Development and evaluation of optimization based data mining techniques for analysis of brain data

Mahdi Zarei

Principal Supervisor: Dr Zari Dzalilov
Associate Supervisor: Assoc.Prof. Adil Baghirov

Thesis is submitted in total fulfilment of the
requirement for the degree of Doctor of Philosophy

School of Applied and Biomedical Sciences
Faculty of Science and Technology
Federation University Australia
PO Box 663
University Drive, Mount Helen
Ballarat, VIC 3353, Australia.

Abstract

Neuroscience is an interdisciplinary science which deals with the study of structure and function of the brain and nervous system. Neuroscience encompasses disciplines such as computer science, mathematics, engineering, and linguistics. The structure of the healthy brain and representation of information by neural activity are among most challenging problems in neuroscience. Neuroscience is experiencing exponentially growing volumes of data obtained by using different technologies. The investigation of such data has tremendous impact on developing new and improving existing models of both healthy and diseased brains.

Various techniques have been used for collecting brain data sets for addressing neuroscience problems. These data sets can be categorized into two main groups: resting-state and state-dependent data sets. Resting-state data is based on recording the brain activity when a subject does not think about any specific concept while state-dependent data is based on recording brain activity related to specific tasks.

In general, brain data sets contain a large number of features (e.g. tens of thousands) and significantly fewer samples (e.g. several hundred). Such data sets are sparse and noisy. In addition to these problems, brain data sets have a few number of subjects. Brains are very complex systems and data about any brain activity reflects very complex relationship between neurons as well as different parts of the brain. Such relationships are highly nonlinear and general purpose data mining algorithms are not always efficient for their study.

The development of machine learning techniques for brain data sets is an emerging research area in neuroscience. Over the last decade, various machine learning techniques have been developed for application to brain data sets. In the meantime, some well-known algorithms such as feature selection and supervised classification have been modified for analysis of brain data sets. Support vector machines, logistic regression, and Gaussian Naive Bayes classifiers are widely used for application to

brain data sets. However, Support vector machines and logistic regression algorithms are not efficient for sparse and noisy data sets and Gaussian Naive Bayes classifiers do not give high accuracy.

The aim of this study is to develop new and modify the existing data mining algorithms for the analysis brain data sets. Our contribution in this thesis can be listed as follow:

1. Development of new algorithms:

   1.1. Development of new voxel (feature) selection algorithms for Functional magnetic resonance imaging (fMRI) data sets, and evaluation of these algorithms on the Haxby and Science 2008 data sets.

   1.2. Development of new feature selection algorithm based on the catastrophe model for regression analysis problems.

2. Development and evaluation of different versions of the adaptive neuro-fuzzy model for the analysis of the spike-discharge as a function of other neuronal parameters.

3. Development and evaluation of the modified global k-means clustering algorithm for investigation of the structure of the healthy brain.

4. Development and evaluation of region of interest (ROI) method for analysis of brain functional-connectivity in healthy subjects and schizophrenia patients.

Statement of authorship

This thesis contains no work extracted in whole or in part from a thesis, presented for another degree or diploma except where explicit reference is made. No other persons work has been relied upon or used without due acknowledgment in the main text and bibliography of the thesis.

Signed: _____ M. Zarei

Date: 03/08/2015

Acknowledgement

First of all, I would like to express my gratitude to my principal supervisor Dr. Zari Dzalilov and my associate supervisor associate Professor Adil Baghirov. I was supported by their guidance, advises, enthusiasm and comments and not only in the sense of study and research. I have been extremely lucky to have supervisors who responded to my questions and queries so promptly.

I am thankful to our co-author in the paper Dr. Mikail Rubinov (University of Cambridge, UK) for his advice and help during these years.

I would like to acknowledge the support from Federation University Australia, the School of Applied and Biomedical Sciences and research office for the financial support.

I appreciate my mother's and siblings' moral support and encouragement for doing of the research in Australia.

This dissertation is dedicated to my father who encourages me in my life and research. I lost him, but I will do my best to make him proud.

Dedication

This thesis is dedicated to my father,

Badr Zarei.

List of publications

Selected publications

1. A Bagirov, M. Zarei, Z. Dzalilov. Voxel selection algorithm for fMRI data sets using hyperrectongles, NeuroEng 2014, the 7th Australian Workshop on Computational Neuroscience Date: 2-3 February, 2014

2. A Bagirov, M. Zarei, Z. Dzalilov. Optimization challenges in resting state fMRI data analysis, Recent Advances on Optimization, July 24-26, 2013

3. A Bagirov, M. Zarei, Z. Dzalilov. Feature selection algorithm for fMRI data sets using hyperrectangles to improve performance of classifiers, (under review)

4. M. Zarei, Z. Dzalilov, M. Rubinov, A. Anderson. Functional connectivity differences between men and women in healthy brains and schizophrenia, (under review)

Conference presentations

1. M. Zarei, A. Bagirov, Z. Dzalilov. Optimization challenges in resting state fMRI data analysis, 3rd European Conference on Computational Optimization 2013, Chemnitz, Germany

2. M. Zarei, A. Bagirov, Z. Dzalilov. Analysis of fMRI data using machine learning algorithms, 2012 Annual Research Conference, Ballarat, November 8, 2012, Ballarat, Australia

3. M. Zarei, A. Bagirov, Z. Dzalilov. Selecting informative Voxels of fMRI data sets using clustering and Neural Networks, Australian Mathematical Society 56th Annual Meeting, University of Ballarat, September 24-27, 2012, Ballarat, Australia

Contents

List of abbreviations

- Adaptive neuro fuzzy inference system (ANFIS)

- Akaike information criterion (AIC)

- Analysis of Functional NeuroImages (AFNI)

- Artificial Neural Network (ANN)

- Blood Oxygen Level-Dependent (BOLD)

- Center for Biomedical Research Excellence (COBRE)

- Cerebrospinal Fluid (CSF)

- Independent Component Analysis (ICA)

- Component-based noise Correction method (CompCor)

- Diffuse Optical Imaging (DOI)

- Diffusion Tensor Imaging (DTI)

- Dorsolateral Prefrontal Cortex (DLPFC)

- Dynamic evolving neural-fuzzy inference system (DENFIS)

- Electroencephalography (EEG)

- Event-related optical signal (EROS)

- FMRIB Software Library (FSL)

- function Magnetic resonance imaging (fMRI)

- Gaussian Naive Bayes (GNB)

- General linear model (GLM)

- genetic for lateral tuning and rule selection of linguistic fuzzy system (GFS.LT.RS)

- subtractive clustering and fuzzy c-means (SBC)

- Hybrid neural Fuzzy Inference System (HyFIS)

- inferior frontal junction (IFJ)

- Magnetic resonance imaging (MRI)

- Magnetoencephalography (MEG)

- MultiVariate Pattern Analysis (MVPA)

- MultiVariate Pattern Analysis in Python (PyMVPA)

- matrix laboratory (MATLAB)

- Positron emission tomography (PET)

- prefrontal cortex (PFC)

- Principal component analysis (PCA)

- region of interest (ROI)

- Single-photon emission computed tomography (SPECT)

- Resting state fMRI (rsfMRI)

- Statistical parametric mapping (SPM)

- Support Vector Decomposition Machines (SVDMs)

- Support Vector Machines (SVMs),

- Waikato Environment for Knowledge Analysis (WEKA)

- Wang and Mendel (WM)

Introduction

Understanding the human brain is one of the greatest challenges of 21st century science. Neuroscience as an interdisciplinary science deals with anatomical, functional, medical, molecular, cellular, developmental, evolutionary and computational aspects of the brain and nervous system. It encompasses various range of disciplines like computer science, mathematics, engineering and linguistics. Problems of the structure of a healthy brain and how the neural activity represents information are among most challenging in neuroscience. Neuroscience is experiencing exponentially growing volumes of data based on different technologies for healthy and diseased brains. Such data has tremendous impact on developing new and improving existing models of both healthy and diseased brains.

Various techniques like magnetic resonance imaging, functional magnetic resonance imaging, diffuse optical imaging, event-related optical signal, diffusion tensor imaging, electroencephalography, magneto encephalography, positron emission tomography, single-photon emission computed tomography are used for collecting brain data sets [8, 50, 66, 69, 81, 197, 201]. These data sets can be categorized into two main groups; resting-state and state dependent data sets. Resting-state data is based on recording the brain activity when subject does not think about specific concept, but state-dependent data is based on recording brain activity in that is related to some inputs [55, 155].

In general, brain data sets contain a large number of features (tens of thousands) and significantly less samples (several hundred). Such data sets are sparse and noisy. In addition to these problems, brain data sets have a few number of subjects.

The brain is a very complex system and a data about any brain activity reflects very complex relationship between neurons as well as different parts of the brain. Such relationships are highly nonlinear and general purposed data mining algorithms are not always efficient to study such data sets.

Neuroscience Research Questions

As mentioned before the problems of the structure of a healthy brain and how the neural activity represents information are among the most challenging in neuroscience. The study of such problems can be achieved by answering the following questions:

1. Is there any difference in neural activities when a subject thinks about different concepts? Application of data mining techniques can help to answer this question.

2. Are neural representations similar across people when they think about the same stimulus? To answer this question, it is important to study different subjects under the same conditions.

3. What kind of information can be extracted from fMRI data with available technologies? Study from [138] demonstrates that it is possible to extract different ideas such as words when subject think about them.

4. How to find the best model for a healthy brain?

5. What is the main difference in functional-connectivity between healthy and unhealthy brains?

Research aims

Over the last decade various machine learning techniques have been applied to study brain data sets. The development of machine learning techniques for brain data sets is an emerging research area in neuroscience. Results show that traditional feature selection, supervised classification and clustering algorithms are not efficient for brain data sets due to the high nonlinearity of functional connectivities between different parts of the brain. Recently, some well known algorithms for feature selection and supervised classification in data mining have been modified for brain data sets. Generally classifiers work better with less features, but choosing this number of features is a complicated process. Our knowledge of how to best implement feature selection for fMRI data is still preliminary [75]. Activity, Accuracy, Searchlight Accuracy, Analysis of Variance and Stability methods are commonly used for feature selection of fMRI data [139]. Support vector machines, logistic regression, and Gaussian Naive Bayes classifiers are widely used for application to brain data sets. However,

Support vector machines and logistic regression algorithms are not efficient for sparse and noisy data sets and Gaussian Naive Bayes classifiers do not give high accuracy.

The aim of this study is developing of new and modification of existing data mining algorithms for the analysis of large-scale, spars and noisy brain data sets. New voxel (feature) selection algorithm for fMRI data sets and feature selection algorithm for regression analysis problems have been developed in this thesis. In order to investigate the structure of healthy brain the global k-means clustering algorithm is applied for a resting-state fMRI data. Region of interest method developed and evaluated for analysis of brain functional-connectivity in healthy subjects and schizophrenia patients.

Significance

Many data sets are collected using different technologies to analyze brain's functions. Also numerous algorithms developed to answer some open questions in this field, but development of more efficient and accurate algorithms for obtaining more accurate results is still important. Designing new models will answer many questions in neurolinguistics and in detecting psychiatric illnesses, amongst others. Designing brain-computer interfaces and brain-to-brain communication computer based models are other potential applications of results obtained in this thesis.

Analyzing the different brain data sets using graph theoretical, statistical and data mining algorithms based on different scenarios can answer many questions about structure of healthy brain and the areas of the brain that are involved in different brain disorders.

Contributions

The main contributions of this work are:

- Development of new voxel (feature) selection algorithms for fMRI data sets. Haxby and Science 2008 data sets are used for evaluation of these algorithms.

- Development of new feature selection algorithm based on the catastrophe model for regression analysis problems.

- Development and evaluation of different versions of the adaptive neuro-fuzzy model for the analysis of the spike-discharge as a function of other neuronal parameters.

- Development and evaluation of the modified global k-means clustering algorithm for investigation of the structure of the healthy brain.

- Development and evaluation of region of interest (ROI) method for analysis of brain functional-connectivity in healthy subjects and schizophrenia patients.

Outline of the thesis

This thesis is divided into six chapters as follows.

- In Chapter 1 we present the literature on related work. First we describe different brain imaging technologies and present the data sets that analyzed in this thesis. Then we briefly describe some of commonly used algorithms on supervised classifiers, dimension reduction, complex networks.

- The proposed algorithms are discussed in Chapter 2. First we explain our voxel (feature) selection algorithm that is based on overlap between voxels' activation levels. Then the application of this algorithm to Haxby and Science 2008 data sets is discussed. Another algorithm is feature selection algorithm for regression analysis problems that finds the most relevant feature based on the catastrophe model. The effectiveness of the proposed algorithm is tested using Breast cancer, Parkinson Telemonitoring and Slice locality data sets.

- In Chapter 3 we developed and evaluated different versions of the adaptive neuro-fuzzy model for the analysis of the spike-discharge. The spike-discharge is considered as a function of other neuronal parameters..

- The well-known Modified Global k-means algorithm is applied to Oxford data and the obtained results are shown in Chapter 4. It is demonstrated that this clustering algorithm can be used for finding the structure of the healthy brains.

- In Chapter 5 Region of Interest (ROI) method is developed and evaluated for analysis of brain functional-connectivity. Two groups of healthy controls and schizophrenia patients from CO-BRE data is used for this purpose.

