method for instance to diagnosis the mental disorders as in paper by Sidhu [50]. He used KPCA to distinguish between the individuals with Attention-Deficit Hyperactivity Disorder and healthy ones.

Kernel PCA Algorithm:

As in PCA method we assume a set of n samples $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ where each sample is a d-dimensional vector and the sample mean is subtracted from each of the sample vectors.

Step 1:

Each d-dimensional sample vector $\mathbf{x}_k$ is projected into $\varphi(\mathbf{x}_k)$ point in M-dimensional feature space by nonlinear transformation $\varphi(\mathbf{x})$.

$$\varphi : R^d \rightarrow f^M$$

The kernel function is defined as:

$$k(\mathbf{x_i}, \mathbf{x_j}) = \varphi(\mathbf{x_i})^T \varphi(\mathbf{x_j}) \tag{2.10}$$

$\mathbf{x}_i, \mathbf{x}_j$ are sample vectors. There are different types of kernels for example the Gaussian kernel function is in the form of:

$$k(\mathbf{x_i}, \mathbf{x_j}) = exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{0.1}\right) \tag{2.11}$$

Step 2:

This time the conventional PCA is applied in the feature space to compute the covariance matrix. Subsequently, the eigenvectors of covariance matrix and their corresponding eigenvalues and principal components are computed using the kernel function [21].
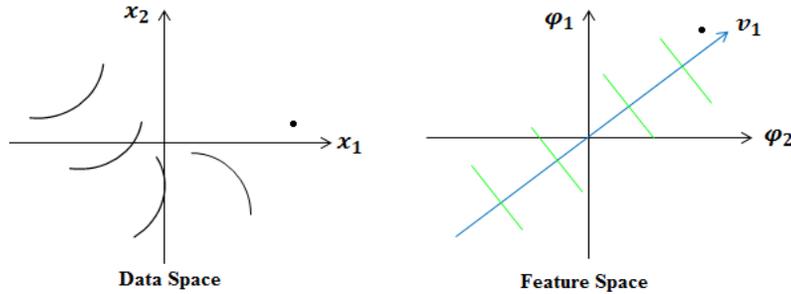
Figure 2.8: Kernel PCA [7]

The Figure 2.8 shows the nonlinear transformation of sample vectors onto feature space. The blue vector $\mathbf{v}_1$ represents the first principal component extracted after the PCA is performed in feature space. Since the nonlinear PCs cannot be shown in data space by a vector, the corresponding linear projections onto PCs is shown in feature space as green lines.

C. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is one of the most popular dimensionality reduction algorithms [48, 51] first developed by Robert Fisher [52] in 1936 for "taxonomic classification" [53]. The purpose of using this method is determining project directions on which data points that belong to the same class become close to each other while the data points belonging to different classes become far from each other. In other words it seeks an optimal linear subspace required for discrimination [54]. The optimal way to compute projection axes can be applying an eigen-decomposition on the scatter matrices of the given training data [55]. It finds the projection directions in a way that the between-class variance S_B is maximized relative to the within-class variance S_W. Subsequently; the dimension of the data is reduced by projecting the data to these directions, which are called discriminant variables. Discriminant variables applied as inputs to different classification methods, like k -nearest neighborhood (k-NN), and support vector machines (SVM) [56].

LDA has been widely applied to pattern recognition [36] problems such as text processing [57], face recognition [53, 58]. Baker and Nayar have developed a two class linear discriminant pattern rejection theory [59]. Cui, Swets, and Weng [60] applied LDA for recognizing hand motions.

LDA algorithm (2-Class):

Assume a set of n sample vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ where each sample is a d-dimensional vector. If we consider a two-class problem, a subset of n_1 sample vectors are labeled as first class c_1 and a subset of n_2 sample vectors are labeled as second class c_2 [48].

Step1:
The mean vectors of two classes are computed.

$$m_1 = \frac{1}{n_1}\sum_{\mathbf{x}\in c_1}\mathbf{x} \quad , \quad m_2 = \frac{1}{n_2}\sum_{\mathbf{x}\in c_2}\mathbf{x} \tag{2.12}$$

Step2:
The scatter matrices of two classes are computed.

$$\mathbf{S}_1 = \sum_{\mathbf{x}\in c_1}(\mathbf{x}-m_1)(\mathbf{x}-m_1)^t \quad , \quad \mathbf{S}_2 = \sum_{\mathbf{x}\in c_2}(\mathbf{x}-m_2)(\mathbf{x}-m_2)^t \tag{2.13}$$

Step3:
The within-class scatter matrix and the between class scatter matrix is defined as:

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 \tag{2.14}$$

$$\mathbf{S}_B = (m_1 - m_2)(m_1 - m_2)^t \tag{2.15}$$

Step4:
So the Fisher criterion function (Rayleigh Quotient) is defined as:

$$J(\mathbf{w}) = \frac{\mathbf{w}^t\mathbf{S}_B\mathbf{w}}{\mathbf{w}^t\mathbf{S}_w\mathbf{w}} \tag{2.16}$$

Where the vector $\mathbf{w}$ represents the direction of the line onto which the samples are projected.

Step5:
In order to obtain an optimal separation between two classes the $J(.)$ function should be maximum and the vector $\mathbf{w}$ that can optimize the criterion function should satisfy $\mathbf{S}_B\mathbf{w} = \gamma\mathbf{S}_w\mathbf{w}$, where γ is a constant value.

If the $\mathbf{S}_w$ is non-singular the eigenvalue problem is defined as:

$$\mathbf{S}_w^{-1}\mathbf{S}_B\mathbf{w} = \gamma\mathbf{w} \tag{2.17}$$

where $\mathbf{S}_B\mathbf{w}$ represents the direction of distance between mean of the samples $(m_1 - m_2)$ so the eigenvalue problem can be rewritten as a linear function:

$$\mathbf{w} = \mathbf{S}_w^{-1}(m_1 - m_2) \tag{2.18}$$

By this way the vector **w** that maximizes the criterion function is calculated. The equation 2.17 maps d-dimensional input space onto one-dimensional space.

Step6:
Finally the dimensionality of any sample $\mathbf{x} \in R^d$ is reduced by formula $\mathbf{x}_r = \mathbf{wx}$ where $\mathbf{x}_r \in R$, is the *one*-dimensional vector.
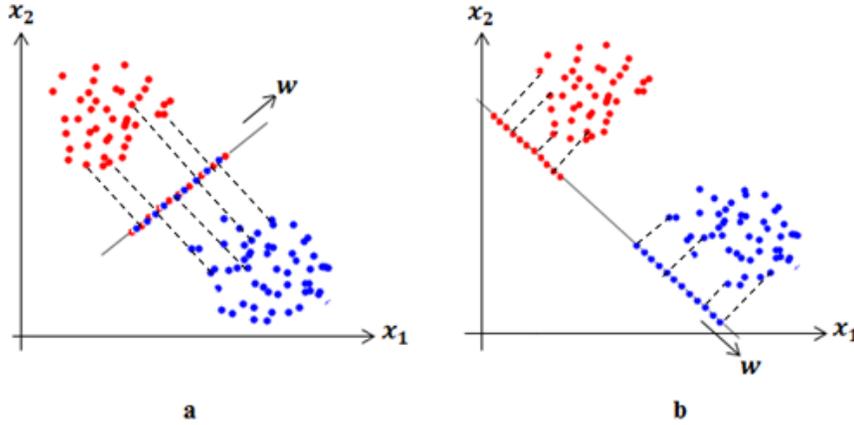
Figure 2.9: Projection of samples onto lines with different directions in LDA.
(a) The worst direction of **w**. (b) The best direction of **w**.

Figure 2.9 shows the projection of two classes of samples onto lines with different direction*s* **w**. The right-hand figure (b) represents the best discrimination between classes.

Generalized LDA (C-class):

LDA consists of c-1 discriminating functions. Duly the d-dimensional input space embeds onto $(c$-1) dimensional space where d$\geq c$. Also the vector **w** is converted to transform matrix **W** whose columns are eigenvectors, in this case criterion function will have the form:

$$J(\mathbf{W}) = \frac{\mathbf{w}^t S_B \mathbf{w}}{\mathbf{w}^t S_W \mathbf{w}} \tag{2.19}$$

In this way, the eigenvalues γ_i will be found as roots of the polynomial for eigenvectors $\mathbf{w}_i$:

$$\mathbf{S}_B \mathbf{w}_i = \gamma_i \mathbf{S}_w \mathbf{w}_i \tag{2.20}$$

$$|\mathbf{S}_B - \gamma_i \mathbf{S}_w| = 0 \tag{2.21}$$

$$(\mathbf{S}_B - \gamma_i \mathbf{S}_w)\mathbf{w}_i = 0 \tag{2.22}$$

18

Since there are only (*c*-1) independent matrices in between-class scatter matrix $\mathbf{S}_B$, there will be (*c*-1) nonzero eigenvalues.

Regularized LDA:

A critical condition to stabilize LDA solution is that the within-class scatter matrix should be nonsingular [54]. But this condition does not hold in the case of HDLSS data, in which dimensions of feature vectors are larger than the sample size $n \ll d$, which occur frequently in real-world problems such as gene expression analysis, face recognition and medical imaging. For HDLSS data, scatter matrices are singular so LDA is not applicable [56].

Some supplementary preprocessing methods have been introduced for this problem. The most preferred one is to apply dimensionality reduction in two steps. For instance, first step includes known methods as PCA or SVD. Swets and Weng[56], Belhumeur [53] and Torkkola [57] have implemented PCA or SVD before LDA for face recognition and document classification, respectively. Another way [55] to deal with the singularity of scatter matrix is to apply regularization, by adding some constant values to the diagonal elements of $\mathbf{S}_W$ as $\mathbf{S}_W + \alpha\mathbf{I}$, for some $\alpha > 0$, so scatter matrix become nonsingular. This approach is called Regularized Discriminant Analysis (RDA) [61, 62].

2.5.3 Classification

The second step of pattern recognition is the classification step where the feature vectors obtained from fMRI data from models of previous step is used as the input. Each classification method consists of two phases. A classifier is trained using training set in first phase and then the performance of the classifier is evaluated by testing the new input data called test data in the second phase. In following sections types of classifiers will be described.

2.5.3.1 Support Vector Machine

Support Vector Machine is a statistical pattern recognition classifier based on Vapnik statistical learning theory [63]. The SVM algorithm separates classes by proposing a decision hyper surface. The minimum distance from the separating boundary to the closest training samples is called margin. Support vectors are the training feature vectors located on the margin, and since classifying of these vectors is very difficult theoretically, they can determine the position of the hyper surface.

What makes SVM an exclusive method is its capability in selecting most important training samples for classification and using kernel functions with nonlinear boundaries for fine scaling high dimensions [64].

The SVM method has been used in many classification studies. Specially, in recent years it is widely applied in fMRI data analysis [65] to discriminate mental patients from healthy control subjects using the structural and functional differences of their brain [66-69]. SVM seeks statistical patterns of labeled fMRI training data sets helpful for identifying between brain regions, and select discriminating features [70] then map these patterns to a subject's cognitive states. Fan and Shen [67] used a classifier based on SVM to discriminate schizophrenia patients from healthy controls. Newly, resting-state functional connectivity data of the brain is used as classification components. Zhu [71] used rs-FC to distinguish the hyperactivity disorder and attention-deficit patients from normal controls.

In high dimensional low sample size (HDSS) data such as fMRI data, using a kernel classifier is preferred instead of a linear classifier [23, 66, 72].

Linear SVM Algorithm:

Assume a linearly separable training data set of n samples $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ where each sample is a d-dimensional vector with corresponding target classes $\{c_1, \dots, c_n\}$ where $c_n \in \{+1, -1\}$ [48].

Step 1:
A linear model of a 2-class classification problem is defined as [21, 48]:

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \tag{2.23}$$

Where $\mathbf{x}$ denotes the input vector, $\varphi(\mathbf{x})$ is a fixed transformation vector, $\mathbf{w}$ is the weight vector, b is a scalar parameter. The sign of $y(\mathbf{x})$ determines to which class the test data is belong.

Step 2:
Find $\mathbf{w}$ and b parameters for linear model such that:

$$\begin{cases} y(\mathbf{x_k}) > 0 & if\ c_k = +1 \\ y(\mathbf{x_k}) < 0 & if\ c_k = -1 \end{cases} ,\ k = 1, \dots, n \tag{2.24}$$

Step 3:
Since there are many solutions for $\mathbf{w}$ and b parameters, the svm uses margin to find the parameters that create minimum generalization error. By definition "the margin is the smallest distance between decision boundary and any of the sample vectors" which is calculated by $\dfrac{y(\mathbf{x})}{\|\mathbf{w}\|}$.

The Figure 2.10 shows the perpendicular distance between decision boundaries defined by $y(\mathbf{x}) = 0$ and sample point x. In the other hand, only the solutions that lead the correct classification ($c_k y(\mathbf{x}_k) > 0$) are considered, so the margin is rewritten as:

$$\frac{c_k\ y(\mathbf{x}_k)}{\|\mathbf{w}\|}$$

20
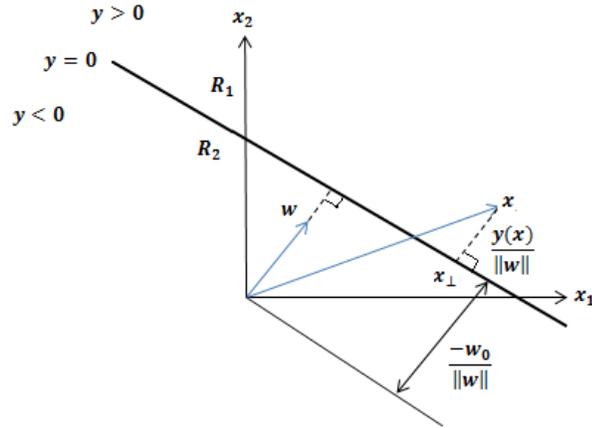
Figure 2.10: Linear SVM geometry

The Figure 2.10 shows the geometry of linear svm for two-class regions R_1, R_2. The vector w is perpendicular to the hyper surface shown with bold line and the shift of the hyper surface from the origin is handled by w_0 parameter. The orthogonal distance between the hyper surface and any sample point x is called margin.
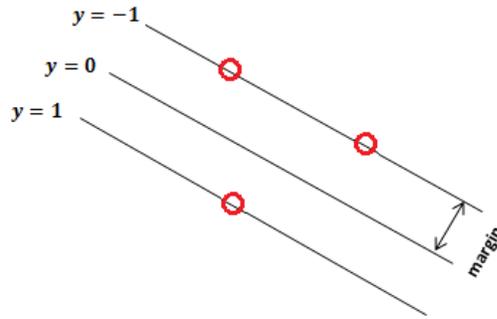
Figure 2.11: Linear SVM

The Figure 2.11 shows the margin and support vectors. The position of the particular hyper surface is decided by support vectors.

Step 4:

In order to find a particular hyper surface, the margin is maximized by optimizing $\mathbf{w}$ and b parameters.so considering equation 2.22 and margin definition the maximum margin is calculated:

$$arg\ max_{\mathbf{w},b}\left\{\frac{1}{\|\mathbf{w}\|}min\{c_k(\mathbf{w}^T\varphi(\mathbf{x})+b)\}\right\},\quad k=1,...,n \qquad (2.25)$$

Solving the equation 2.24 gives the solution for maximum margin. Since the maximum margin equation is complicated to solve, it is simplified by assuming that the data points next to the hyper surface (support vectors) satisfy $c_k(\mathbf{w}^T\varphi(\mathbf{x})+b)=1$, in this condition all data points will satisfy the $c_k(\mathbf{w}^T\varphi(\mathbf{x})+b)\geq 1$, so the maximum margin will have the form:

$$maximum\ margin=\ arg\ max_{w,b}\left\{\frac{1}{\|\mathbf{w}\|}\right\}\sim arg\ min_{\mathbf{w},b}\|\mathbf{w}\|^2 \qquad (2.26)$$

It seems that the parameter b does not contribute in optimization of the margin but it is determined inevitably during optimization since "changes to $\|\mathbf{w}\|$ be compensated by changes to b".

Step 5:

After the parameters w and b are maximized, decision is made such that if samples satisfy $c_k y(\mathbf{x}_k)>0$ for all k to they are classified correctly.

Nonlinear SVM Algorithm:

Most of the time it is much more complex than the linear svm can be used. In such cases the nonlinear svm is preferred to implement in which the $\varphi(\mathbf{x})$ transformation vector of equation $y(\mathbf{x})=\ \mathbf{w}^T\varphi(\mathbf{x})+b$ is nonlinear.

2.5.3.2 Gaussian mixture Model

In pattern recognition phase the GMM is a well-known classification algorithm based on clustering [73] that has been frequently applied to various fields like speech detection, instrument classification, image segmentation [74]. This technique specially has been used in medical diagnosis applications like neurological disorders as Alzheimer's or schizophrenic disease [75]. Some studies reported that GMMs have been applied to brain imaging applications and fMRI data analysis for modeling the fMRI images or characterize activation paradigms from fMRI data to discover voxels related to BOLD activation regions [73]. Also using GMMs, the grey matter distributions of the brain are approximated so the brain image can be modeled [76]. In addition, Gaussian mixtures are used as a framework to analyze the resting-state activation regions of the brain.
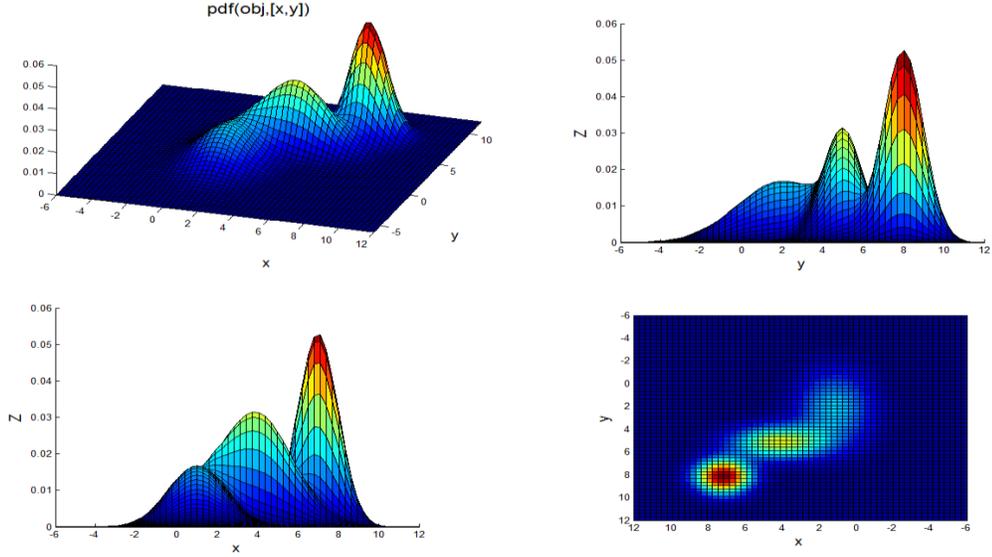
Figure 2.12: Gaussian Mixture Model

The Gaussian mixture model which is often used for data clustering is composed of the weighted combination of the Gaussian components to model the distribution of patterns. By assuming that the Gaussian density components and their weights are known also the parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_M\}$ and the label of classes are unknown therefore the probability density function for samples is defined as [48]:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^{M} w_i \, p(\mathbf{x}|w_i, \boldsymbol{\theta}_i) \tag{2.27}$$

$$\boldsymbol{\theta} = \{w_i, \mathbf{C_i}, \boldsymbol{\mu_i}\}, \qquad i = 1, 2, \dots, M \tag{2.28}$$

Where $p(\mathbf{x}|\boldsymbol{\theta})$: Gaussian mixture model or mixture density
$p(\mathbf{x}|w_i, \boldsymbol{\theta}_i))$: Gaussian components
w_i: Weight of Gaussian components

$\boldsymbol{\theta}$, is a set of parameter vectors of mixture model includes $\mathbf{C}_i$, covariance matrix of components that can be full rank or diagonal, $\boldsymbol{\mu}_i$, mean vector of components and w_i weight of Gaussian components i. M represents the total number of components. If the parameter vector $\boldsymbol{\theta}$ is known, the GMM can be decomposed into its components so the classes can be discriminated using a classifier. So, the aim is to estimate the $\boldsymbol{\theta}$ using the samples. First of all the likelihood of the data set $\mathbf{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ of n samples is defined as:

$$p(\mathbf{D}|\boldsymbol{\theta}) = \prod_{k=1}^{n} p(\mathbf{x}_k|\boldsymbol{\theta}) \tag{2.29}$$

The maximum likelihood estimate of the value $\boldsymbol{\theta}$ is a parameter called $\hat{\boldsymbol{\theta}}$ that maximizes likelihood. In the same time $\hat{\boldsymbol{\theta}}$, maximizes the logarithm of the $p(\mathbf{D}|\boldsymbol{\theta})$, such that:

$$l(\boldsymbol{\theta}) = \ln(p(\mathbf{D}|\boldsymbol{\theta}) \tag{2.30}$$

$$\hat{\boldsymbol{\theta}} = arg_{\boldsymbol{\theta}}max(l(\boldsymbol{\theta})) \tag{2.31}$$

Expectation maximization is an iterative algorithm that finds the $\hat{\boldsymbol{\theta}}$ parameter by iteratively estimation of log-likelihood. In the each iteration of the EM algorithm, there are two steps, the expectation step or the E-step and the maximization step or the M-step. The EM algorithm can be described as following steps:

Step 1: Input data $\mathbf{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$
Step 2: Initialize model parameter $\boldsymbol{\theta}^0$ and convergence criterion T
Step 3 (E-step): Compute expectation value of the log-likelihood $l(\boldsymbol{\theta})$ of the data for θ^i

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^i) = E[\ln p(\mathbf{D}|\boldsymbol{\theta})|\boldsymbol{\theta}^i] = E[l(\boldsymbol{\theta})] \tag{2.31}$$

Here $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^i)$ is the function of $\boldsymbol{\theta}$ with $\boldsymbol{\theta}^i$ assumed fix.
Step 4 (M-step): Estimate $\boldsymbol{\theta}^{i+1}$ such that maximizes the log-likelihood
Steps 3 and 4 iterate until $Q(\boldsymbol{\theta}^{i+1}|\boldsymbol{\theta}^i) - Q(\boldsymbol{\theta}^i|\boldsymbol{\theta}^{i-1}) \leq T$ then the parameter $\boldsymbol{\theta}^{i+1}$ which is called $\hat{\boldsymbol{\theta}}$ is found such that maximizes the log-likelihood function.

