Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

DIFFERENTIATING SCHIZOPHRENIC PATIENTS FROM HEALTHY CONTROL; APPLICATION OF MACHINE LEARNING TO RESTING-STATE FMRI

by
Hossein Ebrahimi Nezhad

In recent years, one analysis approach that has grown in popularity is the use of machine learning algorithms to train classifiers to decode stimuli, mental states, behaviors and other variables of interest from fMRI data. Most of these studies focus on fMRI low frequency oscillations. This study focuses on the amplitude of low-frequency fluctuations (ALFF) and fractional amplitude of low-frequency fluctuations (fALFF). A Voxel-wise analysis is performed on the whole brain for two groups of subjects. A machine learning algorithm is applied to two independent groups of subjects (a total of 160 healthy control and schizophrenic subjects) to classify Schizophrenia subjects from healthy control. Kendall tau rank correlation coefficient is also used to dominate most important voxels (features). This study is done on three datasets: a) fALFF b) mALFF dataset and c) combination of mALFF and fALFF.

The results show that using the combination dataset improves the classification and demonstrates that machine learning algorithms can extract new information from a resting state image of schizophrenia which can help in diagnosing and treating schizophrenic patients in the future. Future studies can focus on testing these algorithms on different modalities and moreover on different physiological disorders.

DIFFERENTIATING SCHIZOPHRENIC PATIENTS FROM HEALTHY CONTROL; APPLICATION OF MACHINE LEARNING TO RESTING STATE FMRI

by
Hossein Ebrahimi Nezhad

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Biomedical Engineering

Department of Biomedical Engineering

January 2014

APPROVAL PAGE

DIFFERENTIATING SCHIZOPHRENIC PATIENTS FROM HEALTHY CONTROL; APPLICATION OF MACHINE LEARNING TO RESTING STATE FMRI

Hossein Ebrahimi Nezhad

Dr. Bharat Biswal, Dissertation Advisor — Date
Professor of Biomedical Engineering, NJIT

Dr. Xin Di, Committee Member — Date
Research Assistant Professor of Biomedical Engineering, NJIT

Dr. Zhi Wei, Committee Member — Date
Associate Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Hossein Ebrahimi Nezhad

Degree: Master of Science

Date: January 2014

Undergraduate and Graduate Education:

- Master of Science in Biomedical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2014

- Bachelor of Science in mechanical Engineering,
  Shahid Bahonar University, Kerman, I. R. Iran, 2010

Major: Biomedical engineering

This thesis is dedicated to:
The four pillars of my life: God, my lovely wife, and my parents.
Without you, my life would fall apart.

I might not know where God will take me in the life's road,
but I strongly believe you, God, are the only one who
knows what's best for me and you are the one guiding me
every single step I take.

Hoda, you are everything for me, without your love and
confidence I would not be able to make it.

Maman, you have given me so much, thanks for your faith
in me, and for teaching me that I should never surrender.

Baba, you always told me to "reach for the stars." I think I
got my first one.

We made it…

ACKNOWLEDGMENT

I would like to express the deepest appreciation to my thesis advisor, Dr.Bharat Biswal, who guided me all the way through a new field in BME. Without his supervision and constant help, this project would not have been possible.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr.Xin Di and Dr. Zhi Wei, for their encouragement, insightful comments, and hard questions.

My gratitude also goes to the Brain Connectivity Lab. There are not enough words to describe your excellent work. Suril, and Ruyi: you made this study happen. You were there to help no matter what time or day of the week. . Thank you so much for accepting me into your home.

The most special thanks goes to my best partner and friend, my lovely wife. Hoda, you gave me your unconditional support and love through all this long process. “Dooset daram!”

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1. Preliminary Research

In the last decade there have been a growing number of research interests in decoding brain activities using different neuroimaging modalities such as EEG (electroencephalography), fMRI.

fMRI (functional Magnetic Resonance Imaging) is a type of neuroimaging procedure which uses the Magnetic Resonance Imaging technology to measure the blood oxygen level dependent (also known as BOLD) to map the neural activity of the brain's functions.

In the last few years the use of machine learning methods and pattern recognition has improved these field of studies; they are especially done providing a better understanding of psychiatric disorders such as schizophrenia, major depression, Alzheimer's disease, bipolar disorder and autistic spectrum disorder(Orrù, Pettersson-Yeo et al. 2012). Some of these studies introduce modern and novel machine learning algorithms to improve the result of the previous studies (Rasmussen 2006, Pereira, Mitchell et al. 2009, Shoeb, Kharbouch et al. 2011).

These studies can be divided to three main categories: (a) Diagnostic studies which compare neuroimaging data of two main groups, healthy control group (HCs) and patients, to investigate the diagnostic potential e.g., (Shen, Wang et al. 2010, Costafreda, Fu et al. 2011) (b) Studies on predicting a disease which compares different brain scans in a follow-up period of time e.g., (Zhang, Wang et al. 2011, Koutsouleris, Borgwardt et al. 2012) and (c) Prediction of treatment response studies e.g., (Gong, Wu et al. 2011).

The main aim of most of fMRI studies is to investigate in the brain functions on resting state. These studies have pointed were highly synchronous between the right and left primary motor cortices at resting state (Biswal, Zerrin Yetkin et al. 1995). Biswal et al. used the root mean square to measure the LFO amplitude. More recently Zang et al. has introduced methods to examine the standard deviation of the BOLD signal in the sub-0.1 Hz band, which is referred to as ALFF and fractional ALFF(Yang, Long et al. 2007, Yu-Feng, Yong et al. 2007, Zou, Zhu et al. 2008).

Previously, studies mainly focused on using basic statistical methods to study schizophrenia disorder, which couldn't help due to the complexity and dissconnectivity in the entire brain for fMRI images for this disease(Heckers, Goff et al. 2002, Lawrie, Buechel et al. 2002, Bluhm, Miller et al. 2007, Yin, Hu et al. 2008). On the other hand, in order to have a more accurate study on schizophrenia disorder from neuroimaging data, a group of studies has been done which use machine learning methods (Pereira, Mitchell et al. 2009). Many other studies use modern dimensional reduction methods to have a better classification, methods such as principle component analysis (PCA) and independent component analysis(Jafri, Pearlson et al. 2007, Pagani, Salmaso et al. 2009). The main idea behind these reduction methods is to convert high dimensional space data to subspace data with fewer dimensions. The reduced data is then used to classify the healthy control and schizophrenic subjects.

To classify two groups of subjects many different machine learning methods have been used, recently. it has been proved that SVM is the most optimum method to be used for classification of healthy control and other mental disorders using their neuroimaging data (LaConte, Strother et al. 2005). Of different machine learning methods the most

popular technique in the neuroimaging literature is SVM(Vapnik 2000). This method is used in this study and later on this method will be explained more in details.

1.2. Objective

The main aim of this study is to do a classification between schizophrenia and healthy control subjects using machine learning methods from their fMRI images, to find the most important voxels in the classification. Feature reduction methods will also be used to have a smaller data to do the classification using reduced dataset; furthermore, the accuracy change will be observed to see if feature selection is going to help the classification or not. This study will do the steps on three kinds of data, i) mALFF ii) fALFF and iii) combination of mALFF and fALFF dataset. Another independent group of subjects will also be used to validate our results.

It is hypothesized that feature reduction will help the classification and using SVM method will give a reliable accuracy. Moreover combined dataset will give a better accuracy for the classification.

1.3. Outlines

This study is done on two independent group of subject, the first group contains 74 healthy and patient subjects and the second contains 86. The same algorithm will be applied for both groups to validate the results. Machine learning and its different approaches toward a good classification is introduced in Chapter 2. In Chapter 3 the dataset and the methods used on the dataset will be discussed, and finally in Chapter 4 the results and discussion will be given.

CHAPTER 2

MACHINE LEARNING

Over the past two decades Machine Learning has become one of the most important concepts of IT and Data analysis. With the enormous increase in data these days it isn't unbelievable that this smart data analysis is becoming one of the mainstays of data mining.

In this chapter, machine learning and its different kinds will be introduced. After that, some basic statistical methods commonly used will be given. Finally, the main machine learning concepts used in this project will be pointed out.

2.1. Introduction to Machine Learning

In 1959, Arthur Samuel defined machine learning as a "Field of study that gives computers the ability to learn without being explicitly programmed."**(Simon 2013)** He was able to write a checkers playing program. What he did was, he had to program for it to play 1000's of games against itself. The checkers playing program learns over time what are good board positions and what are bad board positions by only watching tons of different games.

One of the most recent definitions of machine learning was defined by Tom Mitchell. He defines machine learning by saying that "A well posed learning problem is defined as follows. He says:

"A computer program is said to learn from experience E, with respect to some task T, and some performance measure P, if its performance on T as measured by P improves with experience E."(Mitchell 1997)

For the checkers playing example the experience "**E**", will be the experience of having the program play 1000's of games against itself. The task **"T"**, will be the task of playing checkers. And the performance measure **"P"**, will be the probability that it wins the next game of checkers against some new opponent.

There are several different types of machine learning algorithms. The main two types are supervised and unsupervised learning. There are other minor kinds such as reinforcement learning and recommender systems. Two main types will be discussed.

2.1.1 Supervised Learning

In a short definition, in supervised learning the computer is been taught how to do something; whereas, in unsupervised learning it learns by itself. In supervised learning or learning with teacher the data is a sample of input-output patterns. In which, the algorithm will go through the dataset to find the function between the input and output (Camastra and Vinciarelli 2008). In other words in this type of learning the computer will be trained by a labeled dataset, called the training set or training sample, the goal is to find a deterministic function that maps any input to an output that can predict future input output observations to distinguish the two classes of subjects. This study is an example of supervised learning.

Depending on the type of the outputs, supervised learning is divided into classification and regression learning:

2.1.1.1 Classification Learning. If the output space has a structure whether two elements of the output are equal or not, it is called classification learning. Each output

element is called a class and the algorithm that finds the classification is called the classifier. The goal in this kind of problems is to assign new inputs to one of a number of discrete classes or categories. This problem is used for pattern recognition tasks.

2.1.1.2 Regression. If the output space has a structure of values of continuous variables, then the learning algorithm is known as the regression problem. A good example for this kind is estimating the value of physical measures such as temperature in a section of a thermoelectric plant(Camastra and Vinciarelli 2008).

2.1.1 Unsupervised Learning

If the data is only a sample of objects without defining which group they belong to (the dataset is not labeled), and the labeling has to be done by the algorithm (known as clustering) the problem is known as unsupervised learning. In unsupervised learning there is not a teacher to teach. In this type of learning a training sample of dataset (e.g. images) is given to the algorithm with the goal to extract some structure from the dataset. If some structure exists in training data, it would find a short description of data. The most common way is to specify a similarity between any pairs of objects. If two objects share same structures, they can be clustered in the same group. This idea underlies clustering algorithms that form a great subclass of unsupervised algorithms.

