Functional Magnetic Resonance Image (FMRI) Based Brain Activity Classification

Mehran Sahandi Far

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
July 2013
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

Certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Computer Engineering.

Assoc. Prof. Dr. Muhammed Salamah
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Hakan Altınçay
Supervisor

Examining Committee

1. Prof. Dr. Hakan Altınçay

2. Prof. Dr. Hasan Kömürcügil

3. Asst. Prof. Dr. Ahmet Ünveren

ABSTRACT

Functional Magnetic Resonance Imaging (FMRI) is a powerful tool to predict the current activity in the human brain. With the help of machine learning tools, the cognitive function of the human brain can be automatically classified into two or more states by analyzing the FMRI images. In this thesis, the main goal is to design an automated system to predict whether a given subject is viewing a picture or a sentence. A dataset of six subjects is considered for this purpose. Two classification schemes, namely support vector machines (SVM) and nearest neighbor classifier (NN) are used. Due to the high dimensionality of the FMRI data, feature selection is generally considered. In order to reduce the feature dimensionality, four reduction methods, namely region of interest (ROI), N-most active voxels, ROI average and N-most active voxels within ROI are studied. Both subject dependent and subject independent experiments are conducted where the former studies the categorization problem separately for each subject and the latter does not use the tested subject during training.

Experimental results have shown that SVM provides better scores compared to NN approach and selecting N-most active voxels within the ROI provided the best scores, verifying the importance of applying feature selection in this domain.

Keywords: Functional Magnetic Resonance Imaging, Cognitive State Decoding, Feature Selection, Support Vector Machines, Nearest Neighbor Classifier

ÖZ

Fonksiyonel Manyetik Rezonans Görüntüleme (FMRG) insan beyninde şu andaki aktiviteyi tahmin etmek için güçlü bir araçtır. Makineye dayalı öğrenme araçları yardımıyla insan beyninin bilişsel fonksiyonu iki veya daha fazla durumdan birine FMRG görüntüleri incelenerek otomatik olarak sınıfladırılabilmektedir. Bu tezde, esas amaç bir kişinin bir resime veya bir cümleye bakmakta olduğunu ayırt edecek otomatik bir sistem tasarlamaktır. Bu amaçla altı kişi içeren bir verikümesi kullanılmıştır. Destek vectör makinaları (DVM) ve enyakın komşu (EK) olmak üzere iki sınıflandırıcı kullanılmıştır. FMGR verisinin yüksek boyutlu olmasından dolayı genellikle öznitelik seçimi uygulanmaktadır. Öznitelik boyunun azaltılması için İlgi Alanı (İA), N-En aktif voxel, İA ortalama ve İA içerisindeki N-En aktif voxel olmak üzere dört farklı yöntem denenmiştir. Kişiye bağlı ve kişiden bağımsız deneyler yapılmış olup, ilkinde sınıflandırma promlemi her kişi için ayrı olarak çalışılmış, ikincisinde ise test edilen kişi eğitme verisi içinde yer almamıştır.

Deneysel sonuçlar DVM yaklaşımının EK'ya göre daha başarılı sonuçlar verdiğini, İA içerisindeki N-En aktif voxel seçiminin de eniyi başarımı sağladığını göstermiş ve bu alanda öznitelik seçmenin önemini onaylamıştır.

Anahtar Kelimeler: Fonksiyonel Manyetik Rezonans Görüntüleme, Bilişsel Durum Çözümleme, Öznitelik Seçme, Destek Vektör Makinaları, Enyakın Komşu Sınıflandırıcısı

I would like to dedicate this thesis to my lovely parents

For their endless love and supports.

ACKNOWLEDGMENTS

There are a number of people without whom this thesis might not have been written, and to whom I greatly indebted.

I would like to cordially thank my dear supervisor Prof. Dr. Hakan Altınçay for his guidance, support and encouragements throughout this study. I sincerely thank Asst. Prof. Dr. Adnan Acan for his all useful classes and compassionate help. Also, I would like to foreordain committee members of my thesis defense for their helpful comments.

I would like to extend my gratitude to dear Lord who has always been there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 FMRI Data Analysis

The brain is the most fascinating and the least understood organ in the human body. For centuries, the brain has been an unknown entity and scientists have pondered how it works. Proving that the brain is the main controller of the human body was the first step for further studies. In the 17th century, Thomas Willis proposed that each region of the brain has a specific function and, in the 19th century, Gall introduced two important facts. Firstly, the brain is responsible for moral, intellectual and all physiological activities of the human body. Secondly, different regions of brain response to different activities. These two facts led scientists to design a brain map which contains each area with its responsibilities. The studies in 19th century resulted in a better understanding of the mysteries of the brain by using electrical currents to stimulate the cortex of animal and human brain's mapping [1]. These maps were not complete and precise, but they were later completed in the 20th century by using new methods such as cortical stimulation by employing electrodes [1, 2]. The invention of brain imaging technologies towards the end of the century was a milestone in this area. Methods such as *Positron Emission Tomography* (PET) [3] and *Magnetic Resonance Imaging* (MRI) [4] have opened a new window for deeper understanding of the brain. Although MRI was not the first method

used in the brain and neuroscience studies, it became the most popular method in recent years [4].

MRI is a technique which uses a strong nuclear magnetic field to produce a visual representation of internal body structures and biological tissues such as bones, organs or soft tissues [4, 5]. The MRI scanner can detect properties and distinguish between tissue types using different pulse sequence types where a pulse sequence is a series of changes in a magnetic field gradient and oscillating electromagnetic fields that allow the MRI scanner to create a sensitive images of the human body [4, 3]. For example, MRI can reveal differences between the grey level and the white level of the brain matter. MRI produces high resolution images which help detecting disorders and problems in the other parts of body as well.

Functional MRI (FMRI) uses MRI technology to measure changes in the blood oxygen level and evaluate metabolic changes in the brain over time [5]. FMRI is a series of MRI images which are taken over a period of time with lower resolution than normal MRI images [4]. These images measure neuron activation changes during a stimuli. FMRI plays a major role in human brain studies and brain mapping, which is generally seen as a safe, fast and reliable method [4, 5, 6].

Since the beginning of the FMRI studies, analysis of its massive data is questioned. Using methods such as clinical and human experts based analysis are time consuming and not reliable. On the other hand, machine learning methods are generally considered as the most logical solution to the problem of analyzing this type of data. Therefore, machine

learning methods played important roles in recent FMRI data analysis studies and researchers focused on developing models and software to carry out this task.

When FMRI was first discovered, most of the studies are done on individual voxel basis. After a while, analysis of multi box activation patterns during particular stimulus has become increasingly common. This is known as *multi voxels pattern analysis* (MVPA) [7]. Using this approach, representation of a particular activity within the brain can be obtained. For instance, graphs of relative timing of an activation within a particular part of the brain or network diagram that shows the functional relation among many different regions also can be used in brain disorder studies [4]. Moreover, the task which is performed by a subject referred as brain reading can be detected [7]. It also provides answers to unlimited number of questions about how the brain works. The FMRI technology and its combination with other brain study models also helped the researchers in psychological studies [8].

The aforementioned applications of FMRI data analysis need the use of pattern classification technology. Pattern classification science corresponds to labeling unseen data as one of the previously known classes or groups [9]. A *pattern class* denotes a group of objects or data which they have same or common properties. Each object is characterized by a set of features where each feature element represents a distinguishing property of the objects. Each object used in classification is known as a *sample* and collection of obtained samples is named as a data set [10, 11].

The solution of a pattern classification problem includes two main stages, *training* and *testing* [10]. The main goal of training is *model estimation* where the decision boundaries in the feature space are computed. The test phase corresponds to computing the most likely class that the unseen data may belongs [9, 11].

Both training and testing may include *pre-processing* which aims to remove noise and normalize the data. In practice, the features computed may not be jointly best-fitting. In such a case, *feature selection* where the redundant information is removed is applied in order to simplify the computations and obtain more discriminative feature vectors [12].

In the case of FMRI classification, class stands for the kind of stimuli shown or the task asked during the image acquisition phase. Features are the neuron activation levels during each stimulus. A sample is the data recorded for each stimulus in the form of FMRI images. Different samples that are recorded in FMRI data acquisition process are represented as vectors of large number of elements [6, 13, 14].

1.2 Objectives

In this study, three important tasks in FMRI data analysis are considered. Evaluation of different machine learning algorithm to study their relative performance on FMRI data is addressed. Dimensionality reduction to select an informative feature subset to improve the classifier accuracy and reduce the computational complexity and, comparing subject independent and dependent models to show the freedom of brain activity prediction from subjects are also considered.

Two well-known classification algorithms are used to predict the stimulus type in the brain, namely *support vector machine* (SVM) [15] and *nearest neighbor classifier* (NN) [16]. Their average performance over different experiments are computed and compared.

We applied several dimensionality reduction techniques such as *region of interest* (ROI) and N-most active voxels to reduce the dimensionality of feature vectors which is one of the principal problems in FMRI data analysis.

The experiments conducted can be categorized into two broad groups: *subject dependent* where each subject is studied independently and *subject independent* where different subjects are used in training and test phases. For both experiments, feature reduction algorithms are applied.

The rest of this thesis is organized as follows: Chapter 2 presents a literature review including information on brain structure and anatomy, introduction to FMRI and BOLD, FMRI data classification and feature selection. Chapter 3 presents information on the classification of pre-processed data set. Chapter 4 presents experimental results in simulation studies. Chapter 5 provides the conclusions drawn and information about future studies on this topic.

Chapter 2

LITERATURE REVIEW

2.1 The Anatomy and Structure of the Brain

The human brain is divided into two hemispheres, left and right and into four different regions, each of which has its own special functionality. These regions are frontal lobe, parietal lobe, temporal lobe and occipital lobe [1] which are shown in Figure 1. Although the functional organization of the brain is poorly understood, the primary functions of the aforementioned areas are known [17, 18]. For example, the primary visual cortex is located in the occipital lobe. It deals with the reception and explanation of visual signals which are mapped on the left hemisphere from the right retina. Temporal lobe is responsible for auditory action in the human brain. It has the same structure as visual cortex where the left side ear mapped on the right side. Basic functions of other lobes of the brain are also known today [1, 19]. The location of these functions are shown in Figure 2. Some basic relations of these parts also have been studied and referred as the functional connectivity of brain regions [18].