Chapter 1

Literature review

Understanding the organization of the brain and determining how mental representations map onto patterns of neural activity are among the greatest challenges facing 21st century science. Neuroscience as an interdisciplinary science deals with anatomical, functional, medical, molecular, cellular, developmental, evolutionary and computational aspects of the brain and nervous system. It encompasses various range of disciplines like computer science, mathematics, engineering and linguistics. Exponentially growing volumes of data based on different technologies for healthy and diseased brains has tremendous impact on developing new and improving existing models of both healthy and unhealthy brains.

Different technologies are used for measuring brain activity and its structure in order to collect brain data sets to address the neuroscience problems. These data sets can be arranged into main groups: resting-state and state-dependent data sets. Resting-state data is based on recording the brain activity when subject does not think about any specific concept, but state-dependent data is based on recording brain activity in that is related to some inputs.

Generally, brain data sets contain a large number of features and significantly less samples. Therefore such data sets are noisy and sparse. Brain is a very complex system and very complex relationship between neurons and different parts of the brain can be reflected by a data about brain activity. Such relationships are highly nonlinear and general purpose data mining algorithms are not always efficient to study such data sets.

Recently, various machine learning algorithms have been applied to study brain data sets. The development of these algorithms for brain data sets is an emerging research area in neuroscience. Over

the last decade, some well-known feature selection and supervised classification in data mining have been modified for brain data sets. Usually classifiers work better with less features, but choosing this number of features is a difficult process. Our knowledge of how to best implement feature selection for brain data is still preliminary [75]. Activity, Accuracy, Searchlight Accuracy, Analysis of Variance and Stability methods are commonly used for feature selection of fMRI data [139]. Support vector machines, logistic regression, and Gaussian Naive Bayes classifiers are widely used for application to brain data sets. However, Support vector machines and logistic regression algorithms are not efficient for sparse and noisy data sets and Gaussian Naive Bayes classifiers do not give high accuracy.

In this chapter we present the literature on the brain data sets based on different brain imaging techniques and different data analysis algorithms including dimension reduction, classification, regression analysis, clustering, neuro-fuzzy and complex networks. First, we describe different brain imaging technologies and present the data sets that analyzed in this thesis. Then we briefly describe commonly used algorithms including supervised classifiers, dimension reduction and complex networks.

1.1   Brain imaging techniques

Different technologies are used for measuring of brain activity and it's structure. In this section we present some of them that are widely used in the neuroscience.

- Functional magnetic resonance imaging measures brain activity by detection fluctuating of oxygenation and flow that occurs in response to neural activity in blood [81]. FMRI is a powerful tool for the brain functions studies. It is a technique for obtaining three-dimensional images related to activity in the brain through time. FMRI Scanners can measure the changes in the blood magnetic resonance that called the Blood Oxygen Level-Dependent (BOLD) signal. The smallest unit measured by BOLD fMRI is called a voxel. FMRI's spatial resolution is around 2mm and it's contrast can be performed on more than 100,000 voxels [128].

- Magnetic resonance imaging is used for studying of brain anatomy and static structure of the brain matter. It is also called nuclear magnetic resonance imaging and has many applications in different areas of medicine.

- Diffuse optical imaging is a 3d imaging technique that measures the changes in oxygenated and deoxygenated hemoglobin [50].

- Event-related optical signal is a scanning technique based infrared light and measures changes of the cortex [66].

- Diffusion tensor imaging is another non-invasive technology that characterizes the three-dimensional diffusion of water as a function of spatial location [8, 21].

- Electroencephalography records electrical activity along the scalp by measuring the voltage fluctuation within neurons [126].

- Magnetoencephalography records magnetic fields in the brain. It's temporal resolution is 1ms and spatial resolution is around 2 mm [69].

- Positron emission tomography by detecting gamma rays into the body provides a three-dimensional image of the functional process [197].

- Single-photon emission computed tomography uses gamma rays and provides a three dimensional model of the brain [201].

Research in this thesis relates to state-dependent and resting-state data. State-dependent data is based on recording brain's activities during special task [211] while resting-state data reflects the brain activity of the subjects that do not do any task [27].

1.2 Brain data sets

Many data sets which tested in this thesis are based on fMRI technique. This technique is widely used for analyzing brain activities because it has very high spatial resolution and each voxel in fMRI image has about two millimeters width. Thinking process is very fast and sometimes subject thinks about many concepts in one second, whereas in fMRI we can get only one image per second. Usual method for collecting state-dependent data with functional magnetic resonance imaging technique is presenting stimuli to subjects, instruct them to think about stimuli and capturing their brain images when they think. Haxby, Science2008 and StarPlus are state-dependent data sets while COBRE and Oxford data sets are resting state data sets.
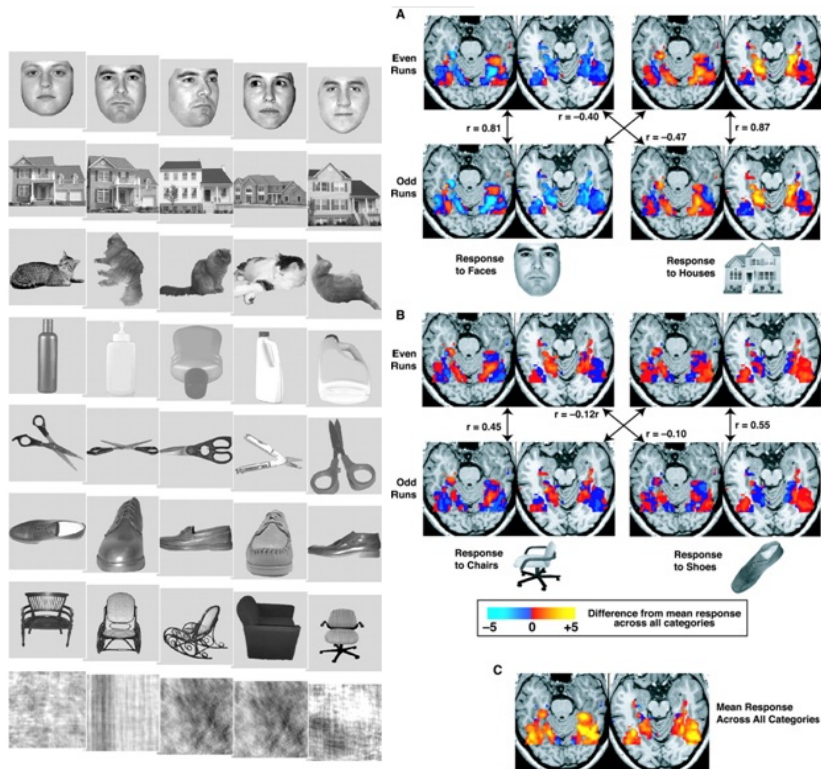
Figure 1.1: Haxby data

Resting-state fMRI (rsfMRI) data is another type of fMRI data that has been used in many applications in the neuroscience area. Study of resting-state of the brain can help us to better explain its functional connectivities.

1.2.1 Haxby data

Haxby data was collected by Haxby et al. [48] to study the face and object representation in human ventral temporal cortex. The temporal lobe is a region of the cerebral cortex of the brain that involved in processing of semantics in the both speech and the vision. In each trail subjects looked at greyscale images of eight object categories. Each image was shown for 500msec and brain fMRI data were recorded every 2.5s and trail were ran 12 times for every subject. Figure 1.1 shows the material and categories of stimuli in this data.

1.2.2 Science data

This data was collected by Mitchell et al. [118] and it is based on capturing brain's image when subjects think about various stimuli. These stimuli can be an easily-distinguishing words such as

Categories	Exemplars
body parts	leg, arm, eye, foot, hand
furniture	chair, table, bed, desk, dresser
vehicles	car, airplane, train, truck, bicycle
animals	horse, dog, bear, cow, cat
kitchen utensils	glass, knife, bottle, cup, spoon
tools	chisel, hammer, screwdriver, pliers, saw
buildings	apartment, barn, house, church, igloo
part of building	window, door, chimney, closet, arch
clothing	coat, dress, shirt, skirt, pants
insects	fly, ant, bee, butterfly, beetle
vegetables	lettuce, tomato, carrot, corn, celery
man made objects	refrigerator, key, telephone, watch, bell
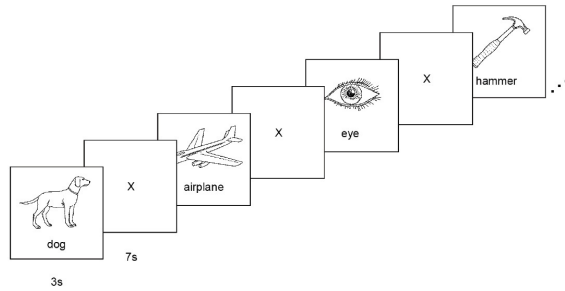
Figure 1.2: Schematic depiction of presentation timing
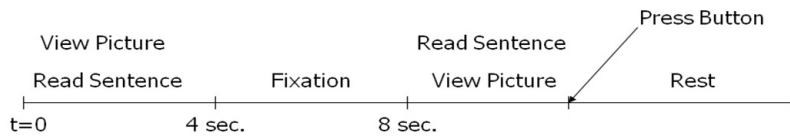
Figure 1.3: Collecting Starplus data

dog and *airplane*. Nine people participated in their project and the stimuli were line drawing of 60 concrete objects from 12 semantic categories with five samples in each category. All set of stimuli were presented six times to subjects in a different random order. Each stimuli presented 3s, followed by a 7s rest period (figure 1.2).

This data is similar to Starplus data [188] that used by researchers to study the cognitive states of subjects in a sentence-picture verification task [153]. In each trail, a person instructed to look at a pairs of sentence and picture and decides about matching of them. First stimulus, that can be a sentence or picture, presented to person, then after four second the stimulus replaced by a blank screen. Four second later on the next stimulus is presented and every 500msec images collected.

1.2.3 COBRE data

COBRE data set that is available on [184]. This data contains raw anatomical and functional MR data from 72 patients with Schizophrenia and 75 healthy controls, ranging in age from 18 to 65 years old. Resting fMRI, anatomical MRI, phenotypic data for every participant including: gender, age, handedness and diagnostic information are released. Some papers by the Center for Biomedical Research Excellence (COBRE) group published based on this data [33, 112, 169]

1.2.4 Oxford data

This data is from 1000 Functional Connectomes Project. It is available on [193] and it is used in this project to study the modularity in the human brain. It contains the data of 22 healthy brains. Figure 4.6 shows the structure of the input data of the subject 02248 after preprocessing and dimension reduction which contains 1008 brain regions with 166 brain activity. Last three columns are the 3 dimensional coordinates for each region.

1.2.5 Cerebral Cortex of the Cat

The Cerebral Cortex of the Cat data is based on the single neuron response to direct sensorimotor cortex stimulation in cats [73]. Neuronal responsiveness of each of the four paws to strong cortical surface stimulation was assessed for understanding facilitatory and inhibitory modulation of the wide-field neurons by small-field neurons. This data is publicly available on [187].

1.3 Algorithms and methods

Brains are very complex systems. Technologies for collecting brain data are still not perfect, that is why the data sets about brain and brains' activities are high-dimensional, noisy and sparse. Conventional machine learning algorithms like support vector machines and logistic regression are not efficient for analysis such data sets. The developments of new algorithms for selection of informative features in the brain data sets is very important. In this thesis, we introduce new feature selection algorithms for analysis of brain data sets. The overlapping feature selection algorithm and feature selection based on catastrophe model are helpful for selection group of the most informative features. It was proved by results of application of these methods to Haxby, Science 2008 and Parkinson's

disease data sets. The proposed algorithms are able to reduce significantly the number of features in Haxby and Science 2008 data sets. improving classification accuracy of most classifiers like support vector machine, Bayes networks, Logistic regression. In this chapter, we demonstrate some of the well-known supervised classification, clustering and complex networks algorithms and the results of application of the addressed classifiers for Haxby and Science 2008 data sets. Then we describe and recommend some publicly available packages and libraries analysis of large-scale brain data sets.

1.3.1 Dimension reduction

Principal component analysis: Principal component analysis (PCA) maps a set of data points to linearly uncorrelated variables called as principal components. Suppose a vector x with p random variables is given. Also α_k is a vector of p constants and $\alpha_k' = \sum_{j=1}^{p} \alpha_{kj} x_j$. PCA iteratively finds linear function of x, $\alpha_1' x$ with maximum variance, then finds another linear function of x, $\alpha_2' x$ that is uncorrelated with $\alpha_1' x$ maximum variance [196].

Independent component analysis: Independent component analysis decomposes a multivariate signal into independent non-gaussian signals. A given data $x_i(t)$ is modelled using hidden variables $s_i(t)$, as

$$x_i(t) = \sum_{j=1}^{m} a_{ij} s_j(t), i = 1...n \tag{1.1}$$

or

$$X = AS \tag{1.2}$$

where a_{ij} is called *mixing matrix* and $s_i(t)$ are called *independent components* or *source signals* [85].

1.3.2 Voxel (feature) selection

Generally, classifiers work better with less features, but choosing this number of features is not always easy (figure 1.6). Our knowledge of how to implement feature selection for fMRI data is still preliminary [75]. Generally, voxel selection algorithms can be categorized as follows [134].

1. Embedded (learner dependent)

   Support Vector Decomposition Machines (SVDMs) is one representative of such algorithms.