For clustering a fixed number of clusters should be given to the algorithm, the aim is to find a grouping of the objects such that similar objects belong to the same cluster. The theory behind this clustering is that the similarities of the objects in one cluster are

much greater than the similarities among objects from different clusters(Camastra and Vinciarelli 2008). In this article mathematics behind the clustering won't be discussed.

In addition to clustering algorithms in unsupervised learning, there are algorithms that reduce the number of dimensions, called dimensionality reduction methods (DRM). These techniques are important because, first of all the space to store the data is Important and DRM technique is a great method to reduce the storage space. The second reason in using DRM is that, dimension reduction will improve the speed of algorithms and reduce the computation time.

2.2 Support Vector Machine

Support Vector Machine (SVM) is a specific type of supervised ML method that aims to classify data points by maximizing the margin between classes in a high-dimensional space(Vapnik 2000). The idea of support vector machine which is known as the most powerful and most common algorithm in supervised learning comparing with linier regression and neural network was first introduced by Vladimir Vapnik (Vapnik 2000) in the 1990s. In this method the optimal separation hyperplane is calculated using labeled training to classify the two classes of the dataset. For example consider that you have two different classes y_1 and y_2 and for each object you have different features; let's say $x_1, x_2, x_3, \ldots, x_n$ In order to tell apart objects belonging to y_1 and y_2, one might consider any number of descriptive, quantifiable features of these two objects. The main question which can be answered using SVM is: can all be assembled in a vector of values used to describe y_1 andy_2. SVM will use the given labeled data to compute a separation hyperplane to place y_1 in one side of this hyperlane and $y_2$on the other side. The purpose

of this section is to introduce the theory and equations behind Support Vector Machines and to show how it can be applied to larger datasets and machine learning problems.

2.2.1 Basics and Equations

A vector for each subject will be defined; where each dimension (feature) represents some characteristic of each of the subjects we wish to study. Let's define:

$$\vec{x} \equiv \{ x_1 , x_2 , x_3 , \dots . , x_N \} \tag{2.1}$$

$$y \equiv \pm 1$$

In this study, N is the number of voxels in the brain for each subject and for the parameter y, we define subjects as healthy control whenever y = +1 and y = -1 will be defining the patients. So each subject is defined by an N dimension vector $\vec{x}$ and the group they belong to is defined by y.

Imagine a dataset of different subjects is given whose classes are defined (+1 or -1). Suppose we are able to plot these subjects in an N dimensional space in the normal way. Now the goal is to find the hyperplane between these two classes. Figure 4.1 is a simple example of plotting a two dimensional dataset. As it can be seen, we have to find a unique plane of separation.

The equation for an N dimension hyperplane is defined as:

$$f(\vec{x}) = \vec{w} . \vec{x} - b \tag{2.2}$$

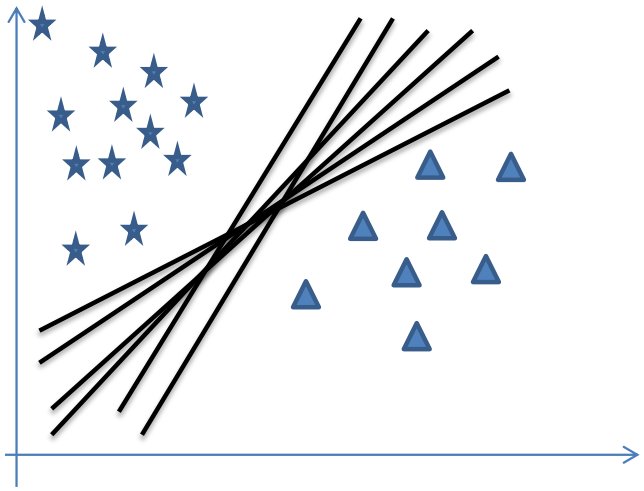


Figure 4.1 Supervised machine learning problem with two classes - as it is shown many planes can separate the two classes.

Where $\vec{w}$ is the normal vector for the plane and b is its offset from the origin. To find the learned function the $\vec{w}$ and b has to be defined afterwards by computing the $f(\vec{a})$ for an unknown vector and defining its sign we can predict which class it belongs to.

We also know the perpendicular distance r from a plane to a point ($\vec{z}$) is computed by:

$$r = \frac{\vec{w} . \vec{z} - b}{|\vec{w}|} \tag{2.3}$$

If we consider $\vec{w}$ as a unit vector, however, a more convenient method for our future calculations is simply to compute $\vec{w}$ such that the value of the learned function is ±1 at the margin. This fixes the scale, and the overall width of the margin is then:

$$d = \frac{2}{|\vec{w}|} \tag{2.4}$$

Our desire is to minimize the length of $\vec{w}$. The direction of $\vec{w}$ specifies the orientation of the separating hyperplane, b displaces it from the origin to its proper absolute position, and the magnitude of $\vec{w}$ sets the width of the margin.

To ensure none of the data points exceed the margin, the following linier constraints are required on each of the N training points.

$$y_i(\vec{w} \cdot \vec{z} - b) \geq 1 \tag{2.5}$$

Putting all the equations together, we can assemble a Lagrangian to minimize this problem:

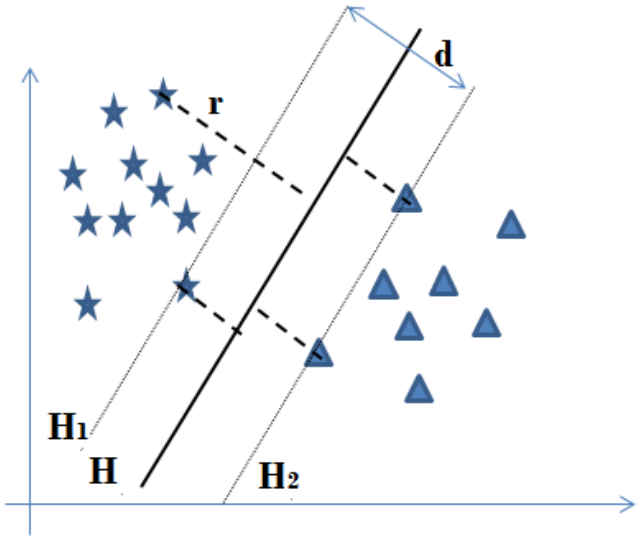


Figure 2.2 SVM geometry.

$$L = \frac{1}{2}\vec{w} \cdot \vec{w} - \sum_{i=1}^{N} \alpha_i y_i(\vec{w} \cdot \vec{x_i} - b) + \sum_{i=1}^{N} \alpha_i \tag{2.6}$$

In which the first term will maximize the margin, and the subsequent N terms ensure the separation of the two classes in our training set. Notice the use of Lagrange multipliers ($\alpha_i's \geq 0$) to enforce these linear constraints. This approach is common in solving quadratic optimization problems such as we have here. A dual Lagrangian can be constructed by taking the partial derivatives with respect to the primal variables $\vec{w}$ and b and noting they are zero at optimality. This leads to the following relations:

$$\frac{\partial L}{\partial \vec{w}} = 0 \Rightarrow \vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x_i} \tag{2.7}$$

$$\frac{\partial L}{\partial \vec{w}} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \tag{2.8}$$

Substituting these back into our original Lagrangian, we get the dual Lagrangian

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x_i} . \vec{x_j} \tag{2.9}$$

This is the complete formulation of our problem in terms of α's. After computing α's we are able to recover and compute $\vec{w}$ from eq. 2.7, and b from the following:

$$b = \vec{w} . \vec{x_k} - y_k \tag{2.10}$$

For all of the training vectors, as long as its corresponding $\alpha_i \neq 0$ .Using the new formulation can re-express the learned function in terms of the α's.

$$L_D = \sum_{i=1}^{N} \alpha_i y_i \vec{x_i} . \vec{z} - b \tag{2.11}$$

It has to be mentioned again, the learning function depends on the training points which have a $\alpha \neq 0$. The points whose $\alpha = 0$ hold the decision plane and are called the support vectors.

2.2.2 Soft Margin (constant C in SVM)

In finding the hyper plane sometimes some noises! appear, noises like pointes that are mislabeled in the training set, to avoid these noises we need to reformulate our equation to allow some pointes to cross the margins so the eq. 2.5 changes to:

$$y_i(\vec{w} . \vec{z} - b) \geq 1 - \varepsilon_i \tag{2.12}$$

$$\varepsilon_i \geq 0 \tag{2.13}$$

This $\varepsilon_i$will allow some of the training points enter the margin. So the Lagrangian eq. will change into:

$$L = \frac{1}{2} \vec{w} . \vec{w} + C \sum_{i=1}^{N} \varepsilon_i - \sum_{i=1}^{N} \alpha_i y_i (\vec{w} . \vec{x_i} - b) + \sum_{i=1}^{N} \alpha_i (1 - \varepsilon_i) \tag{2.14}$$

The term C is a new term which will tend to cap the error. As we are minimizing L, and both C and ε_i are non-negative the Lagrangian equation can be re arranged like this:

$$L = \frac{1}{2} \vec{w} . \vec{w} - \sum_{i=1}^{N} \alpha_i y_i (\vec{w} . \vec{x_ı} - b) + \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N} \varepsilon_i (C - \alpha_i) \tag{2.15}$$

In order to constrain L, we note what happens to the objective function when α's are allowed to grow larger then C. In that case, the sum in the last term can be driven to -∞ by letting the ε's become arbitrarily large. To avoid this situation, it is required that the α's are bound within the range $0 \leq \alpha \leq C$. This result follows naturally by taking:

$$\frac{\partial L}{\partial \varepsilon_i} = 0 \Rightarrow \alpha_i = C \tag{2.16}$$

These new equations are solving the same problem as before, except it has a Lagrange multiplier C which controls the rigidity of the margin. To summarize the equations:

$$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \vec{x_ı} . \vec{x_j} \tag{2.17}$$

$$0 \leq \alpha \leq C \tag{2.18}$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \tag{2.19}$$

As it will be shown in the next chapters the constant C plays an important role in the machine learning algorithm. By changing this constant parameter the accuracy (will get to its definition) will change.

For most of the research this parameter has to be randomly chosen, and the C with which a higher accuracy is occurred will be chosen. Usually the different values for this parameter are chosen from [0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, inf].

2.3 Cross Validation

To have a better understanding of the performance of the pattern extracted by machine learning it is advised to do a cross validation. Cross validation is an approach by which the dataset is divided to two main subsets; training set and test set or validation set. The algorithm uses the training set to train and extract the pattern. Then to evaluate the pattern is tested on the validation set.

The question which pops out is; how should the dataset be divided? Is just dividing the dataset to two parts, a training set and a testing set enough, or should it be divided in more parts? There are different ways to divide the training set; Hold out method, random subsampling, k-fold cross validation and leave one out.