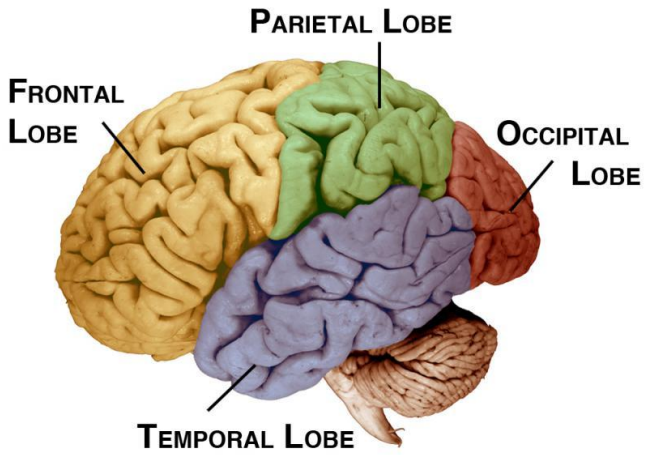

Figure 1. Different Lobes of Brain Separated by Colors [20]

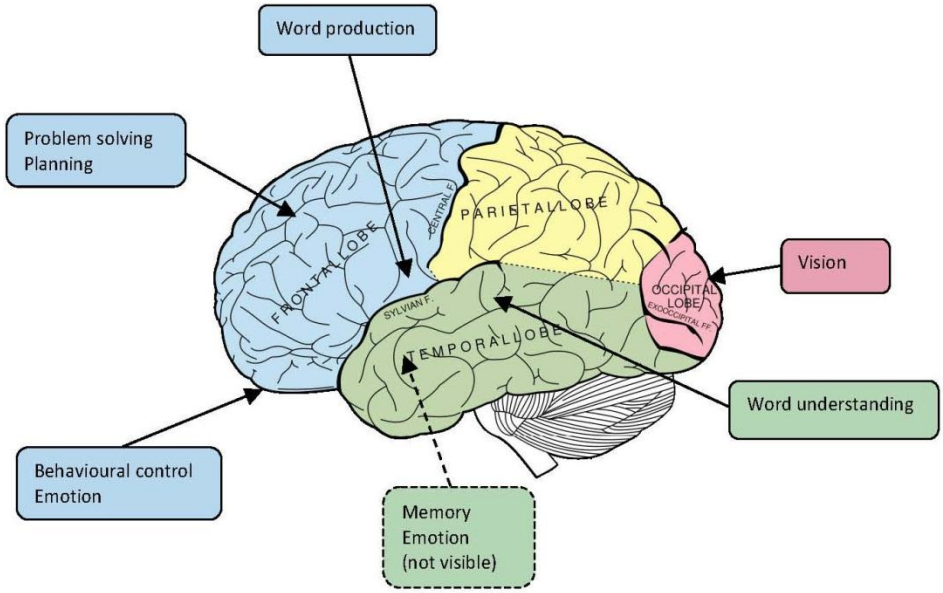

Figure 2. Approximate Locations of Primary Sensory Areas in the Human Brain [21]

Figure 2 illustrates the locations of primary sensory areas in the brain. Despite our knowledge about each region's main duty, it is incorrect to assume that only one part of the brain is responsible for each function since different parts may cooperate during even a simple work [18].

2.2 Functional Magnetic Resonance Imaging

Although functional imaging did not start with FMRI, its advantages made FMRI one of the most widely used techniques for observing brain activities [22, 23]. The most commonly used method before FMRI was PET. It is based on tracing the radioactivity flowing with blood to the brain [1, 3]. As the active parts of the brain need more oxygen, the blood flow which contains the radioactive material injected into vessels will be increased in those parts. The sensors on the PET machine record the changes in the brain. The amount of activation is measured and active parts of the brain are identified. Although PET is still in use, its disadvantages such as the side effects, safety issues of radioactive injection, price of materials and low speed of image acquisition process make it an unfavorable option for brain studies [3, 4].

FMRI only uses oxygen level changes in the brain which is a natural process. It is a noninvasive method and it has fewer side effects on subjects. It can be repeated as much as needed which is impossible in the PET due to safety reasons [4, 24]. Moreover, it benefits from changes in blood oxygenation level. These changes are really fast, which starts in seconds after the stimuli. Consequently, FMRI can measure the changes in a fraction of seconds [5, 25, 26].

The FMRI scanner produces time series of MRI data. The data obtained in FMRI sessions are normally 2D images of different slices of the brain which form 3D images in combination. The slices may be taken in three perpendicular planes named as axial, coronal and sagittal [1, 22]. These planes are shown in Figure 3.

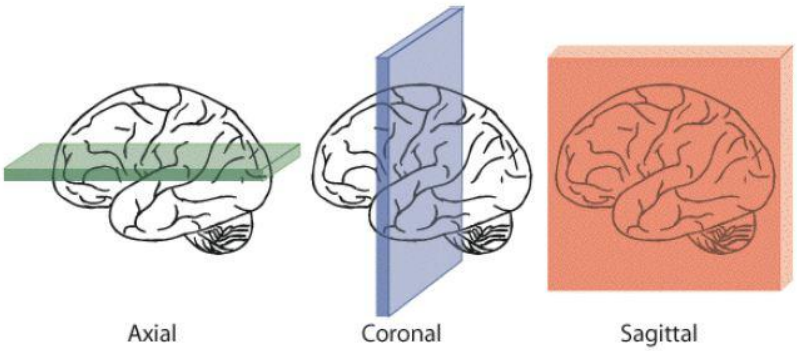

Figure 3. Different Planes of Sections in the Brain Used In the MRI Imaging Method [22]

The smallest addressable part in an FMRI image is a three-dimensional rectangular cuboid that is referred as a *voxel* [26]. MRI scanners take a snapshot in every 2-3 seconds. Imaging over a period of time during a particular activity represents neural changes in the brain due to the stimuli [1, 4, 5]. Figure 4 illustrates the images recorded during an auditory stimuli.

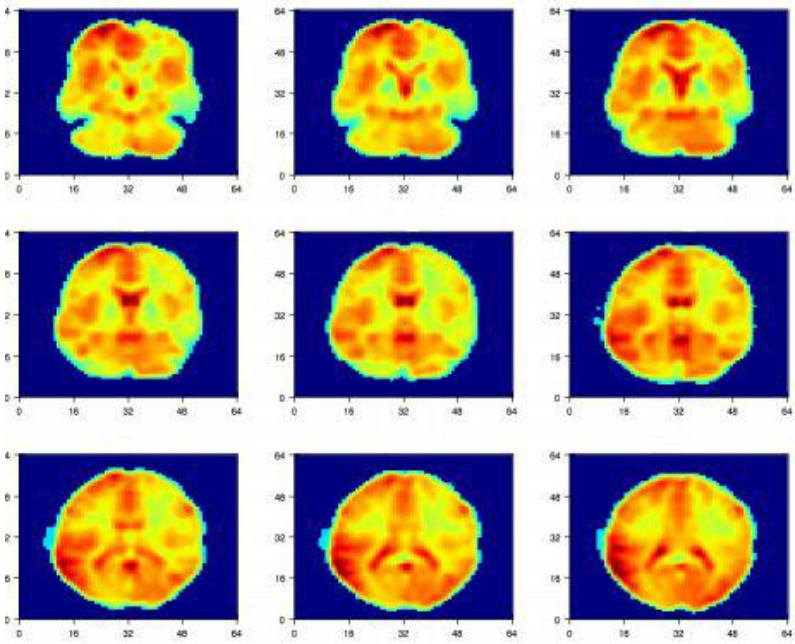

Figure 4. FMRI Data Collected While a Person Listened to an Auditory Stimuli. Cyan and Yellow Areas Show the Lowest Intensity of Activities where Orange and Red Represent the Highest Intensity [27]

2.3 Blood Oxygenation Level Dependent (BOLD) Signals

The hemoglobin molecule in the blood has magnetic properties which allows detecting whether or not it is bound to oxygen [4] and the amount of blood flowing through a particular part of the brain depends on the local neural activity. FMRI scanners employ the changes in the oxygen levels in the blood to generate blood oxygen level dependent (BOLD) response based images [4, 25, 26].

The increase in blood flow to its peak value takes approximately 4-5 seconds and it does not return to the baseline before at least 15-20 seconds [5, 6, 25]. The delay of hemodynamic response time in comparison to stimulus time should be considered in

experimental designs. Figure 5 shows the BOLD signal responses from four different individuals during a particular activity.

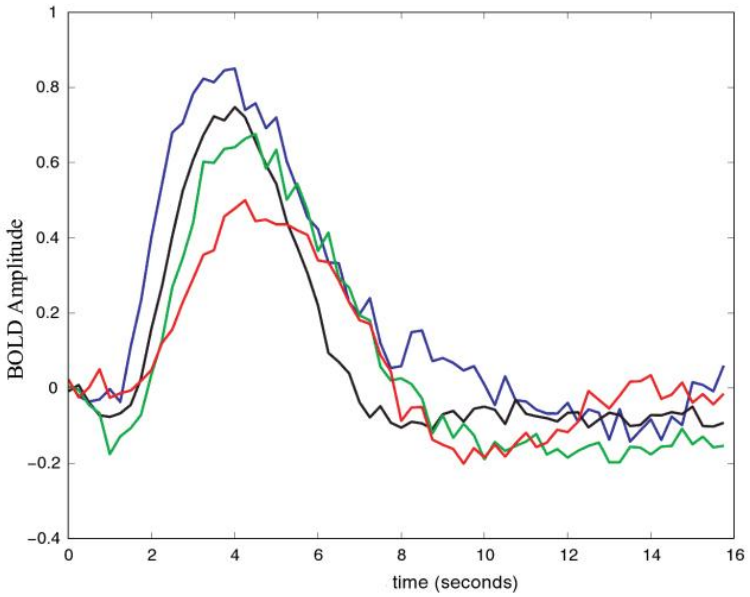

Figure 5. BOLD Signals for a Particular Activity. Each Line is the Data for a Different Individual [6]

A BOLD signal is generally described in terms of its peak height, time to peak, width, initial dip and post stimulus undershoot value as illustrated in Figure 6. Peak height (*H*) is the maximum value of the neuron's hemodynamic response during the stimulus which is related to the amount of activity in the neurons. Time to peak *(TP*) is the time that the BOLD signal takes within the stimulus onset to reach its peak value. Finding the suitable *TP* time is important for labeling each signal's category and defining the fixation period signal value. Width (*W*) is the time calculated for a BOLD signal between reaching to it's the peak value and the falling to the fixation period value. Initial dip (*ID*) appears before the BOLD signal rises in the form of descending signal which may accrue due to early oxygen consumption. As it is not a strong signal, in some cases it can be ignored. Post

stimulus undershoot (*PSU*) is the negative reflection of BOLD signal which appears after this signal reached to its peak and come back to rest mode. PSU has smaller absolute amplitude compared to the amplitude of BOLD at TP [6, 28].