2. Filter (learner independent)

   For example it can be done by averaging over spatial/temporal dimensions or using multiple hypothesis testing for active voxels (e.g. t-test)

3. Wrapper (wraps around an induction algorithm)

As mentioned before, fMRI data sets are high dimensional with more than 20,000 voxels and several hundred samples. In such data sets feature selection is an important step to avoid an over-fitting of a classifier. One approach to feature selection is to limit the analysis to specific anatomical regions. For example, Haxby et. al. [75] in their study of visual object processing, restricted their data analysis to the ventral temporal cortex. Another feature selection approach is the computation of univariate (voxel-wise) statistics ([75], [35]). The challenge in this approach is the presence of very large space of voxel sets. Common fMRI feature selection methods include the following [139]:

- Activity

  This method scores a voxel by the difference in its mean activity level during each task condition versus a baseline condition, as measured by a t-test.

- Accuracy

  This method is based on accuracy of prediction in Gaussian Bayesian classifier for each sample and requires performing a cross-validation within the training set for each voxel.

- Searchlight Accuracy

  This method is similar to Accuracy, but instead of using the data from a single voxel it uses the data from the voxel and it's immediately adjacent neighbors in three dimensions [102].

- Analysis of Variance

  Analysis of Variance looks for voxels where there are reliable differences in mean value across conditions.
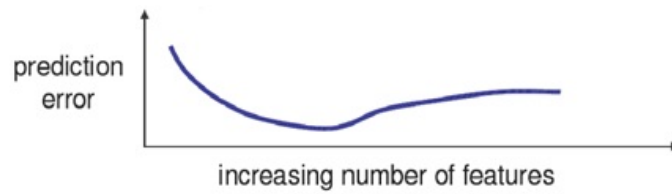
Figure 1.4: Prediction error of classifier vs. increasing number of features

Figure 1.5: Stimuli in Haxby data

- Stability

  Stability is based on consistently of voxels' reaction to the various conditions across cross-validation groups in the training set.

Figure 1.5 shows the results of classification of 8 categories experimental data (see Table 1.3.2). Here stimuli are photographs of objects in 8 categories (faces, houses, cats, bottles, scissors, shoes, chairs, scrambled) and classification task is prediction of the category.

Table 1.1: Accuracy of classification with selecting 200 voxels with the 4 methods and all voxels [138]

	GNB	Log.Reg.
activation	85%	88%
accuracy	86%	90%
searchlight accuracy	84%	88%
weight range	93%	92%
all cortex voxels	35%	43%

Table 1.2 shows the results of classification of ten categories experimental data. Here classification task is prediction of the exemplar. One can see that the Searchlight Accuracy algorithm produces better accuracy than other algorithms. The reason is that neighbor voxels have similar activity and classifier weights for these voxels have similar magnitude. Therefore the Searchlight Accuracy gives good results because it takes into account spatial locality of voxels.

13

High Bias
Low Variance

Low Bias
High Variance

Test Sample

Training Sample

Low

High

Prediction Error
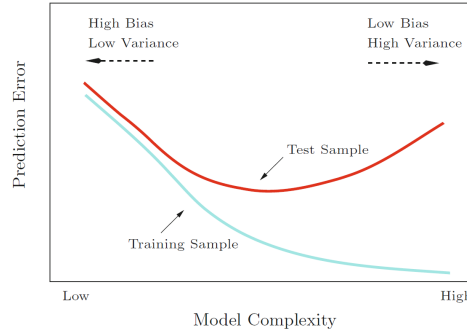
Model Complexity

Figure 1.6: Test and training error as a function of model complexity [74]

Table 1.2: Accuracy of classification with selecting 400 voxels with the 4 methods and all voxels [138]

	GNB	Log.Reg.
activation	70%	58%
accuracy	72%	70%
searchlight accuracy	90%	92%
weight range	72%	45%
all cortex voxels	23%	43%

1.3.3 Supervised classification algorithms

In this section we review the various Machine Learning classifiers that are used for analysis of fMRI data. Generally we prefer the classifiers that have reasonable tradeoff between training set and test set (see figure 1.6). Support Vector Machines (SVMs), Logistic Regression, Nearest Neighbor Method, Gaussian Naive Bayes are commonly used in fMRI data analysis. For different category of fMRI data different classifier can be used. For example, for analyzing the state-dependent data like Science data Gaussian Naive Bayes is more popular than others. The main reasons for choosing GNB for these data are: easily implementing and being robust to noise, having acceptable performance with a small number of data points without requiring prior dimensionality reduction step.

Linear classifiers: Linear classifiers apply linear functions to separate classes. A linear classifier with a set of weights $w = (w_1, \dots, w_v)$ applied to a sample x that has v features as follows:

$$xw = x_1 w_1 + \dots + x_v w_v \tag{1.3}$$

For two classes A and B,, this classifier predicts class A if $xw > 0$ or class B if $xw < 0$ [139].
Logistic Regression: Logistic Regression is a function approximation algorithm that uses training

14

data for direct estimation of $P(class_k|x)$ [116] and fits data to Logistic Function for prediction of an event [139]. It models $P(class_k|X)$ as

$$P(class_k|x) = \frac{exp(x^{'}w_k + w_{0,k})}{1 + \sum_{j=1}^{k-1} exp(x^{'}w_j + w_{0,j})} \tag{1.4}$$

for classes $1, ..., k-1$ and

$$P(class_k|x) = \frac{1}{1 + \sum_{j=1}^{k-1} exp(x^{'}w_j + w_{0,j})} \tag{1.5}$$

for class k. It fits parameters w_j by solving the following optimization problem:

$$max \sum_{i=1}^{n} log(P(y_i|X_i)) \tag{1.6}$$

where X_i is the ith sample and y_i is its label.

Linear Support Vector Machines: Linear Support Vector Machines is another discriminative classifier that learns a discriminant w by solving the following problem:

$$min \left\| w \right\|_2^2 + \lambda \sum_{i=1:n} h(y_i x_i^{'} w) \tag{1.7}$$

where h is loss function and $y_i \in \{-1, 1\}$. Minimizing the first term will maximize hyperplane margin and minimizing of second term will penalize misclassification of data.

Gaussian Naive Bayes: Gaussian Naive Bayes (GNB) is based on Bayes rule [133]:

$$\mathbb{P}(Y|X) \propto \mathbb{P}(X|Y)\mathbb{P}(Y) \tag{1.8}$$

where $X \in \Re^J$ is the example and $Y \in \{0, 1\}$ is the class label. The likelihood of the ith sample for a feature j using a normal Gaussian is:

$$X_{i,j}|Y = c \sim N(\theta_j^{(c)}, \sigma_j^2(c)) \quad i = 1, ..., N, \tag{1.9}$$

$$\mathbb{P}(X|Y = c) = \prod_{j=1}^{J} \mathbb{P}(X_j|Y = c). \tag{1.10}$$

The most probable value of Y can be calculated as follow:

$$Predicted - class = argmax_c \mathbb{P}(Y = c) \prod_{j=1}^{J} \mathbb{P}(X_j | Y = c) \tag{1.11}$$

$$= argmax_c \mathbb{P}(Y = c) \prod_{j=1}^{J} N(\hat{\theta}_j^{(c)}, \hat{\sigma}_j^{2(c)}) \tag{1.12}$$

$$\hat{\theta}_j^{(c)} = \frac{1}{\sum_{i=1}^{N} \delta(Y_i = c)} \sum_{i=1}^{N} \delta(Y_i = c) X_{ij} \tag{1.13}$$

$$\hat{\sigma}_j^{2(c)} = \frac{1}{\sum_{i=1}^{N} \delta(Y_i = c)} \sum_{i=1}^{N} \delta(Y_i = c)(X_{ij} - \hat{\theta}_j^{(c)})^2 \tag{1.14}$$

$$\mathbb{P}(Y = c) = \frac{1}{N} \sum_{i=1}^{N} \delta(Y_i = c) \tag{1.15}$$

where $\hat{\theta}_j^{(c)}$ and $\hat{\sigma}_j^{2(c)}$ are sample mean and variance for the feature j and class c and $\mathbb{P}(Y = c)$ is class frequency. Here $\delta(.)$ is the indicator function ($\delta : A \rightarrow \{0, 1\}$ where $\delta(x) = 1$ if $x \in A$ and $\delta(x) = 0$ otherwise).

Hierarchical Bayesian Model: The GNB can be trained using data from only one subject under assumption that there are no variations across subjects, but Hierarchical Bayesian Model supposes that the individual θ_j are related by some distribution, and if suppose that θ_j are all drawn from a common normal distribution, then we will have:

$$X_{ij} | \theta_j \sim N(\theta_j, \sigma^2) \tag{1.16}$$

$$\theta_j \sim N(\mu, \tau^2) \tag{1.17}$$

μ and τ^2 are called $hyperparameters$ for the model and θ_j are related by some distribution. In this mode, the aim is to find the best estimate of θ_j not only by given data, but also using prior information about its distribution ($\theta_j | \mu, \tau^2, X$) and if $\forall j, \sigma_j^2 = \sigma^2$ then:

$$\hat{\theta}_j = \frac{\frac{N}{\sigma^2} \bar{X}_{\cdot j} + \frac{1}{\tau^2} \mu}{\frac{N}{\sigma^2} + \frac{1}{\tau^2}} \tag{1.18}$$
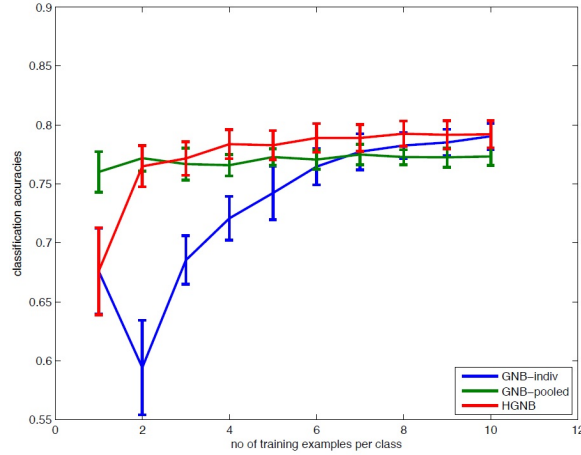
Figure 1.7: Classification accuracies VS the number of training samples for Starplus dataset [157]

$$\hat{\mu} = \frac{1}{J} \sum_{j=1}^{J} \bar{X}_{\cdot j} \qquad (1.19)$$

$$\hat{\tau}^2 = \frac{1}{J} \sum_{j=1}^{J} (\bar{X}_{\cdot j} - \hat{\mu})^2 \qquad (1.20)$$

In this model, variance σ^2 must be estimated from the data. Rustandi et. al. [157] applied the hierarchical GNB classifier to Starplus and Two categories data sets. Two categories dataset consists of fMRI activations of 6 subjects, where in each trial, each subject looked at a word that belong to one of two categories, and had to think about the properties of the word. Figure 1.7 shows the classification accuracies vs number of training samples for the hierarchical GNB and the two reference methods (GNB-indiv and GNB-pooled) for Starplus data set. Figure 1.8 illustrates results of classification accuracies versus number of training samples for the hierarchical GNB and the two reference methods for Two categories dataset.

Feature Sharing Classifier: An assumption that all the parameters θ_j are drawn from the same distribution is one disadvantage of standard hierarchical Bayes model. Palatucci et. al. [133] developed an algorithm based on HGNB and improved its accuracy. Consider two variables that are perfectly correlated while the parameters distributions are significantly different. If we suppose parameters for these two variables are drawn from a common normal distribution, the estimation of μ, τ and θ_j will not be reliable. As mentioned before, fMRI is a three dimensional image of brain activity, so each voxel has a time series for activation (Figure 1.9).