2.3.1 Hold Out Method

Hold out method is the simplest way to divide the dataset by which the dataset is divided to two sets called the training set and testing set. The function behind this method is to

find the pattern using the training set and testing the pattern using the unseen testing set (the algorithm hasn't seen the outputs of the testing set during the training). As it is advised when dividing the data set using this method it is better that the data set to be divided to a 70% training set and a 30% testing set. One important notice; keep in mind when dividing two different groups of dataset, let's say group A and group B, it is better to have 70% of A objects and 70 % of the B objects in training and the rest 30% objects of each group has to be placed in the testing set.

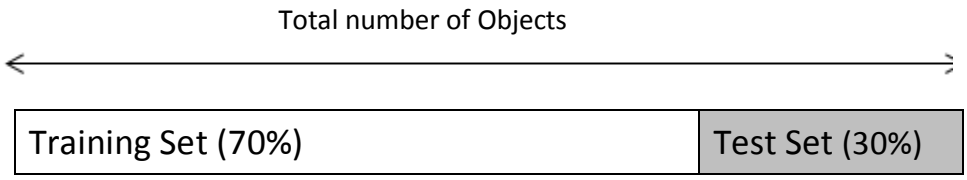


Figure 2.3 Hold Out Method.

2.3.2 Random Subsampling

Another way of splitting the data is Random subsampling by which the data is divided into randomly selected (fixed) number of examples without replacement. The data set will be divided several times and each time it will train using the training set and be evaluated using the testing set. For each data split, we retrain the classifier from scratch with the training examples and then estimate the class loss or the accuracy for the testing set. The true accuracy of the classifier is obtained as the average of each of the accuracies. This separation is much better than the hold out method.

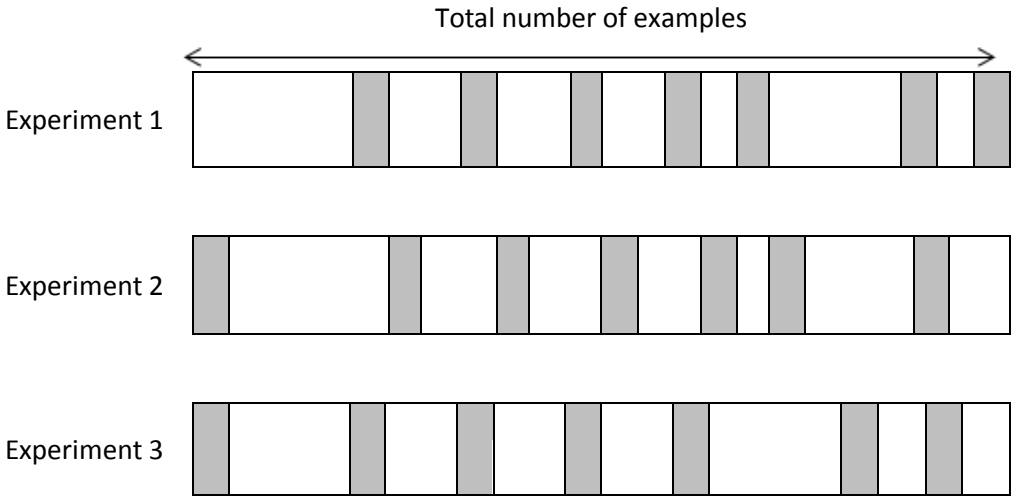


Figure 2.4 Random Subsampling.

2.3.3 K-Fold Cross Validation

In k-fold cross validation the dataset is partitioned into k equal subsets, the training and testing will be done k times and each time one of the k folds will be known as a testing set and all other subjects will be the training set. In each the K training and validation runs a simple hold out will be done on the data set. The overall accuracy or class loss will be the average of K time separation.

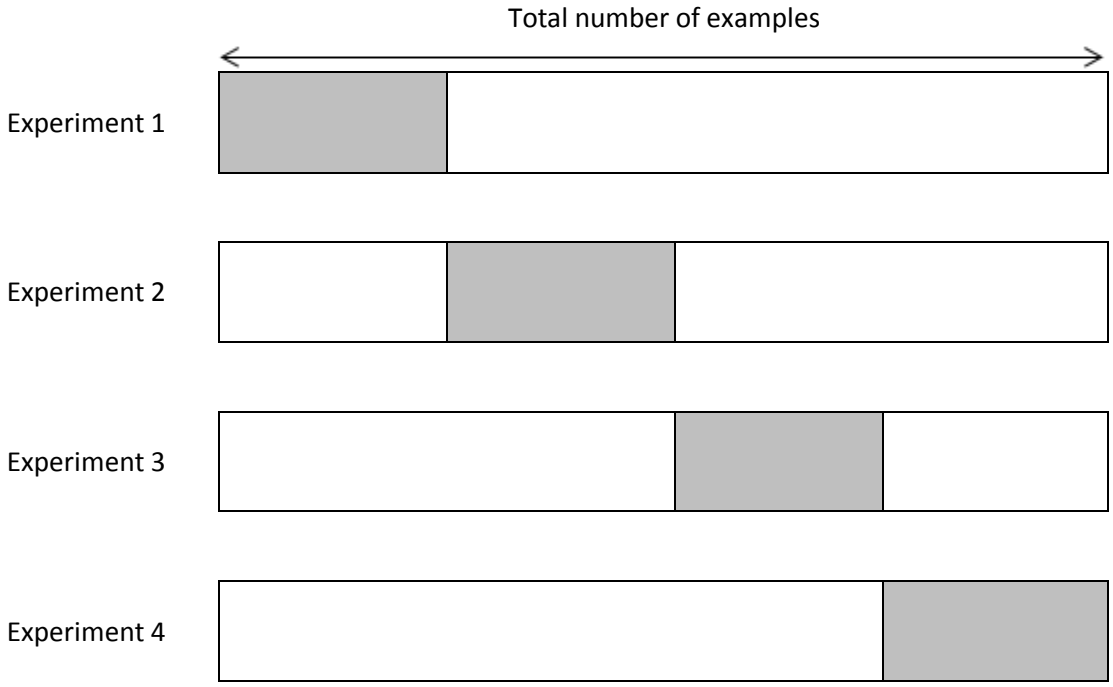


Figure 2.5 4-fold cross validation.

2.3.4 Leave One out Cross-Validation

Leave one out cross validation (LOOCV) is a special case of k-fold cross validation where K is equal to the number of the subjects. In this case in each run a subject will be left out as the testing set and all other subjects will be the training set. The division will be done N times and the accuracy will be calculated as the average accuracy of these divisions.

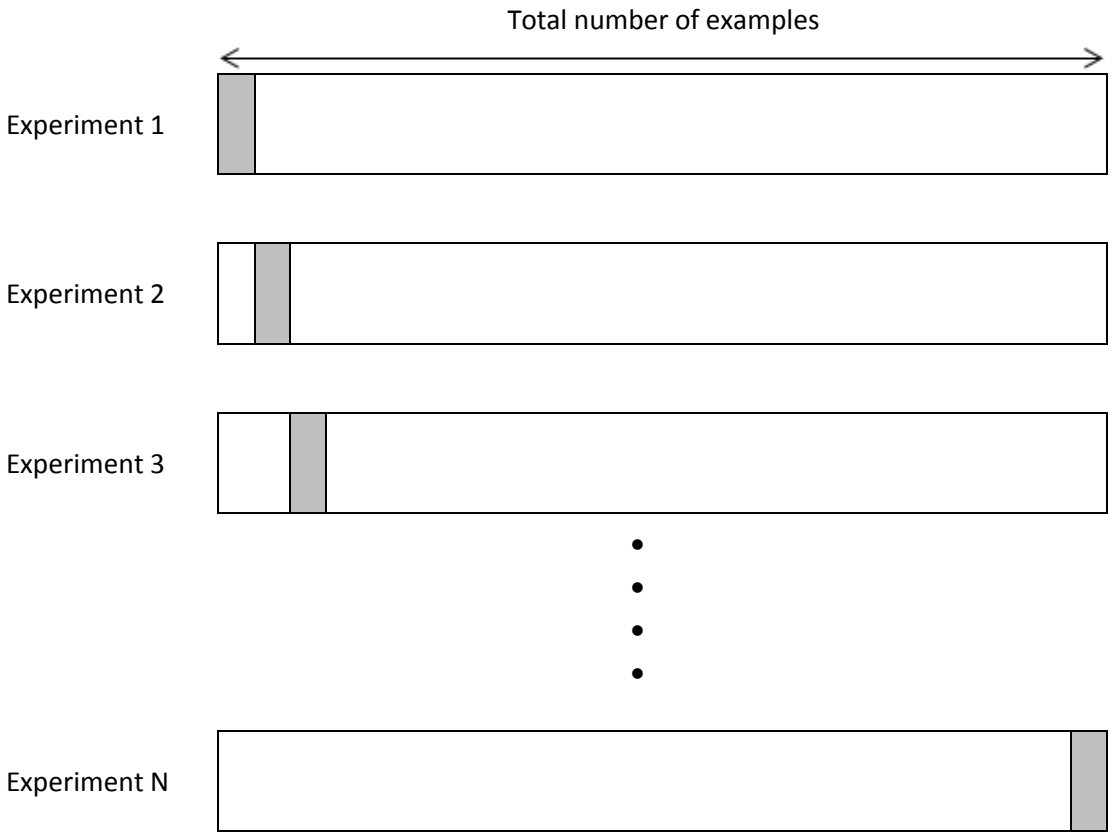


Figure 2.6 Leave One Out Cross validation.

2.4 Accuracy, Sensitivity and Specificity

In machine learning studies some concepts are repeated over and over, especially to evaluate the algorithm some concepts are used to report the evaluation; for this evaluation accuracy, sensitivity and specificity is used which will be explained in this section.

Consider two groups of subjects, so a subject either they belong to the healthy subjects (labeled as +1) or belong to patients (labeled as -1). After the training step, the pattern will be applied to the testing set to predict the subjects in the testing set using the pattern. There will be four conditions for the prediction; i) The subject is a healthy subject and correctly predicted as healthy control (True Positive, TP), ii) the subject is a healthy subject but predicted as a patient (False positive, FP) also known as type I error iii) the subject is a patient but predicted as a healthy subject (False Negative, FN) also known as type II error, and iv) the subject is a patient and correctly predicted as a patient (True Negative, TN). The sensitivity is calculated as shown in Table 2.1

Table 2.1 Sensitivity and Specificity Chart

		Condition	
		Condition positive	**Condition Negative**
Test Outcome	Test Outcome Positive	**True Positive , TP**	**False Positive**, FP Type I error
	Test Outcome Negative	**False Negative, FN** Type II error	**True Negative, TN**
		Sensitivity = $\frac{\sum True\ positive}{\sum Condition\ positive(TP+FN)}$	**Specificity** = $\frac{\sum True\ Negative}{\sum Condition\ Negative(FP+TN)}$

Source: http://en.wikipedia.org/wiki/Sensitivity_and_specificity Accessed on 10/10/2013

And the accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.20}$$

CHAPTER 3

METHODS

In this chapter the methods and algorithm applied to different groups of dataset is discussed. Datasets and Subjects

In This research two different datasets were used, both of them were independent to each other and the same algorithm was applied to each. Details about each dataset are shown in Table 3.1. To be able to refer to each of the dataset, the datasets are named as group 1 and group 2. Group 1 dataset is publicly available at http://cobre.mrn.org/ which and will referred to as COBRE dataset and group 2 was gathered by Long Island Jewish Clinic referred to as TAIWAN dataset.