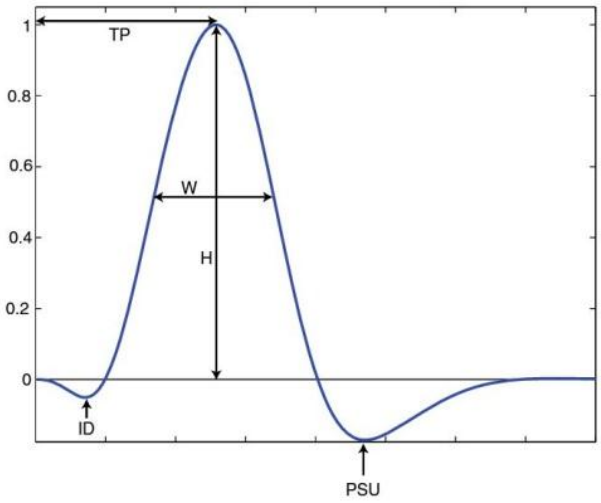

Figure 6. BOLD Signal Characteristics Which can be Described With the Term: Time From Stimulus Starts Until the Peak Time (TP), Height of Signal Response (H), Weight of Signal (W), Post Stimulus Undershoot (PSU) and Initial Dip (ID) [6]

The main components of FMRI data classification are shown in the block diagram below. The data acquisition and preprocessing tasks are presented in the following sections while feature selection and classification are explained in the next chapter.

Figure 7. Block Diagram of FMRI Classification System

2.4 Collection and Processing FMRI Data

2.4.1 Data Acquisition

Data acquisition is the process of using magnetic resonance scanners to take a series of images over time. These machines use strong magnetic fields to align the magnetization on hydrogen or oxygen nuclei which are common elements in the human body. The electromagnetic energy absorbed in the nuclei and emitted after a while are different due to the differences in the amount of hydrogen or oxygen in the organs [1, 24, 26]. Functional MR data in brain activity studies consists of a series of images which shows the location and the amount of activation in the brain [19]. The process of data acquisition may differ with the use of scanner type, power and imaging methods.

2.4.2 Data Preprocessing

FMRI data are highly liable to a number of artifacts from different sources which may lead to unreliable information and add unwanted variation [29]. In particular, movement, subject distraction, scanner system noise or thermal noise may distort the FMRI data [29]. Another source of data variability across subjects is the differences in head shapes and sizes which produce data with different amount of voxels and geometric position of voxels [6, 22]. This are important problems during subject independent experiments. Data variations within a subject may also occur. These variations may be due to several reasons such as the head movements, slice timing and voxels activation levels [1, 4, 6]. Because of these distortions, the FMRI data should be pre-processed before classification. Typically applied operations are in the order given as follows [6]:

1. Quality assurance
2. Distortion correction
3. Slice timing correction
4. Motion correction
5. Spatial smoothing
6. Spatial normalization

2.4.2.1 Quality Assurance

Before starting the preprocessing of the data, it is vital to be sure about the quality of acquired data. This process is performed by automatically detecting systematic patterns of noise. However, most of the time, it is better to have close eyes on the raw data as some problems are easily visible in the images. Extreme scanner artifacts or incompatible subject's data in the FMRI acquisition process should also be checked before applying other steps to prevent further analysis on wrong data [6, 14].

One of the artifacts caused by MRI scanner is spikes which are small changes in the image brightness due to electrical instability as illustrated in Figure 8. Spikes normally appear as stripes across the images. This kind of artifact does not occur in the new generation MRI scanners as they use new methods and technologies to avoid them. However, when they occur, they can adversely affect the analysis results [4, 29].

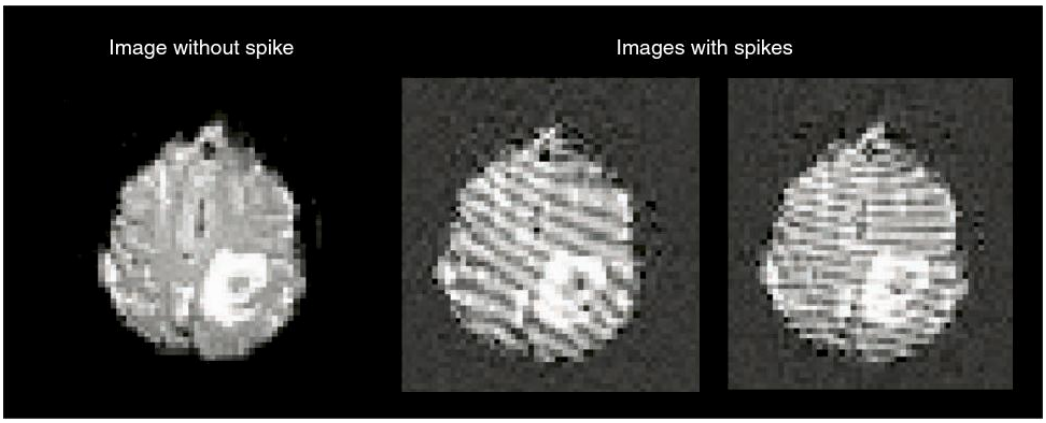

Figure 8. Example of an FMRI Image of a Patient with Tumour Disease with Spike Effect in Images [6]

Another artifact is ghosting which occurs due to periodic movements of subjects such as heartbeat or breathing as shown in Figure 9. Ghosting appears as a dim ghost of brain to each side of the brain area in the images. The recognition of ghosting artifact is difficult and the brightness level should be changed. This artifact may cause to consider the non-brain voxels as informative voxels and also to miss localization of activities. This kind of artifact rarely occurs in new generation scanners. Appearance of this artifact is more possible if we have a stimulus which needs periodic movements which should be solved during data acquisition [6, 30].

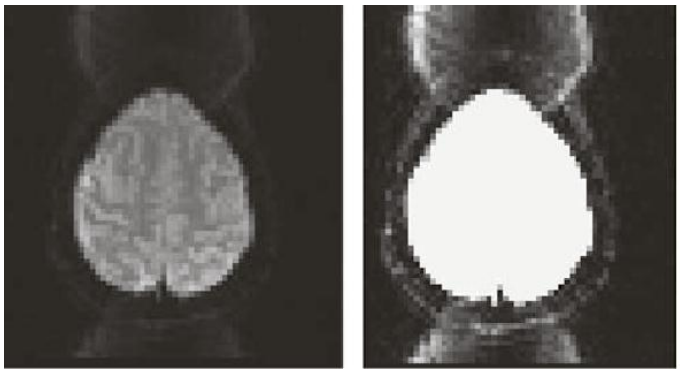
Figure 9. Example of Ghosting Effect in FMRI Images. The Right Panels Shows the Same Image when the Intensity is Reduced [6]

2.4.2.2 Distortion Correction

On the MRI images, the region of the brain where air and tissue meet causes a special kind of artifact known as distortion [4, 31]. This is because of inhomogeneity of tissue and air. The distortion may be in two main forms, namely dropout and geometric distortion.

The dropout effect is seen as a reduced signal in the air-tissue interface areas such as the orbitofrontal cortex [4]. The effect of dropout is presented in Figure 10. This kind of effect can change the neuronal signals in these areas and should be removed. The dropout effect occurs during MRI acquisition and mostly it is impossible to retrieve the original data after the image is taken [6, 31]. They should be avoided during MRI image acquisition. Otherwise, we have to exclude these areas during processing if they are not responsible for a particular task in the study under concern.

Spatial or geometric distortion appears in the form of spatially distorted voxels in the air-tissue interface regions. This may cause unwanted changes in the voxels location for different snapshots. Geometric distortion results in error during spatial information extraction, difficulties to align the images with structural images or destroying voxels information [6, 31].

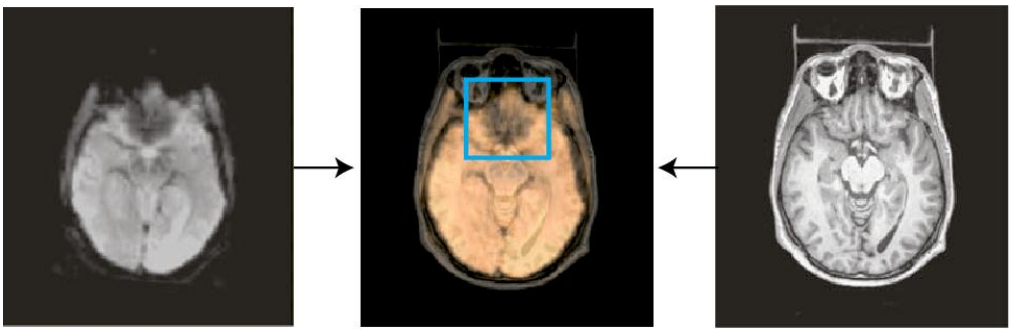

Figure 10. Blue Box Shows the Place where Air and Tissue Meet which Has Dropout Distortion Effect. Images in Two Sides are Combined to Show this Effect [6]

2.4.2.3 Slice Timing Correction

FMRI image acquisition methods use different techniques for slicing. Some methods take an image slice by slice in sequence and some others use the interleaved acquisition method [4, 32]. In this way, every other slice is acquired sequentially and followed by the next half of the remained slices. It usually takes odd slices in one sequence and the even slices in the next sequence as shown in Figure 11. This may lead to different timing sources for different acquisition sequences [32]. This timing difference gives rise to difficulties in analysis, as most of the FMRI data analysis assume the data with the same timing source [14, 32].

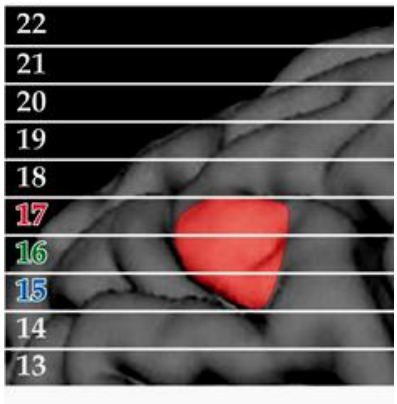

Figure 11. Axial Slice of Brain Imaging in FMRI Method [2]

In order to solve this problem, interpolation is generally applied. Considering one slice as a reference image, all other slice's data are interpolated in the same timing order as illustrated in Figure 12. This method works best when the time between slices is higher than the changes in the data [6, 32].