17
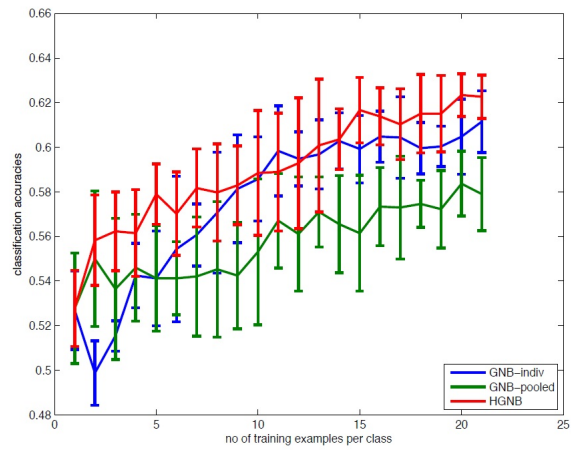
Figure 1.8: Classification accuracies vs number of training samples for Two categories dataset [157]
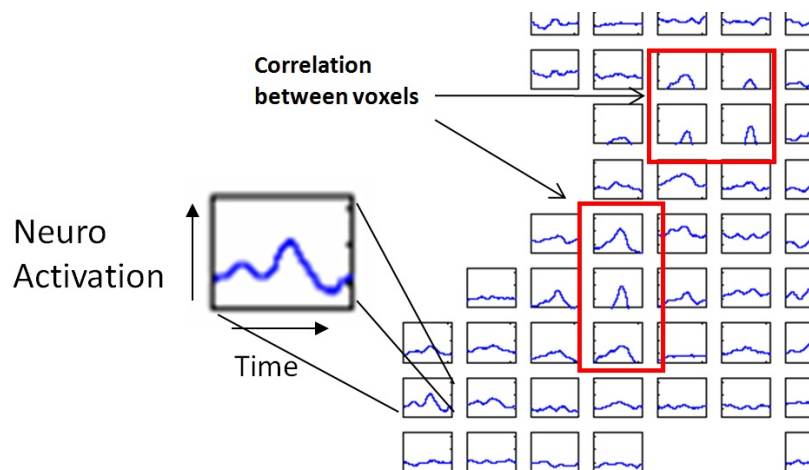
Figure 1.9: Time series of neural activation in visual cortex [133]

Niculescu et. al. [125] showed that there there is a strong correlation between most of voxels. Palatucci et al. [133] used linear mapping to estimate features for a single voxel. The Feature Sharing Classifier is based on the Hierarchical Bayesian Model but unlike the latter the parameters can be drawn from different distributions. Assume we have two random variables, X and Y , parameterized by θ_X and θ_Y. The $m_{X \to Y}(\theta_X)$ be of a parameter transformation function that maps parameters of variable X to those of variable Y. The set of all other variables is denoted by $\mathbb{C}_j$. Let $G_j = |\mathbb{C}_j|$ be the number of variables in that set.

$$\hat{\theta}_j = \frac{\frac{N}{\sigma^2}\bar{X}_{\cdot j} + \frac{1}{\tau^2}\hat{\mu}}{\frac{N}{\hat{\sigma}^2} + \frac{1}{\hat{\tau}^2}} \tag{1.21}$$

$$\hat{\mu}_j = \frac{1}{G_j} \sum_{g=1}^{G_j} m_{g \to j}(\bar{X}_{\cdot g}) \tag{1.22}$$

$$\hat{\tau}_j^2 = \frac{1}{G_j} \sum_{g=1}^{G_j} (m_{g \to j}(\bar{X}_{\cdot g}) - \hat{\mu}_j)^2 \tag{1.23}$$

$$\hat{\sigma}_j^2 = \frac{1}{G_j} \sum_{g=1}^{G_j} m'_{g \to j}(S_g^2), \tag{1.24}$$

where S_g^2 is sample variance for feature g. Now suppose we want to estimate parameters for a single feature for class k by the feature sharing classifier. The following steps provide calculation of these parameters:

1. For each voxel, compute the sample mean over N training samples;

2. Perform a linear regression with each neighbor for each voxel;

3. For each voxel-timepoint, compute estimates of the hyperparameters;

4. Smooth sample mean with these hyperparameters.

Support Vector Decomposition Machines: Support Vector Decomposition Machines (SVDM) is another classifier that developed by Pereira et. al. [137]. They combined the goals of dimensionality reduction and classification into a single objective function, and presented an efficient alternating-minimization algorithm for optimizing this objective. For a learning problem, suppose there are
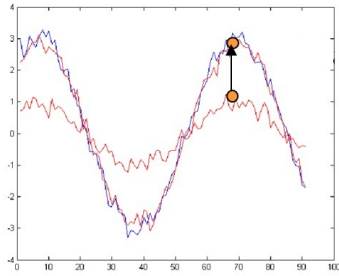
Figure 1.10: Transforming features into space of the other before computing hyperparameters [133]
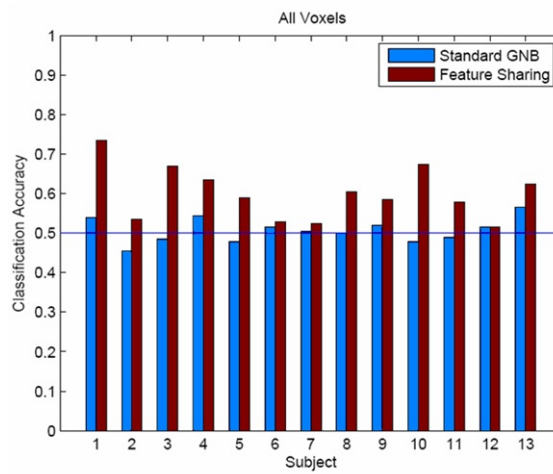
Figure 1.11: Accuracies of the standard Gaussian Naive Bayes classifier and the Feature Sharing classifier for 13 human subjects with two training samples per class [133].
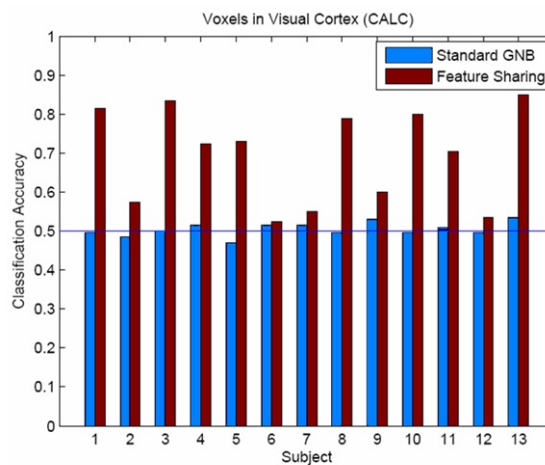
Figure 1.12: Accuracies of the standard Gaussian Naive Bayes classifier and the Feature Sharing classifier using only voxels in the Visual Cortex [133].

n data samples $x_i \in R^m$ with labels $y_i \in \{0, 1\}$. The goal of the singular value decomposition algorithm is to find a representation of training data as a product of lower-rank matrices.

$$X_{n \times m} = \begin{pmatrix} x_1(1) & x_1(2) & ... & x_1(m) \\ x_2(1) & x_2(2) & ... & x_2(m) \\ ... & ... & ... & ... \\ x_n(1) & x_n(2) & ... & x_n(m) \end{pmatrix} \qquad (1.25)$$

$$X_{n \times m} \approx Z_{n \times l} W_{l \times m} \qquad (1.26)$$

$$min_{Z,W} \|X - ZW\|_{Fro}^2 \qquad (1.27)$$

Pereira et. al. [137] formulated problem as follows:

$$Min_{Z,W,Q} \|X - ZW\|_F^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{k} max(0, \mu - Y_{ij}[ZQ]_{ij}) \qquad (1.28)$$

subject to: $Z_{i,1} = 1$, $Z_{i,2:end} <= 1, i = 1, 2, ..., n$ and $\|Q_{:,j}\|^2 <= 1, i = 1, 2, ..., l; j = 1, 2, ..., k$ ($\lambda > 0, u > 0, l \in Z^{++}$).

Here $X_{n \times m} \in R^{n \times m}, Y_{n \times m} \in \{1, -1\}^{n \times m}, Z \in R^{n \times l}, W \in R^{l \times m}, Q \in R^{l \times k}$

Information captured by classifiers Next we discuss application of some of above described classifiers in brain data sets.

Distinguishing semantic categories of stimuli by classifiers.

Now we would like to understand which kind of information is captured by classifiers. When one classifier can classify brain activities and distinguish them, we can claim that it could get a meaning of stimuli. But this is not enough to decode processes of brain, because fMRI can get combination of all processes. Mitchell et al. [118] developed a model for prediction of brain activity when subjects think about concrete nouns. First question of Mitchell et al. was distinguishing differences between brain images when people think about different stimuli. Figure 1.13 shows classification accuracy for subjects thinking about *tool* or *building*. Each bar is related to classification accuracy of one subject. This result illustrates that we can train classifier to distinguish semantic categories of stimuli.

To answer the question about possibility of distinguishing semantic meanings of stimuli by classi-
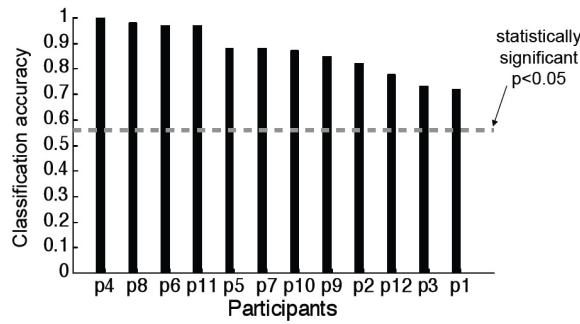
Figure 1.13: Classification accuracy for subjects which think about *tool* or *building*
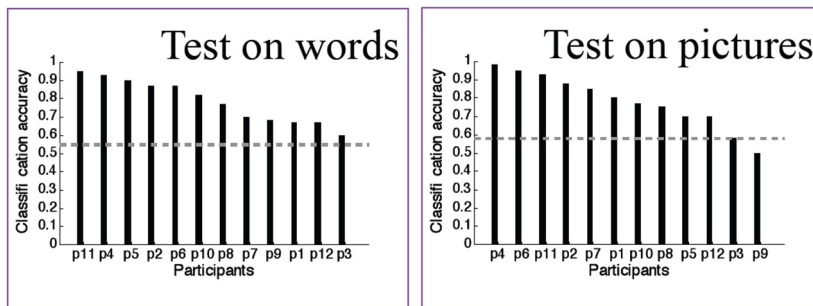
Figure 1.14: Training classifiers when presenting words, then decode for category of picture stimuli

fiers, we can train a classifier when subject thinks about pictures and later use this classifier to decode fMRI activity for words. Figure 1.14 shows classification accuracy when presenting English words to decode category of words stimuli and also presenting words, to decode category of picture stimuli [162]. These results are almost the same if instead of English words, Portuguese words are used to train classifiers [161]. It means that classifier can capture some semantic structure of stimuli.

Brain representations similarity across people

One of the challenging question in neuroscience area is about similarity of neural activation between different subjects when they think about the ideas. To clarify this, we can train classifier on one group of subjects and test it on another group. Figure 1.15 shows the results of training of classifiers on a group of people, and using it for new person. Here black bars are classification accuracy for one group of participants and white bars show classification accuracy for the new one. Again we can conclude that neural representations on different subjects are almost the same.
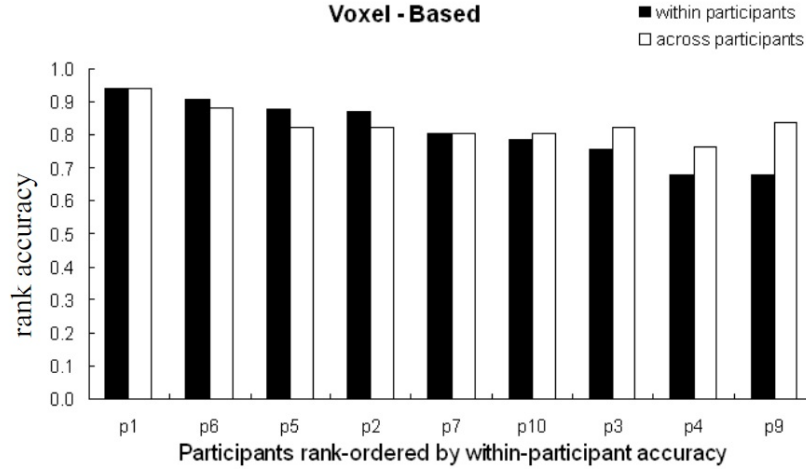
Figure 1.15: Training of classifiers on a group of people, and using it for new person

1.3.4 Regression analysis

In this subsection, we discuss about regression analysis and it's application to analyze the brain data.

The aim of regression analysis is finding the relation between variables and fitting a function to a data to see how the one group of variables vary as a function of another group [186]. There are three groups of variables that we can consider in regression analysis; unknown parameters (β), independent variables (X) and dependent variable (Y). It means that Y is related to X and β and a regression model can be described as follows:

$$Y \approx f(X, \beta). \tag{1.29}$$

There are different regression analysis algorithms. Here we describe some of them that are more common in the different applications.

Logistic regression: The general form of linear equation that is used in linear regression analysis is as follow [205]:

$$y = b_0 + b_1^T x \tag{1.30}$$

where b_0 and vector $b_1 \in R^n$ are constants, x is the independent variable, and y is the dependent variable.

23

Multiple regression: Generally, the approximation 1.29 is explained as $E(Y|X) = f(X, \beta)$ that can be linear or non-linear. A multiple regression model can be formulated as follow:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \tag{1.31}$$

where x_{ij} is the ith observation on the jth independent variable, β_i is the regression intercept and ε_i is the residual.

Neural networks for regression analysis: Neural networks is a common method for solving different machine learning problems. As a supervised learning model, it applies weight vector w and minimizes the sum of square residuals. Searching the maximum likelihood weights is based on gradient descent approach. The weight updating in a neural networks is as follow:

$$\Delta w_{ij}(t+1) = \gamma \Delta w_{ij}(t) + \alpha \delta_i(t) x_j(t) \tag{1.32}$$

where $\Delta w_{ij}(t)$ is the weight value difference between nod i and j at time t, $\alpha \in (0, 1]$ is the learning rate and $\gamma \in (0, 1]$ is the momentum parameter [13, 26].

Application of regression analysis algorithms in brain data analysis: Different regression analysis algorithms have been used for analysing various brain data. One of the famous problem in neuroscience area is epileptic seizures analysing. Generally, epileptic data is based on electroencephalography (EEG) technology that provides appropriate temporal resolution of brain signals [126] and regression analysis is an effective method for analyzing this type of brain data. Subasi et al. analyzed EEG signals using wavelet transform and classification by logistic regression and artificial neural network to support a physician in the diagnosing process [173]. Tomioka et al. [179] regularized logistic regression used for classification of single trial electroencephalography. It's shown that using regression analysis can improve the sensitivity of predictor and classify patients correctly with respect to epilepsy diagnosis [49]. In [121] an association between precentral gyrus connectivity structure and autism spectrum disorders is investigated using logistic regression and showed the strength of connectivity within and between distinct functional subregions of the precentral gyrus which related to autism spectrum disorders diagnosis.