Table 3.1 Subject and Dataset Information

Group		Total	N	Age (SD)	% female	Dataset size			number of voxels	
						X	Y	Z	total	After taking out the zeros
Group 1	SCH	74	28	40.67 (13.29)	50	61	73	61	271633	63915
	HC		46	34.34 (11.02)	50					
Group 2	SCH	86	43	21.40 (5.38)	50	53	63	46	153594	52815
	HC		43	21.40 (5.38)	50					

For each subject, the mALFF and fALFF of their resting state fMRI was calculated(Zou, Zhu et al. 2008) and each of these datasets was combined with other subjects and finally was fed to the algorithm in addition to these two dataset for each subject fALFF and mALFF was combined and after combining with other subjects it was fed to the algorithm.

3.1 Data Acquisition and Image Preprocessing

The Resting state Images for both groups were acquired with a 3.0 Tesla scanner (EPI). With a plane matrix of 64x64 and an acquisition voxel size of 3x3x3 mm (TR=2sec TE=2sec, with 150 time points). During the scan, participants were instructed to rest. Image processing was done using AFNI software (http://afni.nimh.nih.gov) on each of the subject, Motion correction was done with respect to the mean image. To have the ALFF (amplitude of low frequency fluctuations) dataset for each subject, the Power containing the frequency band between 0.01 to 0.1 Hz was calculated using the Fourier transform(Biswal, Zerrin Yetkin et al. 1995). Then the square root was calculated at each frequency of the power to derive amplitude between the frequency ranges, Average of amplitude between 0.01 to 0.1 Hz to derive ALFF for each of the voxel (Zou, Zhu et al. 2008).

Dividing the amplitude between 0.01 to 0.1 Hz with the amplitude of whole frequency band is done to derive fALFF (Fractional ALFF) and finally each of the subjects ALFF/fALFF map transformed into MNI 3mm standard space using linear registration and to have the dataset ready for the machine learning steps each

ALFF/fALFF 3D map in standard space was converted in to 1D file (using the reshape function in MATLAB).

3.1. Algorithm Steps

As mentioned in Chapter 1, the main goals of this study are to, 1) Rank the features according to their importance, and 2) Evaluate the ranking using supervised machine learning methods. To reach these goals the algorithm is divided to three main parts which will be explained separately. i) Preparing the dataset ii) Ranking the voxels and reducing number of features using the ranked voxels iii) Applying a supervised machine learning to evaluate the feature reduction.

3.2.1 Preparing the Dataset

The aim of this step is to label the dataset according to their group they belong to, arrange the datasets, scale the features and finally take out the voxels which don't contain any information (Figure3.3).

3.2.1.1 Labeling the Subjects. As this is a supervised learning study, the subjects have to be. Subject's label depends on which group they belong to (schizophrenia or healthy control). For this study I labeled the schizophrenia patients -1 and healthy control subjects are labeled to +1.

3.2.1.2 Subjects and Features Arrangement. After labeling the subjects, there will be an N×M metrics (N: number of subjects and M: number of features) in which the value of each voxel is the mALFF or the fALFF value. A schematic of the arrangement of the dataset is shown in figure 3.2.

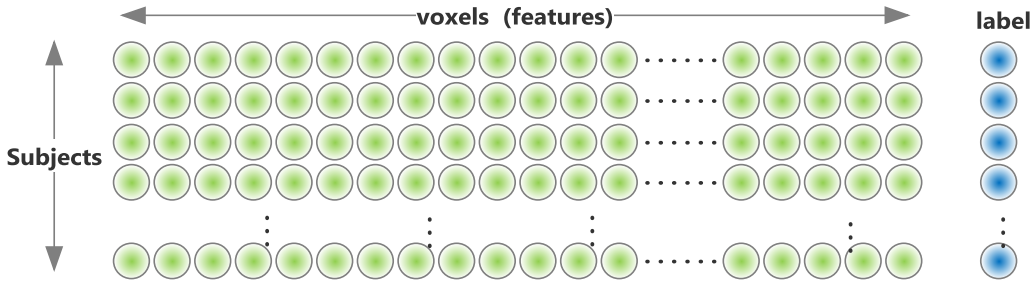


Figure 3.2 mALFF and fALFF dataset model.

Notice: the Xth feature is representing the same voxel for the entire subjects.

As mentioned before the machine learning steps was applied to mALFF, fALFF and combination of mALFF and fALFF data, to create the combination dataset, simply the mALFF and fALFF are added together for each subject. Which means instead of M feature there will be $2 \times M$ features (Figure 3.3).

Scaling Each Voxel

One other step done on the preparing the dataset is to scale the features. This is done to have a normal distribution for each feature. Let's say x_i^j represents the voxel value of the i^{th} feature of the subject number j ($i \in (1:M)$ and $j \in (1:N)$). The scaling voxel value (Z_i^j) is calculated using:

$$\boldsymbol{Z_i^j = \frac{x_i^j \times \mu_i}{\sigma_i}} \qquad (3.1)$$

In which μ_i is the mean of the i^{th} feature across the subjects and σ_i will be the standard deviation of the i^{th} feature across the subjects.

3.2.1.3 Taking out Voxels with a Value Equal to Zero. The brain fMRI images usually contain voxels with a zero value which represent outside of the brain and the brains skull. For a machine learning study, these voxels (feature) won't be representing anything but will interrupt the machine learning calculation which must be taken out. Using the mask dataset which is the same size of each subject's brain image and contains zeros and ones (ones are the voxels of the inner brain) the index of the outer brain is extracted and these values will be taken out from the dataset. The index will be saved for the reconstructing step which we'll explain in section 3.2. . The new number of voxels, after taking out the zeros is shown in Table 3.1.

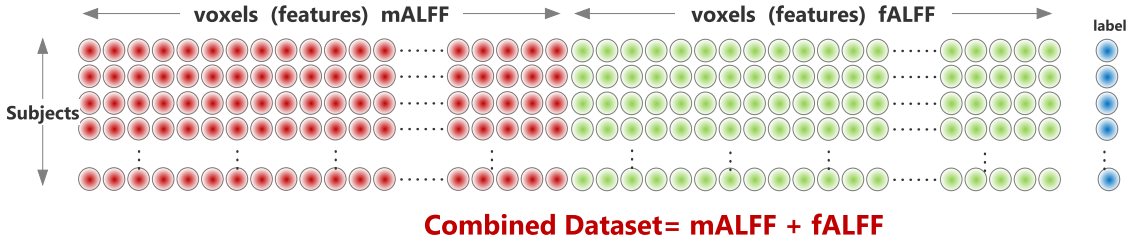


Figure 3.2 Combined dataset model.

Notice: when combining the dataset it is important to keep in mind that the mALFF data set has to be added to its own fALFF.

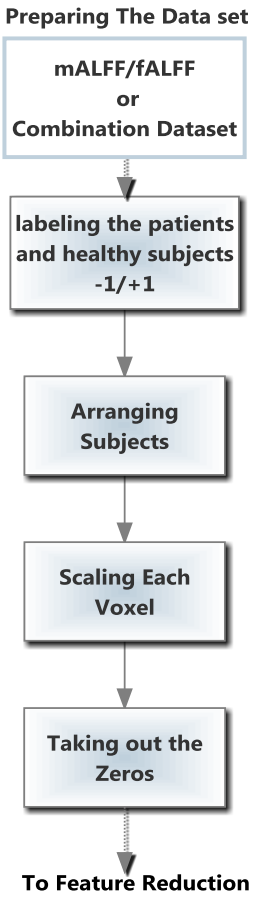


Figure 3.3 Flowchart- dataset preparations.

3.2.2 Features Ranking and Reduction

After preparing the dataset the dataset is ready for the next step which is feature reduction, and will be done using feature ranking

3.2.2.1 Feature Reduction. In machine learning studies especially when the dataset is too large to have a better classification it is advised to do a feature reduction and get rid of voxels which are misleading the classification, or by having those voxels in the dataset will not help on the classification. One of the purposes of this study is to reduce the unwanted features and rank the feature according to their importance. The result will be a smaller dataset and we assume by choosing the top ranked features we will have a better

accuracy and further more we can observe the voxels which are playing a more important role in the machine learning classification.

It is not exaggerating to say there is tons of feature selection and feature ranking methods. On this study there is only one of these methods used. To evaluate if the proper method is used, the machine learning step will be applied to see whether the ranking method is improving the accuracy or not.

Here, Kendall tau rank correlation coefficient is used to rank the features(Kendall 1948) which provides a distribution to measure the relevance of each feature represented by functional connectivity to classification. Suppose there are **m** samples in the control group and **n** samples in the patient group. Let x_i^j denote the voxel value of feature **i** of the **jth** samples and y_i^j denote the class label of this sample (+1 for controls and −1 for patients). The Kendall tau correlation coefficient for feature i can be defined as:

$$\tau_i = \frac{n_c - n_d}{m \times n} \tag{3.2}$$

In which n_c is the number of concordant pairs and n_d is the number of discordant. For a two class dataset and for a two observation of $\{x_i^j, y_i^j\}$ and $\{x_i^k, y_i^k\}$, it is a concordant pair when:

$$sgn\left(x_i^j - x_i^k\right) = sgn\left(y_i^j - y_i^k\right) \tag{3.3}$$

And it is a discordant pair if:

$$sgn\left(x_i^j - x_i^k\right) = -sgn\left(y_i^j - y_i^k\right) \tag{3.4}$$

Thus, if the τ_i has a positive value it means the i^{th} feature decreases in the patient group compared to the healthy subjects and vice versa for a negative τ_{ii}(Shen, Wang et al. 2010).

After calculating all features according to the absolute value of their Kendall coefficient value, the features will be ranked and a new dataset will be created. The index of the features will be saved for reconstructing the dataset.

3.2.2.2 Reconstructing the Brain Images. To observe the regions of the brain after creating the dataset using their τ_i value, a single vector will be created the same size as the entire features and the top ranked feature will have a maximum value and as the rank decreases their value will be decreased to, at the end there will be a vector with a range of (1-(total number of features)). The results are shown in the next chapter.