2.6 Leave-One-Out Cross-Validation (LOO-CV)

Cross validation is a method to estimate the generalization performance of a classifier. This method divide data samples into two sets: first, the training set that trains the classifier. Second, test set that estimate the generalization error. The generalization of cross validation is k-fold cross validation where samples are randomly split into k equal parts. The classifier is trained k times so that within each iteration, k-1 parts are used to train the classifier and the remaining part is left out as test set to test the validation of the classifier and find the generalization error. Then the mean of k errors obtained from k iterations gives us performance estimation of the classifier.

The special case of k-fold cross validation where the k equals the number of samples is called Leave-one-out cross-validation (LOO-CV), so that this method trains the classifier k times in each iteration, all the samples except one are used as training set and the excluded one is used as test set to calculate generalization error. Reviewing literature , The LOO-CV technique is used frequently in disease detection and other classification problems for instance, Shen et al. [23] used this method to evaluate the performance of the classifier in discriminative analysis of schizophrenia disease. Zhu et al. [77] also implemented this approach to detect attention deficit/ hyperactivity disorders of brain at resting state.

CHAPTER 3

MATERIALS AND METHODS

This chapter consists of three main sections. In the first section the fMRI data collection and preprocessing steps which consist of image enhancement and registration, ROI analysis, functional connectivity analysis are discussed. In the second section the feature extraction methods including dimensionality reduction and similarity measurement methods are discussed. In the last section the classification methods as SVM and GMM used to discriminate between classes are discussed in detail.

3.1 FMRI Data Collection and Analysis

3.1.1 Data Collection

The fMRI data was collected by Prof. Dr. Metehan Çiçek's neuroimaging and brain research team with 1.5 Tesla Siemens Magnetom Symphony Maestro Class MRI system (Siemens, Erlangen Germany)'' at Integra Imaging Center, Ankara [18].

T1-weighted functional scans with high resolution were obtained applying echo-planar sequence in axial plane. The functional imaging parameters are as follows: time to repeat (TR) = 4940 ms, time to echo (TE) = 36 ms, slice number = 36, slice thickness = 5 mm, matrix size= 64×64, field of view (FOV) = 224, flip angle (FA)= $90^{°}$.To keep equilibrium of signal, the first four images of all functional sequences were removed [78]

3.1.2 Participants

The total number of 24 right-handed volunteers with minimum age of 18 years took part in resting-state fMRI data collection. 12 (six male and six female) of participants were healthy individuals similar in terms of sex and education level (control group) and the other12 (six male and six female) participants had Obsessive Compulsive Disorder (OCD) last of 6 months to7 years. Eight OCD patients had cleaning obsession, three of them had checking obsession and one was with damaging obsession. Neither neural system nor psychiatric disorder was detected in control group attendants and no psychiatric disorder was recognized in patients other than OCD [2, 18].

3.1.3 Experimental Process During Data Acquisition

Participants were placed in the MRI scanner. The original experiment included four functional sequences and in each run participants were given a figure for seconds and requested to perform imagination (imagine the figure), suppression (omit the figure) and erasing (erase the figure in mind) tasks and between these tasks they rest with closed eyes and think freely which is called free imagination . At the end of the experiment, 300 fMRI images were obtained that only 60 of them contained free imagination images [18]. Resting-state functional connectivity (rsFC) analysis performed on the resting (free imagination) data

3.2 Preprocessing

3.2.1 Image Enhancement and Registration

fMRI data was preprocessed using "SPM8 software (Welcome Department of Cognitive Neurology, London, UK)" in a MATLAB context (Math works, Sherborn, Mass., USA). The applied preprocessing steps [2] consists of:

1. Slice timing
2. Realignment
3. Co-registration
4. Normalization
5. Smoothing

Each step will be explained in detail in the following.

Step 1. Slice-timing Correction
This step corrects the delays or advances that happen in the acquisition timing order of different slices by fitting all slices to a standard slice so that activity of each point will be the same for all slices in time.
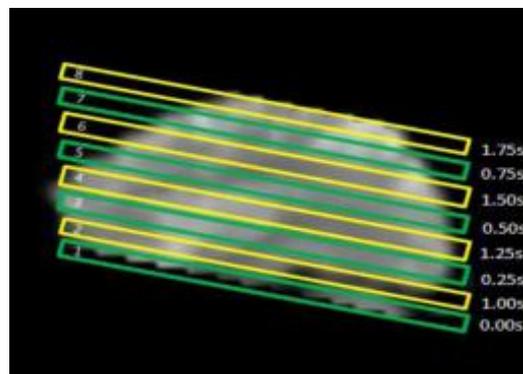
Figure 3.2: Slice timing. According to acquisition times in the right-hand the slices are acquired in 1, 3,5,7,2,4,6,8 order [2]

Step 2. Motion Correction

The subject's head movement during the fMRI scans causes the position and coordinate of the brain change in functional images. The realignment method reduce the misalignment among functional images by realigning the FMRI time series to a standard coordinate so that all functional images will have the identical coordinate as shown in Figure 3.2. By result the source of the signal in all voxel time-series will be same.
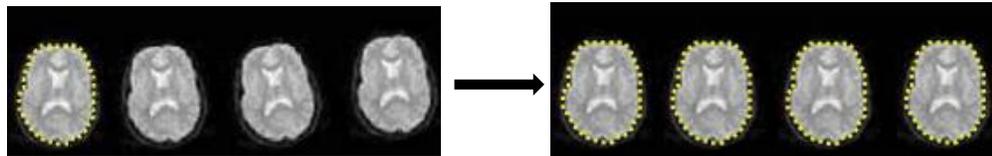
Figure 3.3: Realignment [66]

Step 3. Co-registration

In this step, the functional images of low-spatial resolution are aligned with structural fMRI images of high resolution.

Step 4. Normalization

The brain image of individuals differs from each other in terms of size and shape. However, there should be a reference for brain images to be able to compare them in aspect of functional activity. So, all brain images (functional and structural images) must be mapped onto a template brain image as shown in Figure 3.3 in other words spatially normalized to template image whose reference is Talairach and Tournoux [79] so that these normalized images can be compared with each other.
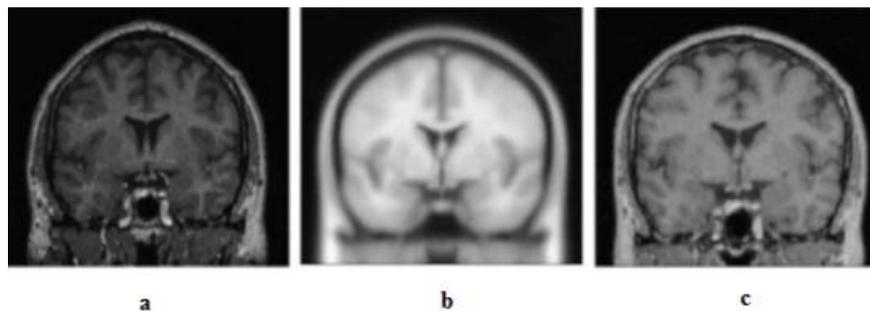
Figure 3.4: Normalization: (a) High resolution co-registered image, (b) Template image, (c) Resulted normalized image

Step 5. Smoothing

This method increases signal-to-noise ratio by reducing the noise. The functional image convolves with a Gaussian kernel in order to remove the noise. The result is a blurred image as shown in Figure 3.4.

In this study the 9-mm full-width half-maximum (FWHM) Gaussian kernel is used for smoothing.
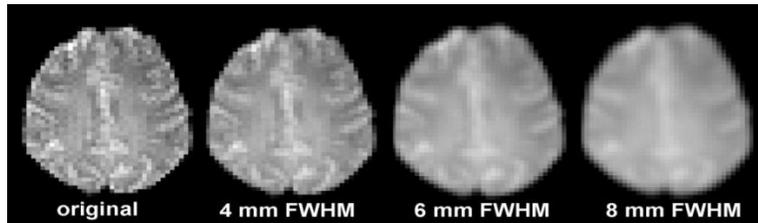
Figure 3.5: Smoothing

3.2.2 ROI Analysis

As explained in section 2.4 parts of DMN of the brain such as PCC, LIPL, RIPL regions in which presence of functional abnormalities is proved in OCD are considered as ROIs. The functional correlations or disassociations between mentioned ROIs and other voxels of the brain are calculated as explained in following section.

3.2.3 Functional Connectivity Analysis of FMRI data

Resting-state functional MRI analysis was conducted with CONN software [10, 80] which is a toolbox for functional connectivity estimation. Using CONN toolbox, we can analyze different types of the resting-state functional connectivity such as within ROIs (LIPL, RIPL and PCC) called ROI to ROI or between ROIs and other brain voxels (ROI or seed region to voxel).

The CONN toolbox needs some basic information to analyze the functional connectivity (FC) of the resting-state fMRI data including:

1. Functional data
2. Structural data
3. ROI definitions: Region of Interest (ROI) is a selected portion of voxels of an (functional) image recognized for specific task.

The input data of CONN is the preprocessed 3D fMRI data of size $91 \times 109 \times 91$ and the output data is supposed to be 3D FC data of the same size that represents the functional connectivity between PCC or LIPL or RIPL ROIs and all brain voxels as correlation coefficient values calculated by Pearson's correlation coefficient formula.

28

The successive steps are followed for functional connectivity analysis of preprocessed fMRI data using CONN toolbox as given below:

Step 1: Setup
During this step the sequential parts are defined:
- **Basic:** the experiment information including participant and data acquisition information.
- **Functional :** functional images are entered
- **Structural:** structural/anatomical images are entered
- **ROIs:** The required regions of interest are imported.
- **Condition:** the experimental condition (imagination, suppuration, erasing, resting) and its related durations and onsets are defined.
- **Covariates:** variables that can affect the experiment such as movement parameters, behavioral measures.
- **Options:** choose which analysis type you want to do(ROI_ROI, voxel-voxel, seed-voxel)
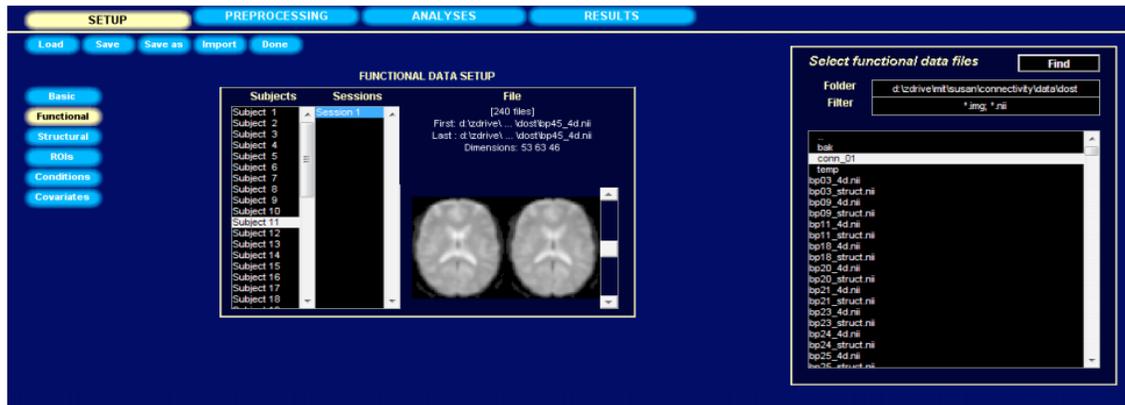
Figure 3.6: The first step of rsFC analysis using CONN toolbox [73]

Step 2: Band Pass Filtering
The functional data is filtered by using a band-pass filter to remove very low frequency (0.004 Hz < f < 0.08 Hz) distracts such as BOLD oscillations and motion parameters that affect voxels and ROIs.

Step 3: First Level Analysis (Functional Connectivity Analysis)
The sources of interest (seeds) are defined and selected afterwards a selected type of connectivity analysis (ROI-ROI or voxel-voxel or seed-voxel) is applied to all subjects using the selected sources. Eventually, at the end of the first level analysis correlation coefficients are computed by Pearson's correlation coefficient method and the connectivity (correlation) map

with corresponding coefficients is constructed for each subject and ROI in resting-state condition.
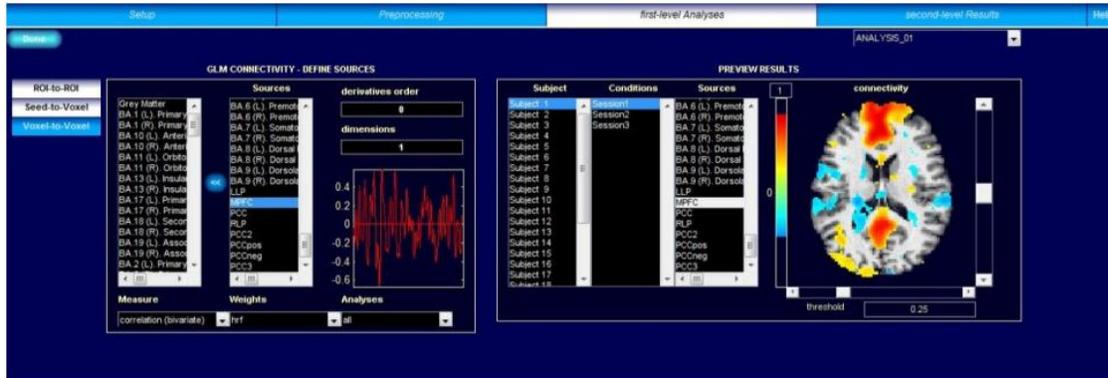
Figure 3.7: First level analysis and Functional connectivity results [73]

At the end of functional connectivity analysis the FC images provided for PCC, LIPL and RIPL ROIs are obtained as inputs for successive processes described in the following.

3.3 Determination of Training and Test Sets by Double- Leave One Out-Cross Validation (D-LOO-CV)

Two nested loops were created for LOO-CV process that we called D-LOO-CV where in the first (outer) LOO loop one healthy sample was excluded and in second (inner) LOO loop one patient sample was excluded. The main goal of such approach is to produce training and test sets with more samples since there are small numbers of samples obtained by single LOO in contrast with large number of features which may not be enough to train the classifier in classification phase.

Using only LOO-CV approach, there will be 22 train samples (11 healthy and 11 patient) for train the classifier and 2 test samples for each LOO iteration. In order to raise the number of training samples we generate the train and test sets as different combinations of healthy and patient samples called D-LOO-CV algorithm given in Figure 3.8. Using D-LOO-CV resulted in 12 (1×12) test samples and 132 (11×12) healthy and 132 patient train samples.

```
FOR each healthy sample, h
        FOR each patient sample, p
                Define F(k, h, p), as the D-LOO feature vector for subject k where h ∈ H, p ∈ P
                are left out for test.
                Determine training and test set such that:
                        Train(s) = {(F(k, h, p)|k ≠ s; (s = h)or(s = p); s ∈ S}
                            Healthy Test(s) = {(F(k, h, p)|k = s; s = h}
                Where both the train and test set are matrices of size m × n, where m is the
                number of   subjects and n is the number of features.
        END FOR
END FOR
```

Figure 3.8: Double Leave-One-Out algorithm for healthy subjects

The same algorithm is true for determining patient test set except that the order of loops should be changed and the patient test set would be:

$$Patient\ Test(s) = \{(\mathbf{F}(k, h, p)|k = s;\ s = p\} \tag{3.1}$$

If we assume healthy samples $H = \{h_i\,|\,i = 1, ...,12\}$ and patient samples $P = \{p_j\,|\,j = 1, ...,12\}$ then the test and train sets are selected as follows by using D-LOO_CV:

Table 3.1: Applying D-LOO-CV to generate train set and healthy test set

D-LOO-CV		
Iteration	Test set for healthy subject h_i	Healthy and patient subjects in Train set
1	$h = h_i, p = p_1$	$k \in \{H - h_i\} \cup \{P - p_1\}$
. . .	. . .	. . .
12	$h = h_i, p = p_{12}$	$k \in \{H - h_i\} \cup \{P - p_{12}\}$

Table 3.1 shows the D-LOO-CV in which only one healthy test sample and 11 healthy train and 11 patient samples are selected. After 12 iterations, there will be 12 healthy test samples and 132 healthy and 132 patient samples in training set in total. So by using D-LOO-CV instead LOO-CV we could increase the number of test samples from 1 to 12 and train samples from 22 to 264. Note that patient samples are useless while generating healthy test samples and they are only removed from train set to balance the number of healthy and patient samples in training set.

The patient test samples were obtained using the D-LOO-CV as shown in Table 3.2.

Table 3.2: Applying D-LOO-CV to generate train set and patient test set

D-LOO-CV		
Iteration	Test set for patient subject p_j	Healthy and patient subjects in Train set
1	$p = p_j, h = h_1$	$k \in \{H - h_1\} \cup \{P - p_j\}$
. . .	. . .	. . .
12	$p = p_j, h = h_{12}$	$k \in \{H - h_{12}\} \cup \{P - p_j\}$

3.4 Feature Extraction Methods

As demonstrated in Figure 1.1 two approaches were considered in this study to extract the significant features; the first approach was subsampling followed by dimensionality reduction of the 3D FC data using PCA, KPA, LDA methods which will be discussed in section 3.4.1. The second approach for feature extraction is get use of similarity measurement methods correlation, dot product and cosine similarity which will be described in detail in section 3.4.2.

3.4.1 Dimensionality Reduction Methods

As mentioned before, the dimensionality reduction methods reduce the dimensionality of the data using linear or nonlinear techniques and find out feature vectors that help classifier to separate the data optimally. In present study, we investigated dimensional reduction methods explained in chapter 2 previously. In high dimensional low sample size data like fMRI data where the number of the subjects are smaller than the number of the features, the dimension reduction methods may not give the desired results due to problems such as singularity, insufficient retention space and cost of computation. One way to overcome this problem is reducing the number of features in two steps. In first step, subsample the 3D FC data and then use the dimensional reduction methods. The other way to settle high-dimensionality problem is to employ feature extraction methods such as similarity measure (cosine similarity, dot product similarity, correlation) techniques.

It is important to note that we are only interested in previously discussed LIPL, RIPL and PCC ROIs since it is claimed in previous studies [2, 18] that there is functional connectivity differences between OCD and healthy groups. According to the third step of fMRI data analysis (first-level analysis) the resulted high-dimensional data is the functional correlation coefficient values between ROIs and other voxels of the brain. So, there is three high-dimensional FC data (PCC to voxel, LIPL to voxel, RIPL to voxel) to be reduced in dimensionality.

3.4.1.1 Steps of Dimensionality Reduction Method

In this section the general steps followed in each dimensionality reduction method that were used in this study are listed with brief explanations and definitions about each step.
The feature vectors from 3D FC data are extracted through four steps:

1. Input 3D FC data
2. Subsample the FC data
3. Determine training and test sets using two nested loops called double-LOO-CV.
4. Extract features using dimensionality reduction method (PCA, KPCA, LDA)

These steps will be described in detail in which the following notations are used:

$k = 1, ..., 24$ represents the number of subjects
$H = \{h_1, h_2, ..., h_{12}\}$ represents healthy samples.
$P = \{p_1, p_2, ..., p_{12}\}$ represents the patient samples.
$S = \{H\} \cup \{P\}$ represents all samples
$\mathbf{I}(k)$ represents the data vector of subject k which can also called FC image.

Step 1. Input the 3D FC data
Here the FC data $\mathbf{I}(k)$ derived from first-level analysis of CONN program was used as input data for feature extraction.

Step 2. Subsample the FC data
Here each $\mathbf{I}(k)$ vector was filtered with a 3D Gaussian and then the filtered $\mathbf{I}(k)$ subsampled with sample size $n = 10$ such that:

$$subsample[\mathbf{I}(k), n]$$

Step 3. Determine Training and Test Sets Using Double-LOO
This step was explained completely in section 3.2.

Step 4. Extract Features Using Dimensionality Reduction Method
In this step feature vectors are extracted by either reducing the dimensionality of train and test sets using PCA, KPCA and LDA methods or similarity measurement methods. The resulted low-dimensional outputs will be used as inputs for train the classifier and test its performance in classification phase.

In following sections the subsampling and dimensionality reduction approaches applied to resting-state functional connectivity data, prior to classification of the data, will be described.

3.4.1.2 Principal Component Analysis (PCA)

The PCA method as introduced in previous chapter, is a feature reduction method that orthogonally projects the subsampled data (here original data) from high-dimensional input space into a low-dimensional new space with new coordinates called principal components (PCs) so the projected data will have less number of features compared with original data.