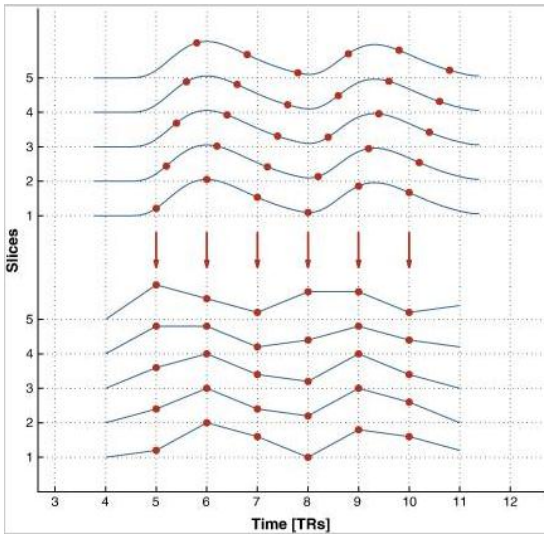

Figure 12. Slice Timing Correction .The Top Diagram Shows Time Difference (Red Points) for Different Slices and Bottom Diagram is obtained after Applying Slice Timing Correction [32]

2.4.2.4 Motion Correction

Since the early days of FMRI, head movement has been one of the serious problems. Different facilities such as restraints (face mask, vacuum cushion and padding) are used to keep the subject's head steady and fixed during FMRI sessions as shown in Figure 13. Subjects are given the structures and trained to keep their head with less possible movement. However, even the best subjects have head movements during the image acquisition. Because of this, motion detection and correction are absolutely vital in pre-processing steps [6, 33].

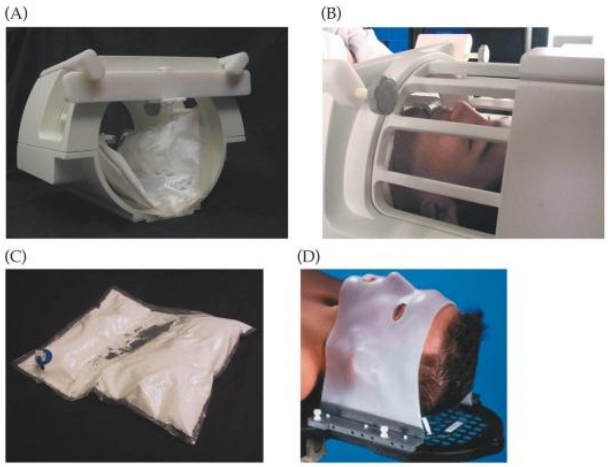

Figure 13. Various Methods and Materials Used for Head Movement Control During FMRI Scans. (A) face mask, (B) vacuum, (C) cushion and (D) padding [4]

Head movement effects can be classified into two classes. Firstly, head movement can cause mismatch of the location of voxels in time series images which is referred as bulk-motion. This effect usually occurs near the edges of the brain, where the informative voxels of the brain and empty voxels which are not part of the brain meet [34]. The sudden movement of the head leads to sudden changes in the value of voxels which are placed on the edge [33, 34]. Figure 14 illustrates the bulk-motion effects.

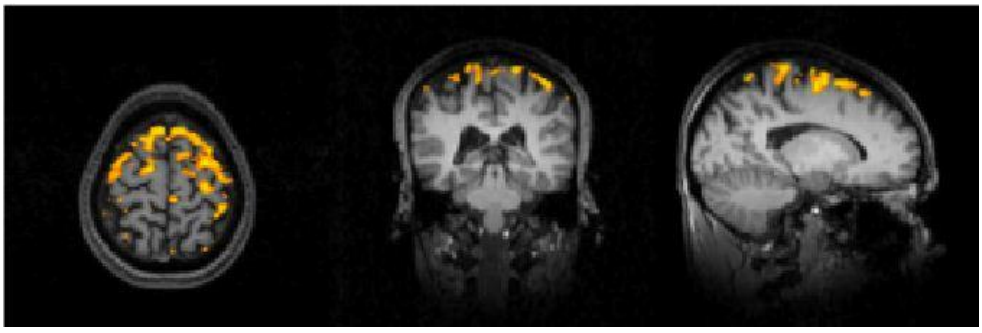

Figure 14. Bulk-Motion Effect on the Edges which Appears as a Ring of Positive or Negative Changes. This Effect is shown in Different Planes. Axial (Left), Coronal (Middle) and Sagittal (Right) [6]

Bulk-motion correction assumes that the movements are just transformation of the body, where the position of the head changed along a particular axis, but the shape of the head is the same as reference images [33, 34]. It is based on realigning the FMRI images to a reference image. The most common types of image transformation are translation, rotation, scaling and shearing [6, 33]. Figure 15 shows the effect of each transformation technique on a given image. The effect of rotation at different angles on FMRI image is illustrated in Figure 16.

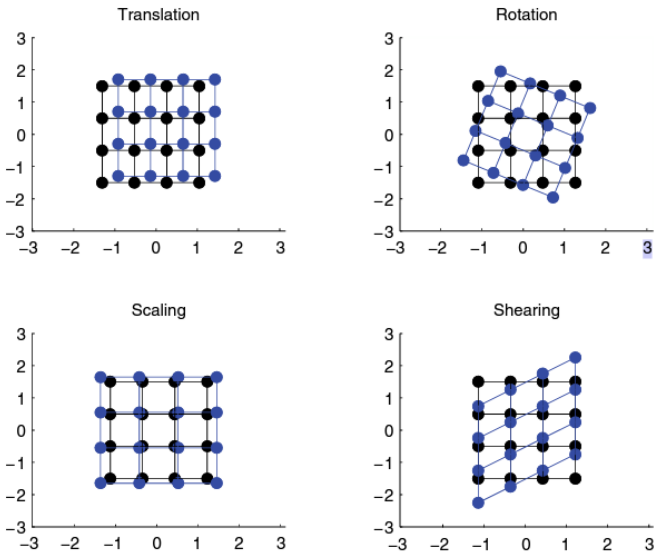

Figure 15. Image Transformation Methods such as Translation, Rotation, Scaling and Shearing [6]

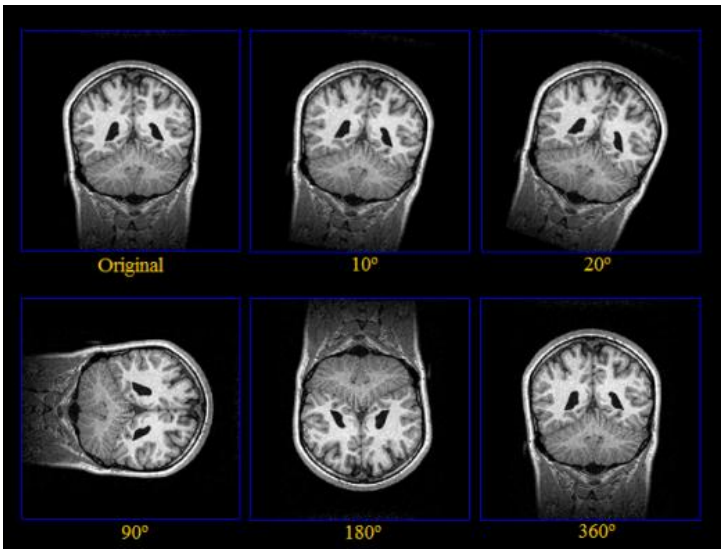

Figure 16. FMRI Image Rotation with Different Degrees for Bulk-motion Movement Correction

The second type of head movement effect on the FMRI images is spin-history effects which appears as disruption effect on the FMRI images [6, 35]. This kind of artifact is due to the proton movements from different slices near the target voxels, which causes receiving an unexpected signal that carries inaccurate information. In the case of interleaved acquisition model it can appear in a form of dark and bright stripes [6, 4, 35]. The effect of this artifact is shown in Figure 17.

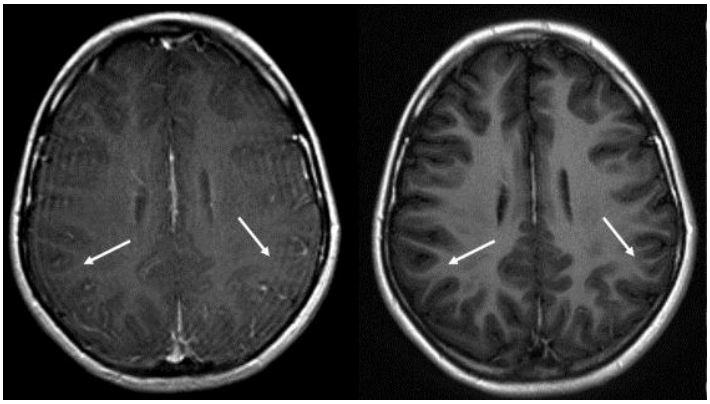
Figure 17. Spin History Effect in the Case of Head Movement [35]

This class of artifact cannot be corrected with normal motion correction models and needs more complicated algorithms such as a spin-history correction [34, 35]. In practice, the data is deleted if the movement is more than 1-2 voxels. Alternatively, a threshold in millimeters may be defined to identify the movements which need to be corrected [6]. As these movements are normally undeniable, various methods are used to correct the movement effects on the images.

2.4.2.5 Spatial Smoothing

Spatial smoothing corresponds to replacing the intensity of each voxel with the average of the neighboring voxels and the voxel under concern. In frequency domain, this corresponds to applying low-pass filter to remove high-frequency information [1]. The most common method for applying spatial smoothing is the convolution of the image with a Gaussian filter [1, 36]. The effect of spatial smoothing can be seen in Figure 18.

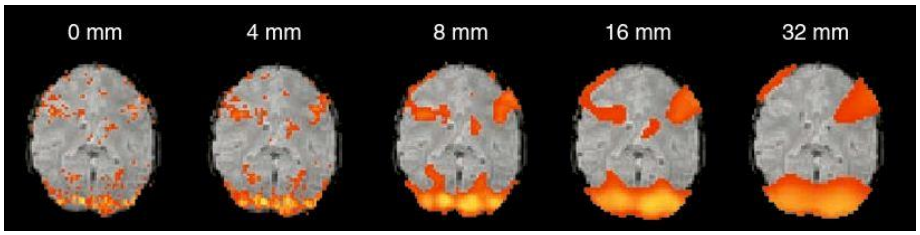

Figure 18. The Effect of Spatial Smoothing. Active regions can be Easily Detected when Smoothing is increased. The Amount of Smoothing is given in Millimeters [6]

2.4.2.6 Spatial Normalization

FMRI image classification may be subject dependent or independent. In subject dependent classification, the activation of brain for a given subject is explored to study various disabilities, planning for surgery or decoding activation areas [4, 6, 36]. On the other hand, subject independent experiments are to generalize and extend analysis across

different subjects. This requires employing an alignment scheme to compensate data variability due to head shape and size in different subjects [7]. Transforming the FMRI data into a common space is known as spatial normalization [36]. The first standard space designed this purpose is by Talariach as illustrated in Figure 19 [37]. By employing different landmarks, Talariach defined a boundary box that specifies the location of the brain in each dimension.