3.2.3 Machine Learning Steps

After ranking the features to evaluate the reduction applied to the dataset, a machine learning classification was used for different number of the top ranked features. As mentioned on Chapter 2 there is infinity different machine learning approaches for a classification on a dataset. Some of the approaches with different parameters were examined and the best was chosen. In this chapter the machine learning algorithm and the parameter used on this step will be explained.

3.2.3.1 Dividing the Dataset. As mentioned in chapter 2, the machine learning has to be first applied to the training set which is known as the training phase and after the classification pattern is extracted the pattern has to be validated using the testing dataset, the testing phase. Different kinds of cross validation will be explained.

For this project I divided the dataset to a training set containing 70% of the subjects and tested the pattern on the testing set which was containing the rest 30% of subjects. A 10-fold cross validation was applied on the training set.

3.2.3.2 Subject Shuffling. On the dividing step, the subjects are divided to two groups

(training and testing group), depending on the order of the subjects the accuracy will change. To overcome this problem the algorithm is applied to the dataset 1000 times and each time the subjects are shuffled. At the end the accuracy, sensitivity and specificity was calculated by averaging them over 1000 run.

3.2.3.3 Machine Learning Parameters. Different machine learning methods with different parameters was used to classify the two groups of subjects. For the final results I tried to pick the one with best accuracy result on the datasets. For this study the Support vector machine classifier with a soft margin equal to 0.001 and a linear kernel was chosen to classify the subjects.

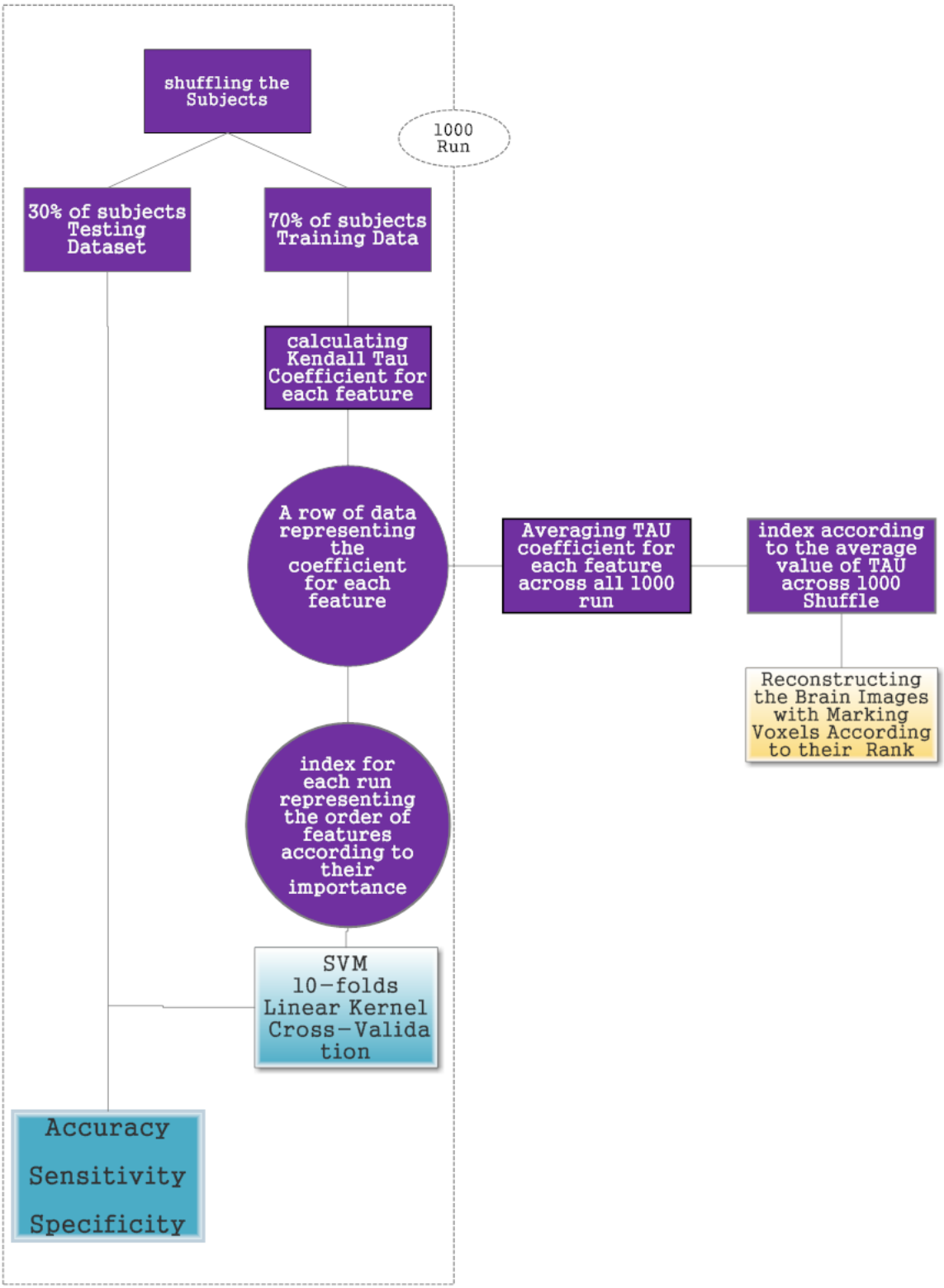


Figure 3.4 Flowchart - feature ranking (purple blocks), image reconstruction (in yellow) and machine learning (the blue blockes) steps.

CHAPTER 4

RESULTS AND DISSCUSSION

In this chapter, the results will be summarized and discussed. The results are summarized into two parts: 1) the accuracy of the machine learning step with different numbers of features and 2) reconstructed brain images which will illustrate the most important voxels according to their rank. The direction of these kinds of studies will also be explained.

4.1 Accuracy, Sensitivity and Specificity

As explained in Chapter 3 on the last step, the machine learning is applied to the dataset with different number of features. For each dataset the algorithm was applied for eight different numbers of features (top ranked voxels). The algorithm was applied for top 200,500, 1000, 3000,6000,15000, 30000 and all voxels. Each time the accuracy, sensitivity was calculated. Table 4.1 summarizes the result for 1000 shuffle.

Table 4.1 summarizes the accuracy, sensitivity and specificity for different features selected for the first group and Table 4.2 is the same result for group 2. To have a better understanding of the results, these two tables are plotted in Figures 4.1 and 4.2.

Table 4.1 Accuracy, Sensitivity and Specificity for different number of features selected for the algorithm (Group1).

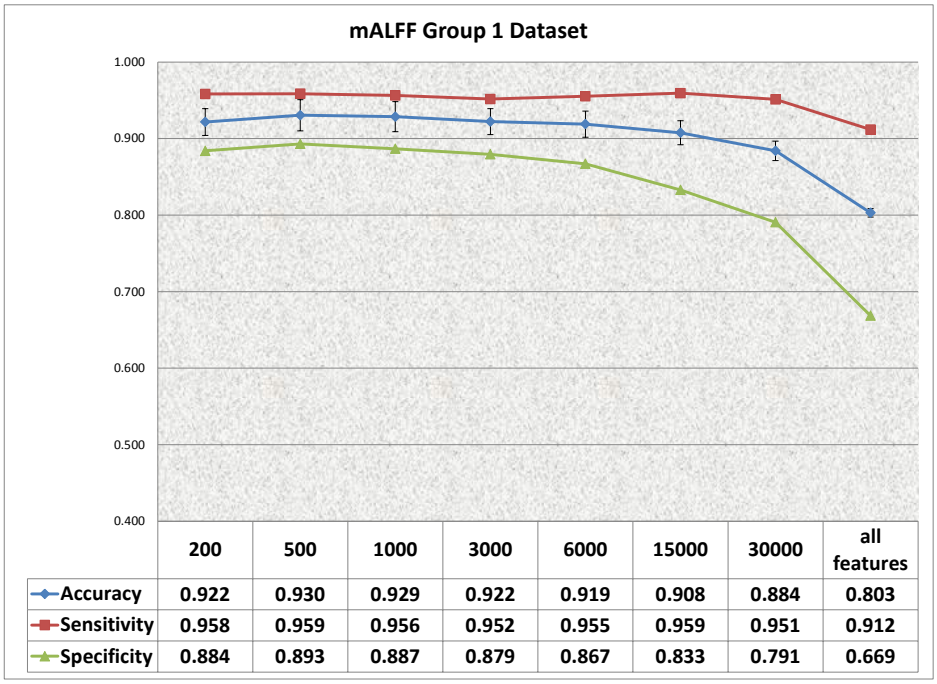
			Number of features selected							
			200	**500**	**1000**	**3000**	**6000**	**15000**	**30000**	**63915**
mALFF Group 1 Dataset	**accuracy**	Mean	0.778	0.802	0.793	0.784	0.776	0.773	0.757	0.694
		Std	0.065	0.062	0.064	0.065	0.068	0.068	0.071	0.08
	Sensitivity	Mean	0.823	0.839	0.828	0.818	0.82	0.841	0.846	0.815
		Std	0.065	0.062	0.064	0.065	0.069	0.068	0.071	0.08
	Specificity	Mean	0.728	0.759	0.75	0.739	0.716	0.68	0.634	0.531
		Std	0.065	0.062	0.064	0.065	0.069	0.068	0.071	0.08
fALFF Group 1 Dataset	**accuracy**	Mean	0.804	0.79	0.774	0.72	0.646	0.604	0.606	0.747
		Std	0.065	0.069	0.069	0.064	0.067	0.068	0.068	0.074
	Sensitivity	Mean	0.888	0.858	0.846	0.794	0.72	0.678	0.684	0.847
		Std	0.084	0.085	0.088	0.102	0.114	0.114	0.118	0.097
	Specificity	Mean	0.69	0.695	0.671	0.614	0.54	0.497	0.494	0.601
		Std	0.156	0.152	0.155	0.162	0.156	0.149	0.151	0.165
Combined Group 1 Dataset	**accuracy**	Mean	0.889	0.888	0.868	0.824	0.797	0.788	0.796	0.831
		Std	0.053	0.055	0.06	0.062	0.058	0.065	0.065	0.062
	Sensitivity	Mean	0.929	0.925	0.903	0.855	0.831	0.844	0.864	0.91
		Std	0.064	0.065	0.073	0.086	0.085	0.088	0.091	0.075
	Specificity	Mean	0.832	0.831	0.818	0.78	0.748	0.705	0.694	0.709
		Std	0.111	0.112	0.119	0.133	0.132	0.146	0.153	0.154

Table 4.2 Accuracy, sensitivity and Specificity for three datasets of group 2.