The PRINCOMP command from statistics toolbox of the matlab was used to apply PCA method on subsampled FC data. The eigenvalues and eigenvectors were obtained. Afterwards eigenvalues are ranked in decreasing order and the eigenvectors corresponding to first M eigenvalues whose sum constitutes 90 percent of the total sum of all eigenvalues were chosen. The subsampled data was projected to the selected eigenvectors resulting in principal components. The feature vectors made of PCs are then used as inputs for classification.

---

1. **Input** 3D FC data
2. **FOR** each ROI
3.   Subsample the FC data
4.   Determine training and test sets using D-LOO-CV algorithm and 2 of 24 subjects were excluded.
5.   Extract features using PCA method:
            a. Obtain the eigenvalues and corresponding eigenvectors for the training set. 22 eigenvalues and eigenvectors were obtained in this step.
            b. Calculate the percentage of partial sum of eigenvalues to the sum of all the eigenvalues ($\sum \gamma$).

$$\sum \gamma = \sum_{i=1}^{22} \gamma = \gamma_1 + \gamma_2 + \cdots + \gamma_{22}$$

$$\frac{\sum_{i=1}^{22} \gamma}{\sum \gamma} \%$$

         c. Choose first *M* eigenvectors such that the percentage calculated in step 5.b is about 90% and constitute matrix *A* whose columns are chosen eigenvectors.

$$\frac{\sum_{i=1}^{M} \gamma}{\sum \gamma} \% \sim 90\%$$

         d. Obtain feature vectors with reduced dimension as:

$$F(k, h, p) = A\ (subsample[\mathbf{I}(k), n], M)$$

Figure 3.9: The feature extraction algorithm for each OI using PCA method

3.4.1.3 Kernel Principal Component Analysis (KPCA)

As stated in Chapter 2 the KPCA is a nonlinear expansion of PCA to settle nonlinear problems. Nonlinear mapping of the high-dimensional original data into low-dimensional feature space is performed using nonlinear kernel functions. In this thesis study, The KPCA algorithm presented in the paper by Bernhard and Solma [81] were used. Since the KPCA is an extended version of the PCA, all the steps followed for the PCA also were identically applied to KPCA with only difference that the original data is mapped to feature space using nonlinear kernels such as Gaussian kernels, polynomials, etc. Again before applying KPCA, data is subsampled in a same way due to the same reason discussed in the previous section.

Gaussian kernel function is used in this study for nonlinear mapping in KPCA.

$$e^{-(|x-c|^2)/2t^2} \tag{3.2}$$

where t is the parameter of Gaussian related to its width and its default value equals 1, c is the center and $|x - c|^2$ is Euclidian Distance of x to c.
Then the vectors constructed in train and test sets are projected by eigenvector matrix into low-dimensional feature space.

3.4.1.4 Linear Discriminant Analysis (LDA)

PCA approach and its extensions find components that are helpful for data characterization although these components may not give further information useful for classifying the data. So, PCA finds components needed for data representation. However, LDA looks for important directions or components for data classification. When a data is projected from a d-dimensional space into a one dimension space, despite the data is compact and well-separate, it becomes extremely overlapped and confounded combination which is very difficult to discriminate. To resolve this problem, the line should be oriented around the space until the best direction is found for maximum separation and this is what a linear discriminant method does.

The process of dimensionality reduction using LDA is same as PCA method except that in LDA method the eigenvalues and eigenvectors are obtained then by solving the eigenvalue problem,

$$\mathbf{w} = \mathbf{S}_w{}^{-1}(m_1 - m_2) \tag{3.3}$$

The direction of vector $\mathbf{w}$ for the projection of the data was determined in this way and finally the subsampled train and test sets were projected by eigenvector matrix onto one- dimensional feature space in determined direction.

3.4.2 Similarity Measurements

3.4.2.1 Cosine Similarity Measure

As declared earlier, the cosine similarity measure is the *cosine* measure of the angle between the two vectors of d-dimensions [82] which was defined in equation 2.3. In present research, the cosine similarity is preferred as a feature extraction method to measure the distance between each sample and the mean cluster of the other samples. Methodologically, this approach computes the cosine value of the angle between the two vectors formed from the functional connectivity matrices of each subject sample and means of healthy and patient subjects in the training sets.

In cosine similarity measurement procedure the feature vectors in training and test sets were determined using the D-LOO-CV algorithm. Hence the double leave one out feature vector is defined as:

$$\mathbf{F}(k,h,p) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} \cos\left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{H-h-k}\right) \\ \cos\left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{P-p-k}\right) \end{bmatrix} \tag{3.4}$$

where f_1, f_2 represent the first and second components of feature vector. The $mean\ \mathbf{I}_{H-h-k}$ is the mean of all the healthy subjects excluding healthy test (h) and kth subjects and the $mean\ \mathbf{I}_{P-p-k}$ is the mean of all the patient subjects excluding patient test (p) and kth subjects such that:

$$mean\mathbf{I}_{H-h-k} = \frac{\sum\limits_{\substack{i\in H \\ i\neq h,k}} \mathbf{I}(i)}{n_{H-h-k}} \quad , \quad mean\mathbf{I}_{P-p-k} = \frac{\sum\limits_{\substack{i\in P \\ i\neq p,k}} \mathbf{I}(i)}{n_{P-p-k}} \tag{3.5}$$

Here n_{H-h-k} is the number of healthy subjects excluding healthy test subject (h) and kth subject where $h \in H, k \in S$ and $h \neq k$. Also n_{P-p-k} is the number of patient subjects excluding patient test subject (p) and kth subject where $p \in P, k \in S$ and $p \neq k$.

Note that the $mean\mathbf{I}_{H-h-k}$ and $mean\mathbf{I}_{P-p-k}$ were used instead of $mean\ \mathbf{I}_{H-h}$ and $mean\ \mathbf{I}_{P-p}$ to settle bias problem caused by kth subject which is included in $mean\ \mathbf{I}_{H-h}$ and $mean\ \mathbf{I}_{P-p}$.

Considering equation (3.4), the cosine similarity can be interpreted as a feature reduction method since it has reduced the high dimensional data to a simple two dimensional vector which is much easier to interpret.

The simple algorithm of implementation of cosine similarity method to 3D FC data is shown in Figure 3.10.

---

1. **Input 3D FC data**
2. **FOR** each ROI
3.   Determine training and test sets using D-LOO-CV explained in Figure 3.8 in which the feature vector is set as:

$$\mathbf{F}(k,h,p) = \begin{bmatrix} \cos\left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{H-h-k}\right) \\ \cos\left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{P-p-k}\right) \end{bmatrix}$$

4.   The resulting two dimensional feature vectors in training and test sets will be used will be used in classification algorithm.
5. **END FOR**
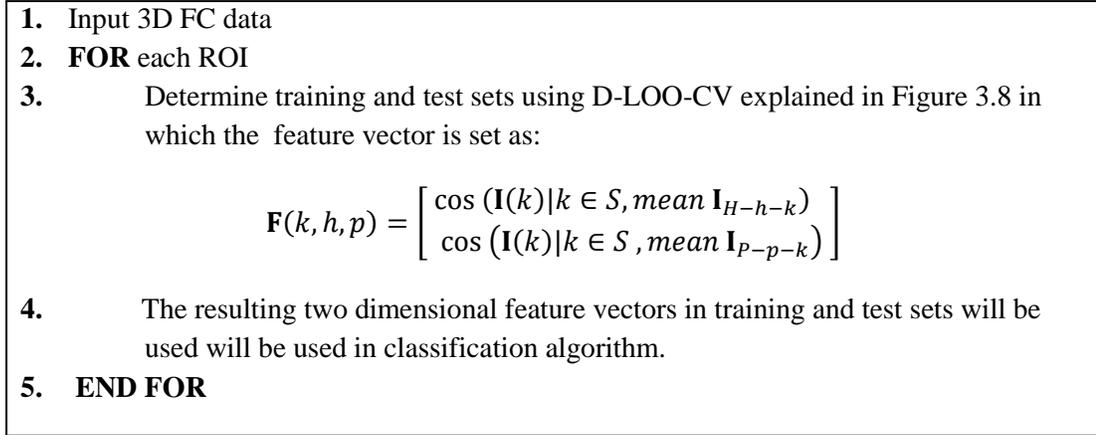
Figure 3.10: The cosine similarity algorithm for a healthy subject

3.4.2.2 Dot product Similarity measure

As mentioned in Chapter 2 the dot (inner or scalar) product is the simple form of the similarity measure between two vectors. In present work, the dot product similarity considers the similarity between a data vector of a subject k, $\mathbf{I}(k)$, and the mean vector of the healthy and patient subjects. This method was applied as a feature reduction method and reduced the number of features to two features. The first feature is the dot product of an arbitrary subject k and the mean of healthy subjects and the second feature is the dot product of the arbitrary subject $k \in S$ and the mean of patient subjects.

$$\mathbf{F}(k,h,p) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} dot\ \left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{H-h-k}\right) \\ dot\ \left(\mathbf{I}(k)|k \in S, mean\ \mathbf{I}_{P-p-k}\right) \end{bmatrix} \tag{3.6}$$

The dot product algorithm is same as cosine algorithm which was shown in Figure 3.10 and only the $\mathbf{F}(k,h,p)$ was changed using dot product similarity measure.

3.4.2.3 Correlation Similarity Measure

As the two previous methods, correlation similarity was applied to FC data in order to extract correlation features between a sample and the mean of other samples. The only difference between dot product and correlation is that in correlation method the mean value was subtracted from the sample and the mean of other samples, so that correlation feature vectors were produced such that:

$$\mathbf{F}(k,h,p) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} corr\left((\mathbf{I}(k) - \overline{\mathbf{I}(k)})|k \in S, (mean\,\mathbf{I}_{H-h-k} - \overline{mean\,\mathbf{I}_{H-h-k}})\right) \\ corr\left((\mathbf{I}(k) - \overline{\mathbf{I}(k)})|k \in S, (mean\,\mathbf{I}_{P-p-k} - \overline{mean\,\mathbf{I}_{P-p-k}})\right) \end{bmatrix} \quad (3.7)$$

$$\overline{\mathbf{I}(k)} = \frac{1}{d}\sum_{i=1}^{d}\mathbf{I}_i \qquad (3.8)$$

$$\overline{mean\,\mathbf{I}_{H-h-k}} = mean(mean\,\mathbf{I}_{H-h-k}) = mean\left[\frac{1}{n_{H-h-k}}\sum_{\substack{i \in H \\ i \neq h,k}}\mathbf{I}(i)\right] \qquad (3.9)$$

$$\overline{mean\,\mathbf{I}_{P-p-k}} = mean(mean\,\mathbf{I}_{P-p-k}) = mean\left[\frac{1}{n_{P-p-k}}\sum_{\substack{i \in P \\ i \neq p,k}}\mathbf{I}(i)\right] \qquad (3.10)$$

,

3.5 Classification Analysis

Using either dimensionality reduction methods or similarity measurements we could obtain low-dimensional feature vectors containing significant information necessary for classification. In order to discriminate between healthy and patient subjects, we used classification algorithms such as Support Vector Machine (SVM) and Gaussian Mixture Model (GMM) described in detail at further sections. For similarity measurements, it is also possible to decide on the class of subjects by a simple comparison of similarities as explained in the following

3.5.1 Decision by Comparison of Similarities

Let $\mathbf{F}(k,h,p) = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ be feature vector for a test sample k generated either by cosine, dot product or correlation similarity measurements. The class of the test sample k can be decided simply by comparison of f_1 and f_2 such that:

$$\begin{cases} if \;\; f_1 > f_2 \\ \quad k \in H \\ Otherwise \\ \quad k \in P \end{cases} \qquad (3.11)$$

3.5.2 Support Vector Machine

As mentioned in previous sections, the purpose of using SVM is to find a decision hyper surface that separates two classes as well as possible. Linear SVM classifier was not suitable for separation of the classes in the fMRI data so the nonlinear form of the SVM was preferred to discriminate patients from healthy subjects. The frequently used kernel functions in svm classification are linear kernel, Gaussian radial basis function and polynomial kernel function.

Linear: $K(x_i, x_j) = x_i^T x_j$

Radial Basis Function (RFB): $\exp(-\gamma \|x_i - x_j\|^d \quad , \gamma > 0$

Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d \quad , \gamma > 0$

where γ, r and d are kernel parameters.

According to [83], The RBF kernel function was preferred to linear and polynomial kernel functions to use as a kernel function in the SVM method since the linear kernel is the special case of the RBF kernel so its performance is same as the RBF kernel and the polynomial kernel function cause higher computational complexity than RBF kernel also for large orders of the polynomial kernel, infinite kernel values may happen [84]. Since various kernel functions with different parameters cause different classification results so the selection of right kernel and parameter adjustment are essential steps to find the appropriate kernel function with maximum separation results. In this study, the RBF kernel with parameter $\gamma = 7$ was found to be the most appropriate value.

Also the margin could be adjusted by selection of appropriate penalty parameter called m where $m > 0$, so that the SVM classifier could make maximum separation [84, 85].

SVM classification algorithm was applied to low-dimensional features extracted by PCA, KPCA and LDA from subsampled FC data and also two features were extracted by cosine similarity, dot product similarity and correlation measures. As discussed before, The SVM algorithm consists of two phases; first phase is training the SVM classifier using train set and second phase is testing the classifier using new data set (test set). Train and test sets were generated by D-LOO-CV approach that consists of two nested loops. The separating hyper surface and support vectors were acquired then the new data was tested and the accuracy rate of the classifier was calculated as percentage. Finally, the performance of the SVM classifier was estimated by taking the average of percentages for healthy and patient test subjects separately.

3.5.3 Gaussian Mixture Model

As stated previously, the GMM is a probability density function that is composed of the weighted summation of the (Gaussian) density components and the EM algorithm is commonly use to estimate the parameters of Gaussian components.

In this thesis, the distribution of healthy and patient samples taking part in the training set were modeled separately by a Gaussian mixture with a preset number of mixture components but unknown parameter values. The iterative algorithm EM was applied in order to find the optimum parameter $(\widehat{\boldsymbol{\theta}})$ for each component of the Gaussian mixture model. In order to classify a new sample in the test set as healthy or patient, the value of the probability density function (PDF) at this sample point should be calculated using parameters obtained from both healthy and patient samples in training data.

If the p_H which is the value of the pdf of healthy subjects at the test sample is greater than the p_p which is the value of the pdf of the patient subjects at the test sample and it is classified as healthy otherwise the test sample is classified as patient. The process can be summarized as follows:

1. Input the training and test data
2. Fit Gaussian mixture models separately on healthy and patient samples of training data set.
3. Expectation-Maximization algorithm to find the optimum parameters $\widehat{\boldsymbol{\theta}}$ and coefficients $\widehat{w}$ for healthy GMM (GMM_H) and patient GMM (GMM_P).
4. Calculate probabilities of the test data $s \in S$ for being healthy and being patient using the parameters and coefficients of the healthy and patient GMMs such that:

$$p_H\left(s|\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}\right) = \sum_{i=1}^{M} w_{i,H} \cdot p(s|w_{i,H}, \boldsymbol{\theta}_{i,H}) \tag{3.12}$$

$$p_p\left(s|\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}\right) = \sum_{i=1}^{M} w_{i,p} \cdot p(s|w_{i,p}, \boldsymbol{\theta}_{i,p}) \tag{3.13}$$

   where M is the number of Gaussian components in the mixture, $w_{i,H}$ is the weight of component i in GMM_H and $w_{i,p}$ is the weight of component i in GMM_P.
5. The class of the test sample s is decided as:

$$decision\ (s) = \begin{cases} if & p_H\left(s|\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}\right) > p_p\left(s|\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{w}}\right) \\ & s \in H \\ & otherwise \\ & s \in P \end{cases} \tag{3.14}$$

CHAPTER 4

EXPERIMENTAL RESULTS

This chapter consists of three main sections. In the first section, the results related to the functional connectivity (FC) between three regions of interest (PCC, RIPL, LIPL) and other voxels of the brain are presented. The second section has discussed and compared the results of feature extraction methods that are used to reduce the dimensionality of FC data. In the third section the classification results obtained using the features defined in Chapter3 are considered.

4.1 ROI-Voxel Functional Connectivity

The resting-state functional connectivity was estimated using CONN toolbox in which the functional correlation coefficients were calculated by Pearson's correlation method. Figures 4.1-3 exhibit the resting-state functional connectivity between chosen ROIs among PCC, LIPL and RIPL and all brain voxels. The brighter voxel means the higher connectivity value for that voxel.
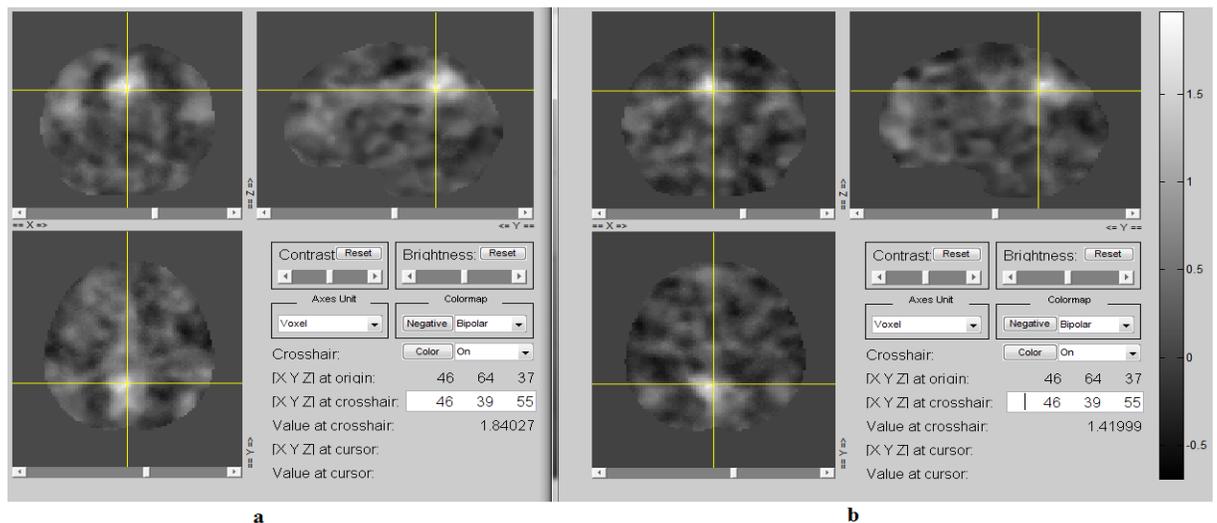
Figure 4.1: (a) rsFC between PCC ROI and all brain voxels of a healthy sample. (b) rsFC between PCC ROI and all brain voxels a patient sample.
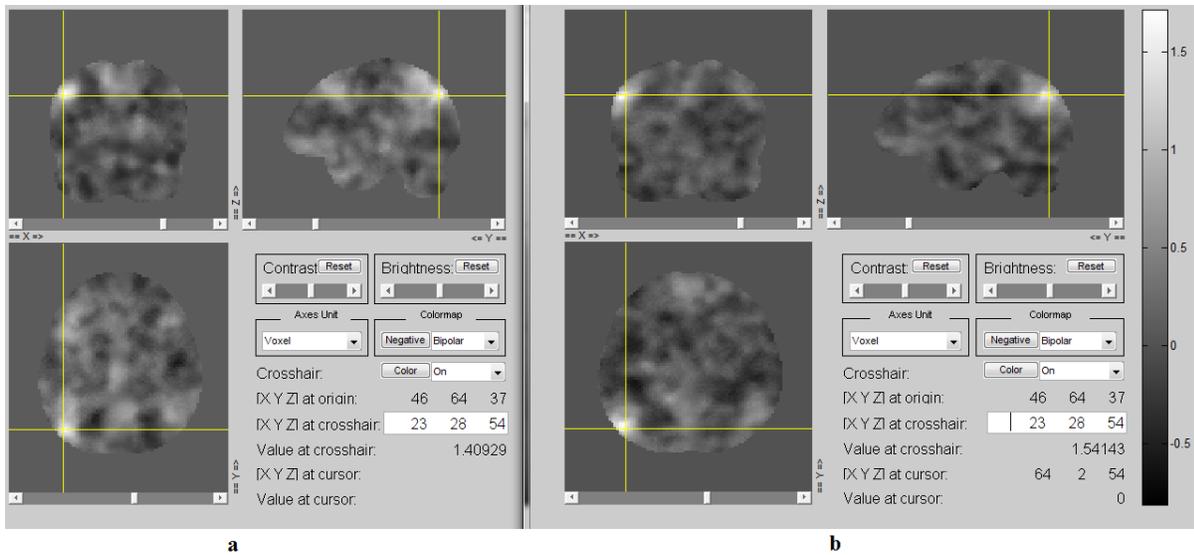
Figure 4.2: (a) rsFC between LIPL ROI and all brain voxels of a healthy sample. (b) rsFC between LIPLROI and all brain voxels of a patient sample.
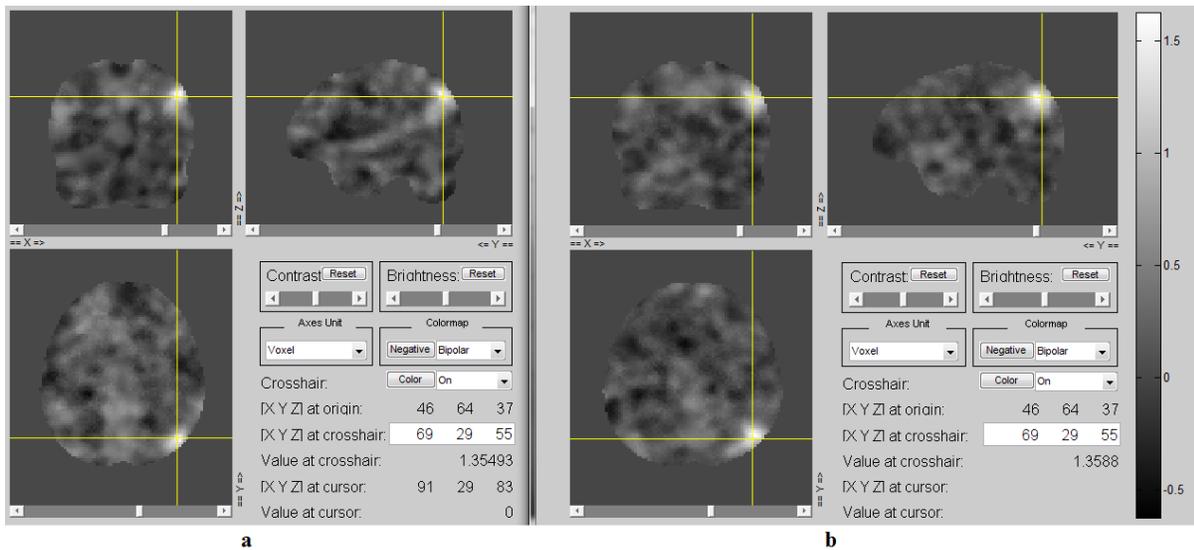
Figure 4.3: (a) rsFC between RIPL ROI and all brain voxels of a healthy sample. (b) rsFC between RIPL ROI and all brain voxels of a patient sample.