Recently, researchers used regression analysis algorithms for decoding and uncovering the information in the brain [1, 119]. Miyawaki et al. [1] used logistic regression in a model for reconstructed

visual images and decoding fMRI activity.

1.3.5 Neurofuzzy model

This model is based on artificial neural networks and fuzzy logic by using capabilities of both models [180]. In the other words it applies neural networks structure and at the same time uses *if-then* rules in fuzzy logic. Also, it uses prior knowledge to compute membership function. Different algorithms, including the backpropagation algorithm, can be used to learn such neural networks.

Adaptive neuro fuzzy inference system (ANFIS) is a well-known neurofuzzy system that implement a Sugeno fuzzy system and uses a t-norm and differentiable membership function [86, 177].

For given two inputs x_0 and y_0 and corresponding linguistic labels A_i and B_i, each neuron in the first layer of neurofuzzy model transmit crisp signal to the next layer in accordance with the following equations:

$$A_i(u) = exp\left[-\frac{1}{2}\left(\frac{u - a_{i1}}{b_{i1}}\right)^2\right]$$

and

$$B_i(u) = exp\left[-\frac{1}{2}\left(\frac{u - a_{i2}}{b_{i2}}\right)^2\right],$$

where $\{a_{i1}, a_{i2}, b_{i1}, b_{i2}\}$ is the parameters set. Second layer is responsible for fuzzification and each neuron in this layer determines the fuzzy degree of received crisp input:

$$\alpha_1 = A_1(x_0) \times B_1(y_0) = A_1(x_0) \wedge B_1(y_0)$$

and

$$\alpha_2 = A_2(x_0) \times B_2(y_0) = A_2(x_0) \wedge B_2(y_0).$$

Every node in the third layer calculates the ratio of the ith rules firing strength to the sum of all rules firing strengths:

$$\beta_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$

and

$$\beta_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2}.$$

Layer 4 or output membership layer combine all its inputs by using the fuzzy operation union:

$$\beta_1 z_1 = \beta_1 (a_1 x_0 + b_1 y_0)$$

and

$$\beta_2 z_2 = \beta_2 (a_2 x_0 + b_2 y_0).$$

The last layer is responsible for Defuzzification:

$$o = \beta_1 z_1 + \beta_2 z_2.$$

1.3.6   Complex networks and brain data analysis

Brain can be considered as a complex network for information transfer through interconnected regions [166, 167]. Although, complex networks' structure can be considered as a simple graph consisting of nodes and the edges between them, the measures that are used for analyzing such networks are not simple and trivial. On the other hand, there are some measures that make the complex networks different from simple graphs, random graphs and lattices [6]. Scale-free networks [17] and small-world networks [204] are two classes of complex networks that are widely used in the different applications. Small-world networks have a high level of clustering and a short average node-to-node distance [204]. Scale-freeness means that in the networks most of the nodes have only a limited number of connections, but a small number of so called hub-nodes have a large number of connections which are holding the network together [16].

Anatomical structure of a healthy brain has a hierarchical organization characterized by very low clustering of high-degree nodes [22]. This feature can be used for comparison of the structure of healthy brain and a brain with abnormality or disorder. For example local-distance is increased in cortical-cortical activity in autistic children [82]. A graph theoretical analysis was implemented for investigation of modularity and local connectivity in childhood-onset schizophrenia in [10]. The authors examined the network that obtained from fMRI data and showed that local connectivity and modularity is disrupted in childhood-onset schizophrenia.

1.4 General machine learning software for analyzing the brain data

There are some publicly available packages and libraries that are used for brain data analysis. In this section we describe some of them.

Analysis of Functional NeuroImages: Analysis of Functional NeuroImages (AFNI) is a package of computer programs for analysis and visualization of three-dimensional human brain functional magnetic resonance imaging (FMRI) which was developed at the Medical College of Wisconsin beginning in 1994. It can color overlay neural activation maps onto higher resolution anatomical scans [42]. It is available on [182].

FMRIB Software Library: FMRIB Software Library (FSL) is a general library of analysis tools for functional, structural and diffusion MRI brain imaging data [87] (see Figure 1.16). It contains many tools for Functional and structural MRI, Diffusion MRI and GLM that are listed in the table 1.4. The FEAT, MELODIC, FLOBS and SMM packages are for analyzing of functional MRI, but BET, SUSAN, FAST, FLIRT, FUGUE and SIENA packages are for structural MRI. Also, FDT and TBSS are for low-level diffusion parameter reconstruction and probabilistic tractography and Tract-Based Spatial Statistics of Diffusion MRI. This software can be downloaded from [185].

Table 1.3: FSL tools

Functional MRI	Structural MRI	Diffusion MRI: FDT	GLM / Stats: GLM general advice	Other: FSLView
FEAT	BET	TBSS	Randomise	Fslutils
MELODIC	FAST	EDDY	Cluster	Atlases
FABBER	FIRST	TOPUP	FDR	Atlasquery
BASIL	FLIRT & FNIRT	Dual Regression	SUSAN	FUGUE
	FSLVBM	FLOBS	Mm	Miscvis
	SIENA & SIENAX		MCFLIRT	POSSUM
	fsl_anat			

Figure 1.16: The FSL logo [87]

Statistical parametric mapping: Statistical parametric mapping (SPM) is a software for the statistical analysis of functional imaging data [59]. Using statistical techniques, it examines differences in brain activity recorded during functional neuroimaging experiments. It is available on [190].

MultiVariate Pattern Analysis (MVPA) in Python (PyMVPA): MultiVariate Pattern Analysis in Python (PyMVPA) is a Python package for statistical learning analyses of large data sets [70]. It provides a framework for data import and export and different machine learning algorithms like classification, regression, feature selection. It is powerful software for neuroimaging domain and is eminently suited for such data sets. PyMVPA is free software and available on [194].

Figure 1.17: The PyMVPA logo

The R Project for Statistical Computing: R is an environment for statistical computing and graphics that contains a comprehensive libraries of machine learning and statistical analysis applications (see Figure 1.18). It's a free software and available on [195]. It contains different packages like AnalyzeFMRI, fmri and dcemriS4 for analysis brain data [108, 176, 206].

Figure 1.18: The R logo

MATLAB (matrix laboratory): MATLAB is an interactive environment for numerical computation, visualization, and programming that developed by MathWorks [191]. It provides varieties of toolboxes for signal processing, image processing, control systems, test and measurement, and computational biology [191]. Different toolboxes in MATLAB environment like CONN [207], cPPI [53], and Brain Connectivity Toolbox [156] are available for analyzing brain data.
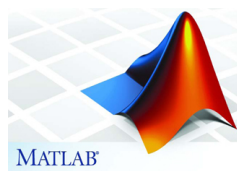
Figure 1.19: The MATLAB logo [191]

Waikato Environment for Knowledge Analysis (WEKA): WEKA - Waikato Environment for Knowledge Analysis - is a collection of machine learning algorithms for data mining tasks. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization

[110]. A WEKA Interface to solve classification and prediction problems for fMRI Data is introduced in [146]. It is publicly available on [189].
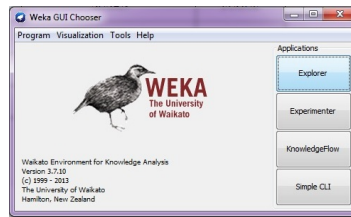
Figure 1.20: The WEKA logo

1.5 Conclusion

In this chapter, we presented the literature review on related work. First we described different brain imaging technologies and presented the data sets that analysed in this thesis. Then we briefly demonstrated some of commonly used algorithms for dimension reduction, feature selection, supervised classifiers, regression analysis, Neuro-fuzzy and complex networks. We showed results in using Haxby and Science 2008 data sets obtained using classifiers like Support Vector Machines, the Logistic Regression and the Gaussian Naive Bayes. Different software and packages based on the famous algorithm are developed and we introduced some of them like ANFI, FSL, SPM, PyMVPA and the environment like MATLAB, R and WEKA.

Chapter 2

Feature selection algorithms for brain data sets

In this chapter we introduce feature selection algorithms for brain data sets. Overlapping feature (voxel) selection algorithm uses the n-dimensional rectangles for approximation of classes in the given subset of voxels and computes overlaps between classes. Voxels or groups of voxels providing smallest overlaps are identified as the most informative voxels.

The second algorithm based on Catastrophe model and it removes the irrelevant features in a data set. The importance of a feature is based on its fitting to the Catastrophe model. Breast Cancer and Parkinson Telemonitoring data sets used to evaluate the model.

2.1 Feature selection algorithm for fMRI data sets using hyperrectangles

Functional magnetic resonance imaging (fMRI) technique measures brain activity by detection fluctuating of oxygenation that occurs in response to neural activity in blood [81]. It is a technique for obtaining three dimensional images related to activity in the brain through time. FMRI scanners can measure the changes in the blood magnetic resonance that called the Blood Oxygen Level-Dependent (BOLD) signal. The smallest unit measured by BOLD fMRI, is called a voxel. FMRI's spatial resolution is around 2mm and it is a powerful technique in studies of cognitive processes in healthy brain and it's contrast can be performed on more than 100,000 voxels [127].

The development of machine learning techniques for fMRI data sets is an emerging research area in neuroscience. Over the last decade various machine learning techniques have been applied to study fMRI data sets. In general, fMRI data sets contain a large number of features (tens of thousands) and significantly less samples (several hundred). Therefore such data sets are sparse. Furthermore such data sets contain noise. Conventional supervised classification algorithms are not always applicable to such data sets. Therefore the developments of algorithms for finding informative features in such data sets is very important. Recently, some well known feature selection algorithms have been modified for fMRI data sets [25, 45, 103, 117, 130, 140, 161]. Limiting the analysis to specific anatomical regions [75], univariate feature selection [75, 117, 144] and multivariate feature selection [102, 140] are different ways of selecting informative features (voxels) in fMRI data.

One way to select the set of informative features from the data is to use multivariate feature selection statistics to reduce the complexity of the whole brain data analysis [127]. For example, F-test or t-test to find the most active voxels or the most discriminant voxels and considering them as the set of informative voxels are two usual methods for feature selection in fMRI data sets [45, 117, 120]. The main challenge in the univariate feature selection is the possibility of losing some informative features [127]. In [45, 68, 99], the multivariate feature selection algorithms were applied to find significant features in the whole brain data. For example, the searchlight approach that developed in [102] uses local spatial information of a voxel to select features. Dimension reduction is another efficient method to select informative features that many researchers applied to the voxels in different regions of the brain [103, 120, 132].

In this section we introduce an algorithm for voxel selection in fMRI data sets. This algorithm uses hyperrectangles to approximate classes in such data sets and overlaps between different hyperrectangles. A voxel or a group of voxels providing least overlaps between classes is considered as a most informative voxels. We apply the proposed algorithm to the well-known Haxby data set. The number of voxels in the available version of this data set has been already reduced by another feature selection algorithm [71, 75]. However, the proposed algorithm is able to significantly reduce the number of voxels in this data set improving classification accuracy of most classifiers.

The rest of the section is organized as follows. In Section 2.1.1 we introduce hyperrectangles to approximate classes and compute overlaps between them. We describe the voxel selection algorithm in Section 2.1.4. Computational results and their discussions are given in Section 2.1.5. Section 2.3

concludes the section.

2.1.1 Definition of overlaps

In this section we define one-dimensional and multi-dimensional overlaps between different stimuli using activity levels of voxels. Overlaps can be defined between two classes as well as between a given class and the rest of a data set. We start with the definition of overlaps between two classes.