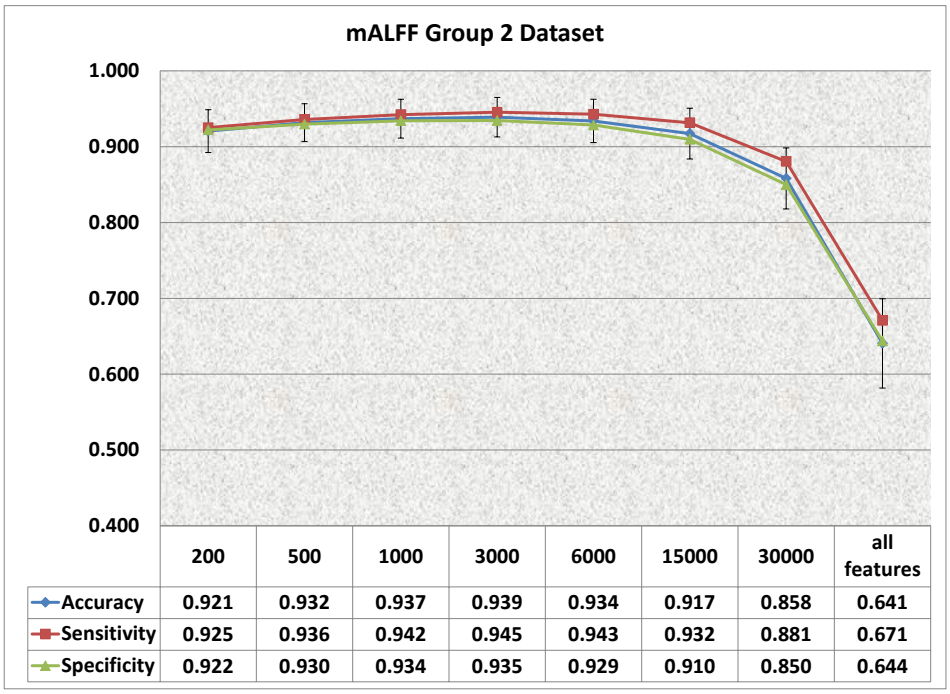
			Number of features selected							
			200	**500**	**1000**	**3000**	**6000**	**15000**	**30000**	**52815**
mALFF dataset Group 2	**Accuracy**	Mean	0.764	0.789	0.792	0.795	0.78	0.762	0.703	0.536
		Std	0.051	0.048	0.046	0.047	0.048	0.052	0.061	0.071
	Sensitivity	Mean	0.79	0.807	0.808	0.818	0.8	0.78	0.729	0.567
		Std	0.073	0.069	0.066	0.062	0.069	0.082	0.103	0.139
	Specificity	Mean	0.746	0.776	0.779	0.775	0.765	0.752	0.692	0.531
		Std	0.083	0.074	0.074	0.076	0.08	0.084	0.102	0.134
fALFF Dataset Group 2	**Accuracy**	Mean	0.698	0.665	0.607	0.458	0.398	0.342	0.338	0.465
		Std	0.062	0.061	0.064	0.066	0.064	0.062	0.062	0.071
	Sensitivity	Mean	0.796	0.744	0.665	0.446	0.374	0.32	0.322	0.472
		Std	0.087	0.097	0.12	0.142	0.142	0.141	0.142	0.153
	Specificity	Mean	0.617	0.598	0.561	0.483	0.436	0.38	0.373	0.485
		Std	0.122	0.114	0.117	0.12	0.119	0.121	0.124	0.137
Combined Dataset Group 2	**Accuracy**	Mean	0.847	0.845	0.847	0.849	0.835	0.796	0.755	0.726
		Std	0.06	0.058	0.055	0.056	0.058	0.063	0.065	0.071
	Sensitivity	Mean	0.938	0.916	0.912	0.885	0.868	0.835	0.8	0.775
		Std	0.059	0.067	0.07	0.078	0.085	0.096	0.107	0.116
	Specificity	Mean	0.767	0.78	0.79	0.821	0.811	0.766	0.723	0.693
		Std	0.108	0.102	0.106	0.104	0.104	0.11	0.116	0.124

As it is shown in the two last tables and was expected when mALFF and fALFF are combined, there will be more information (the features are doubled) for each subject, having more information on subjects and ranking the top features will give us more relevant features and training and testing to the top ranked features the accuracy in combined dataset is more, compared with mALFF and fALFF dataset.

As it is shown in Figures 4.1, 4.2 and 4.3, as the number of features is increasing the accuracy is decreasing which shows the feature with a lower rank, are misleading features for the classification and when they are added to the dataset the accuracy will decrease.

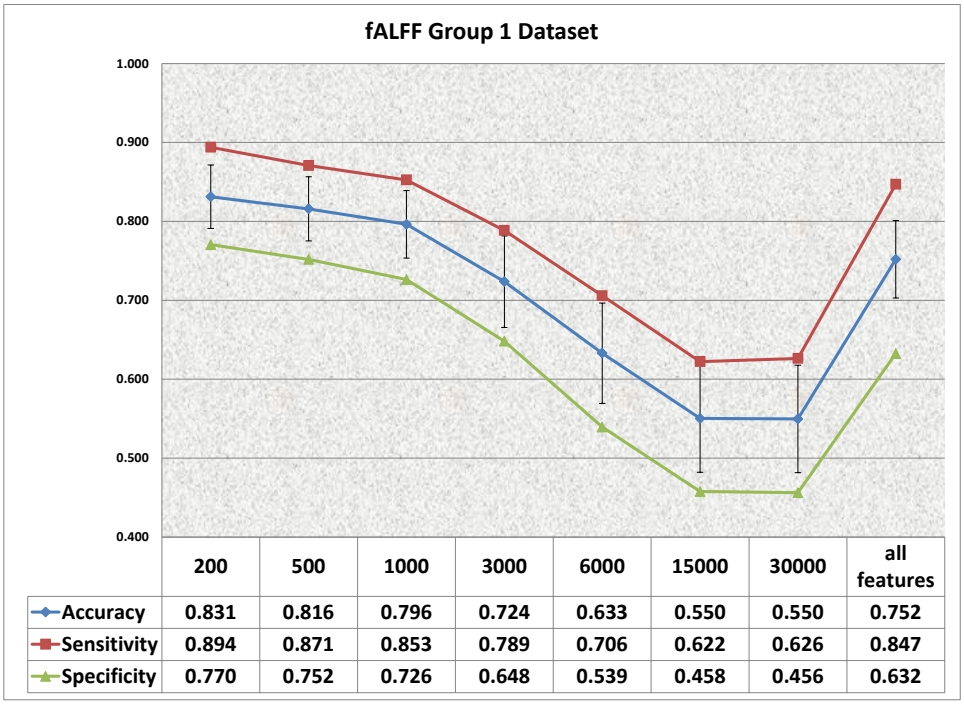


	200	500	1000	3000	6000	15000	30000	all features
Accuracy	0.922	0.930	0.929	0.922	0.919	0.908	0.884	0.803
Sensitivity	0.958	0.959	0.956	0.952	0.955	0.959	0.951	0.912
Specificity	0.884	0.893	0.887	0.879	0.867	0.833	0.791	0.669

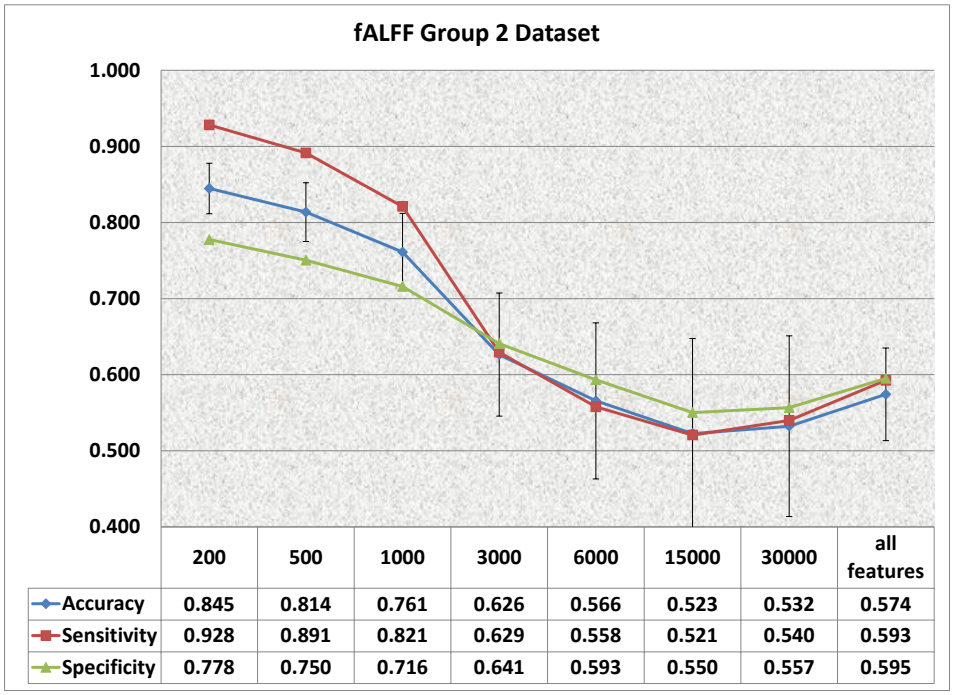


	200	500	1000	3000	6000	15000	30000	all features
Accuracy	0.921	0.932	0.937	0.939	0.934	0.917	0.858	0.641
Sensitivity	0.925	0.936	0.942	0.945	0.943	0.932	0.881	0.671
Specificity	0.922	0.930	0.934	0.935	0.929	0.910	0.850	0.644