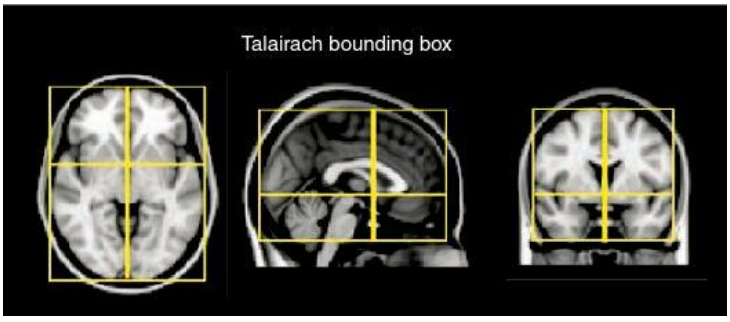

Figure 19. Talairach Bounding Boxes with Landmarks Connection [37]

Talariach also developed a brain atlas which provides a good guidance to the location of brain structures in a coordinate space. It can be used to find each activation place and interpret the results [42].

Chapter 3

CLASSIFICATION OF PRE-PROCESSED FMRI DATA

Feature selection is a crucial step in FMRI classification problem. Since they involve high dimensional data that is on the order of thousands [38]. Improving the classifier performance and decreasing the computational cost are two main aims of feature selection [26, 38, 43]. We used the feature selection methods listed below individually and combined them with each other to evaluate their relative performance

- Regions of Interest (ROI).
- N-most active voxels.
- Regions of Interest (ROI) averaging.
- N-most active voxels within ROI.

3.1 Regions of Interest

ROI is a well-known feature selection method in FMRI image classification. It is well-known that different stimuli can fire neurons of different regions. For instance, some parts are mainly involved in visualization while others are related to hearing [38, 39, 45, 46]. Because of this, various ROI are defined based on clinical experiments by experts. The ROI which are used in this thesis are named as CALC, LIPL, LT, LTRIA, LOPER, LIPS, and LDLPFC [1]. Figure 20 shows the eight different axial slices of whole brain without

considering ROI. Figure 21 shows the effect of applying these regions on a same axial slices of whole brain image.

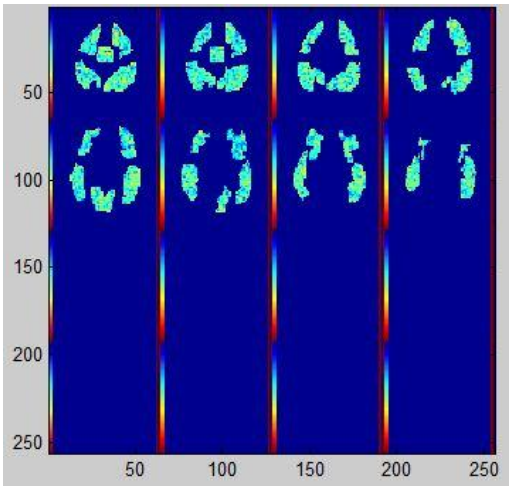

Figure 20. Axial Slices of whole brain without Considering ROI

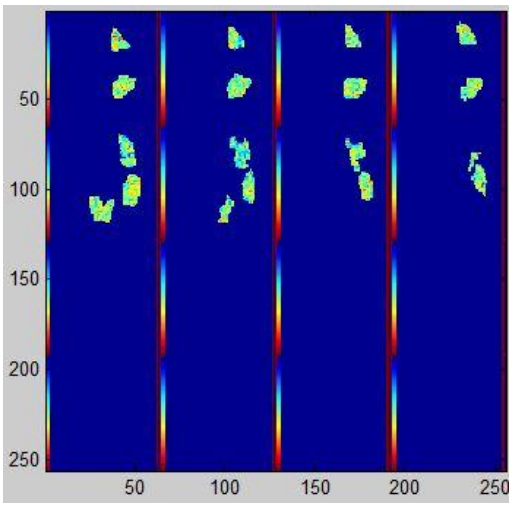

Figure 21 . Axial Slices with Applying ROI

Since only axial slices of brain are presented in figures above, the effects of applying some ROIs cannot be seen clearly.

3.2 ROI Average

In this method, the mean activation value of the voxels within each ROI is calculated. The average values are used as the elements of the feature vectors [39, 40]. It is obvious that the number of features is the same as the number of ROI that are used. This approach is more useful when data is collected from different sources such as various subjects.

3.3 N-most Active Voxels

In this approach, the voxels are ranked in terms of the activation level. N-most active voxels are then selected to define feature vectors. Voxel selection is based on measuring the ability of each voxel to distinguish the class under concern from the fixation period. t-test is generally used for this purpose [10]. The voxels achieving the largest t statistic are selected.

3.4 N-most Active Voxels in ROI

For uniform selection of voxels from different ROI, N-most active voxels are selected from each ROI and they are concatenated to form the feature vectors.

3.5 Classification

Having constructed the feature vectors, classifiers to separate the given FMRI data into two or more classes can be generated. In FMRI studies, both linear and nonlinear classifiers are used. SVM due to its robustness to high dimensional features and NN due to its success in various domain are explored in this study.

3.5.1 Support Vector Machine

In FMRI classification, SVM is considered as one of the most robust methods for classification [43]. Its insensitivity to the high dimensionality of data has made SVM appropriate for the FMRI data analysis [15, 43].

Basically, SVM is designed for two-class classification, and it can be easily extended to C-class data sets. In this case, C(C-1)/2 individual SVM classifiers are trained for each pair of classes, to build a multiclass SVM classifier [41].

Different optimizations and kernels are developed for SVM [41], however, since most FMRI related researches gained better answers with the linear kernel [39, 40, 43], in this thesis linear kernel is used for training SVM as illustrated in Figure 22. In the case that the samples of different classes are distributed in a way that a linear border cannot classify them into different groups, a nonlinear kernel is used. As shown in Figure 23, hyperplanes are not in a linear form.

Assume that there two different classes denoted by -1 and +1 respectively. Let x denote the feature vectors and w denote the coefficient vector of the linear boundary. If the data is separable with linear hyperplane that is defined as $w.x - b = 0$, it means that all samples that belong to the class -1 are located in the region defined by $w.x - b < 0$. Similarly, the data which belong to class +1 are located in the region which is defined by $w.x - b > 0$. Hence, infinite non-optimal solutions can be obtained for the mentioned hyperplane to select the optimal one, SVM employs two hyperplanes such that

there is no training data between them and the distance between these parallel hyperplanes (i.e., $\frac{2}{||w||}$) is maximized [15, 41]. These two hyperplanes can be written as follows:

$$w.x - b = 1$$

and

$$w.x - b = -1$$

Obviously, in order for $\frac{2}{||w||}$ to get maximized the value of w should be minimum.

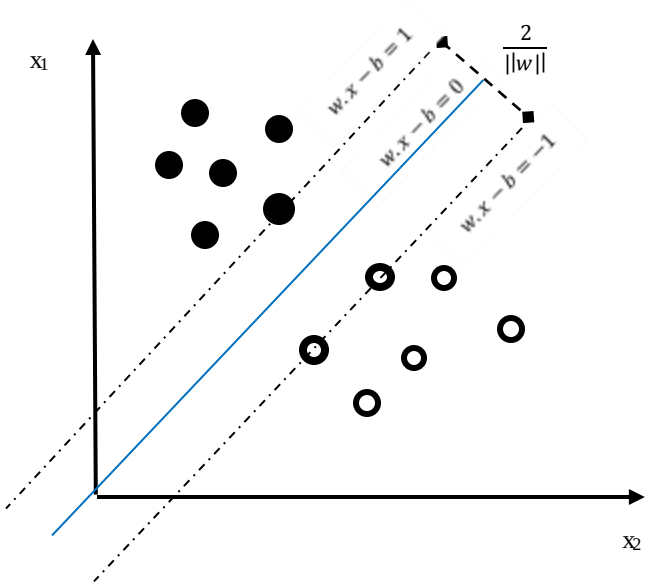

Figure 22. Hyperplanes in SVM with Linear Kernel

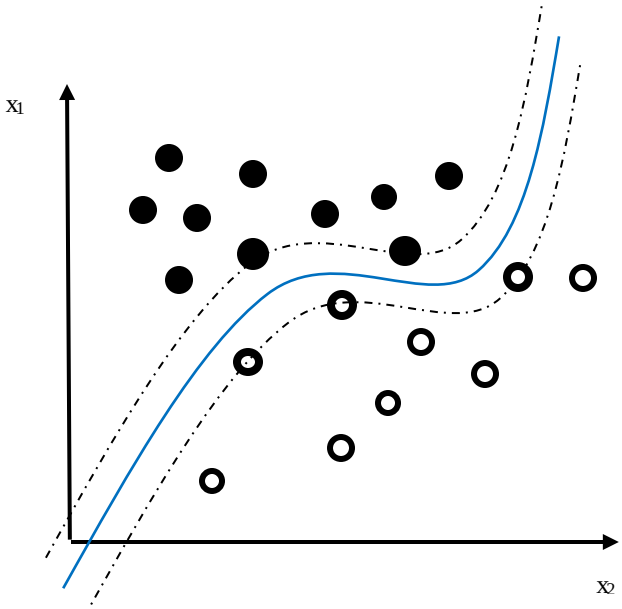

Figure 23. Hyperplanes in SVM with Nonlinear Kernel

3.5.2 Nearest Neighborhood Classifier

Nearest Neighborhood classifier (NN) is a simple but highly efficient method in the field of pattern recognition. It computes the test data point's distance from all training samples and selects the nearest neighbor having the lowest distance. Then it identifies the category of test data as the label of the nearest neighbor [9, 16, 42].

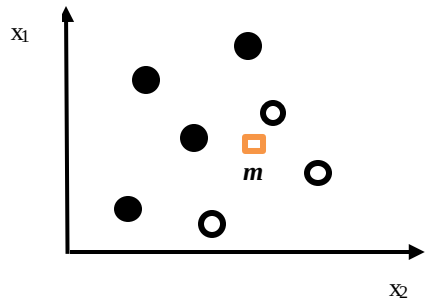

Figure 24. Illustration of the Operation of NN Rule

In Figure 24, assume that *m* denotes the test data. NN rule computes its distance to the training samples and assigns the label of the closest to test data which is ' O'. In the case of a noisy training set, NN rule can lead to large number of misclassifications.