2.1.1.1 Binary univariate overlaps.

Suppose we are given a data set A which contains $m \geq 2$ classes, n_i samples in the i-th class and p voxels. We denote by d_{kj}^i the j-th brain activity value for the k-th sample in the i-th class, where $i = 1, \ldots, m, \; j = 1, \ldots, p, \; k = 1, \ldots, n_i$. This is illustrated in Figure 2.1. Then we introduce the following numbers:

$$a_{ij}^{min} = \min_{k=1,\ldots,n_i} d_{kj}^i, \;\; a_{ij}^{max} = \max_{k=1,\ldots,n_i} d_{kj}^i,$$

$$j = 1, \ldots, p, \; i = 1, \ldots, m.$$

Here a_{ij}^{min} and a_{ij}^{max} are the minimum and maximum activation value for the j-th voxel in the i-th class, respectively. Then the j-th voxel in the i-th class can be identified by a segment $[a_{ij}^{min}, a_{ij}^{max}]$. We call this segment the activation level segment of the j-th voxel in the i-th class. For a given voxel $j = 1, \ldots, p$ and two different classes i and l we define:

$$c_{1j}(i, l) = \max(a_{ij}^{min}, a_{lj}^{min}), \;\; c_{2j}(i, l) = \min(a_{ij}^{max}, a_{lj}^{max}),$$

$$e_{1j}(i, l) = \min(a_{ij}^{min}, a_{lj}^{min}), \;\; e_{2j}(i, l) = \max(a_{ij}^{max}, a_{lj}^{max}).$$

It is clear that the interval $[e_{1j}(i, l), e_{2j}(i, l)]$ contains activation levels of the j-th voxel of all samples from classes i and l and the interval $[c_{1j}(i, l), c_{2j}(i, l)]$, if is not empty, contains samples from both classes. Overlaps for the j-th voxel between these two classes can be defined either using the length of both intervals or the number of samples whose activation level of the j-th voxel are in these intervals.
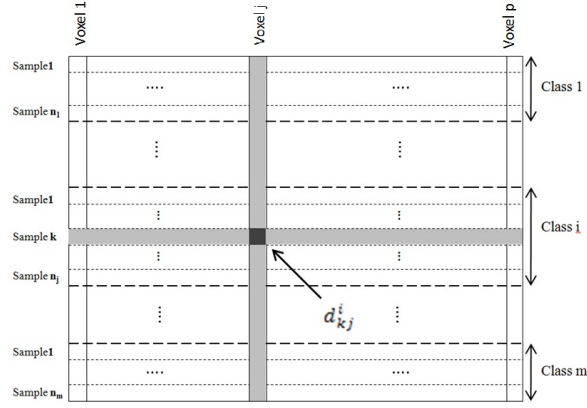
Figure 2.1: Brain activity value

2.1.1.2 The use of the length of intervals.

Consider

$$b_1(i,l) = \max\left\{0, c_2(i,l) - c_1(i,l)\right\}, \; b_2(i,l) = e_2(i,l) - e_1(i,l).$$

One can note that $b_1(i,l) = 0$ if the activation level segment of the j-th voxel in classes i and l either has no intersection or their endpoints coincide. Always $b_2(i,l) \geq 0$ and $b_2(i,l) = 0$ if and only if $a_{ij}^{max} = a_{lj}^{max} = a_{ij}^{min} = a_{lj}^{min}$. Consider the following number

$$z = \left(a_{lj}^{min} - a_{ij}^{min}\right)\left(a_{ij}^{max} - a_{lj}^{max}\right).$$

If $z \geq 0$ then either

$$[a_{lj}^{min}, a_{lj}^{max}] \subseteq [a_{ij}^{min}, a_{ij}^{max}]$$

or

$$[a_{ij}^{min}, a_{ij}^{max}] \subseteq [a_{lj}^{min}, a_{lj}^{max}].$$

In particular, if $b_2(i,l) = 0$ then $z = 0$.

The number

$$O_{il}^j = \begin{cases} 1, & z \geq 0, \\ \frac{b_1(i,l)}{b_2(i,l)}, & \text{otherwise.} \end{cases}$$

is said to be *the overlap of the j-th voxel between classes i and l*. Note that this type of overlaps were

also considered in [15].

2.1.1.3 The use of the number of samples.

Overlaps can also be defined using the number of samples in the interval $[c_{1j}(i,l), c_{2j}(i,l)]$. Consider the following sets:

$$Q_t = \left\{ k = 1, \ldots, n_t : \ c_{1j}(i,l) \le d_{kj}^t \le c_{2j}(i,l) \right\}, \ t = i, l.$$

Let $q = |Q_i| + |Q_l|$ where $|Q_t|$ is the cardinality of the set Q_t, $t = i, l$. Then the number

$$O_{il}^j = \begin{cases} 1, & z \ge 0, \\ \frac{q}{n_i + n_l}, & \text{otherwise.} \end{cases}$$

is said to be *the overlap of j-th voxel between classes i and l*.

It is clear that $O_{il}^j = O_{li}^j$, $O_{il}^j \in [0,1]$ and $O_{ii}^j = 1$ for any $j = 1, \ldots, p$ and $i, l = 1, \ldots, m$. Thus, we can define the following $m \times m$ matrix for the voxel j:

$$O^j = \begin{pmatrix} 1 & O_{12}^j & O_{13}^j & \ldots & O_{1m}^j \\ O_{21}^j & 1 & O_{23}^j & \ldots & O_{2m}^j \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ O_{m1}^j & O_{m2}^j & O_{m3}^j & \ldots & 1 \end{pmatrix}.$$

O^j is a symmetric matrix.

2.1.2 One-Vs-All univariate overlaps

For a given class $i \in \{1, \ldots, m\}$ and voxel $j \in \{1, \ldots, p\}$ we define

$$\bar{a}_{ij}^{min} = \min_{l=1,\ldots,m, l \ne i} a_{lj}^{min}, \ \bar{a}_{ij}^{max} = \max_{l=1,\ldots,m, l \ne i} a_{lj}^{max},$$

$$\bar{c}_{1j}(i) = \max \left(a_{ij}^{min}, \bar{a}_{ij}^{min} \right), \ \bar{c}_{2j}(i) = \min \left(a_{ij}^{max}, \bar{a}_{ij}^{max} \right),$$

$$\bar{e}_{1j}(i) = \min \left(a_{ij}^{min}, \bar{a}_{ij}^{min} \right), \ \bar{e}_{2j}(i) = \max \left(a_{ij}^{max}, \bar{a}_{ij}^{max} \right),$$

$$\bar{b}_1(i) = \max\left\{0, \bar{c}_2(i) - \bar{c}_1(i)\right\}, \quad \bar{b}_2(i) = \bar{e}_2(i) - \bar{e}_1(i),$$

$$\bar{z} = \left(\bar{a}_{ij}^{min} - a_{ij}^{min}\right)\left(a_{ij}^{max} - \bar{a}_{ij}^{max}\right).$$

$$Q_{ti}^0 = \left\{k = 1, \ldots, n_t : \ \bar{c}_{1j}(i) \le d_{kj}^t \le \bar{c}_{2j}(i)\right\}, \ t = 1, \ldots, m,$$

$$\bar{Q}_i = \bigcup_{t=1}^{m} Q_{ti}^0, \ \ \bar{q} = |\bar{Q}_i|, \ \ n = \sum_{i=1}^{m} n_i.$$

We can define the overlap between the class i and the rest of the data set by

$$\bar{O}_i^j = \begin{cases} 1, & \bar{z} \ge 0, \\ \frac{\bar{b}_1(i)}{\bar{b}_2(i)}, & \text{otherwise.} \end{cases}$$

or by

$$\bar{O}_i^j = \begin{cases} 1, & z \ge 0, \\ \frac{\bar{q}}{n}, & \text{otherwise.} \end{cases}$$

Then we can define a vector of overlaps for a given voxel j as follows:

$$\bar{O}^j = \left(\bar{O}_1^j, \ldots, \bar{O}_m^j\right).$$

2.1.3 Multi-dimensional overlaps

A hyperrectangle $B = [a, b]$, $a, b \in R^n$ in n-dimensional space R^n is defined as follows:

$$B = \left\{x \in R^n : \ a_i \le x_i \le b_i, \ i = 1, \ldots, n\right\}. \tag{2.1}$$

Assume that we are given two hyperrectangles $B_1 = [a^1, b^1]$ and $B_2 = [a^2, b^2]$. Their intersection is empty if and only if there exists at least one $i \in \{1, \ldots, n\}$ such that either $b_i^1 < a_i^2$ or $b_i^2 < a_i^1$. In other words the intersection of B_1 and B_2 is empty if and only if

$$\max_{i=1,\ldots,n} \max\left\{a_i^2 - b_i^1, a_i^1 - b_i^2\right\} > 0.$$

This means that B_1 and B_2 have an intersection if and only if:

$$\max_{i=1,\ldots,n} \max\left\{a_i^2 - b_i^1, a_i^1 - b_i^2\right\} \le 0.$$

Then we get that $a_i^1 \leq b_i^2$ and $a_i^2 \leq b_i^1$ for all $i \in \{1, \ldots, n\}$ which implies that $\max\{a_i^1, a_i^2\} \leq \min\{b_i^1, b_i^2\}$ for all $i \in \{1, \ldots, n\}$. Two hyperrectangles do not intersect if and only if $\max\{a_i^1, a_i^2\} > \min\{b_i^1, b_i^2\}$ at least for one $i \in \{1, \ldots, n\}$.

The intersection of two hyperrectangles B_1 and B_2 is also hyperrectangle and it can be described as follows:

$$B_{12} = [\alpha, \beta], \ \alpha, \beta \in R^n$$

where $\alpha_i = \max\{a_i^1, a_i^2\}$ and $\beta_i = \min\{b_i^1, b_i^2\}$, $i = 1, \ldots, n$.

2.1.3.1 Binary multi-dimensional overlaps.

First we define the multi-dimensional overlaps between two classes i and l, $i, l \in \{1, \ldots, m\}$. Let $J = \{j_1, \ldots, j_n\} \subset \{1, \ldots, p\}$, $0 < n \leq p$ be a subset of voxels. Then the group of voxels J in the class t can be identified by the following n-dimensional hyperrectangles:

$$B_t^J = [x^t, y^t], \ x^t, y^t \in R^n, \ x_k^t = a_{tj_k}^{min}, \ y_k^t = a_{tj_k}^{max},$$

$$k = 1, \ldots, n, t = i, l.$$

Let $B_{il}^J = B_i^J \bigcap B_l^J$. We define the multi-dimensional overlaps using the number of samples in hyperrectangles B_{il}^J. Consider the set

$$Q_t^J = \{k = 1, \ldots, n_i : \ u^k = (d_{kj_1}^t, \ldots, d_{kj_n}^t) \in B_{il}^J\}.$$

Let $q = |Q_i^J| + |Q_l^J|$. Then

$$O_{il}^J = \begin{cases} 1, & B_i^J \subseteq B_l^J \text{ or } B_l^J \subseteq B_i^J, \\ \frac{q}{n_i + n_l}, & \text{otherwise.} \end{cases}$$

It is again clear that $O_{il}^J = O_{li}^J$, $O_{il}^J \in [0,1]$ and $O_{ii}^J = 1$ for any $J \subset \{1, \ldots, p\}$, $J \neq \emptyset$ and $i, l = 1, \ldots, m$. Thus, we can define the following $m \times m$ matrix for the subset of voxels J:

$$O^J = \begin{pmatrix} 1 & O_{12}^J & O_{13}^J & \ldots & O_{1m}^J \\ O_{21}^J & 1 & O_{23}^j & \ldots & O_{2m}^J \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ O_{m1}^J & O_{m2}^J & O_{m3}^J & \ldots & 1 \end{pmatrix}.$$

O^J is a symmetric matrix.

One example of two-dimensional binary overlaps for voxels 1 and 2 is given in Figure 2.2. In this figure two classes are illustrated by "+" and "∘", respectively. These classes are approximated by (hyper)rectangles shown by dash lines. Overlap of these two rectangles is shaded area.
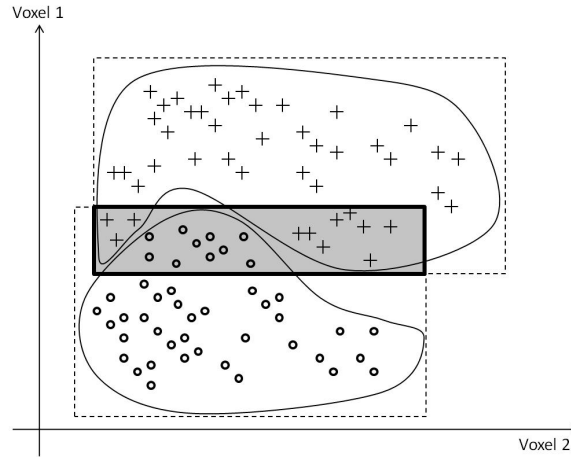
Figure 2.2: Binary two dimensional overlaps

2.1.3.2 One-Vs-All multi-dimensional overlaps.

We can define overlaps between a given class $i \in \{1, \ldots, m\}$ and the rest of the data set for a subset of voxels J in a similar way as in the case of univariate overlaps. First we define the following hyperrectangles:

$$B_i^J = [x^i, y^i], \ x^i, y^i \in R^n, \ x_k^i = a_{ij_k}^{min}, \ y_k^i = a_{ij_k}^{max},$$

$$\bar{B}_i^J = [\bar{x}^i, \bar{y}^i], \ \bar{x}^i, \bar{y}^i \in R^n, \ \bar{x}_k^i = \bar{a}_{ij_k}^{min}, \ \bar{y}_k^i = \bar{a}_{ij_k}^{max},$$

$$\bar{B}^0 = B_i^J \bigcap \bar{B}_i^J.$$

$$Q_t^{J0} = \left\{ k = 1, \ldots, n_t : \ u^k = (d_{kj_1}^t, \ldots, d_{kj_n}^t) \in \bar{B}_0 \right\},$$

$$\bar{Q}_i^J = \bigcup_{t=1}^{m} Q_t^{J0}, \ \ \bar{q} = |\bar{Q}_i^J|.$$

Then we can define the overlap for the subset J between the class i and the rest of the data set by

$$\bar{O}_i^J = \begin{cases} 1, & B_i^J \subseteq \bar{B}_l^J \text{ or } \bar{B}_l^J \subseteq B_i^J, \\ \frac{\bar{q}}{n}, & \text{otherwise.} \end{cases}$$

Then we can define a vector of overlaps for a given subset of voxels J as follows:

$$\bar{O}^J = \left(\bar{O}_1^J, \ldots, \bar{O}_m^J \right).$$

2.1.4 Computation of informative voxels

In this section we consider three different algorithms to compute informative voxels. It is clear that a voxel or a group of voxels with a small overlap are better candidates to separate different stimuli. Let

$$I(n) = \{ J \in \{1, \ldots, p\} : \ |J| = n \}, \ \ 0 < n \leq p$$

be a set of all possible subsets which contain n different voxels.