As can be seen in the figures 4.1-3 there are ROI to voxel functional connectivity differences between healthy and patient brains at the same coordinates shown with yellow crossed lines. Also bright regions in above figures demonstrate that there is high ROI to voxel functional connectivity and the correlation coefficient value is positive. As the darker regions mean the ROI to voxel functional connectivity decrease and the correlation coefficient value become lower to negative values.

4.2 Feature Extraction Results

As mentioned in previous chapters, the pattern recognition problem for rsFC data consists of two steps and the first step is related to feature extraction. First the feature extraction results using similarity measurements methods will be discussed and then dimensionality reduction methods (PCA, KPCA and LDA) that have been applied to rsFC data will be presented.

The Table 4.1 shows the marks used to represent healthy and patient samples in training and test sets in resulted figures.

Table 4.1: The markes used to represent healthy and patient samples in training and test sets

	Healthy	Patient
Training Set	× Green	+ Red
Test Set	× Blue	+ Yellow

As shown in the Table 4.1, the crossed green and blue points represent healthy samples in the training and test sets respectively and the plus red and yellow ponits represent the patient traning and test sets respectively.

4.2.1 Similarity Measurement Results

The flow chart given in the Figure 4.6 summarizes the generation of the training and test sets by D-LOO-CV for each ROI which was given in detail in chapter 3.

4.2.1.1 Cosine Similarity Measure Results

The Figure 4.5 shows the distribution of feature vectors obtained by cosine similarity measures in a two-dimensional feature space. The x-axis and y-axis represent in Figure 4.5 are the first and second feature vector components respectively. The healthy and patient samples in train sets and the healthy samples in test set were produced by D-LOO-CV method. As explained previously in section 3.2 there are 132 healthy samples and 132 patient samples in the training set, and 12 samples in the test set. Each sample has (x, y) values in 2D coordinate system where x-axis corresponds to first feature component that is f_1 and y-axis corresponds to second feature component that is f_2. If we want to calculate size of the test and train sets for whole leave-one out healthy subject iterations then they should be repeated 12 times so the healthy test set will be a 144 by 2 (12 ×12× 2) matrix and the healthy and patient training sets will be matrices of size 1584 by 2(132 ×12 2).

Figure 4.4 presents the similarity measure procedure for cosine similarity using D-LOO-CV.
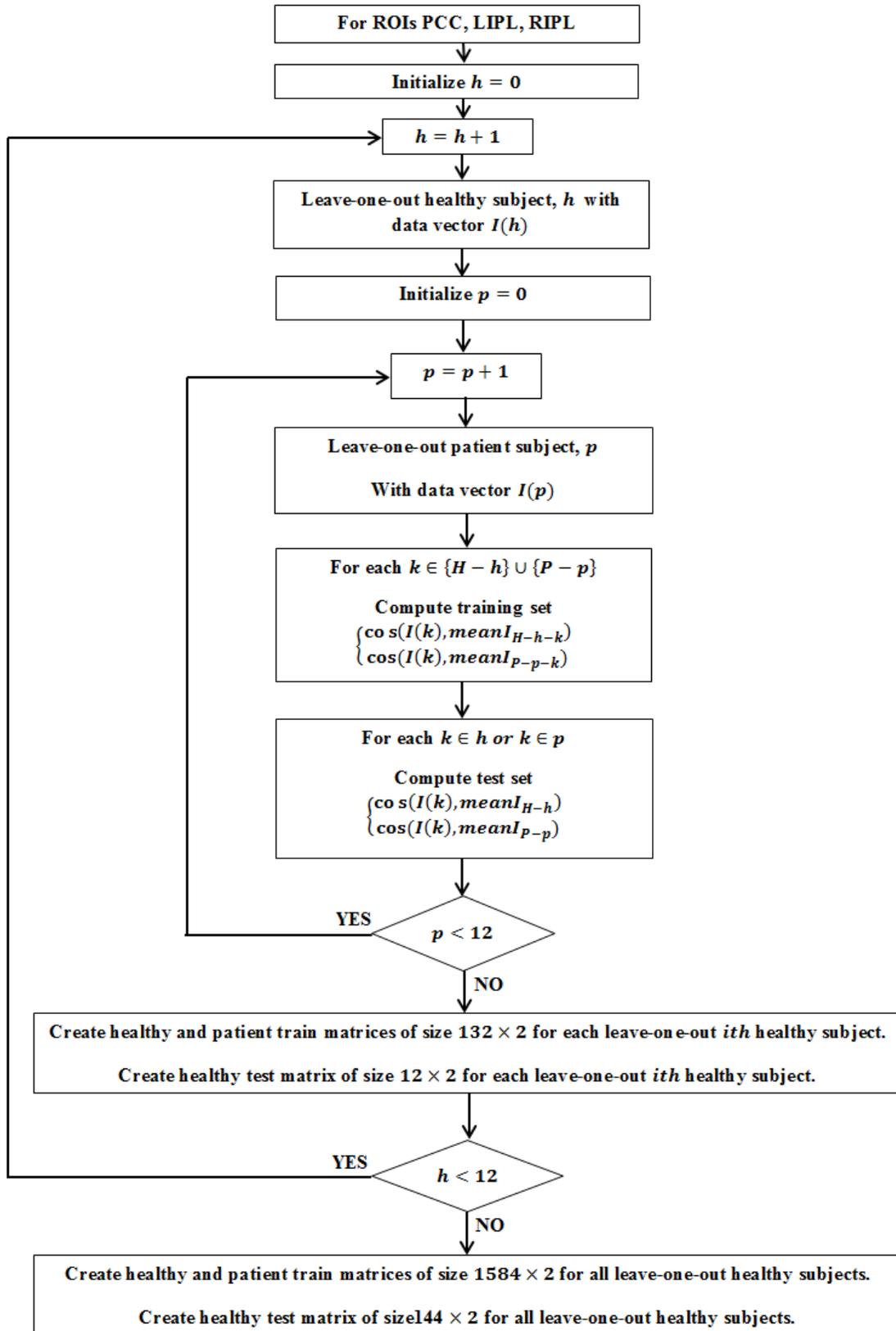
Figure 4.4: Cosine similarity measure flowchart

44

While Figures 4.5 and 4.6 show the distribution of feature vectors obtained by cosine similarity measures using D-LOO-CV for a single healthy subject ($h = 1$) and a single patient subject ($p = 1$) respectively, it is shown for the collection of all healthy subjects ($h = 1, ..., 12$) and patient subjects ($p = 1, ..., 12$) in Figure 4.7.
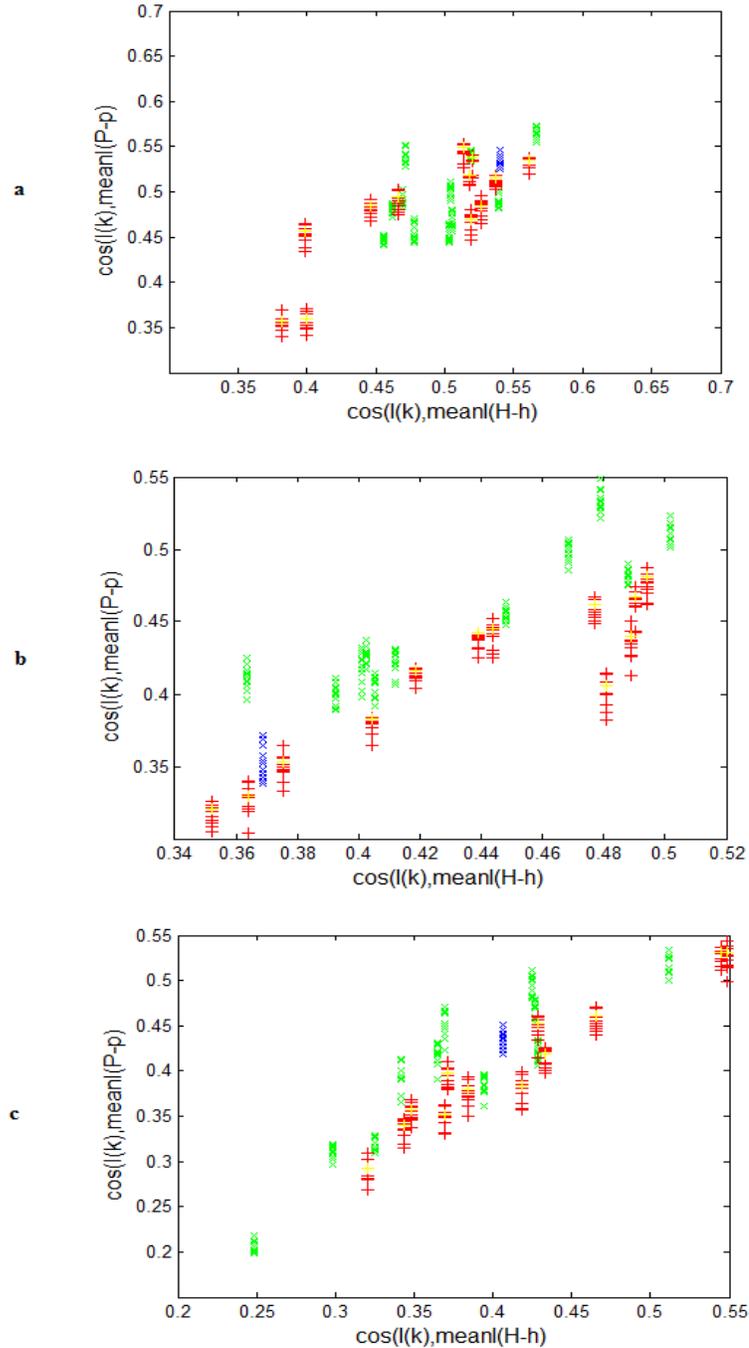
Figure 4.5: Distribution of feature vectors obtained by cosine similarity measures using D-LOO-CV for a healthy subject (h=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
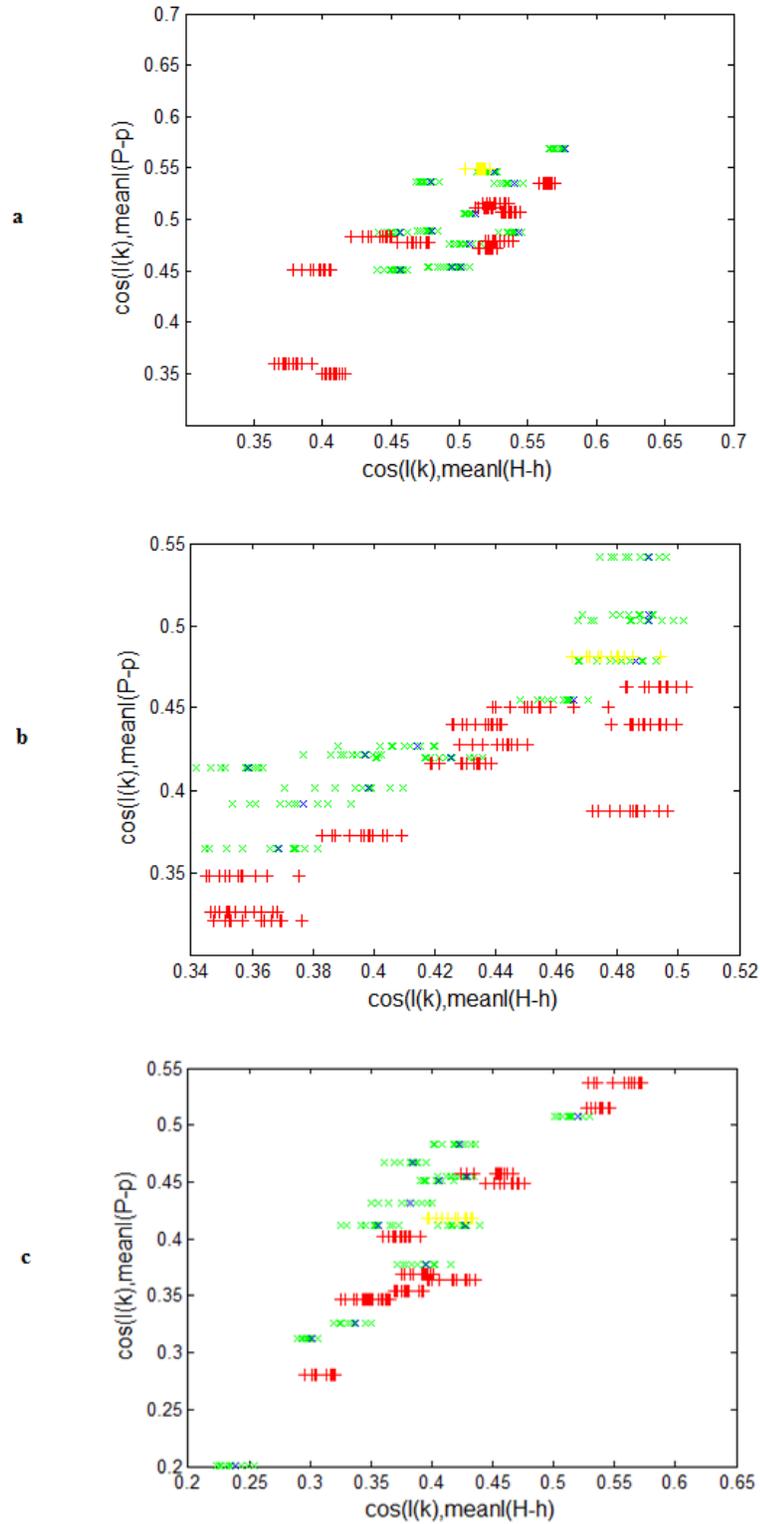
Figure 4.6: Distribution of feature vectors obtained by cosine similarity measures using D-LOO-CV for a patient subject (p=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
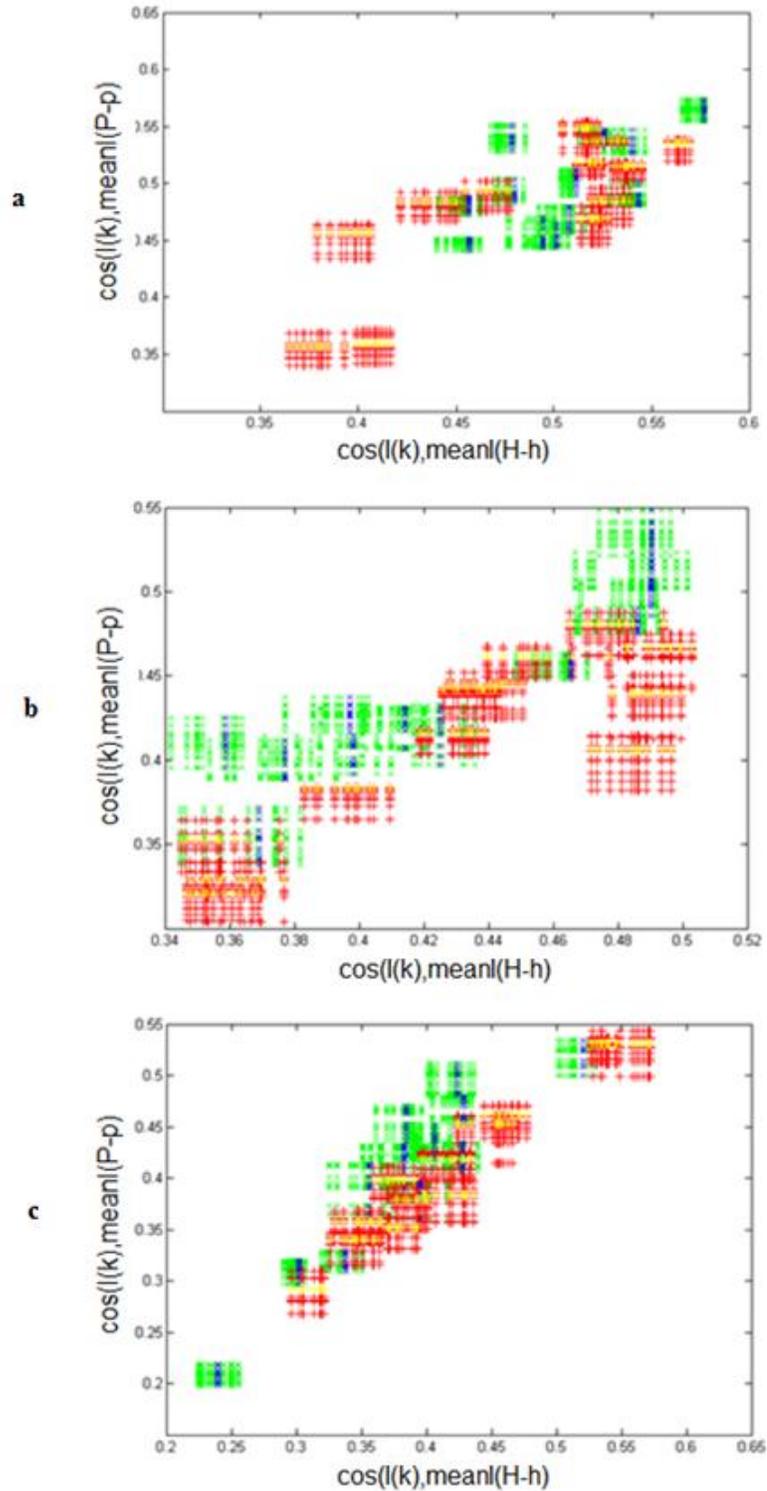
Figure 4.7: Distribution of feature vectors obtained by cosine similarity measures using D-LOO-CV for all healthy and patient subjects in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.