3.6 FMRI Software and Toolboxes

For the processing and classification of FMRI raw data, various software and toolboxes are developed. We listed some of the well-known tool boxes in Table 1. The platform on which they run and information about their accessibility are also given. In this thesis, we used STARPLUS toolbox.

Table 1. FMRI Data Analysis and Visualization Software and Toolboxes Available on the Internet.

Package	Developer	Platform	Licenses
SPM	University collage of London	Matlab	Open source
FSL	Oxford	Unix	Open source
AFNI	NIMH	Unix	Open source
Brain voyager	Brain innovation	IOS, Windows, Linux	Commercial
STARPLUS toolbox	Carnegie Mellon university	Matlab	Open source
Princeton-MVPA toolbox	Princeton Institute	Matlab	Open source

Chapter 4

EXPERIMENTAL RESULTS

4.1 The STARPLUS Data Set Information

In this thesis, we used STARPLUS data set [44, 46] to evaluate different feature selection and classification schemes. In collecting this dataset, a sequence of sentences and simple pictures are shown to the subjects in each trial and they were asked whether the sentence describes the picture correctly or not.

In half of the trials, a picture is shown first and then a sentence. This set of trials is named as *PS* data set. In the rest of the trials, a sentence is shown first, and then a picture which is referred as *SP* data set. The data set contains 54 trials in total where four of them are eliminated as the start and end period. The blank screen is used between two stimuli to clarify the start and the end point of trials. The total number of rest period is 10 trials. The remaining 40 trials are related to the PS/SP data which are used for classification (20 trials from each one).

At first, either picture or sentence is presented for 4 seconds, followed by a blank screen for the next 4 seconds. Then, the second stimulus is presented in the following 4 seconds and, depending on whether the sentence described a picture correctly or not, the subject

must press "yes" or "no". Finally, 15 seconds is used as a rest or fixation period. This procedure is shown in Figure 25 (A).

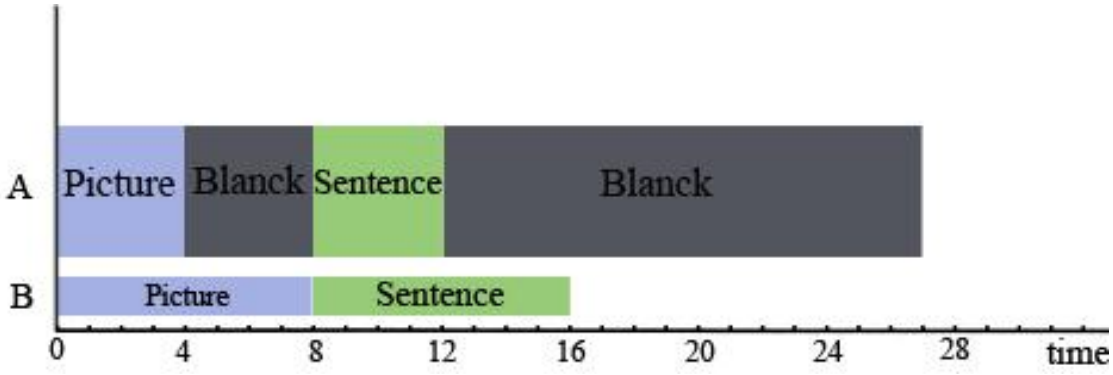

Figure 25. (A) Original Time Sequence of Stimuli for PS Trials, (B) Extended Time Sequence of Stimuli for PS Trials

The FMRI images are collected in every 500ms. Although the stimulation period of each condition is about 4 seconds, 8 seconds intervals of FMRI images are considered as the classifier input as illustrated in Figure 24 (B). The time extension is used to cover the FMRI BOLD signal extension which is about 9-12s beyond the normal brain activation. Because of using 8 seconds for each activity, each trial consists of 16 Images. The mapping defined by classifier can be represented as follows:

$$f : FMRI - \text{sequence}\left[I_{p1}, \ldots, I_{p16}, I_{s1}, \ldots, I_{s16}\right] \longrightarrow \{\text{Picture}, \text{Sentence}\}$$

Where I_{pj} is the voxels value vector in the j^{th} image for the class picture (P), and I_{sj} shows the j^{th} image for the class of sentence (S).

The sequence of images belonging to each class in the trials SP and PS can be represented as follows:

PS: $I_{p1}^{PS}, I_{p2}^{PS}, \ldots, I_{p16}^{PS}, I_{s1}^{PS}, I_{s2}^{PS}, \ldots, I_{s16}^{PS}$

SP: $I_{s1}^{SP}, I_{s2}^{SP}, \ldots, I_{s16}^{SP}, I_{p1}^{SP}, I_{p2}^{SP}, \ldots, I_{p16}^{SP}$

For generating the data for class S, combination of both SP and PS trials is required. This combination is done by using the first part of SP and the second part of PS. The same approach is used for class P where the first part of PS and the second part of SP trials are used. The overall data matrix is shown in Figure 26.

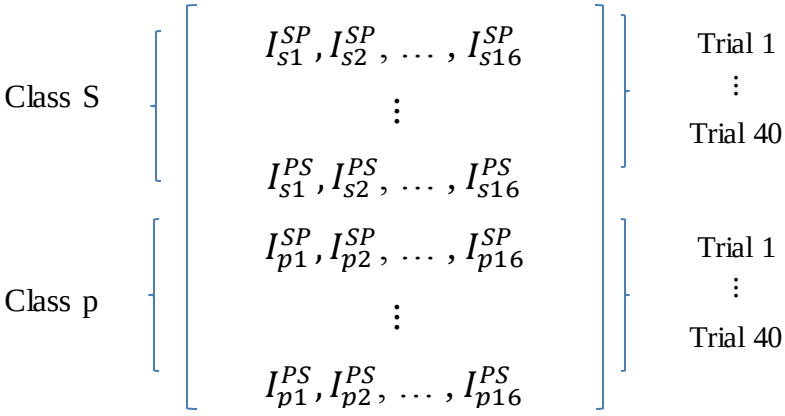

Figure 26. The Overall Training Data for a Given Subject

The total number samples for each class is equal to 40. Hence, the data matrix has 80 rows and 16 columns where each column corresponds to a different snapshot.

Each subject may have different number of voxels as presented in Table 2. The first row represents the subject ID and the second row shows the number of voxels of that subject.

Table 2. The Number of Voxels in each Subject

Subject	**04799**	**05710**	**04820**	**04847**	**05675**	**05689**
No. of Voxels	4949	4634	5015	4698	5135	5062

In this thesis, two different series of experiments are conducted. These are namely subject dependent and subject independent. Two types of subject dependent experiments are considered. In the first type, the images collected from both PS and SP trials are pooled (PS+SP). The second type of experiments corresponds to using PS and SP trials separately (PS/SP)

4.2 Subject Dependent Experiments on PS+SP

In the subject dependent case, 10-fold cross-validation is applied where the data is initially partitioned into 10 folds. Then, leaving each fold out one by one for testing and using the rest for training, the experiments are repeated ten times and the average accuracies of ten folds are recorded. This corresponds to using 72 samples for training and 8 for testing,

The experiments are conducted using all features and using the feature subsets computed by applying four feature selection schemes. Since each feature selection leads to a different number of features, the exact number of features considered in each experiment is also presented. The total number of samples employed in the experiments is 80 (40 samples per class).

4.2.1 No Feature Selection

In this experiment, feature selection is not applied and all voxels are used in constructing the feature vectors. This system is considered as our reference for evaluating the performances of the feature selection schemes employed.

Table 3. Information about the Number of Samples and Features when Feature Selection is Not Applied

Subject	Number of Samples		No. of Voxels	No. of Snapshots	No. of Features (Voxels×Snapshots)
	Class 'S'	Class 'P'			
04799	40	40	4949	16	79184
05710	40	40	4634	16	74144
04820	40	40	5015	16	80240
04847	40	40	4698	16	75168
05675	40	40	5135	16	82160
05680	40	40	5062	16	80992

Table 3 summarizes the characteristics of the classification problem corresponding to each subject. Due to the differences in brain shapes and sizes, each subject has different number of voxels. The average accuracies achieved are presented in Table 4. It can be seen in the table that SVM is superior to NN. This is reasonable since SVM is known to be more robust to high dimensional feature vectors including redundant or correlated information.

Table 4. Classification Performance in Percentage when All Voxels are Employed

Subject	SVM	NN
04799	60	54
05710	74	65
04820	69	60
04847	87	77
05675	69	66
05680	74	51
Average:	72	63

4.2.2 ROI Based Features

In this experiment, 7 ROIs are employed to reduce the number of features. Information about the total number of voxels of each subject in the ROIs under concern are listed in Table 5.

Table 5. Information about the Number of Samples and Features when ROI based Feature Selection is Applied

Subject	Number of Samples		No. of Voxels	No. of snapshots	No. of features (Voxels×Snapshots)
	Class 'S'	Class 'P'			
04799	40	40	1874	16	29984
05710	40	40	1883	16	30128
04820	40	40	1888	16	30208
04847	40	40	1715	16	27440
05675	40	40	2239	16	35824
05680	40	40	2230	16	35680

Table 6 presents the average performances achieved when ROI based features are considered. It can be seen that ROI based features provide better average accuracies compared to the reference system where all voxels are used.

Table 6. Classification Performance in Percentage after Applying ROI based Feature Selection

Subject	SVM	NN
04799	67	75
05710	81	61
04820	71	61
04847	95	86
05675	76	74
05680	81	61
Average:	78	70

4.2.3 N-most Active Voxels in each ROI

In this experiment, the N-most active voxels within each ROI is used to reduce the dimensionality of the feature vectors. The number of features is reduced to No. of Voxels kept × No. of snapshots, leading to 600×16 = 9600 features.

Table 7 presents the average accuracies achieved when N-most active voxels within each ROI based features are considered. It can be seen that, considering N-most active voxels based features, further improvements are achieved.

Table 7. Classification Performance in Percentage after Applying N-most-Active Voxels in each ROI based Feature Selection

Subject	SVM	NN
04799	90	62
05710	94	87
04820	74	54
04847	97	92
05675	90	76
05680	87	81
Average:	89	76

4.2.4 ROI Average

In this experiment, since the average in each ROI is used to define novel features, the total number of features can be computed as 7×16=112.

The average performances achieved when ROI average based features are considered is presented in Table 8. It can be seen that the performance is severely deteriorated. Hence, it can be concluded that averaging based feature selection discards valuable information.