The following algorithms can be used to determine the most informative voxels. In all algorithms we will consider binary and one-vs-all overlaps.

Algorithm 1. The use of minimum overlaps

Binary overlaps. For any $J \in I(n)$ we define the following numbers

$$r_J = \max_{i=1,\ldots,m} \max_{l=i+1,\ldots,m} O_{il}^J$$

and

$$R = \min_{J \in I(n)} r_J.$$

We assume that $R \in [0,1)$. The subset of voxels $J \in I(n)$ is said to be most informative subset if $r_J = R$.

One can take any tolerance $\varepsilon > 0$ such that $\varepsilon \leq 1 - R$ and define a subset of informative voxels

with respect to this tolerance. $J \in I(n)$ is a subset of informative voxels with respect to the tolerance $\varepsilon > 0$ if

$$r_J \leq R + \varepsilon.$$

If $\varepsilon = 0$ we get the most informative voxels and we get all voxels as informative ones if $\varepsilon = 1 - R$. Increasing ε from 0 to $1 - R$ we can get a sequence of subsets with the increasing number of voxels.

One-vs-all overlaps. Here we define

$$\bar{r}_J = \max_{i=1,\dots,m} \bar{O}_i^J, \quad \bar{R} = \min_{J \in I(n)} \bar{r}_J.$$

and assume that $\bar{R} \in [0, 1)$. The subset of voxels $J \in I(n)$ is said to be most informative if

$$\bar{r}_J = \bar{R}.$$

Again we can take any tolerance $\varepsilon > 0$ such that $\varepsilon < 1 - \bar{R}$ and define a subset of informative voxels with respect to this tolerance. $J \in I(n)$ is a subset of informative voxels with respect to the tolerance $\varepsilon > 0$ if

$$\bar{r}_J \leq \bar{R} + \varepsilon.$$

Increasing ε from 0 to $1 - \bar{R}$ we can get a sequence of subsets with the increasing number of voxels, where $\varepsilon = 0$ corresponds to the subset of the most informative voxels and $\varepsilon = 1 - \bar{R}$ corresponds to the whole set of voxels.

Algorithm 2. The use of the sum of overlaps.

Binary overlaps. For each subset $J \in I(n)$ of voxels we compute

$$f_J = \sum_{i=1}^{m} \sum_{k=i+1}^{m} O_{il}^J,$$

and

$$F = \min_{J \in I(n)} f_J.$$

The subset of voxels $J \in I(n)$ is called the most informative if $f_J = F$.

Let $\varepsilon > 0$ be a given tolerance. Then $J \in I(n)$ is called a subset of informative voxels with

respect to $\varepsilon > 0$ if

$$f_J \leq F + \varepsilon.$$

One-vs-all overlaps. Here for given $J \in I(n)$ we compute

$$\bar{f}_J = \sum_{i=1}^{m} \bar{O}_i^J.$$

Let

$$\bar{F} = \min_{J \in I(n)} \bar{f}_J.$$

$J \in I(n)$ is called the subset of most informative voxels if

$$\bar{f}_J = \bar{F}.$$

$J \in I(n)$ is called a subset of informative voxels with respect to $\varepsilon > 0$ if

$$\bar{f}_J \leq \bar{F} + \varepsilon.$$

In both cases $0 \leq \varepsilon < \infty$ and increasing ε from 0 to ∞ we can get a sequence of subsets with the increasing number of voxels, where we get the subset of the most informative voxels if $\varepsilon = 0$ and the set of all of voxels if ε is sufficiently large.

Algorithm 3. The use of the number of well-separated classes.

Binary overlaps. Let

$$\theta = \min_{i=1,\ldots,m} \min_{l=i+1,\ldots,m} O_{il}^J.$$

and $\alpha \in [\theta, 1]$. For the subset $J \in I(n)$ of voxels we define the following set:

$$N_J(\alpha) = \left\{ (i,l) : \ i = 1, \ldots, m, \ l = i+1, \ldots, m, \ O_{il}^J \leq \alpha \right\}.$$

Let

$$N_0 = \max_{J \in I(n)} |N_J(\alpha)|,$$

where $|Q|$ is the cardinality of the set Q. J is called the subset of most informative voxels if

$$|N_J(\alpha)| = N_0.$$

It is clear that $N_0 \leq \frac{m(m-1)}{2}$. Let $q > 0$ be any integer such that $0 \leq q \leq N_0$. $J \in I(n)$ is called a subset of informative voxels with respect to the number q if

$$|N_J(\alpha)| \geq q.$$

One-vs-all overlaps. Let

$$\bar{\theta} = \min_{i=1,\dots,m} \bar{O}_i^J,$$

and $\bar{\alpha} \in [\bar{\theta}, 1]$. For the subset $J \in I(n)$ of voxels we define the following sets:

$$\bar{N}_J(\bar{\alpha}) = \left\{ i \in \{1, \dots, m\} : \ \bar{O}_i^J \leq \bar{\alpha} \right\}.$$

Let

$$\bar{N}_0 = \max_{J \in I(n)} |\bar{N}_J(\bar{\alpha})|.$$

The subset of voxels J is called the most informative if

$$|\bar{N}_J(\alpha)| = \bar{N}_0.$$

It is clear that $\bar{N}_0 \leq m$. Let $\bar{q} > 0$ be any integer such that $\bar{q} \leq \bar{N}_0$. $J \in I(n)$ is called a subset of informative voxels with respect to the number $\bar{q}$ if

$$|\bar{N}_J(\alpha)| \geq \bar{q}.$$

We can compute a sequence of subsets with increasing number of voxels by increasing α $(\bar{\alpha})$ from $\theta(\bar{\theta})$ to 1 and by decreasing $q(\bar{q})$ from $N_0(\bar{N}_0)$ to 0.

It should be noted that for $m = 2$ all algorithms produce the same results. Furthermore, Algorithms 1 and 2 are the same in this case. However, for larger number of classes they may differ and sometimes significantly. If Algorithm 1 determines voxels which are good for separation of a

few classes only, Algorithms 2 and 3 are efficient to find voxels for separation of all classes. Since the most of fMRI data sets contain more than two classes, Algorithms 2 and 3 are more efficient to compute informative voxels in such data sets than Algorithm 1. Therefore in our computational experiments in the next section we will use only these two algorithms. Algorithm 2 tries to find voxels with least overall overlaps and Algorithm 3 finds voxels which are good for separation as many classes as possible.

2.1.5 Computational results

To verify the effectiveness of the proposed algorithm we carried out a number of numerical experiments with Haxby data set [75]. Numerical experiments have been carried out on a PC with Processor Intel(R) Core(TM) i5-3470S CPU 2.90 GHz and 8 GB RAM running under Windows 7.

We use the well-known Haxby data set, available from [75], for calculations. It contains data on face and object representation in human ventral temporal cortex and has 6 subjects viewing 8 categories of stimuli (faces, houses, cats, chairs, shoes, bottles, scissors and scrambled images). In each run subjects passively viewed greyscale images of eight object categories repeated 12 times (see, for details, [75]).

The original Haxby data set contains $64 \times 64 \times 40$ voxels [75, 131]. In [75] significant voxels were selected according to their variance and user defined threshold [71]. Only this data set is publicly available (not the original one). The data set is retrievable from [183]. The number of informative voxels in this data set varies between 307 and 675 for different subjects.

In numerical experiments we apply Algorithm 2 with both binary and one-vs-all overlaps to compute subsets of voxels using the number of samples. Furthermore, we apply this algorithm to find a sequence of subsets with the increasing number of voxels using different values of $\varepsilon > 0$. Then we apply different classifiers from WEKA to compute classification error with different subsets of voxels. See, Subsection 1.4 for details of WEKA.

Classifiers based on completely different approaches are chosen for classification task. These classifiers include:

- BayesNet: Builds a tree, where each node is a random variable and edges between nodes are probabilistic dependencies among variables [57].

- SMO: Implements the sequential minimal optimization algorithm for training a support vector

classifier [142].

- Logistic regression: It calculates optimized values of parameter matrix using a Quasi-Newton Method and finds probability of a class for a given data point [105].

- IBK: K-nearest neighbours classifier. It assigns a data point to a class common amongst its k nearest neighbors where k is the closest training examples in the feature space [4].

- PART: rule-based classifier. It builds a decision tree in each iteration and refines rule set by discarding individual rules to make them work better together rather complex optimization stage [56].

- J48 is the implementation of C4.5 algorithm which is based on decision tree approach [93, 147]

- FLR: the Fuzzy Lattice Reasoning Classifier induces rules in a mathematical lattice data domain such that a rule antecedent corresponds to a lattice interval and a rule consequent is a class label and a lattice interval corresponds to a N-dimensional hyperbox [12, 89].

- FURIA: Fuzzy unordered rule induction algorithm based on RIPPER (repeated incremental pruning to produce error reduction) algorithm [208]. This classifier learns fuzzy rules instead of conventional rules and unordered rule sets instead of rule lists [83].

Results for all subjects are presented in Tables 2.1-2.12. In these tables classification accuracies obtained by classifiers are given. The last line in all tables contains classification accuracy using all voxels. All classification accuracies using subset of voxels which are better than the classification accuracy using all voxels, are in bold font.

Table 2.1: Classification performance for Subject 1 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	43.04	41.60	42.84	39.60	39.53	41.53
5	41.74	43.73	44.83	55.30	46.42	46.90
10	40.36	44.90	47.73	66.32	49.86	48.48
20	40.43	48.42	50.69	74.52	52.62	**54.20**
50	50.00	59.44	**57.23**	86.50	**58.61**	**56.20**
100	52.00	66.46	**59.85**	88.50	56.82	**54.48**
150	**53.65**	71.76	**57.64**	88.71	**60.06**	**54.41**
250	**54.13**	75.21	**56.96**	**89.60**	57.16	**54.96**
300	**53.31**	77.07	**58.75**	82.51	**57.37**	**56.54**
350	52.55	78.17	**58.95**	**89.67**	56.20	**55.65**
400	**52.96**	79.20	**57.58**	89.19	**57.78**	**54.55**
500	**52.82**	81.34	**53.93**	**89.67**	55.17	**53.99**
577	52.82	81.75	52.82	89.33	57.09	52.62

Table 2.2: Classification performance for Subject 2 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	45.39	44.08	46.21	42.42	39.74	43.53
5	46.28	47.31	49.24	54.96	48.55	48.28
10	50.34	53.37	55.65	63.64	**53.86**	52.69
20	51.31	58.06	63.02	70.39	**54.55**	51.45
50	53.99	72.52	71.01	76.93	**55.10**	**53.10**
100	54.96	81.61	71.01	82.51	**56.75**	**53.03**
150	56.82	85.61	73.21	82.85	**55.44**	**52.41**
250	**60.06**	90.01	73.69	86.09	**56.54**	**53.37**
350	**60.67**	90.77	**79.41**	86.23	**54.68**	**54.61**
300	**60.61**	89.53	76.24	85.61	**55.30**	52.34
400	**59.23**	90.91	**83.33**	**87.60**	**53.17**	51.93
464	59.02	91.74	77.75	87.47	52.62	52.14

Table 2.1 presents classification results for Subject 1 using binary overlaps. One can see that IBK classifier achieved best result in this data set and this result is obtained using 350 out 577 voxels. On the same time very good accuracy is obtained using only 250 voxels. These results demonstrate the proposed voxel selection algorithm improves the performance of all classifiers except SMO classifier. Results from Table 2.2 for Subject 2 show that SMO classifier achieves the best classification accuracy on this subject. Although this classifier gets the best accuracy using all voxels its result with the use of 350 voxels is also quite high. For all other classifiers classification accuracy is improved using subset of voxels.

Table 2.3: Classification performance for subject 3 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	40.50	40.50	**40.43**	35.54	31.54	38.50
5	40.56	40.50	**40.98**	40.08	36.64	36.02
10	39.94	40.56	**41.67**	44.21	39.46	37.95
20	40.36	45.18	**43.66**	52.89	42.36	42.08
50	41.80	47.45	**46.28**	65.56	**43.53**	**44.42**
100	41.32	49.24	**45.87**	69.49	**44.97**	**42.84**
150	40.91	51.17	**44.97**	69.77	**44.49**	40.77
250	44.01	52.34	**35.61**	69.63	**43.80**	41.74
307	44.77	53.65	34.23	72.31	41.53	42.63

Results from Tables 2.3 and 2.4 indicate that IBK classifier obtains best accuracy results for Subjects 3 and 4 using binary overlaps. The use of the proposed voxel selection algorithm allows to improve performance of Logistic, PART and J48 classifiers in Subject 3 and all classifiers, except SMO, in Subject 4. Furthermore, the best accuracy is obtained in Subject 4 using 350 voxels out of 675.