4.2.1.2 Dot Product Similarity Measure Results

While Figures 4.8 and 4.9 show the distribution of feature vectors obtained by dot product similarity measures using D-LOO-CV for a single healthy subject (h=1) and a single patient subject (p=1) respectively, It is shown for the collection of all healthy subjects (h=1,…, 12) and patient subjects (p=1,…,12) in Figure 4.10.
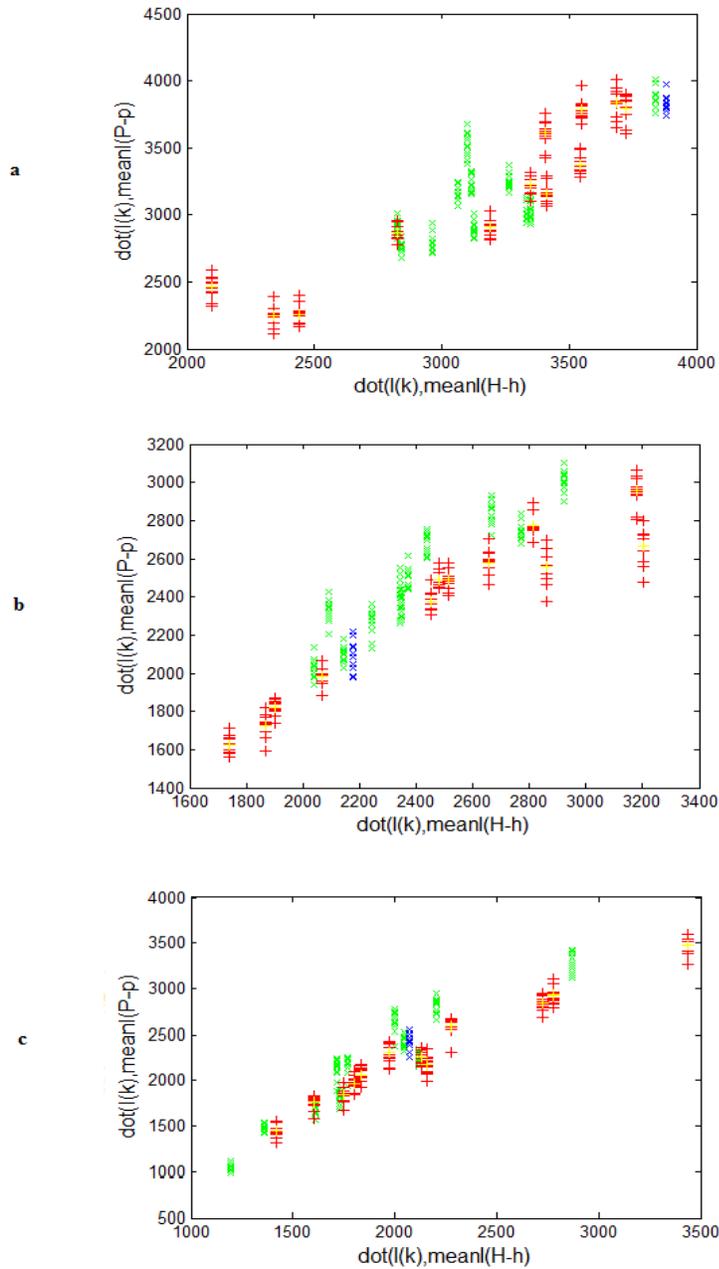
Figure 4. 8: Distribution of feature vectors obtained by dot product similarity measures using D-LOO-CV for a healthy subject (h=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
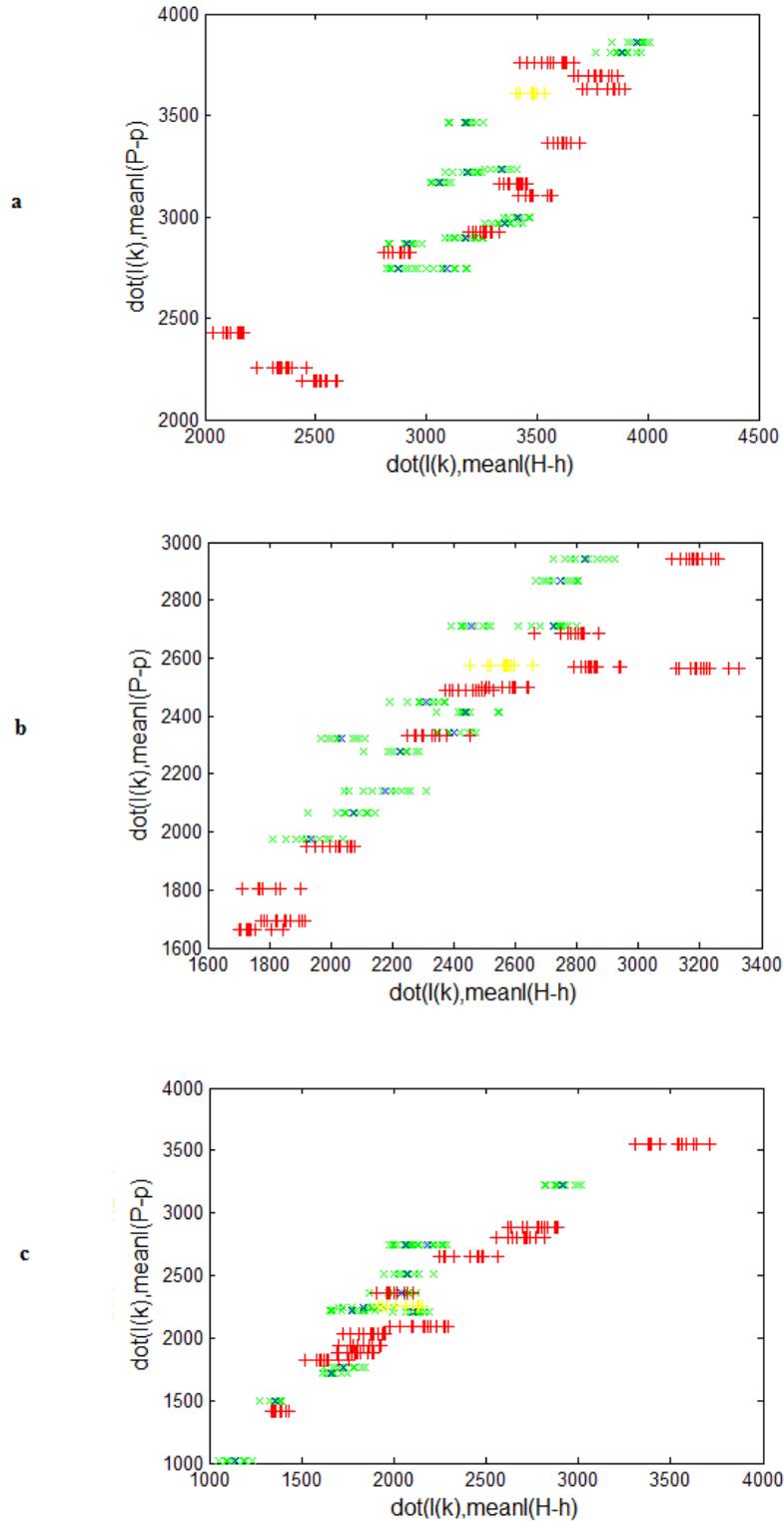
Figure 4.9: Distribution of feature vectors obtained by dot product similarity measures using D-LOO-CV for a patient subject (p=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
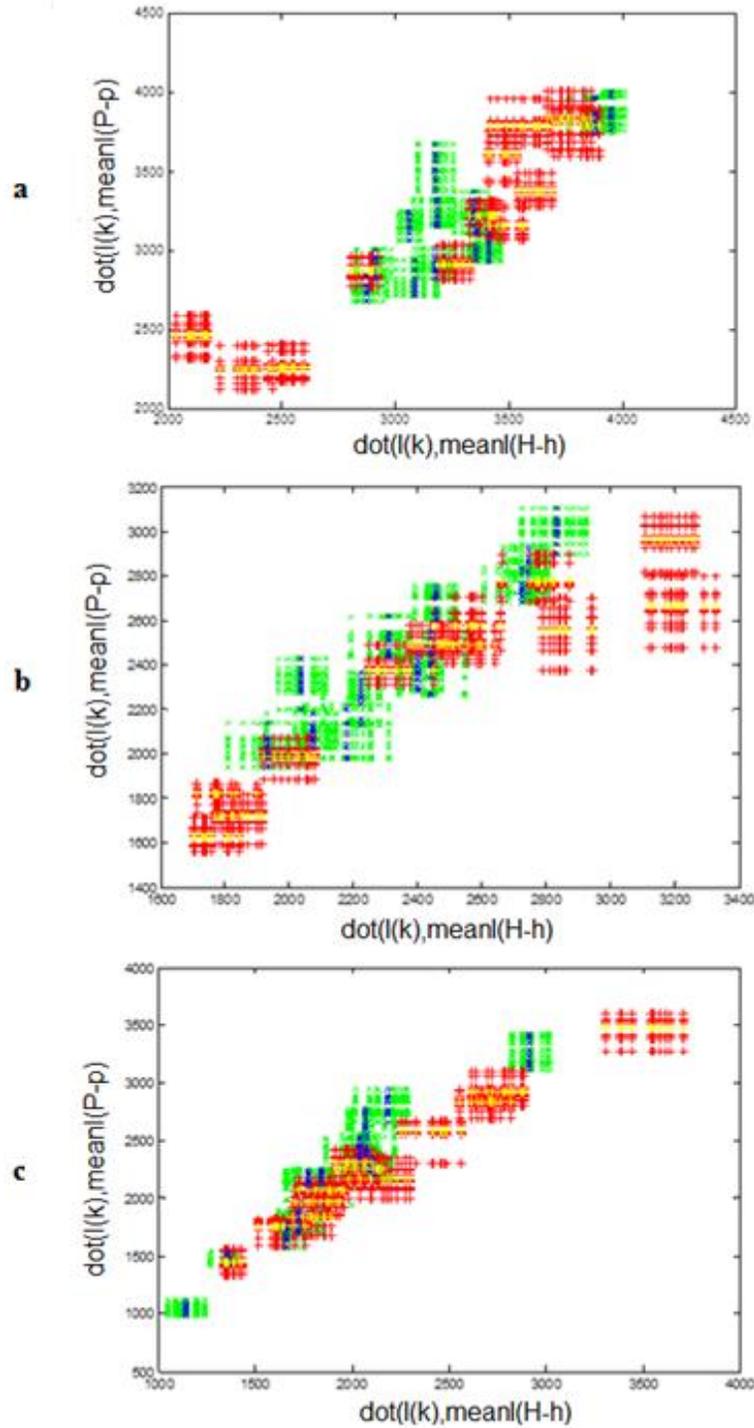
Figure 4.10: Distribution of feature vectors obtained by dot product similarity measures using D-LOO-CV for all healthy and patient subjects in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL

4.2.1.3 Correlation Measure Results

While Figures 4.11 and 4.12 show the distribution of feature vectors obtained by correlation similarity measures using D-LOO-CV for a single healthy subject (h=1) and a single patient subject (p=1) respectively, It is shown for the collection of all healthy subjects (h=1,…, 12) and patient subjects (p=1,…,12) in Figure 4.13.
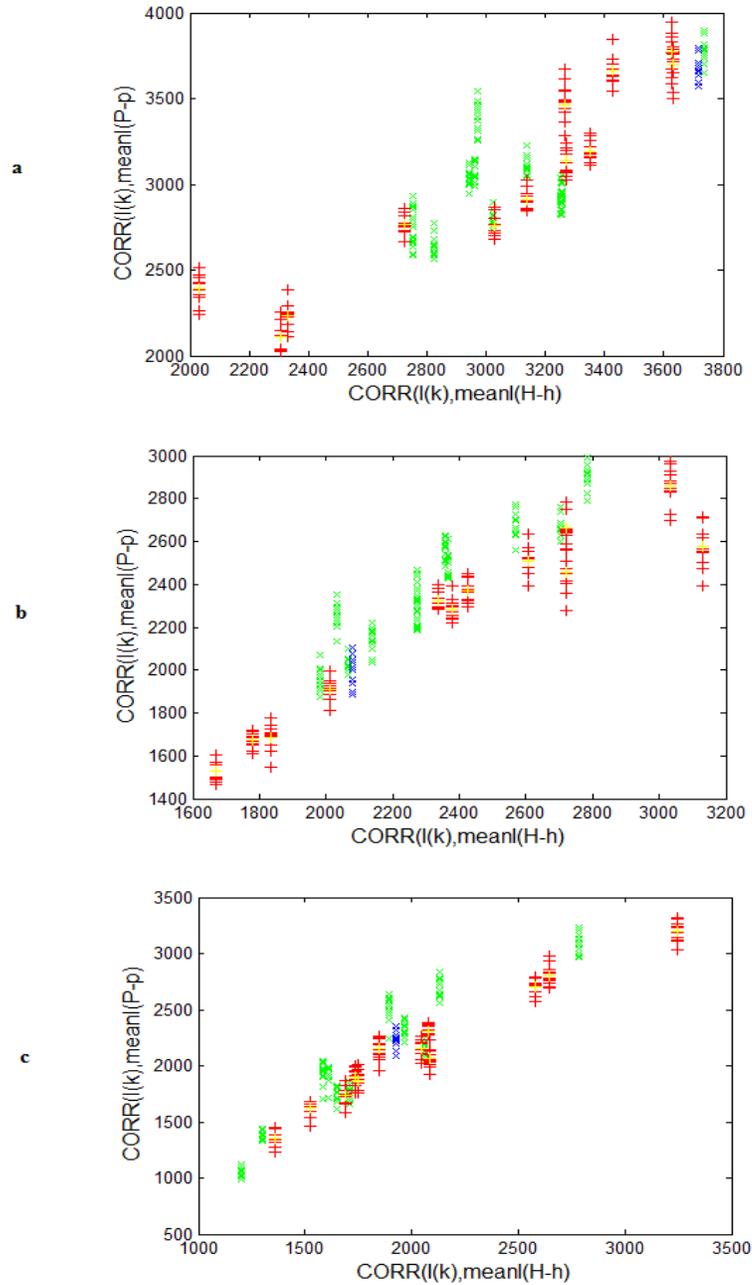
Figure 4.11: Distribution of feature vectors obtained by correlation similarity measures using D-LOO-CV for a healthy subject (h=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
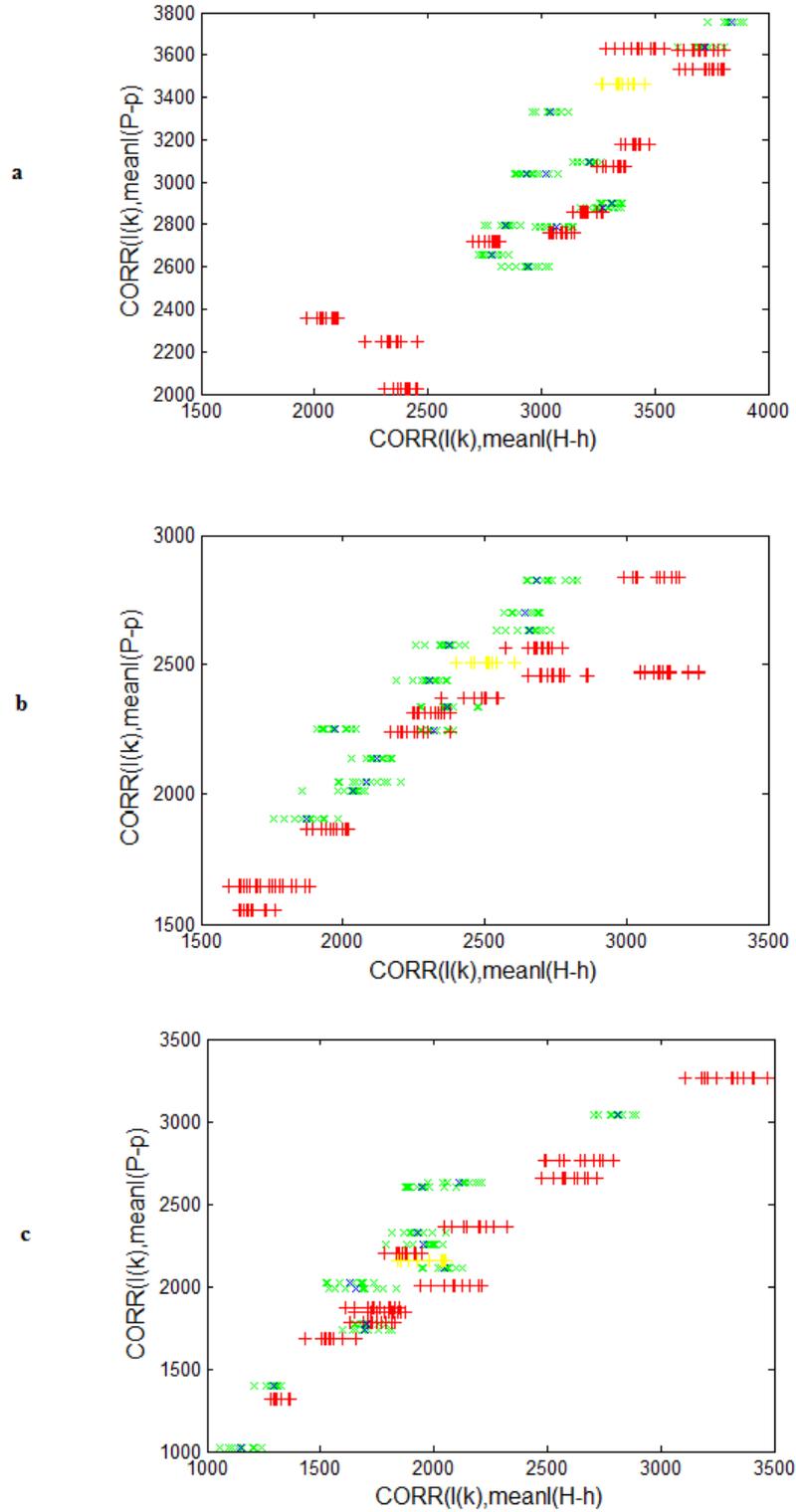
Figure 4.12: Distribution of feature vectors obtained by correlation similarity measures using D-LOO-CV for a patient subject (p=1) in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.
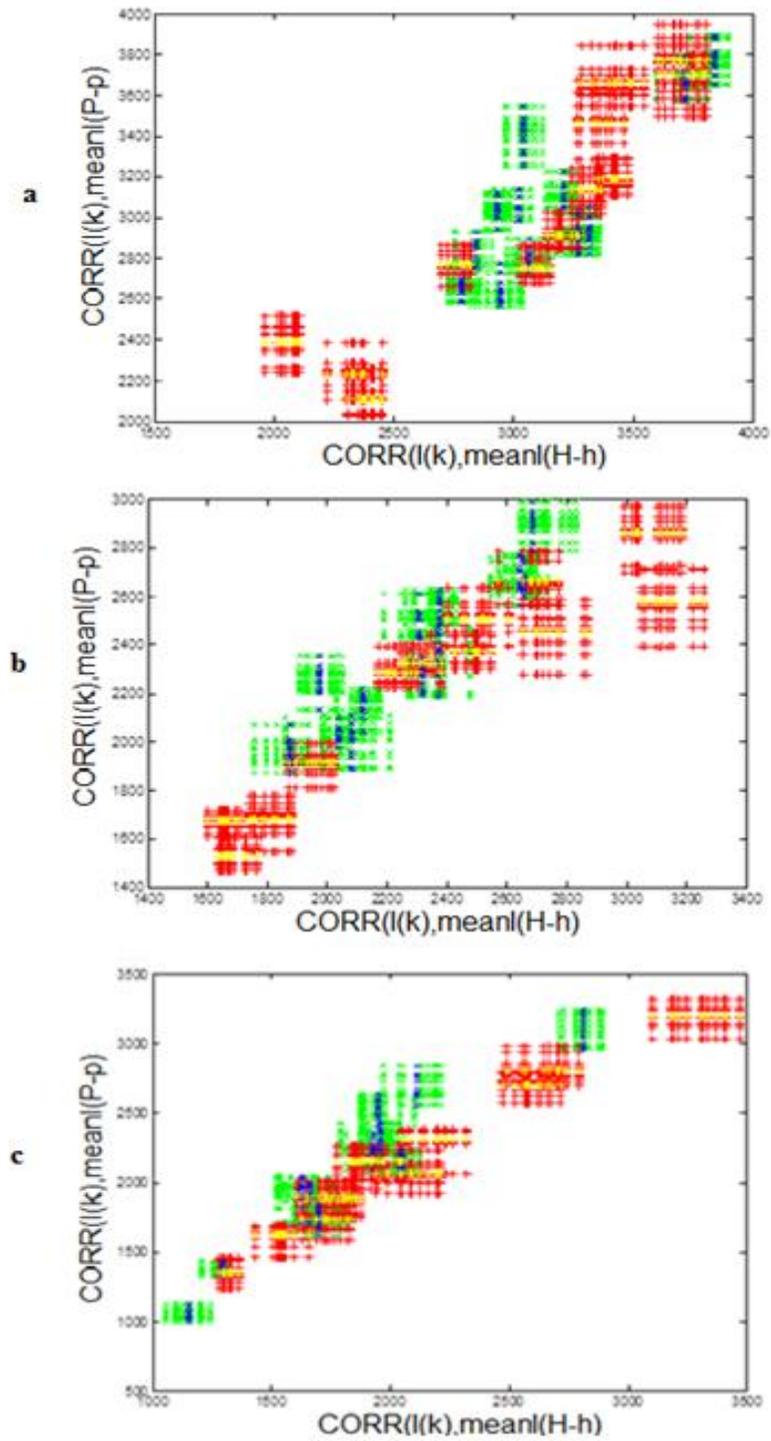
Figure 4.13: Distribution of feature vectors obtained by correlation similarity measures using D-LOO-CV for all healthy and patient subjects in (a) ROI PCC, (b) ROI LIPL and (c) ROI RIPL.

4.3 Results obtained by comparison of similarity measures

As a very simple classification method the first and second feature vector components obtained by similarity measures for subjects in test sets are directly compared. For this purpose the class of the test samples are decided considering the equation (3.11) given in section 3.3.1. As an example the results obtained for cosine similarity is shown in Table 4.2.

Table 4.2: The classes decided considering values of first and second feature vector components obtained by cosine similarity measure for healthy test samples in PCC ROI

Cosine Similarity Method	D-LOO h=1 p=1	D-LOO h=1 p=2	D-LOO h=1 p=3	D-LOO h=1 p=4	D-LOO h=1 p=5	D-LOO h=1 p=6	D-LOO h=1 p=7	D-LOO h=1 p=8	D-LOO h=1 p=9	D-LOO h=1 p=10	D-LOO h=1 p=11	D-LOO h=1 p=12
First feature vector $Cos(I(h), mean\ I_{H-h})$	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540
Second feature vector $Cos(I(h), mean\ I_{P-p})$	0.534	0.538	0.525	0.535	0.531	0.546	0.528	0.528	0.541	0.531	0.528	0.527
Class decided	1	1	1	1	1	-1	1	1	-1	1	1	1
Correctness of result	1	1	1	1	1	0	1	1	0	1	1	1

According to Table 4.2, among 12 healthy samples 10 are classified correctly resulting in $\left(\frac{10}{12}\right) \times 100 \sim 83\%$ accuracy. Note that Table 4.2 shows comparison performance for the first healthy subject ($h = 1$) where Healthy subject and patient subjects ($p = 1, ... , 12$) are excluded using D-LOO-CV.

Classification performance accuracy obtained by simple comparison of similarly measures for healthy and patient test samples in PCC, LIPL and RIPL ROIs and for all similarity measure methods were computed and the results are summarized in Table 4.3.