Table 8. Classification Performance in Percentage after Applying ROI Average Based Feature Selection

Subject	SVM	NN
04799	52	64
05710	82	74
04820	56	59
04847	80	89
05675	50	57
05680	67	60
Average:	63	66

4.2.5 N-most Active Voxels

In this experiment, the N-most active voxels selection is applied to the whole brain data to select those voxels which are more active. The feature size is 600×16=9600.

The average performances achieved when N-most active voxels based features are considered is presented in Table 9. It can be seen that the performance is further deteriorated. Hence, it can be concluded that being active does not mean to convey discriminative information.

Table 9. Classification Performance in Percentage after Applying N-most Active Voxels Based Feature Selection

Subject	SVM	NN
04799	50	51
05710	60	49
04820	61	53
04847	60	55
05675	50	46
05680	59	51
Average:	56	50

4.3 Subject Dependent Experiments on PS/SP

This groups of experiments are also subject dependent. For each subject, the PS and SP data are separated and the experiments are repeated for each set separately. In each set, there are 20 samples from class S and 20 from class P. The number of features is the same as the original dataset. The organizations of the data used are illustrated in Figures 27 and 28, respectively for PS and SP datasets.

In this set of experiments, 5-fold cross-validation is applied. The average accuracies achieved for each subject are recorded.

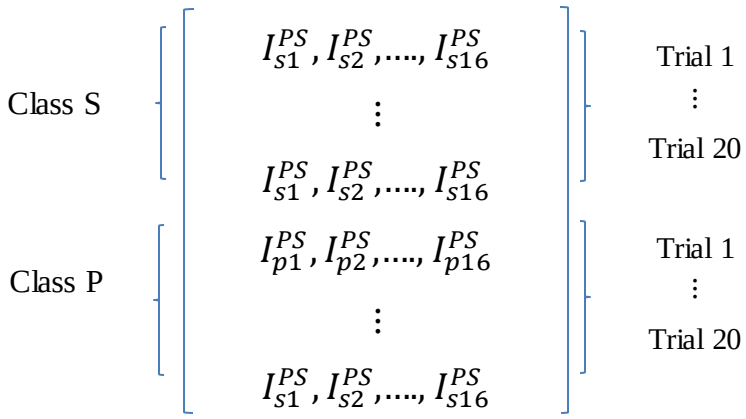

Figure 27. PS Dataset Used in the Experiments

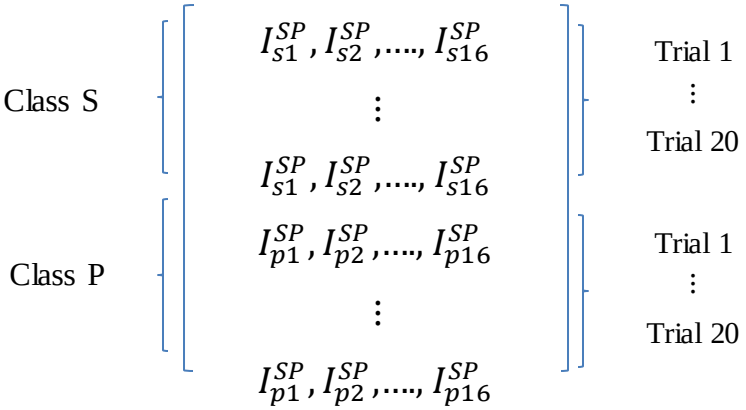

Figure 28. SP Dataset Used in the Experiments

4.3.1 No Feature Selection

In this experiment, feature selection is not applied. This system is considered as our reference for evaluating the performances of the feature selection schemes employed.

Table 10. Information about the Number of Samples and Features when Feature Selection is Not Applied

Subject	Number of Samples		No. of Voxels	No. of snapshots	No. of Features (Voxels×Snapshots)
	Class 'S'	Class 'P'			
04799	20	20	4949	16	79184
05710	20	20	4634	16	74144
04820	20	20	5015	16	80240
04847	20	20	4698	16	75168
05675	20	20	5135	16	82160
05680	20	20	5062	16	80992

Table 10 summarizes the characteristics of the classification problem corresponding to each subject. The average accuracies achieved are presented in Table 11. It can be seen in the table that SVM is superior to NN as before. The accuracies achieved are higher compared to PS+SP. This means that the neural activity patterns of each class depends on the order they are presented to the subjects.

Table 11. Classification Performance in Percentage when All Voxels are Employed

Subject	SP experiments		PS experiments	
	SVM	NN	SVM	NN
04799	87	67	70	72
05710	100	87	95	87
04820	100	82	92	90
04847	97	97	90	80
05675	95	85	95	85
05680	100	80	80	60
Average:	96	84	88	80

4.3.2 ROI Based Features

In this experiment, 7 ROIs are employed to reduce the number of features. Information about the total number of voxels of each subject in the ROIs under concern and the number of selected features are the same as given in Table 5.

Table 12 presents the average performances achieved when ROI based features are considered. It can be seen that ROI based features does not provide consistent improvements in PS/SP experiments.

Table 12. Classification Performance in Percentage after Applying ROI based Feature Selection

Subject	SP experiments		PS experiments	
	SVM	NN	SVM	NN
04799	90	65	65	70
05710	100	95	97	90
04820	97	87	87	80
04847	100	100	92	77
05675	95	80	92	85
05680	100	80	85	55
Average:	97	84	87	77

4.3.3 N-most Active Voxels in ROI

In this experiment, the N-most active voxels within each ROI is used to reduce the dimensionality of the feature vectors. As in the PS+SP experiments, 9600 features are employed.

Table 13 presents the average accuracies achieved when N-most active voxels within each ROI based features are considered. It can be seen that, considering N-most active voxels based features, consistent improvements compared to ROI based features are achieved.

Table 13. Classification Performance in Percentage after Applying N-most-Active Voxels in each ROI based Feature Selection

Subject	SP experiments		PS experiments	
	SVM	NN	SVM	NN
04799	92	75	75	80
05710	100	100	97	92
04820	97	95	92	80
04847	100	100	97	95
05675	95	95	97	87
05680	97	97	80	77
Average:	97	94	91	86

4.3.4 ROI Averaging

In this experiment, since the average in each ROI is used to define novel features, the total number of features can be computed as 7×16=112.

The average performances achieved when ROI average based features are considered is presented in Table 14. It can be seen that the performance is severely deteriorated. Hence, it can be concluded that averaging based feature selection discards valuable information as in PS+SP experiments**.**

Table 14. Classification Performance in Percentage after Applying ROI Average Based Feature Selection

Subject	SP experiments		PS experiments	
	SVM	NN	SVM	NN
04799	87	70	57	57
05710	100	100	90	70
04820	90	95	77	70
04847	97	100	87	77
05675	85	82	77	75
05680	95	95	62	55
Average:	91	89	76	69

4.3.5 N-most Active Voxels

In this experiment, the N-most active voxels selection is applied to the whole brain data to select those voxels which are more active. The feature size is 600×16=9600.

The average performances achieved when N-most active voxels based features are considered is presented in Table 15. It can be seen that the performance is further deteriorated.

Table 15. Classification Performance in Percentage after Applying ROI Average based Feature Selection

Subject	SP experiments		PS experiments	
	SVM	NN	SVM	NN
04799	82	62	60	37
05710	87	72	82	77
04820	90	75	82	65
04847	95	92	75	625
05675	85	75	87	75
05680	97	70	75	55
Average:	89	75	79	64

4.3.6 Summary of the Subject Dependent Results

Table 16 presents the summary of the results achieved in subject dependent experiments. It can be easily seen in the table above that the best accuracies are achieved when N-most Active Voxels in each ROI are used. Another observation is that the accuracies on SP data is higher than that of PS. In the SP data set, the sentence came first and hence subjects have prior expectation about the upcoming image which is the picture described by the sentence. This may cause related areas to be more activated and result in more informative voxels in these regions. It can also be seen that, when SP or PS are separately used, better accuracies are achieved compared to PS+SP.

Table 16. Summary of the Average Accuracies in Percentage Achieved for SP and PS Datasets and PS+SP

Experiment	SP+PS		SP		PS	
	SVM	NN	SVM	NN	SVM	NN
No feature selection	72	63	96	84	88	80
ROI Based Features	78	70	97	84	88	77
N-most Active Voxels in each ROI	**89**	**76**	**97**	**94**	**91**	**86**
ROI Average	63	66	91	89	76	69
N-most Active Voxels	56	50	89	75	79	64

4.4 Subject Independent Experiments

In these experiments, we used the data corresponding to all subjects to study subject independent classification. Anatomical differences in human subject's brain is the biggest obstacle in analyzing of multiple subject's data. This causes a problem to register the voxels in one brain to the precise corresponding location of the others. One solution to this problem is to use geometrical transformation of the different subject's data into the standard anatomical space. However, transformation may lead to noisy and spatially modified voxel values. Alternatively, using ROI averages can solve this problem. Despite of the difference in the brain size or shape, all subjects will have the same numbers of features which is exactly equal to the number of ROIs that is used.

Since there are 6 subjects in our dataset and each subject has 80 samples, there are 480 samples in this classification problem. Subject based cross-validation is employed where in each run, one subject is kept out for testing and the remaining 5 subjects are used for training. Therefore, for each experiment, 80 samples are used for testing and 400 samples are used for the training. Since ROI averages are used, the feature size is 7×16 = 112. Information about the dataset is summarized in Table 17.

Table 17. Characteristics of the Dataset Employed in Subject Independent Experiments

Subject	**No. of Each class sample**		**No. of Voxels**	**No. of snapshots**	**No. of Feature (Voxels×Snapshots)**
	Class 'S'	**Class 'P'**			
All	240	240	7	16	112

Table 18 presents the average accuracies achieved. It can be seen that the performances achieved using subject independence is inferior to subject dependent ones which is as expected. Further studies should be conducted to compute better subject independent features.

Table 18. Classification Performance in Percentage on the Subject Independent Dataset

Subject Independent Experiment	
SVM	**NN**
75	76

4.5 Summary

Designing classifiers with acceptable performance to categorize the data in human brain data is a really important goal. The experiments conducted have shown that the goal is a realistic one. In cases where the subject is fixed, better scores can be achieved. In general, SVM provided better scores compared to NN. Further research should be conducted to develop a better representation for subject independent studies.

Chapter 5

CONCLUSION AND FUTURE WORK

Predicting whether the subject is viewing a picture or a sentence has been the goal behind this study. For this purpose, two classification schemes, namely SVM and NN are used. In majority of our simulations, SVM provided better performance which is mainly due to the robustness of SVM to reasonably high dimensional data. The results achieved clearly indicate the possibility of detecting the cognitive state of the brain.