Figure 4.1 Accuracy, sensitivity and Specificity for mALFF dataset

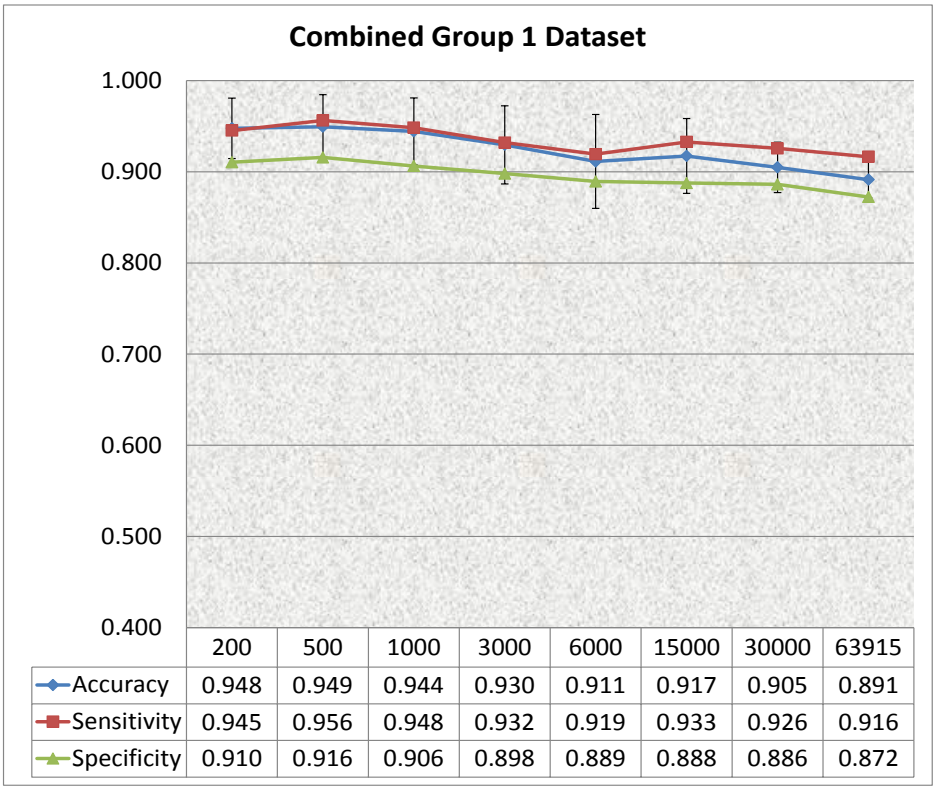


	200	500	1000	3000	6000	15000	30000	all features
Accuracy	0.831	0.816	0.796	0.724	0.633	0.550	0.550	0.752
Sensitivity	0.894	0.871	0.853	0.789	0.706	0.622	0.626	0.847
Specificity	0.770	0.752	0.726	0.648	0.539	0.458	0.456	0.632

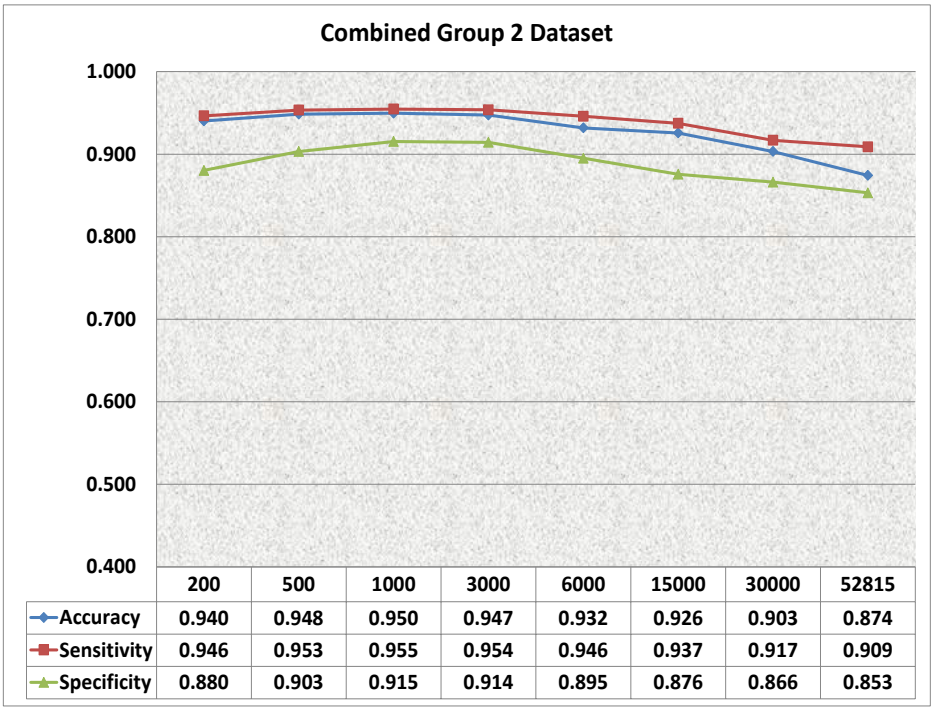


	200	500	1000	3000	6000	15000	30000	all features
Accuracy	0.845	0.814	0.761	0.626	0.566	0.523	0.532	0.574
Sensitivity	0.928	0.891	0.821	0.629	0.558	0.521	0.540	0.593
Specificity	0.778	0.750	0.716	0.641	0.593	0.550	0.557	0.595

Figure 4.2 Accuracy, sensitivity and Specificity for fALFF datasets



	200	500	1000	3000	6000	15000	30000	63915
Accuracy	0.948	0.949	0.944	0.930	0.911	0.917	0.905	0.891
Sensitivity	0.945	0.956	0.948	0.932	0.919	0.933	0.926	0.916
Specificity	0.910	0.916	0.906	0.898	0.889	0.888	0.886	0.872



	200	500	1000	3000	6000	15000	30000	52815
Accuracy	0.940	0.948	0.950	0.947	0.932	0.926	0.903	0.874
Sensitivity	0.946	0.953	0.955	0.954	0.946	0.937	0.917	0.909
Specificity	0.880	0.903	0.915	0.914	0.895	0.876	0.866	0.853

Figure 4.3 Accuracy, sensitivity and Specificity for combined datasets

4.2 Brain Voxels

Following are the top ranked voxels shown in the brain with different thresholds

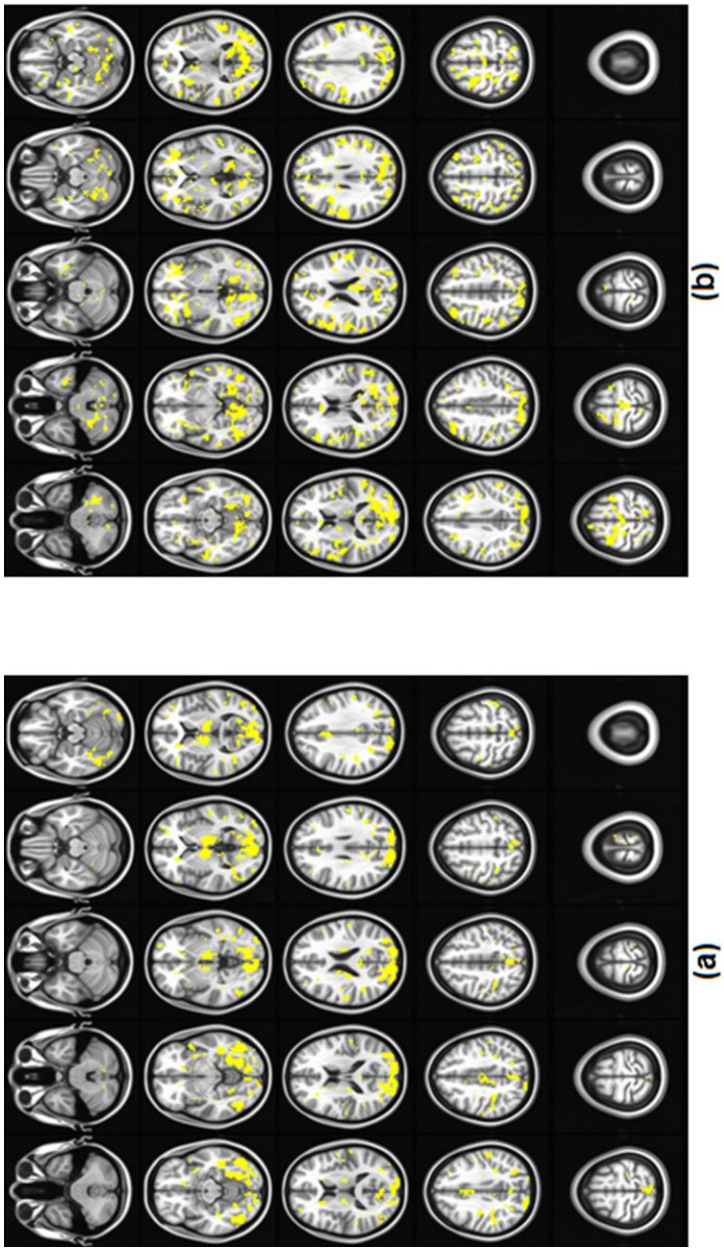


Figure 4.3 fALFF top 5% voxels- (a) group1 (COBRE dataset) (b) Group2 (TAIWAN dataset)

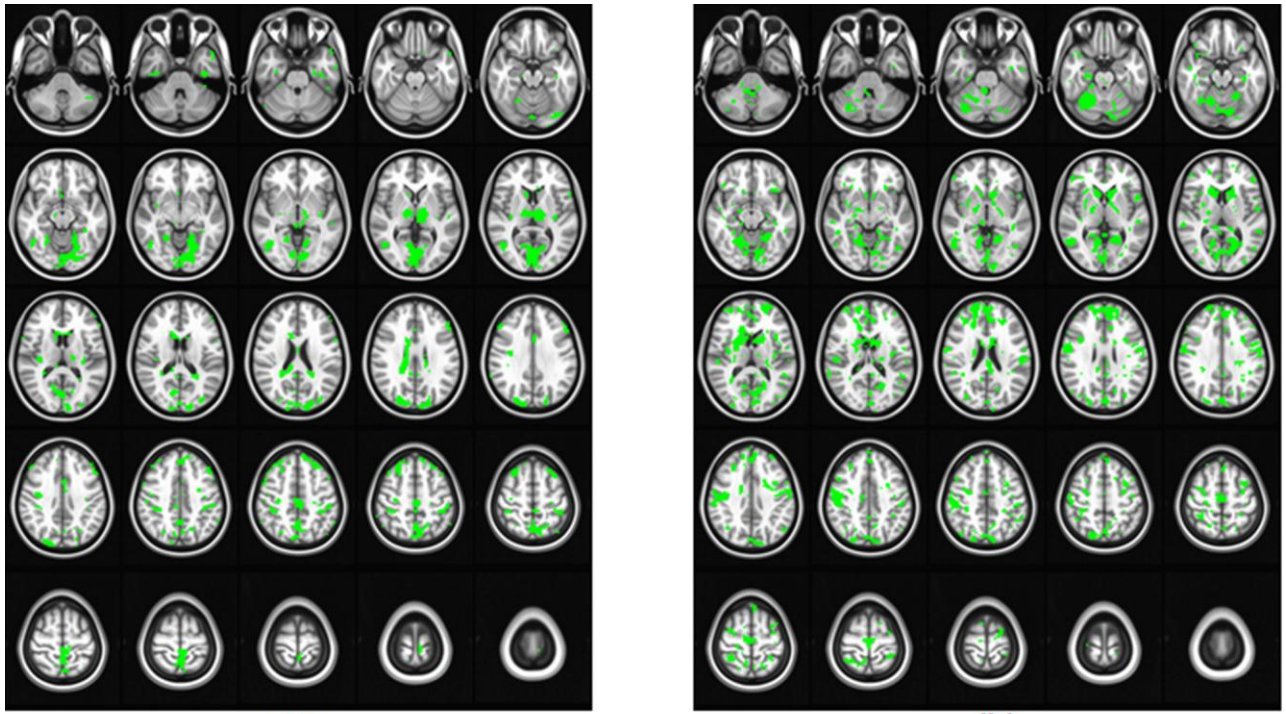


Figure 4.4 mALFF top 5% voxels- (a) group1 (COBRE dataset) (b) Group2 (TAIWAN dataset)