Table 4.3: Classification accuracy results based on comparing similarity measurement values of healthy and patient test sets

Feature Extraction Methods	PCC ROI			LIPL ROI			RIPL ROI		
	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average
Cosine	60%	45%	52.5%	33%	33%	33%	36%	30%	33%
Dot Product	70%	34%	53.5%	26%	31%	28.5%	12.5%	87%	50%
Correlation	70%	34%	53.5%	32%	24%	28%	8%	80%	44%

This table presents the mean of the classification accuracy rate obtained by double leave-one-outing 12 healthy and 12 OCD test samples for PCC, LIPL and RIPL ROIs and cosine, dot product and correlation similarities. It is concluded from this table that even without using any machine learning method to classify two groups, only by comparing the similarity measurement values are encouraging: The patients can be identified correctly by 87% considering the RIPL ROI and the accuracy rate of healthy test samples was 70% considering PCC ROI. But if the

average of both healthy and patient accuracy rates are considered it is concluded that the maximum average rate that can be obtained without using any classification method is 53.5% which is related to PCC ROI while using dot product and correlation methods.

4.4   SVM Classification Results

4.4.1 Parameter Selection for SVM Classification

As mentioned in materials section, the classification analysis was performed on 24 participants (12 healthy and 12 OCD). First, 22 samples (11 healthy and 11 OCD) called training set were used to train the classifier. Then the leave-one-out cross validation was performed 12 times for healthy test sample and 12 times for patient test sample to evaluate the prediction accuracy of the classifier.

The SVM classifier was used to discriminate between healthy and patient groups and as mentioned in chapter 3 the RBF kernel function was preferred to implement to train the SVM classifier.

First of all, the RBF kernel parameter (γ) and margin parameter (m) should be selected such that the best prediction accuracy of test data was obtained.

Table 4.4: Classification accuracy of SVM with RBF kernel for different parameter values of RBF

Feature Extraction Method	Kernel Function Type	Kernel RBF Parameter (γ)	Margin constraints (m)											
			$m = 10^{-1}$			$m = 10^{0}$			$m = 10^{1}$			$m = 10^{2}$		
			D-LOO Healthy	D-LOO Patient	Average	D-LOO Healthy	D-LOO Patient	Average	D-LOO Healthy	D-LOO Patient	Average	D-LOO Healthy	D-LOO Patient	Average
Correlation	RBF	$\gamma = 1$	66%	49%	57.5%	61%	51%	56%	58%	53%	55.5%	61%	52%	56.5%
		$\gamma = 3$	62%	57%	59.5%	65%	59%	62%	68%	49%	58.5%	63%	50%	56.5%
		$\gamma = 5$	52%	48%	50%	69%	60%	64.5%	62%	63%	62.5%	65%	60%	62.5%
		$\gamma = 7$	55%	37%	46%	74%	60%	67%	**75%**	**63%**	**69%**	58%	62%	60%
		$\gamma = 9$	53%	17%	35%	68%	51%	59.5%	74%	63%	68.5%	62%	63%	62.5%

The Table 4.4 shows the average percentage of the classification accuracy for leave-one-out healthy and patient individuals using correlation method to extract the features considering the LIIPL region. It is concluded from the Table 4.4 that the parameters $\gamma = 7$ and $m = 10^1$ get the best prediction for test healthy and patient samples.

4.4.2 SVM Classification Results Using Feature Extraction by Similarity Measure

Table 4.5: Summary of the SVM classification results of D-LOO-CV healthy and patient test sets

Feature Extraction Methods	ROI Types											
	PCC			LIPL			RIPL			All ROIs		
	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average
Cosine Similarity	67%	44%	55.5%	72%	65%	68.5%	58%	70%	64%	51%	67%	59%
Dot Product	75%	58%	66.5%	68%	62%	65%	74%	62%	68%	67%	65%	66%
Correlation	75%	42%	57%	75%	63%	69%	67%	62%	64.5%	70%	60%	65%

Table 4.5 presents mean of SVM classification accuracy of D-LOO-CV of healthy and patient test samples as percentage in D-LOO healthy and patient columns respectively. Also the average of healthy and patient results is presented in average column of the Table. The columns labeled by all ROIs as mentioned earlier relate to classification performance using feature vectors of all ROIs together.

As can be seen in Table 4.5, the best SVM classification accuracy result of D-LOO-CV of healthy test samples was obtained using feature vectors produced by correlation measurement and dot product measures in PCC ROI which is 75%.

The columns under label of all ROIs in Table 4.3 are correspond to classification performance of healthy and patient test subjects when all feature vectors (6) extracted from three ROIs are considered together for classification such that:

$$\begin{bmatrix} PCC\ ROI & LIPL\ ROI & RIPL\ ROI \\ f_1 f_2 & f_1 f_2 & f_1 f_2 \end{bmatrix}$$

The Figure 4.14 shows the distribution of samples for the best D-LOO-CV of healthy test using correlation similarity measures in LIPL ROI and also the hyper surface obtained by SVM.
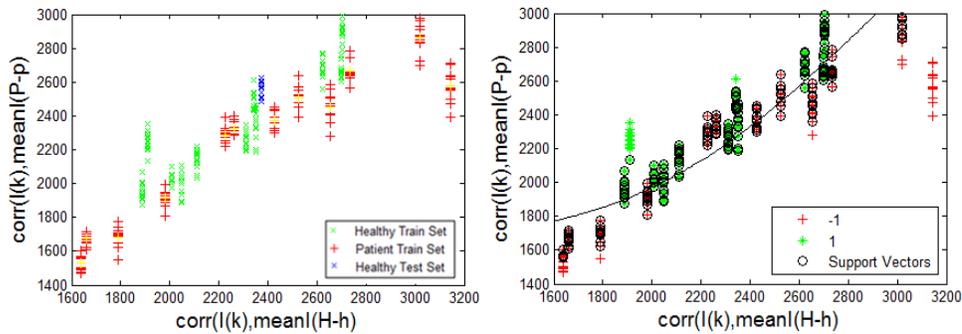
Figure 4.14: The best SVM classification accuracy result (100%) for a healthy test. (Left) distribution of samples. (Right) hyper surface separating classes for correlation similarity measurement in LIPL ROI.

The Figure 4.14 shows the best (100%) svm classification accuracy of D-LOO-CV of a healthy test (h=9) when the correlation similarity measure method was used for LIPL ROI. Support vectors (in circles) and decision hyper surface resulted from training the SVM classifier using training features extracted from correlation measurement considering LIPL ROI are represented in Figure 4.14. The hyper surface separates two classes well and cause the healthy test set, shown as blue crossed points in left-hand scheme, belongs to correct (healthy) class +1 which is shown as green crossed points.

The worst SVM classification accuracy results (0%) of D-LOO-CV of a healthy (h=8) test were obtained using the cosine similarity method in ROI PCC as shown in Figure 4.15.
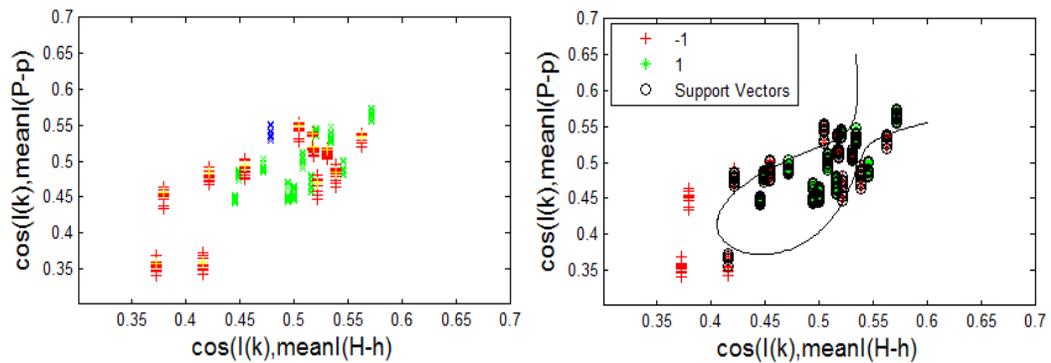
Figure 4.15: The worst SVM classification accuracy result (0%) for a healthy test. (Left) distribution of samples. (Right) hyper surface separating classes for cosine similarity measurement in PCC ROI.

The Figure 4.15 presents the worst hyper surface obtained for separating feature vectors extracted from cosine similarity measurement considering PCC ROI. This hyper surface (shown in Figure4.15 right) has caused the healthy test set (shown as blue crossed points in Figure 4.15 left) belong to incorrect (patient) class -1 (shown as red plus points in Figure 4.15 left).

The same analysis of D-LOO-CV of a patient test samples has done and the results have been inserted in D-LOO patient parts of Table 4.5. In this case, maximum value of the mean of SVM classification accuracy (70%) was achieved by measuring the cosine similarity in RIPL ROI as shown in Table 4.5.
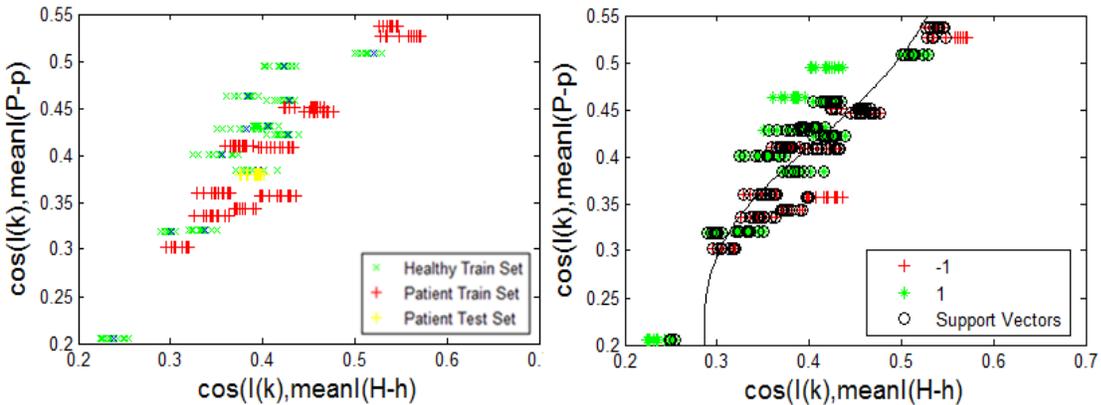
Figure 4.16: The best SVM classification accuracy result (100%). (Left) distribution of samples. (Right) hyper surface separating classes for cosine similarity measurement in ROI RIPL.

The Figure 4.16 shows the best (100%) SVM classification accuracy of D-LOO-CV of a patient test (p=9) when the cosine similarity method was used for RIPL ROI.

Conversely, The Figure 4.17 shows the worst classification performance of D-LOO-CV of a patient test sample (p=6) using correlation method for PCC RIO.
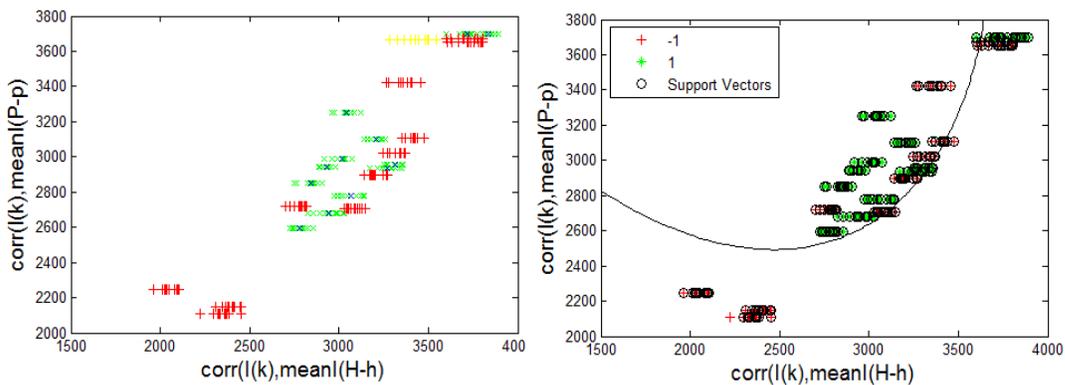
Figure 4.17: The worst SVM classification accuracy result (100%). (Left) distribution of samples. (Right) hyper surface separating classes using correlation similarity measure in PCC ROI.

Considering Table 4.5, the best average of SVM classification accuracy of both D-LOO-CV healthy and patient test samples (69%) was obtained using correlation method for LIPL ROI, which was 75% for healthy subjects and 63% for patient subjects. Distribution of samples and the hyper surface obtained by SVM for both healthy and patient test samples is shown in Figure 4.18.
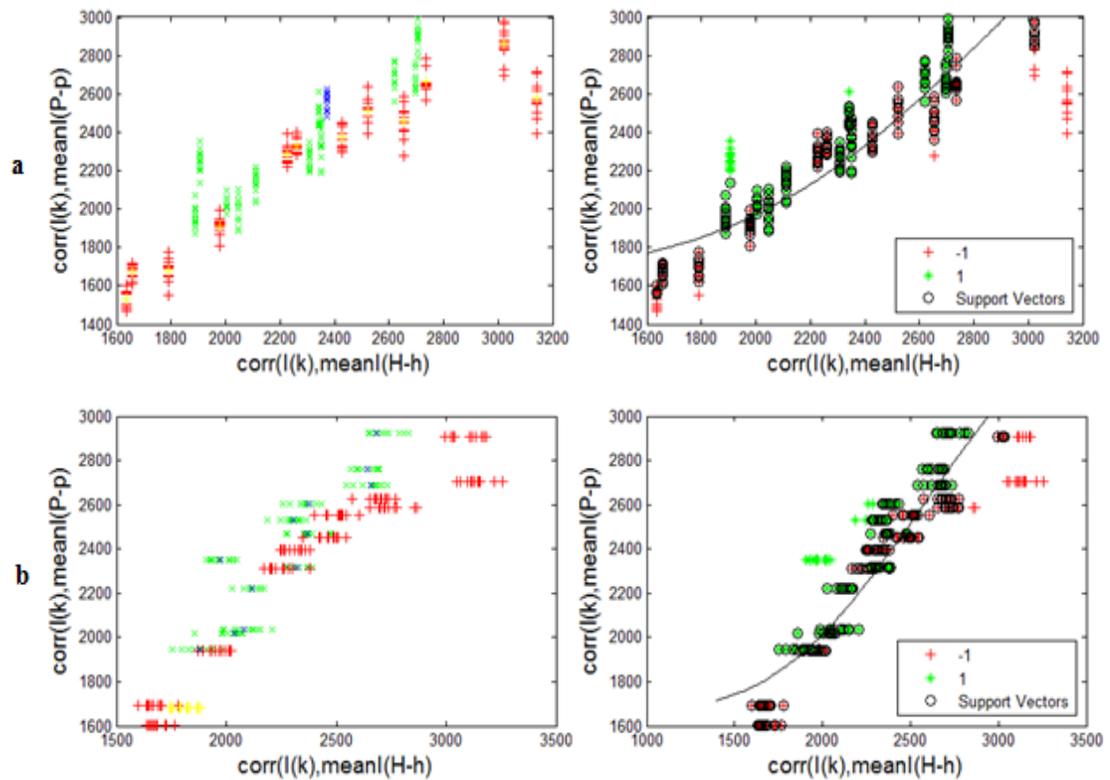
Figure 4.18: The best SVM classification results of D-LOO-CV for (a) healthy and (b) patient test sample

The Figure 4.18 shows the best SVM classification performance of D-LOO-CV of both healthy (h=9, 100%) and patient (p=9, 100%) test samples using correlation method in LIPL ROI.

Conversely, the worst average of SVM classification performance of both healthy and patient test samples (55.5%) was obtained when the features were extracted using the cosine similarity measure in the PCC ROI as shown in Figure 4.19.
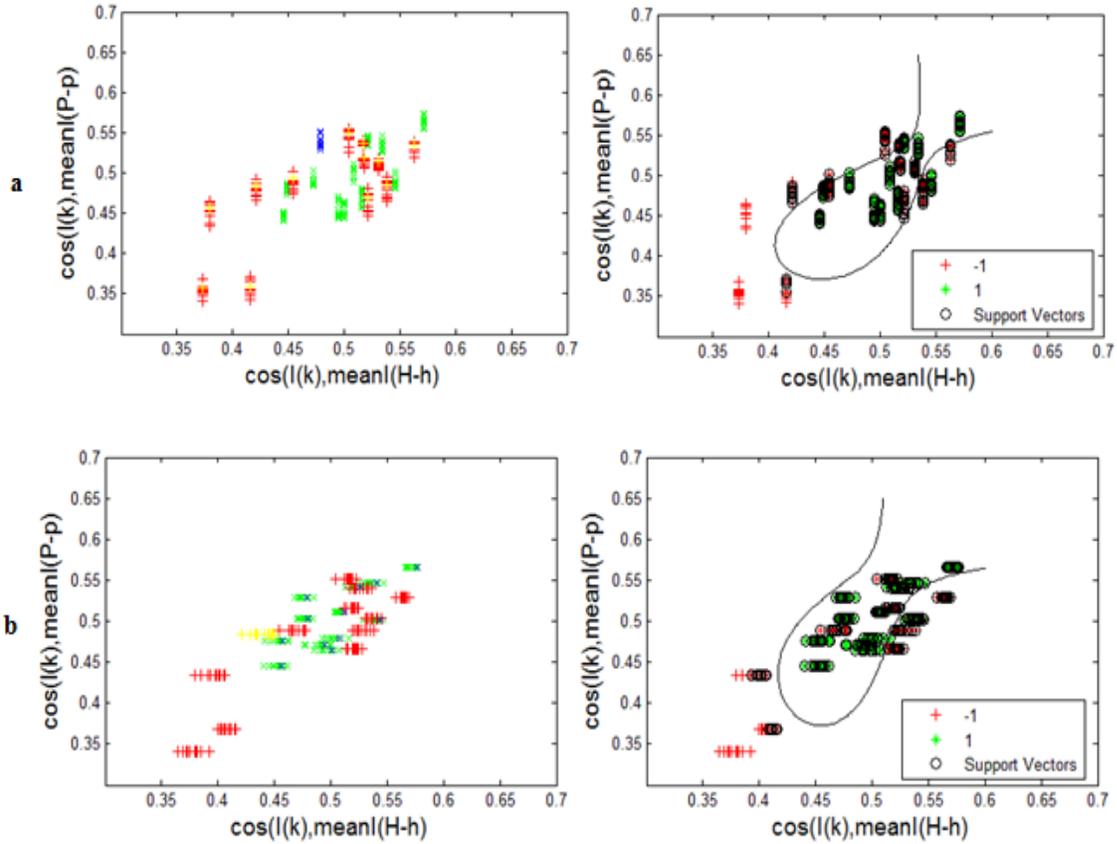
Figure 4.19: The worst SVM classification accuracy results of D-LOO-CV for (a) healthy and (b) patient test samples using cosine similarity measure in PCC ROI.

The Figure 4.19 shows the worst classification performance accuracy of both D–LOO-CV for healthy (h=8, 8%) and patient (p=6, 8%) test samples using cosine similarity in PCC ROI.

Taking mean of the SVM result averages mentioned in Table 4.5 for each similarity measure method in three ROIs, shows that the dot product similarity measure gives the best average classification results for PCC, LIPL and RIPL ROIs as shown in Table 4.6.