In order to study the effect of employing smaller number of features, four dimensionality reduction methods, ROI, N-most active voxels, ROI average and N-most active voxels within ROI are applied. Selecting N-most active voxels within the ROI provided the best scores, verifying the importance of applying feature in this domain.

Subject independent classification is also studied. The best-fitting feature selection scheme from the subject independent experiments which helps to avoid data variation across subjects is considered. Experimental results have shown that FMRI images from an unseen person can be correctly classified with accuracy equal to 75%.

FMRI, due to its advantages, is the focus of brain and cognitive science research. The number of articles published in this area is growing day by day. Using alternative approaches of pattern classification are major topics that should be explored in this area.

In particular, employing other feature selection methods, designing more robust classifiers and improving the pre-processing steps should be considered as future works of this thesis.

REFERENCES

[1] Stuart C. (2009). "Functional MRI: Methods and Applications". *PhD Thesis University of Nottingham, United kingdom*.

[2] Clarke E., Dewhurst K. (1972). "An Illustrated History of Brain Function". *Sandford Publications, Oxford*. ISBN-10: 093040565X.

[3] Pogossian T.M.M., Phelps M.E., Hoffman E.J., Mullani N.A. (1975). "A Positron-Emission Trans Axial Tomography for Nuclear Imaging (PET)". *Radiology*. Vol. 114, Issue 1, pp. 89–98.

[4] Scott A.H., Allen W.S., Gregory M. (2009). "Functional Magnetic Resonance Imaging (2nd edition)". *Sinauer Associates Inc Publisher*. ISBN-10: 0878932860.

[5] Ogawa S., Lee T.M., Kay A.R., Tank D.W. (1990). "Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation". *PNAS*. Vol. 87, Issue 24, pp. 9868-9872.

[6] Russell A.P., Jeanette A.M., Thomas E.N. (2011). "Handbook of Functional MRI Data Analysis". *Cambridge University Press*. ISBN-10: 0521517664.

[7] Norman K.A., Polyn S.M., Detre G.J., Haxby J.V. (2006). "Beyond Mind-Reading: Multi-Voxel Pattern Analysis of FMRI Data". *Trends Cognitive Science*. Vol. 10, Issue 9, pp. 424-30.

[8] Helmut L. (2012). "A Personalized History of EEG–fMRI Integration". *NeuroImage*. Vol. 61, Issue 2, pp. 1056-1067.

[9] Mitchell T.M. (1997). "Machine Learning". *McGraw Hill*. ISBN-10:0070428077.

[10] Jain A.K., Duin R.P.W., Jianchang M. (2000). "Statistical Pattern Recognition: A Review". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 22, Issue 1, pp. 4-37.

[11] Richard O.D., Peter E.H., David G.S. (2001). "Pattern Classification (2nd Edition)". *John Wiley & Sons*. ISBN: 978-0-471-05669-0.

[12] Molina L.C., Belanche L., Nebot A., (2002). "Feature Selection Algorithms: A Survey and Experimental Evaluation". *IEEE International Conference on Data Mining*. pp. 306-313.

[13] Michie D., Spiegelhalter D.J., Taylor C.C. (1994). "Machine learning, Neural and Statistical Classification". *Ellis Harwood Series in Artificial Intelligence*. ISBN-10: 8188689734.

[14] Strother S.C. (2006). "Evaluating FMRI Preprocessing Pipelines". *Engineering in Medicine and Biology Magazine, IEEE*. Vol. 25, Issue 2, pp. 27-42.

[15] Chen P.H., Lin C.J., Schölkopf B. (2005). "A tutorial on V-support Vector Machines". *Appl. Stoch. Models. Bus. Ind*. Vol. 21, Issue 2, pp. 111-136.

[16] Cover T.M., Hart P.E. (1967). "Nearest Neighbor Pattern Classification". *IEEE Transactions on Information Theory*. Vol. 13, Issue 1, pp. 21-27.

[17] Kwong K.K., Belliveau J.W., Chesler D.A, Goldberg I.E, Weisskoff R.M., Poncelet B.P., Kennedy D.N., Hoppel B.E., Cohen M.S., Turner R., (1992). "Dynamic Magnetic Resonance Imaging of Human Brain Activity during Primary Sensory Stimulation". *Proc Natl Acad Sci, U S A*. Vol. 89, Issue 12, pp. 5829-5831.

[18] Meister I.G., Sparing R., Foltys H., Gebert D., Huber W., Töpper R., Boroojerdi B. (2006). "Functional Connectivity between Cortical Hand Motor and Language Areas during Recovery from Aphasia". *Neurological Sciences*. Vol. 247, Issue 2, pp. 165-168.

[19] Mark S.C., Susan Y.B. (1994). "Localization of Brain Function Using Magnetic Resonance Imaging". *Trends Neuroscience*. Vol. 17, Issue 7, pp. 268-77.

[20] http://serendip.brynmawr.edu/exchange/brains/definitions/def-parlobe

[21] http://www.neura.edu.au/frontier/what-frontotemporal-dementia-ftd

[22] Dale P., George J.A., David F., Lawrence C.K., Anthony S.L., James O.M., Williams S.M. (2001). "Neuroscience, (2nd edition)". *Sunderland (MA): Sinauer Massachusetts Associates*. ISBN-10: 0-87893-742-0.

[23] Smith S.M., Dphil M.A. (2004). "Overview of FMRI Analysis". *The British Journal of Radiology*. Vol. 77, Issue 2, pp. 167-175.

[24] Scott A.H., Allen W.S., Gregory M. (2009). "Functional Magnetic Resonance Imaging (2nd edition)". *Sinauer, Massachusetts Associate*. ISBN-10: 0878932860.

[25] Gray D.F. (1982). "Basic Concepts for Nuclear Magnetic Resonance Imaging". *Magnetic Resonance Imaging*. Vol.1, Issue 1, pp. 39-53.

[26] Keith B., Josh C. (2013). "Decoding Neural Events from FMRI BOLD Signal: A Comparison of Existing Approaches and Development of a New Algorithm". *Magnetic Resonance Imaging*. Vol. 31, Issue 6, pp. 967-989.

[27] Imanol G.R. (2011). "FMRI Sonication & Brain Activity Prediction". *Master Thesis, Music Technology Group, Pompeu Fabra Universitat, Spain*.

[28] Zhao X., Wu Y., Guo S. (2007). "Characteristics of FMRI BOLD Signal and its Neurophysiological Mechanism". *Progress in Natural Science*. Vol. 17, Issue 6, pp. 630-638.

[29] Jörn D., Reza S. (2006). "Detecting and Adjusting for Artifacts in fMRI Time Series Data". *NeuroImage*. Vol. 27, Issue 3, pp. 624-634.

[30] Bradley G.G., Hongmei Z., Robert A.B., Mitchell J.R. (2004). "Removal of Phase Artifacts from FMRI Data Using a Stock-well Transform Filter Improves Brain Activity Detection". *Magnetic Resonance in Medicine*. Vol. 51, Issue 1, pp. 16-21.

[31] Jezzard P. (2012). "Correction of Geometric Distortion in fMRI Data." *NeuroImage*. Vol. 62, Issue 2, pp. 648-651.

[32] Sladky R., Friston K.J., Tröstl J., Cunnington R., Moser E., Windischberger C. (2011). "Slice-timing Effects and their Correction in Functional MRI". *NeuroImage*. Vol. 58, Issue 2, pp. 588-594

[33] Freire L., Mangin J.F. (2001). "Motion Correction Algorithms May Create Spurious Brain Activations in the Absence of Subject Motion". *NeuroImage*. Vol. 14, Issue 3, pp. 709-722.

[34] Julian R.M. (2008). "Motion Detection and Correction in Magnetic Resonance Imaging". *PhD Thesis, University of Canterbury, United Kingdom*.

[35] Alibek S., Adamietz B., Cavallaro A., Stemmer A., Anders K., Kramer M., Bautz W., Staatz G. (2008). "Contrast-enhanced T1-Weighted Fluid-Attenuated Inversion-Recovery BLADE Magnetic Resonance Imaging of the Brain: An Alternative to Spin-Echo Technique for Detection of Brain Lesions in the Unseated Pediatric Patient". *Academic Radiology*. Vol. 15, Issue 8, pp. 986-995.

[36] Friston K.J., Ashburner J., Frith C.D., Poline J.B., Heather J.D., Frackowiak R.S.J. (1995). "Spatial Registration and Normalization of Images". *Human Brain Mapping*. Vol. 3, Issue 3, pp. 165-189.

[37] http://www.talairach.org

[38] Li Y., Namburi P., Yu Z., Guan C., Feng J., Gu Z. (2009). "Voxel Selection in FMRI Data Analysis Based on Sparse Representation". IEEE *Transactions on Biomedical Engineering*. Vol. 56, Issue 10, pp. 2439-2451.

[39] Mitchell T.M., Hutchinson R., Niculescu R.S., Pereira F., Wang X., Just M., Newman S. (2004). "Learning to Decode Cognitive States from Brain Images". *Machine Learning*. Vol. 57, Issue 1, pp. 145-175.

[40] Wang X., Hutchinson R., Mitchell T.M. (2003). "Training FMRI Classifiers to Detect Cognitive States across Multiple Human Subjects". *Neural Information Processing Systems.* pp. 465-469.

[41] Christopher J.C.B. (1998). "A Tutorial on Support Vector Machines for Pattern Recognition". *Data Mining and Knowledge Discovery*. Vol. 2, Issue 2, pp. 121-167.

[42] Jennie S., Andy B., Warren P., Donald W. (2004). "Reinforcement Learning and Its Relationship to Supervised Learning" *in Handbook of Learning and Approximate Dynamic Programming*. John Wiley & Sons, Inc., Hoboken, NJ, USA. ISBN-10: 047166054X.

[43] Ludmila I.K., Juan J.R. (2010). "Classifier Ensembles for FMRI Data Analysis: An Experiment". *Magnetic Resonance Imaging*. Vol. 28, Issue 4, pp. 583-593.

[44] Mitchell T., Hutchinson R., Just M., Niculescu R.S., Pereira F., Wang X. (2003). "Classifying Instantaneous Cognitive States from FMRI Data". *American Medical Informatics Association Annual Symposium*. pp. 465-469.

[45] Wang X. (2003). "Using Machine Learning to Detect Cognitive States across Multiple Subjects". *CALD KDD Project Report*.

[46] Wang X., Mitchell T., (2002). "Detecting Cognitive States Using Machine Learning". *Iterim Working Paper*.