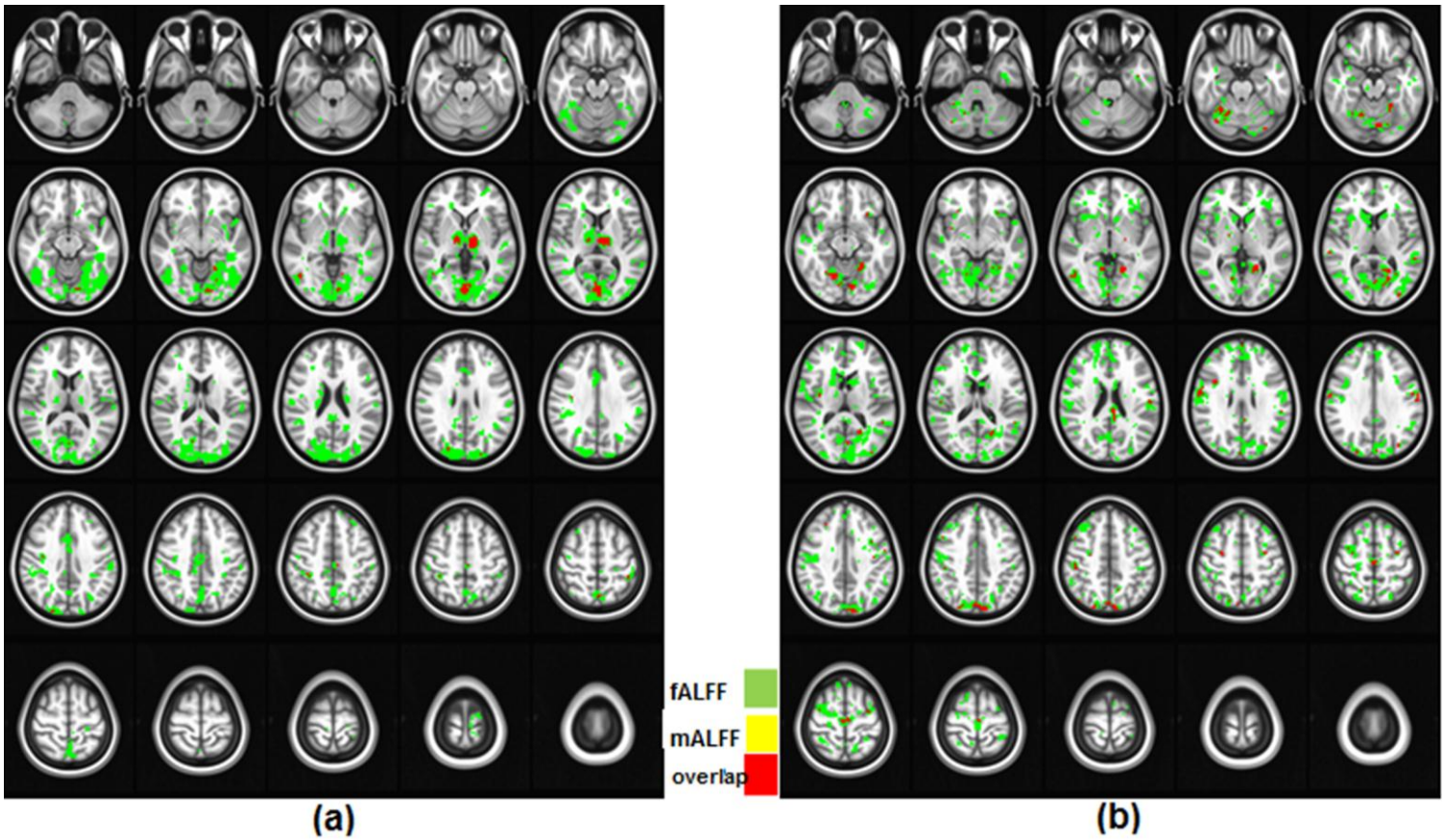


Figure 4.5 Combined top 5% voxels- (a) group1 (COBRE dataset) (b) Group2 (TAIWAN dataset)

4.3 Discussion and Future Directions

Figures 4.3 – 4.5 are illustrating the most important voxels in the brain. As discussed before, it can be reported that having a combined dataset, will be more reliable in classifying the subjects. In mALFF figures for both of the datasets (group1 and group2) the precuneus and visual cortices are playing an important role. For the fALFF dataset additional to these two regions posterior cingulate are marked. Finally the last figure (fig. 4.5) which is illustrating the most important regions in the combined dataset is showing important regions for more regions, the top 5% voxels marked for the combined dataset for both groups is marking superior frontal lobe, lingual gyrus, precuneuse and the visual cortices. And also as shown in the figure mALFF is playing a more important role for the classification (green voxels) comparing with the fALFF voxels (yellow voxels).

The COBRE dataset (group 1) is also showing important voxels in the cerebellum which it's not shown in group 2, which it can be because of the noise in group 2 dataset. The marked region differences between two groups of dataset (group1 and group2) can be because of the differences in group age differences or sex variation differences.

Another approach for these results is to divide the brain to different regions and apply the pattern extracted from the machine learning algorithm on different regions and observe the accuracy. Also future studies can be applied on multi label data set, meaning instead of only labeling the subjects as patients or healthy subjects they can have another label which for example determines if they are male or female or labeling them as old or young subject. This study was a voxel by voxel study done on resting-state fMRI, future studies can focus on other modalities to be compared with each other. In this study, SVM

was used as the classifier with a linear kernel, other classifiers can be applied to see if the accuracy or ranked features change or not.

This study was done in MATLAB using a machine learning toolbox called spider (http://people.kyb.tuebingen.mpg.de/spider/). The same algorithm can be applied to other fMRI datasets to have the same study on different datasets or different disorders. The same study can be done on other modalities and combination of different modalities.

To wrap up the results; in this study, classification of schizophrenic patients and healthy controls was successfully classified and also weighted brain voxels was determined by using mALFF and fALFF calculation of the resting fMRI. The results demonstrate a good performance of the classification algorithm.

REFERENCES

Biswal, B., F. Zerrin Yetkin, V. M. Haughton and J. S. Hyde (1995). "Functional connectivity in the motor cortex of resting human brain using echo-planar mri." Magnetic resonance in medicine **34**(4): 537-541.

Bluhm, R. L., J. Miller, R. A. Lanius, E. A. Osuch, K. Boksman, R. Neufeld, J. Théberge, B. Schaefer and P. Williamson (2007). "Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: anomalies in the default network." Schizophrenia bulletin **33**(4): 1004-1012.

Camastra, F. and A. Vinciarelli (2008). Machine learning for audio, image and video analysis: Theory and applications, Springer, Boston, MA, USA.

Costafreda, S. G., C. H. Fu, M. Picchioni, T. Toulopoulou, C. McDonald, E. Kravariti, M. Walshe, D. Prata, R. M. Murray and P. K. McGuire (2011). "Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder." BMC psychiatry **11**(1): 18.

Gong, Q., Q. Wu, C. Scarpazza, S. Lui, Z. Jia, A. Marquand, X. Huang, P. McGuire and A. Mechelli (2011). "Prognostic prediction of therapeutic response in depression using high-field MR imaging." Neuroimage **55**(4): 1497-1503.

Heckers, S., D. Goff and A. P. Weiss (2002). "Reversed hemispheric asymmetry during simple visual perception in schizophrenia." Psychiatry Research: Neuroimaging **116**(1): 25-32.

Jafri, M. J., G. D. Pearlson and V. D. Calhoun (2007). A maximal-correlation approach using ICA for testing functional network connectivity applied to Schizophrenia. Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on, IEEE.

Koutsouleris, N., S. Borgwardt, E. M. Meisenzahl, R. Bottlender, H.-J. Möller and A. Riecher-Rössler (2012). "Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study." Schizophrenia bulletin **38**(6): 1234-1246.

LaConte, S., S. Strother, V. Cherkassky, J. Anderson and X. Hu (2005). "Support vector machines for temporal classification of block design fMRI data." NeuroImage **26**(2): 317-329.

Lawrie, S. M., C. Buechel, H. C. Whalley, C. D. Frith, K. J. Friston and E. C. Johnstone (2002). "Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations." Biological psychiatry **51**(12): 1008-1011.

Mitchell, T. M. (1997). Machine learning. McGraw-Hill: Boston, MA.

Orrù, G., W. Pettersson-Yeo, A. F. Marquand, G. Sartori and A. Mechelli (2012). "Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review." Neuroscience & Biobehavioral Reviews **36**(4): 1140-1152.

Pagani, M., D. Salmaso, G. Rodriguez, D. Nardo and F. Nobili (2009). "Principal component analysis in mild and moderate Alzheimer's disease—A novel approach to clinical diagnosis." Psychiatry Research: Neuroimaging **173**(1): 8-14.

Pereira, F., T. Mitchell and M. Botvinick (2009). "Machine learning classifiers and fMRI: a tutorial overview." Neuroimage **45**(1): S199-S209.

Shen, H., L. Wang, Y. Liu and D. Hu (2010). "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI." NeuroImage **49**(4): 3110-3121.

Shoeb, A., A. Kharbouch, J. Soegaard, S. Schachter and J. Guttag (2011). "A machine-learning algorithm for detecting seizure termination in scalp EEG." Epilepsy & Behavior **22**: S36-S43.

Simon, P. (2013). Too Big to Ignore: The Business Case for Big Data, Wiley, Hoboken, New Jersey, USA.

Vapnik, V. (2000). The nature of statistical learning theory, Springer-Verlag, New York, USA.

Yang, H., X.-Y. Long, Y. Yang, H. Yan, C.-Z. Zhu, X.-P. Zhou, Y.-F. Zang and Q.-Y. Gong (2007). "Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI." Neuroimage **36**(1): 144-152.

Yin, J., D. Hu and Z. Zhou (2008). "Noisy manifold learning using neighborhood smoothing embedding." Pattern Recognition Letters **29**(11): 1613-1620.

Yu-Feng, Z., H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi and W. Yu-Feng (2007). "Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI." Brain and Development **29**(2): 83-91.

Zhang, D., Y. Wang, L. Zhou, H. Yuan and D. Shen (2011). "Multimodal classification of Alzheimer's disease and mild cognitive impairment." Neuroimage **55**(3): 856-867.

Zou, Q.-H., C.-Z. Zhu, Y. Yang, X.-N. Zuo, X.-Y. Long, Q.-J. Cao, Y.-F. Wang and Y.-F. Zang (2008). "An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF." Journal of neuroscience methods **172**(1): 137-141.