Table 4.6: Mean of the SVM result averages for each similarity measure method in all ROIs

Similarity Measures	PCC ROI average	LIPL ROI average	RIPL ROI average	all ROIs average	Mean
Cosine	55.5%	68.5%	64%	59%	61.75%
Dot product	66.5%	65%	68%	66%	**66.35%**
Correlation	57%	69%	64.5%	65%	63.87%

The Table 4.6 shows that in average the best SVM results are obtained using dot product similarity measure in PCC, LIPL and RIPL ROIs. The Figure 4.20 represents the distribution of the samples and resulting hyper surface by using the cosine, dot product and correlation similarity methods for LIPL ROI, respectively
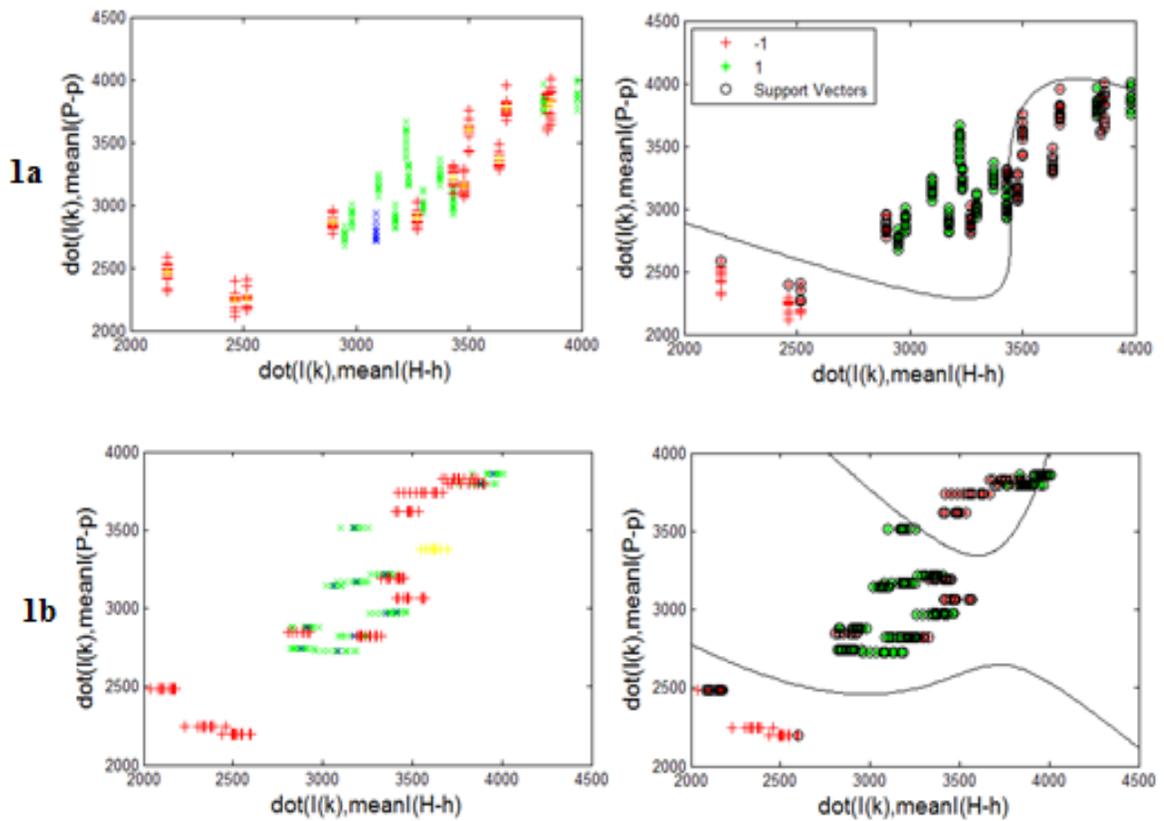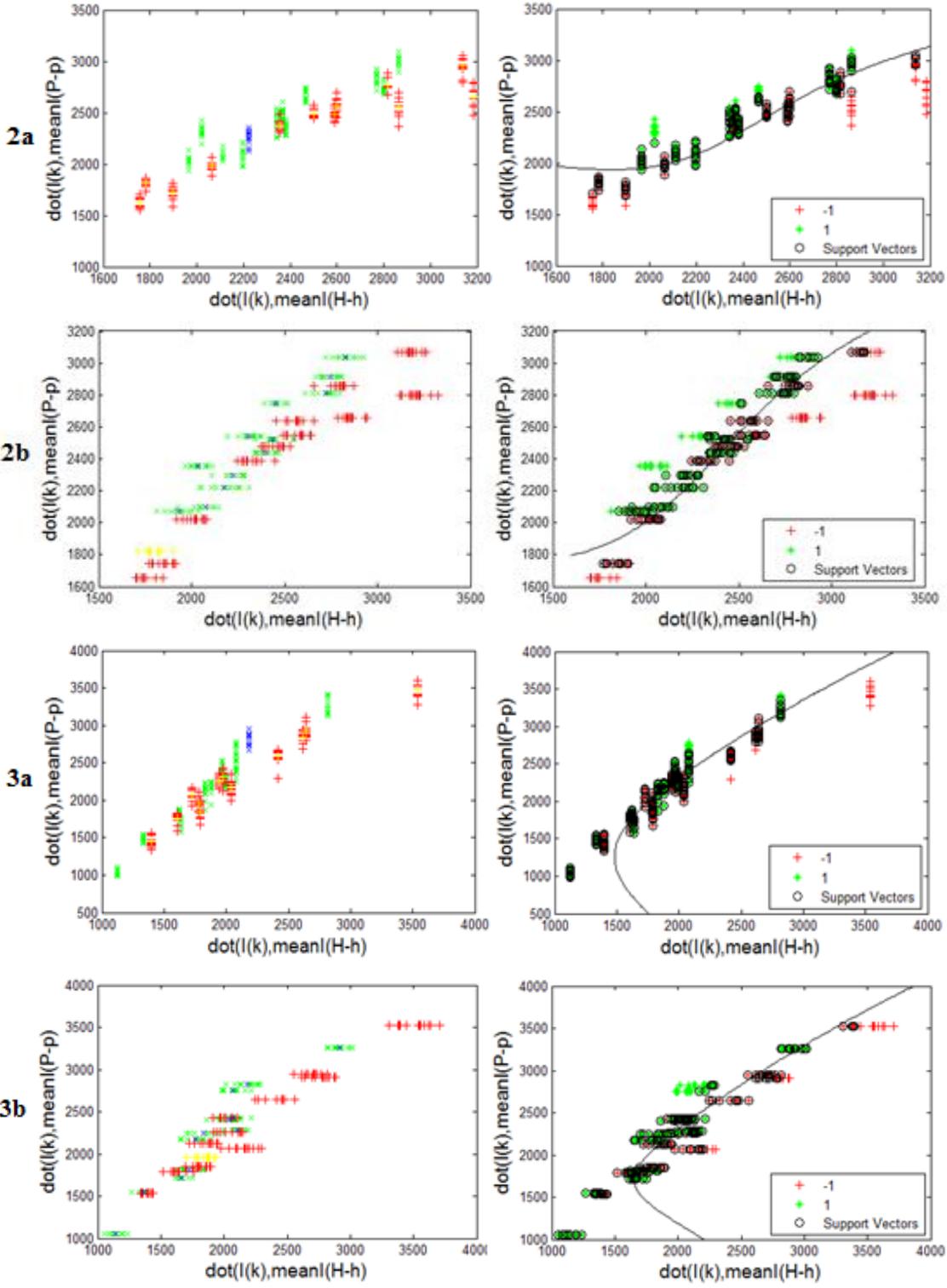
Figure 4.20: The SVM classification results obtained by using dot product similarity method in (1) PCC ROI, (2) LIPL ROI and (3) RIPL ROI. (a), (b) parts relate to D-LOO-CV healthy and patient test samples respectively

Taking mean of the SVM result averages mentioned in Table 4.5 for all similarity measure methods in each ROI, shows that similarity measure methods give the best average classification result for the LIPL ROI as shown in Table 4.7.

Table 4.7: Mean of the SVM results averages for all similarity measure methods in each ROI

Similarity Measures	PCC ROI average	LIPL ROI average	RIPL ROI average	all ROIs average
Cosine	55.5%	68.5%	64%	59%
Dot product	66.5%	65%	68%	66%
Correlation	57%	69%	64.5%	65%
Mean	60%	**67.5%**	65.5%	63%

The Table 4.7 presents that in average all similarity measure methods has obtained best results in LIPL ROI. The Figure 4.21 represents the distribution of the samples and resulting hyper surface by using the cosine, dot product and correlation similarity methods for LIPL ROI, respectively.
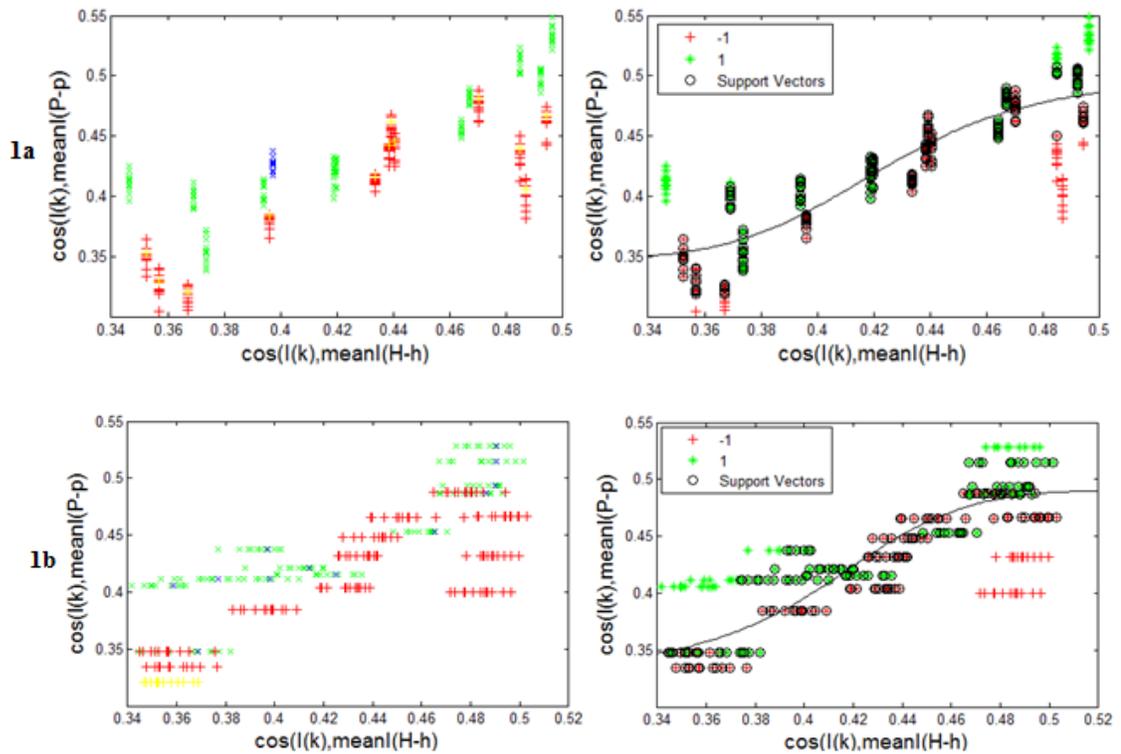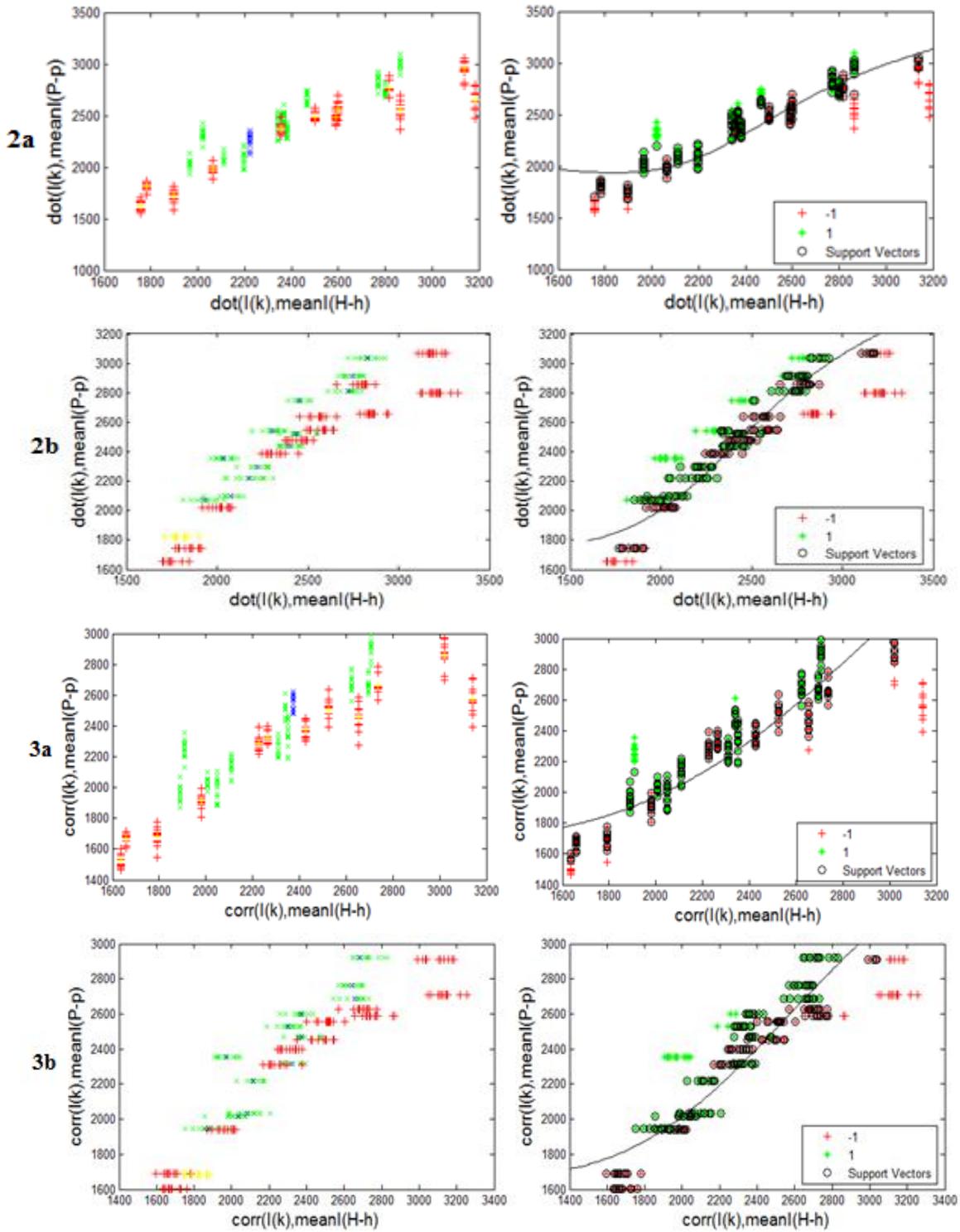
Figure 4.21: The SVM classification results obtained by using (1) cosine (2) dot product (3) correlation similarity methods for LIPL ROI. (a), (b) parts relate to D-LOO healthy and patient test samples respectively

4.4.3 SVM Classification Results Using Dimensionality Reduction Methods

As stated before, since the data had large dimensionality and was complicated too much to classify two categories only using a classification method, it was decided to reduce the dimensionality of the data using linear and nonlinear methods as PCA, LDA and KPCA after subsampling of the data. The SVM classification results obtained after the dimensionality reduction of functional connectivity data was presented in Table 4.8 for double Leave-one-out healthy and patient test sets separately. Also the average of healthy and patient results is presented in the Table.

Table 4.8: Summary of the SVM classification results for D-LOO-CV of healthy and patient test sets

Dimensionality Reduction Methods	ROI Types											
	PCC			LIPL			RIPL			all ROIs		
	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average	D-LOO healthy	D-LOO patient	Average
PCA	71%	40%	55.5%	11%	78%	44.5%	50%	75%	62.5%	50%	58%	52%
Kernel PCA	70%	43%	56.5%	43%	40%	41.5%	60%	50%	55%	60%	42%	51%
LDA	60%	52%	56%	50%	34%	42%	66%	67%	66.5%	67%	42%	54.5%

It can concluded from Table 4.8 that using the PCA method for dimensionality reduction of the rsFC data considering PCC ROI resulted in maximum classification accuracy rate for healthy test samples (71%). The worst classification result of healthy test set was obtained for LIPL ROI (11%).

Table 4.8 also shows that the best SVM classification performance for patient test samples (75%) was obtained using PCA method in RIPL ROI. The worst SVM classification result for patient test samples (34%) was obtained using LDA method in LIPL ROI.

When the average of SVM classification performance of healthy and patient test samples are considered, using the LDA method in RIPL ROI cause the best average of SVM classification results (66.5%). The worst average of SVM classification results relates to LIPL ROI when using the kernel PCA method (41.5%).

Taking mean of the SVM result averages mentioned in Table 4.8 for each dimensionality reduction method in three ROIs, shows that the LDA method gives the best average classification results for PCC, LIPL and RIPL ROIs as shown in Table 4.9.

Table 4.9: Mean of the SVM results averages for all ROI in each similarity measure method

Dimensionality Reduction Methods	PCC ROI average	LIPL ROI average	RIPL ROI Average	all ROIs average	Mean
PCA	55.5%	44.5%	62.5%	52%	53.6%
KPCA	56.5%	41.5%	55%	51%	51%
LDA	56%	42%	66.5%	54.5%	**54.75%**

Taking mean of the SVM result averages mentioned in Table 4.8 for all dimensionality reduction methods in each ROI, shows that all dimensionality reduction methods give the best average classification result for the RIPL ROI as shown in Table 4.10.

Table 4.10: Mean of the SVM results averages for all dimensionality reduction methods in each ROI

Dimensionality Reduction Methods	PCC ROI average	LIPL ROI average	RIPL ROI average	all ROIs average
PCA	55.5%	44.5%	62.5%	52%
KPCA	56.5%	41.5%	55%	51%
LDA	56%	42%	66.5%	54.5%
Mean	56%	42.6%	**61.3%**	52.5%

4.5 GMM results

The Table 4.11 presents the accuracy result by 50 EM iterations for healthy and patient subjects and also their average for features extracted by similarity measures on chosen ROIs. The GMM classification was performed with different mixture components ($k = 2, 4, 6, 8, 10, 12$) among which $k = 2$ got the best results. The Table 4.11 presents the GMM results for $k = 2$.

Table 4.11: Summary of the GMM results trained by EM algorithm for D-LOO-CV healthy and patient test samples

Feature Extraction Method	PCC ROI			LIPL ROI			RIPL ROI			All ROIs		
	D-LOO Healthy Mean	D-LOO Patient Mean	Average	D-LOO healthy Mean	D-LOO patient Mean	Average	D-LOO Healthy Mean	D-LOO Patient Mean	Average	D-LOO Healthy Mean	D-LOO Patient Mean	Average
Cosine	47%	41%	44%	76%	71%	73.5%	49%	62%	55.5%	44%	51%	47.5%
Dot product	64%	64%	64%	60%	65%	62.5%	45%	46%	45.5%	58%	64%	61%
Correlation	63%	45%	54%	43%	50%	46.5%	60%	54%	57%	59%	60%	59.5%

It is concluded from Table 4.11 that the best results obtained using cosine similarity method for LIPL ROI for all of healthy, patient and average cases (76%, 71% and 73.5% respectively). The worst result relates to again using cosine similarity for PCC ROI.

The Figure 4.22 presents the healthy and patient training sets obtained by cosine similarity measure in LIPL ROI and modeled separately by a Gaussian mixture with one mixture component.
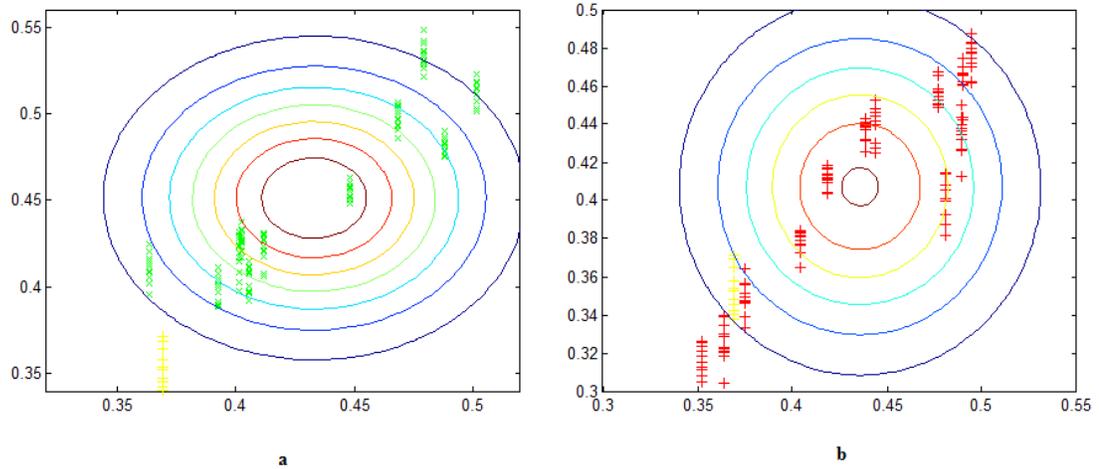


Figure 4.22: Fit GMM on (a) healthy and (b) patient training sets using one mixture component (k=1)

The Figure 4.23 presents the healthy and patient training sets obtained by cosine similarity measure in LIPL ROI and modeled separately by a Gaussian mixture with two mixture components.
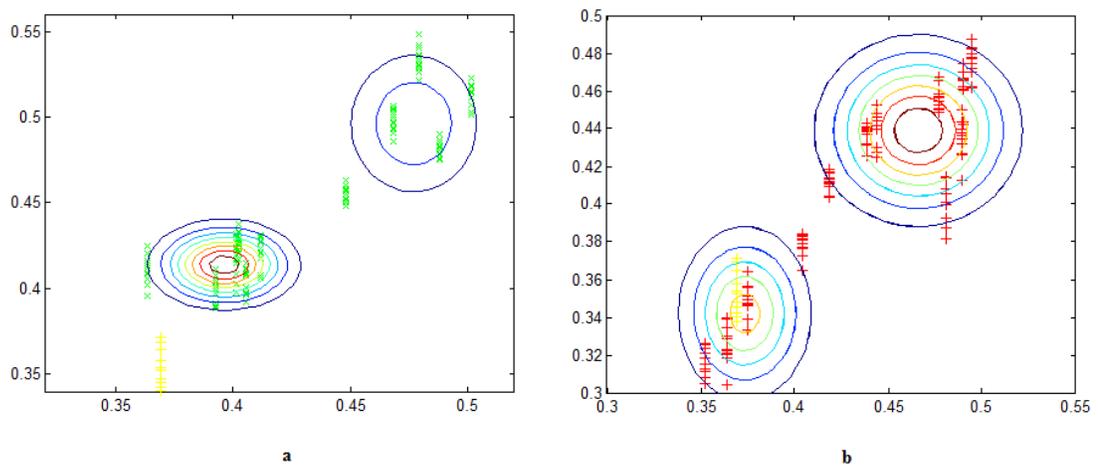

Figure 4.23: Fit GMM on (a) healthy and (b) patient training sets using two mixture components (k=2)

67

CHAPTER 5

CONCLUSION

5.1 Summary

The main purpose of this thesis is the use of the resting-state functional connectivity for discriminative analysis of obsessive compulsive disorder. On the resting-state fMRI data of 12 healthy and 12 patient with OCD. the functional connectivity between three regions of brain (PCC, LIPL, RIPL) and all brain voxels were computed by Pearson's correlation coefficient available in the CONN program, a Matlab toolbox. In order to make a decision between healthy and OCD groups, we considered the functional connectivity as a pattern recognition problem which consists of feature extraction and classification components. Several methods were proposed for each component such that for feature extraction two approaches were examined. In the first approach the FC data was subsampled and then dimensionality reduction methods were applied to reduce the dimensionality. In the second approach the similarity measurement methods were applied on FC data. The feature vectors extracted from the first component were used as inputs for classification part. Using Double leave-one-out cross-validation method the training and test sets were generated from feature vectors. In order to discriminate between OCD and healthy groups the classification methods as support vector machine and Gaussian mixture models were examined. Finally OCD groups were detected according to the classification results.

5.2 Discussions

Through this study four different approaches considered for feature extraction and classification methods to get the best classification results.

Firstly, the similarity measurement methods as correlation, dot product and cosine similarity were used to extract the feature vectors and their classification accuracy was evaluated only by comparing the values of two produced feature vectors. Best was obtained by cosine similarity on PCC ROI (average accuracy 53.5%).

As a second approach, the SVM classifier is used to discriminate healthy and OCD classes through feature vectors obtained by similarity measures. It is observed using the SVM method improved the classification performance with respect to deciding by simple comparison of similarity values. Using the correlation method on the LIPL ROI got the best average SVM classification result (69%) among all similarity measurement methods and all ROIs.

Thirdly, the subsampling followed by dimensionality reduction methods as PCA, KPCA and LDA were used to extract feature vectors for SVM classification. Best average accuracy (66.5%) was obtained by LDA on RIPL ROI. Notice that this result is not better than second approach given above although the approach is more complicated and time consuming.

As the last approach, the combination of similarity measurement as feature extraction method and the GMM as a classifier was used to detect OCD groups. In this case, the cosine similarity method got the best average classification performance (73.5%) of both healthy and patient test samples on LIPL ROI.

In total consideration of four approaches, it is observed that the last approach with cosine similarity followed by GMM on LIPL ROI resulted in the best average classification accuracy rate (73.5%). It is also best for healthy and patient performances.

So the most appropriate method to discriminate OCD from healthy brain is to apply cosine similarity method to extract the features and use GMM as the classifier.

Also comparing the results obtained by simple comparison of similarity measurement values (first approach) with results of using a classifier (second, third and fourth approaches) demonstrates that using SVM or GMM classification methods can enhance detection of the OCD. Another result derived from this research is that the similarity measurement methods were more accomplished than dimensionality reduction methods to extract the proper features from rsFC data.

This study concludes that first, the functional connectivity differences in LIPL and RIPL (especially LIPL) ROIs of OCD brain were most helpful to detect this disease. Second, This thesis supports the research has done by Koçak et al. [2] that suggests there are functional abnormalities in right fronto-parietal networks of the OCD brain including IPL and PCC regions. Also it is important that this study encourages employing pattern recognition algorithms on rs-FC data for detection of OCD, although it is not observable directly in MRI.

5.3 Future work

As the use of machine learning methods for detection of mental disease has not long history, many methods may exist to detect them. In this study the resting-state functional connectivity between three ROIs and all brain voxels were considered while the rs-FC among ROIs or among voxels can be also studied as input data. Moreover, other feature extraction methods rather than similarity measurements or PCA, KPCA and LDA methods can be used to extract features for better discrimination of the OCD from healthy group. For example independent component analysis (ICA), locally linear embedding (LLE) and Adaptive Boosting (AdaBoost) methods may result in better performance. Also other classification methods can be examined to evaluate whether they perform better classification results may enhance the OCD detection. Further, the classification of OCD can be done according to the obsession type of the patients. This research can be generalized for detection of other psychological and mental disorders.

Also, pattern recognition methods can be applied to task-related state of the brain of OCD. Analyze the rs-FC of OCD brain by inciting the OCD disease putting patient in triggering conditions is another work can be focused on in future.

REFERENCES

[1]   M. D. Fox, and M. Greicius, "Clinical Applications of Resting State Functional Connectivity," *Frontiers in Systems NeuroScience*, 2010 March 11.

[2]   O. M. Koçak, E. Kale, and M. Çiçek, "Default Mode Network Connectivity Differences in Obsessive Compulsive Disorders," *Activitas Nervosa Superior*, 2012.

[3]   M. Weygandt, C. R. Blecker, A. Schafer, K. Hacicmack, J. D. Haynes, D. Vaitl, R. Stark, and A. Schienle, "fMRI pattern recognition in obsessive-compulsive disorder," *Neuroimage,* vol. 60, no. 2, pp. 1186-1193, Apr, 2012.

[4]   B. J. Harrison, C. Soriano-Mas, J. Pujol, H. Ortiz, M. Lopez-Sola, R. Hernandez-Ribas, J. Deus, P. Alonso, M. Yucel, C. Pantelis, J. M. Menchon, and N. Cardoner, "Altered Corticostriatal Functional Connectivity in Obsessive-compulsive Disorder," *Archives of General Psychiatry,* vol. 66, no. 11, pp. 1189-1200, Nov, 2009.

[5]   M. Behroozi, M. R. Daliri, and H. Boyaci, "Statistical Analysis Methods for the fMRI Data," *Basic and Clinical NeuroScience,* vol. Volume 2, no. number 4, 10 June 2011.

[6]   R. A. Poldrack, J. A. Mumford, and T. E. Nicholas, *Handbook of Functional MRI Data Analysis*, New York: Cambirdge University Press, 2011.

[7]   K. Friston, and C. Buchel, "Functional Connectivity: Eigenimages And Multivariate Analyses," *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny, eds., pp. 492 - 501, Amesterdam: Elsevier, 2007.

[8]   A. Aertsen, and H. Preissl, "DYNAMICS OF ACTIVITY AND CONNECTIVITY IN PHYSIOLOGICAL NEURONAL NETWORKS," *Nonlinear Dynamics and Neuronal Networks,* vol. 2, pp. 281-301, 1991.

[9]   F. Deleus, and M. M. Van Hulle, "Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis," *Journal of Neuroscience Methods,* vol. 197, no. 1, pp. 143-157, Apr, 2011.

[10]   S. Whitfield-Gabrieli, and A. Nieto-Castanon, "Conn:A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks " *BRAIN CONNECTIVITY,* vol. 2, 2012.

[11]   J. S. Abramowitz, S. Taylor, and D. McKay, "Obsessive-compulsive disorder," *Lancet,* vol. 374, no. 9688, pp. 491-499, Aug, 2009.

[12]   S. E. Stewart, M. C. Rosario, T. A. Brown, A. S. Carter, J. F. Leckman, D. Sukhodolsky, L. Katsovitch, R. King, D. Geller, and D. L. Pauls, "Principal components analysis of obsessive-compulsive disorder symptoms in children and adolescents," *Biological Psychiatry,* vol. 61, no. 3, pp. 285-291, Feb, 2007.

[13]   L. R. Baxter, J. M. Schwartz, K. S. Bergman, M. P. Szuba, B. H. Guze, J. C. Mazziotta, A. Alazraki, C. E. Selin, H. K. Ferng, P. Munford, and M. E. Phelps, "CAUDATE GLUCOSE METABOLIC-RATE CHANGES WITH BOTH DRUG AND BEHAVIOR-THERAPY FOR OBSESSIVE-COMPULSIVE DISORDER," *Archives of General Psychiatry,* vol. 49, no. 9, pp. 681-689, Sep, 1992.

[14]   J. L. Roffman, C. D. Marci, D. M. Glick, D. D. Dougherty, and S. L. Rauch, "Neuroimaging and the functional neuroanatomy of psychotherapy," *Psychological Medicine,* vol. 35, no. 10, pp. 1385-1398, Oct, 2005.

[15]   C. M. Adler, P. McDonough-Ryan, K. W. Sax, S. K. Holland, S. Arndt, and S. M. Strakowski, "fMRI of neuronal activation with symptom provocation in unmedicated patients with obsessive compulsive disorder," *Journal of Psychiatric Research,* vol. 34, no. 4-5, pp. 317-324, Jul-Oct, 2000.

[16]   J. H. Jang, J. H. Kim, W. H. Jung, J. S. Choi, M. H. Jung, J. M. Lee, C. H. Choi, D. H. Kang, and J. S. Kwon, "Functional connectivity in fronto-subcortical circuitry during the resting state in obsessive-compulsive disorder," *Neuroscience Letters,* vol. 474, no. 3, pp. 158-162, May, 2010.

[17]   K. Christoff, A. M. Gordon, J. Smallwood, R. Smith, and J. W. Schooler, "Experience sampling during fMRI reveals default network and executive system contributions to mind wandering," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 106, no. 21, pp. 8719-8724, May, 2009.

[18]   O. M. Kocak, A. Y. Ozpolat, C. Atbasoglu, and M. Cicek, "Cognitive control of a simple mental image in patients with obsessive-compulsive disorder," *Brain and Cognition,* vol. 76, no. 3, pp. 390-399, Aug, 2011.

[19]   N. Hon, R. A. Epstein, A. M. Owen, and J. Duncan, "Frontoparietal activity with minimal decision and control," *Journal of Neuroscience,* vol. 26, no. 38, pp. 9805-9809, Sep, 2006.

[20]   B. Suchan, R. Botko, E. Gizewski, M. Forsting, and I. Daum, "Neural substrates of manipulation in visuospatial working memory," *Neuroscience,* vol. 139, no. 1, pp. 351-357, 2006.

[21]   C. M. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.

[22]   R. C. Craddock, P. E. Holtzheimer, X. P. P. Hu, and H. S. Mayberg, "Disease State Prediction From Resting State Functional Connectivity," *Magnetic Resonance in Medicine,* vol. 62, no. 6, pp. 1619-1628, Dec, 2009.

[23]   H. Shen, L. B. Wang, Y. D. Liu, and D. W. Hu, "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI," *Neuroimage,* vol. 49, no. 4, pp. 3110-3121, Feb, 2010.

[24] O. D. Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition - A survey," *Pattern Recognition,* vol. 29, no. 4, pp. 641-662, Apr, 1996.

[25] Y. Tang, L. F. Wang, F. Cao, and L. W. Tan, "Identify schizophrenia using resting-state functional connectivity: an exploratory research and analysis," *Biomedical Engineering Online,* vol. 11, Aug, 2012.

[26] S. V. Shinkareva, V. Gudkov, and J. Wang, "A Network Analysis Approach to fMRI Condition-Specific Functional Connectivity."

[27] Z. W. Wang, S. K. M. Wong, and Y. Y. Yao, "AN ANALYSIS OF VECTOR-SPACE MODELS BASED ON COMPUTATIONAL GEOMETRY," *Sigir 92 : Proceedings of the Fifteenth Annual International Acm Sigir Conference on Research and Development in Information Retrieval*, pp. 152-160, 1992.

[28] K. Yang, and C. Shahabi, "A PCA-based Similarity Measure for Multivariate Time Series ".

[29] D. D. Garrett, N. Kovacevic, A. R. McIntosh, and C. L. Grady, "Blood Oxygen Level-Dependent Signal Variability Is More than Just Noise," *Journal of Neuroscience,* vol. 30, no. 14, pp. 4914-4921, Apr, 2010.

[30] A. Karnik, S. Goswami, and R. Guha, "Detecting obfuscated viruses using cosine similarity analysis," *AMS 2007: First Asia International Conference on Modelling & Simulation Asia Modelling Symposium, Proceedings*, pp. 165-170, 2007.

[31] A. Suebsing, and Hiransakolwong.Nualsawat, "A Novel Technique for Feature Subset Selection Based on Cosine Similarity," 133, Applied Mathematical Sciences, 2012, pp. 6627-6655.

[32] L. Muflikhah, B. Baharudin, and I. C. Society, "Document Clustering using Concept Space and Cosine Similarity Measurement," *Proceedings of the 2009 International Conference on Computer Technology and Development, Vol 1*, pp. 58-62, 2009.

[33] J. Borge-Holthoefer, Y. Moreno, and A. Arenas, "Modeling Abnormal Priming in Alzheimer's Patients with a Free Association Network," *Plos One,* vol. 6, no. 8, Aug, 2011.

[34] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K. M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with the meanings of nouns," *Science,* vol. 320, no. 5880, pp. 1191-1195, May, 2008.

[35] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation,* vol. 10, no. 5, pp. 1299-1319, Jul, 1998.

[36] R. O. Duda, and P. Hart, *Pattern Classification and Scene Analysis*, First ed., New York: Wiley, 1973.

[37] I. T. Jolliffe, *Principal Component Analysis*, second ed., New York: Springer-Verlag, 2002.

[38] K. I. Diamantaras, and S. Y. Kung, *Principal Component Neural Networks Theory and Applications*, New York: Wiley, 1996.

[39] Y. Xu, D. Zhang, and J. Y. Yang, "A feature extraction method for use with bimodal biometrics," *Pattern Recognition,* vol. 43, no. 3, pp. 1106-1115, Mar, 2010.

[40] Y. Xu, D. Zhang, H. Yang, and J. Y. Yang, "An approach for directly extracting features from matrix data and its application in face recognition," *Neurocomputing,* vol. 71, no. 10-12, pp. 1857-1865, Jun, 2008.

[41] Y. Xu, D. Zhang, F. X. Song, J. Y. Yang, Z. Jing, and M. Li, "A method for speeding up feature extraction based on KPCA," *Neurocomputing,* vol. 70, no. 4-6, pp. 1056-1061, Jan, 2007.

[42] Y. Xu, C. Lin, and W. Zhao, "Producing computationally efficient KPCA-based feature extraction for classification problems," *Electronics Letters,* vol. 46, no. 6, pp. 452-U100, Mar, 2010.

[43] Y. Xu, and D. Zhang, "Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme," *Optical Engineering,* vol. 49, no. 3, Mar, 2010.

[44] L. Zhang, R. Lukac, X. L. Wu, and D. Zhang, "PCA-Based Spatially Adaptive Denoising of CFA Images for Single-Sensor Digital Cameras," *Ieee Transactions on Image Processing,* vol. 18, no. 4, pp. 797-812, Apr, 2009.

[45] F. Song, Z. Guo, and D. Mei, "Feature Selection Using Principal Component Analysis," in International Conference on System Science, Engineering Design and Manufacturing Informatization, 2010.

[46] R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, and J. C. Mazziotta, "Characterizing Distributed Functional Systems," *Human Brain Function*, pp. 107-126, 1997.

[47] N. Ramnani, T. E. J. Behrens, W. Penny, and P. M. Matthews, "New approaches for exploring anatomical and functional connectivity in the human brain," *Biological Psychiatry,* vol. 56, no. 9, pp. 613-619, Nov, 2004.

[48] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Second ed., p.^pp. 117-124, New York: Wiley-Interscience, 2000.

[49] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Processing Letters,* vol. 9, no. 2, pp. 40-42, Feb, 2002.

[50] G. S. Sidhu, N. Asgarian, R. Greiner, and a. R. G. Brown, "Kernel Principal Component Analysis for Dimensionality Reduction in fMRI-Based Diagnosis of ADHD," *Frontiers in Systems Neuroscience*, 2012,November.

[51] k. Fukunaga, *Introduction to Statistical Pattern Recognition*, Second ed.: Academic Press, 1990.

[52] R. A. Fisher, "The Use of Multiple Measurments in Taxonomic Problems," *Annals of Eugenics,* vol. 7, no. 2, pp. 179-188, September,1936.

[53] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, Jul, 1997.

[54] D. Cai, X. He, J. Han, and Ieee, "Spectral regression for efficient regularized subspace learning," *IEEE International Conference on Computer Vision.* pp. 214-221, 2007.

[55] D. Cai, X. F. He, and J. W. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *Ieee Transactions on Knowledge and Data Engineering,* vol. 20, no. 1, pp. 1-12, Jan, 2008.

[56] Z. H. Qiao, L. Zhou, and J. H. Z. Huang, "Effective linear discriminant analysis for high dimensional, low sample size data," *World Congress on Engineering 2008, Vols I-Ii*, pp. 1070-1075, 2008.

[57] K. Torkkola, "Discriminative features for text document classification," *Pattern Analysis and Applications,* vol. 6, no. 4, pp. 301-308, Feb, 2003.

[58] Y. Q. Cheng, K. Liu, J. Y. Yang, Y. M. Zhuang, and N. C. Gu, "HUMAN FACE RECOGNITION METHOD BASED ON THE STATISTICAL-MODEL OF SMALL SAMPLE-SIZE," *Proceedings of the Society of Photo-Optical Instrumentation Engineers (Spie).* pp. 85-95, 1992.

[59] S. Baker, S. K. Nayar, and Ieee, "Pattern rejection," *Proceedings / Cvpr, Ieee Computer Society Conference on Computer Vision and Pattern Recognition.* pp. 544-549, 1996.

[60] Y. T. Cui, D. L. Swets, J. J. Weng, and Ieee, "Learning-based hand sign recognition using SHOSLIF-M," *Fifth International Conference on Computer Vision, Proceedings*, pp. 631-636, 1995.

[61] J. H. Friedman, "REGULARIZED DISCRIMINANT-ANALYSIS," *Journal of the American Statistical Association,* vol. 84, no. 405, pp. 165-175, Mar, 1989.

[62] T. Hastie, A. Buja, and R. Tibshirani, "PENALIZED DISCRIMINANT-ANALYSIS," *Annals of Statistics,* vol. 23, no. 1, pp. 73-102, Feb, 1995.

[63] V. N. Vapnik, *The Nature of Statistical Learning Theory*, second ed., New York: Springer-Werlag, 1995.

[64] J. Mourao-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data," *Neuroimage,* vol. 28, no. 4, pp. 980-995, Dec, 2005.

[65]   D. D. Cox, and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," *Neuroimage,* vol. 19, no. 2, pp. 261-270, Jun, 2003.

[66]   S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. P. Hu, "Support vector machines for temporal classification of block design fMRI data," *Neuroimage,* vol. 26, no. 2, pp. 317-329, Jun, 2005.

[67]   Y. Fan, D. G. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: Classification of morphological patterns using adaptive regional elements," *Ieee Transactions on Medical Imaging,* vol. 26, no. 1, pp. 93-105, Jan, 2007.

[68]   J. D. Haynes, and G. Rees, "Decoding mental states from brain activity in humans," *Nature Reviews Neuroscience,* vol. 7, no. 7, pp. 523-534, Jul, 2006.

[69]   N. Kriegeskorte, R. Goebel, and P. Bandettini, "Information-based functional brain mapping," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 103, no. 10, pp. 3863-3868, Mar, 2006.

[70]   E. Mjolsness, and D. DeCoste, "Machine learning for science: State of the art and future prospects," *Science,* vol. 293, no. 5537, pp. 2051-+, Sep, 2001.

[71]   C. Z. Zhu, Y. F. Zang, Q. J. Cao, C. G. Yan, Y. He, T. Z. Jiang, M. Q. Sui, and Y. F. Wang, "Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder," *Neuroimage,* vol. 40, no. 1, pp. 110-120, Mar, 2008.

[72]   B. Scholkopf, and A. Solma, "Learning Witht Kernels," MIT Press, 2002.

[73]   G. Garga, G. Prasada, L. Gargb, and D. Coylea, "Gaussian Mixture Models for Brain Activation Detection From FMRI Data," *International Journal of Bioelectromagnetism,* vol. 13, no. 4, pp. 255-260, 2011.

[74]   J. Marques, and P. J. Moreno, "A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines ", Compaq Computer Corporation, June,1999.

[75]   F. Segovia, J. M. Gorriz, J. Ramirez, D. Salas-Gonzalez, I. A. Illan, M. Lopez, R. Chaves, C. G. Puntonet, E. W. Lang, and I. R. Keck, "fMRI Data Analysis Using a Novel Clustering Technique," *IEEE Nuclear Science Symposium Conference Record.* pp. 3399-3403, 2009.

[76]   A. Lassl, J. M. G6rriz, J. Ramirez, D. Salas-Gonzalez, C. G. Puntonet, and E. W. Lang, "Clustering Approach for the Classification of SPECT images," in IEEE Nuclear Science Symposium Conference, 2008, pp. 5345-5348.

[77]   C. Z. Zhu, Y. F. Zang, M. Liang, L. X. Tian, Y. He, X. B. Li, M. Q. Sui, Y. F. Wang, and T. Z. Jiang, "Discriminative analysis of brain function at resting-state for attention-deficit/hyperactivity disorder," *Medical Image Computing and Computer-Assisted Intervention - Miccai 2005, Pt 2,* vol. 3750, pp. 468-475, 2005.

[78]   Tx. Yarkoni, J. R. Gray, E. R. Chrastil, D. M. Barch, L. Green, and T. S. Braver, "Sustained neural activity associated with cognitive control during temporally extended decision making," *Cognitive Brain Research,* vol. 23, no. 1, pp. 71-84, Apr, 2005.

[79]   J. Talairach, and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain:3-D Proportional System:An Approach to Cerebral Imaging*, p.^pp. 122: Thieme Medical Pub, 1988.

[80]   S. Whitfield-Gabrieli, and A. Nieto-Castanon, "CONN: FMRI Funcional Connectivity Toolbox Version 13," Massachusetts Institute of Technology, 2011, pp. 1-14.

[81]   B. Scholkopf, A. Smola, and K.-R. Müller, "Kernel Principal Component Analysis," *MIT Press*, pp. 327-352, 1999.

[82]   A. Suebsing, and N. Hiransakolwong, "Feature Selection Using Euclidean Distance and Cosine Similarity for Intrusion Detection Model," *2009 First Asian Conference on Intelligent Information and Database Systems*, pp. 86-91, 2009.

[83]   S. S. Keerthi, and C. J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Computation,* vol. 15, no. 7, pp. 1667-1689, Jul, 2003.

[84]   C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," April 15,2010.

[85]   C. Cortes, and V. Vapnik, "SUPPORT-VECTOR NETWORKS," *Machine Learning,* vol. 20, no. 3, pp. 273-297, Sep, 1995.