Results for Subjects 5 and 6 using binary overlaps are given in Tables 2.5 and 2.6, respectively. In these subjects SMO classifier achieves the best accuracy. The use of the voxel selection algorithm

Table 2.4: Classification performance for subject 4 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	40.36	40.50	39.74	35.26	30.72	32.92
5	38.64	40.50	**40.43**	48.83	40.36	40.56
10	38.43	40.50	**41.39**	57.16	45.73	43.87
20	39.12	42.49	**43.11**	67.98	**49.04**	**50.69**
50	39.94	48.62	**48.76**	77.14	**50.48**	**50.07**
100	43.60	55.17	**53.17**	78.86	**51.45**	**50.21**
150	**45.73**	55.58	**45.11**	80.44	**51.38**	48.07
250	**46.42**	60.61	**45.73**	**78.37**	49.86	46.63
300	**46.76**	62.33	**44.77**	**79.20**	50.55	46.42
350	**46.42**	63.15	**43.80**	**80.58**	48.55	47.66
400	**45.52**	63.91	**43.46**	**80.03**	49.66	47.45
500	**44.28**	64.05	**42.63**	**78.72**	48.76	47.25
675	43.87	66.74	39.53	78.24	48.21	44.63

Table 2.5: Classification performance for subject 5 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	45.87	46.01	47.18	45.66	41.60	47.11
5	**47.59**	46.28	49.10	47.80	43.94	43.04
10	**50.55**	51.93	56.20	53.58	46.42	46.63
20	**52.75**	60.40	**63.29**	60.06	46.69	**47.25**
50	**51.24**	67.70	**69.15**	68.18	**50.76**	**50.55**
100	**50.07**	77.75	**61.36**	73.21	**51.03**	**50.96**
150	**48.48**	79.61	**62.12**	75.07	**50.00**	**49.04**
250	**48.35**	83.95	**61.98**	**77.48**	48.97	49.38
300	**48.14**	84.02	**60.54**	**77.34**	48.97	47.87
350	**48.35**	84.78	**61.98**	**76.58**	48.90	47.66
400	**50.07**	**85.40**	**60.54**	76.17	**48.00**	45.66
423	47.18	85.12	59.71	76.45	47.87	46.69

allows to improve the performance of all classifiers in Subject 5. In this subject SMO classifier produces a good accuracy using 350 voxels. In Subject 6 the performance of all classifiers, except SMO, is improved.

Table 2.6: Classification performance for subject 6 using binary overlap

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	46.63	45.52	47.73	41.32	43.80	44.08
5	46.63	47.25	49.79	45.87	40.98	41.80
10	48.14	52.20	56.06	50.41	44.77	44.42
20	52.48	61.09	**63.64**	59.09	**46.14**	**47.31**
50	52.41	61.91	**68.60**	**61.91**	**47.52**	**46.63**
100	**53.99**	75.83	**61.23**	**65.01**	**48.90**	44.83
150	**54.68**	78.03	**63.50**	**66.12**	**48.90**	44.63
250	**54.55**	81.54	**61.57**	**65.84**	**46.56**	44.49
349	53.65	82.58	59.02	59.64	44.77	45.52

Table 2.7: Classification performance for Subject 1 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	44.77	42.63	44.35	43.11	39.05	42.49
5	44.77	45.18	45.45	52.69	47.66	47.04
10	46.42	48.76	48.97	65.22	51.93	52.34
20	46.01	51.10	51.45	71.69	53.24	52.14
50	50.14	59.99	**60.19**	83.82	**57.30**	**55.44**
100	52.96	65.56	**59.30**	88.02	**58.33**	**54.75**
150	52.34	69.35	**56.61**	87.81	**58.33**	**55.44**
250	**53.58**	76.38	**56.47**	**90.15**	55.79	**54.06**
300	**53.99**	77.27	**58.82**	**90.29**	56.68	**55.03**
350	**53.24**	77.89	**58.61**	**90.01**	58.40	**54.27**
400	**52.96**	78.79	**55.23**	**89.53**	58.47	**54.82**
500	52.69	81.20	**54.61**	89.05	**58.95**	**53.72**
577	52.82	81.75	52.82	89.33	57.09	52.62

Results obtained by the binary overlap algorithm for all subjects from Haxby data are summarised in Figure 2.3. Here Legends show the number of selected voxels and stars indicate the number of voxels in the original data sets. One can see that the IBK classifier obtains the best accuracy and the binary overlap algorithm improves this accuracy.
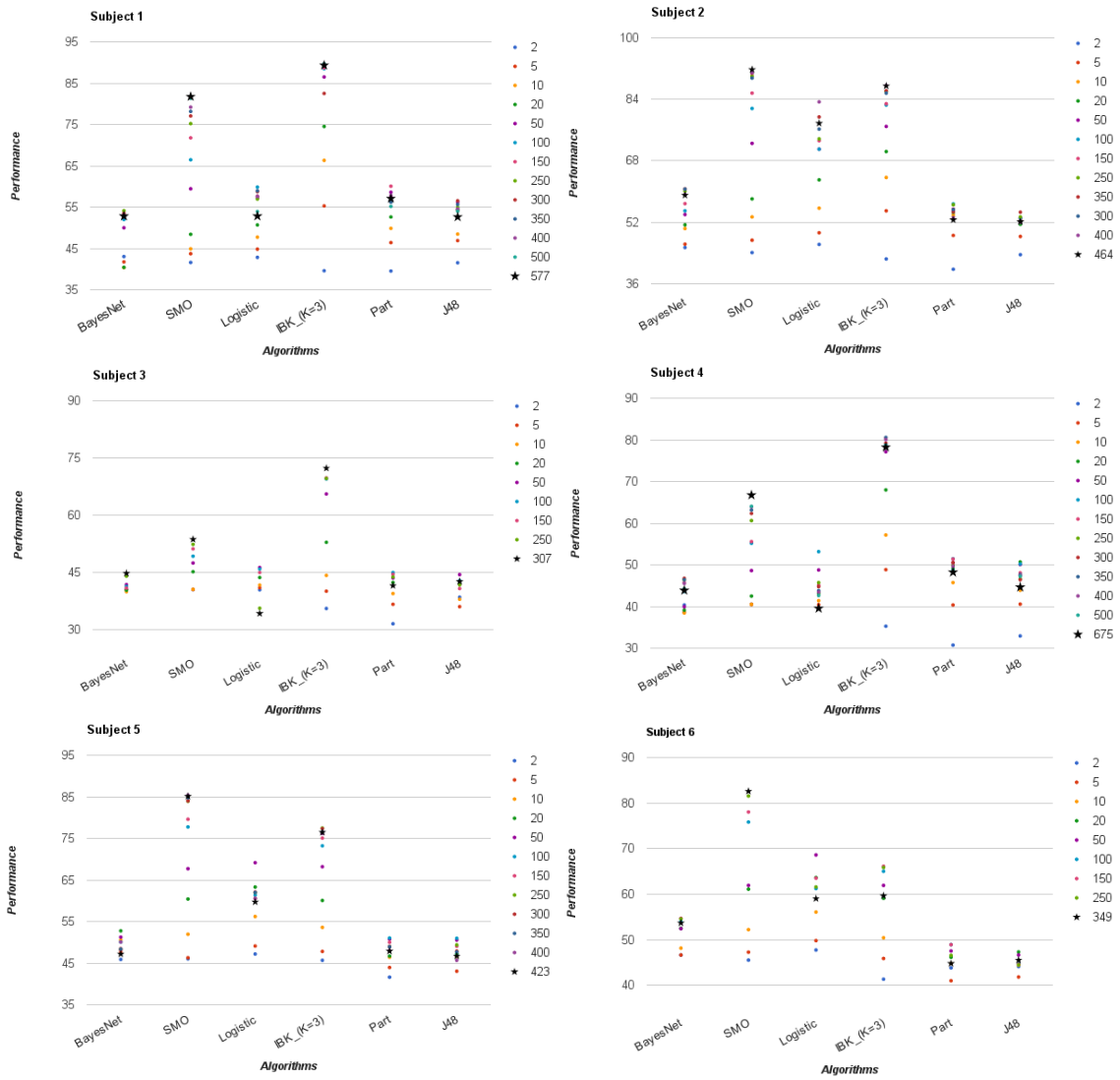
Figure 2.3: Classification performance for Haxby data set using binary overlaps. Horizontal axis indicates the accuracy of classification algorithms and vertical axis shows the classifiers that are used. In the right hand side of the graphs number of selected features are shown.

Results for Subject 1 using one vs all overlaps are given in Table 2.7. The best classification accuracy on this subject was obtained using IBk classifier with 300 voxels. The use of the voxel selection algorithm allows to improve the performance of all classifiers except SMO classifier. The comparison of results from Tables 2.1 and 2.7 show that the use of binary and one vs all overlaps does not lead to any significant difference in classification accuracy for all classifiers. However, note that the best accuracy obtained using one vs all overlaps is slightly better than that of using binary overlaps.

Table 2.8: Classification performance for Subject 2 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	44.56	45.80	45.87	43.04	42.49	43.53
5	48.97	47.11	48.42	51.93	46.97	47.04
10	47.87	50.76	51.24	57.64	48.28	49.31
20	51.79	57.16	58.95	66.94	50.83	51.31
50	55.10	72.31	72.25	77.14	**53.79**	**53.10**
100	57.92	81.68	72.18	84.44	**54.48**	**53.79**
150	**59.23**	85.19	74.38	84.23	**54.82**	**53.24**
250	**59.50**	88.91	74.93	86.36	**55.44**	50.83
300	**59.30**	89.33	77.69	86.16	**53.79**	51.17
350	**59.57**	90.22	**79.55**	86.43	**54.89**	**52.82**
400	**59.23**	90.77	**85.67**	86.98	**54.96**	51.17
464	59.02	91.74	77.75	87.47	52.62	52.14

Results for Subject 2 using one vs all overlaps presented in Table 2.8 show that the best classification accuracy on this subject was obtained using SMO classifier with all voxels, however this classifier achieves a good accuracy using only 350 voxels. One can see that the use of the voxel selection algorithm allows to improve the performance of all classifiers except SMO and IBk classifiers. Comparing results from Tables 2.2 and 2.8 we can conclude that there is no any significant difference in classification accuracy obtained by all classifiers using both the binary and one vs all overlaps.

Table 2.9: Classification performance for Subject 3 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	40.50	40.50	**40.36**	36.64	32.71	38.77
5	40.56	40.50	**41.74**	40.15	33.26	39.67
10	40.56	43.11	**43.66**	43.80	40.36	7.26
20	41.39	45.18	**44.49**	52.96	**41.60**	41.74
50	38.71	46.97	**45.32**	63.84	**45.94**	42.36
100	42.70	49.17	**46.28**	69.90	**45.04**	42.56
150	43.73	51.79	**45.66**	69.35	**44.63**	41.94
250	**44.97**	**53.79**	**36.36**	72.25	**45.11**	**42.98**
307	44.77	53.65	34.23	72.31	41.53	42.63

Table 2.9 contains results for Subject 3 using one vs all overlaps. These results show that the best classification accuracy on this subject was obtained using IBk classifier with all voxels, however this

classifier achieves almost the same accuracy using only 250 voxels. The use of the voxel selection algorithm improves the performance of all classifiers except IBk classifier. Results presented in Tables 2.3 and 2.9 show that the classification accuracy obtained using one vs all overlaps is consistently better than that obtained using the binary overlaps.

Table 2.10: Classification performance for Subject 4 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	41.60	41.60	**42.70**	41.05	38.50	41.05
5	40.98	43.66	**43.80**	47.04	42.15	43.39
10	40.91	43.80	**44.63**	59.23	47.59	**47.87**
20	41.32	45.25	**44.70**	68.94	**49.93**	**48.07**
50	43.32	49.66	**49.66**	76.10	**50.41**	**48.55**
100	**45.59**	53.93	**51.65**	79.82	**50.83**	**48.00**
150	**44.90**	57.44	**47.73**	78.65	**49.79**	**49.10**
250	**46.07**	61.50	**45.87**	79.27	**49.93**	**48.21**
300	**46.14**	63.22	**43.80**	79.89	**50.55**	**46.76**
350	**45.66**	63.15	**45.04**	79.55	**49.24**	**47.38**
400	**44.21**	64.94	**42.91**	79.34	**49.59**	**44.77**
500	**44.70**	66.12	**43.32**	79.55	**48.76**	**45.52**
675	43.87	66.74	39.53	78.24	48.21	44.63

Results for Subject 4 using one vs all overlaps are presented in Table 2.10. On this subject the best classification was obtained using IBk classifier with 300 voxels. The use of the voxel selection algorithm improves the performance of all classifiers except SMO classifier. Furthermore, this algorithm significantly improves the performance of the Logistic classifier. Comparing results from Tables 2.4 and 2.10 we can see that the classification accuracies obtained using binary overlaps are slightly better than those obtained using the one vs all overlaps.

Table 2.11: Classification performance for Subject 5 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	45.04	45.45	46.90	43.39	38.84	43.46
5	46.35	47.38	49.93	49.38	44.56	44.77
10	49.24	53.86	57.44	55.58	48.55	48.42
20	46.28	59.30	**65.01**	61.78	**50.69**	**50.07**
50	48.62	69.01	**69.77**	68.25	**50.28**	**49.04**
100	49.52	76.72	**63.09**	75.07	**50.28**	**48.48**
150	47.73	79.75	**61.43**	75.69	**48.48**	**49.17**
250	47.11	82.51	**61.09**	77.75	**48.28**	**48.90**
300	46.56	83.33	**62.74**	75.69	47.45	**48.00**
350	46.69	84.92	**60.40**	77.41	**49.45**	**47.52**
400	**47.18**	84.99	59.64	**76.93**	46.90	**47.11**
423	47.18	85.12	59.71	76.45	47.87	46.69

Results for Subject 5 using one vs all overlaps are reported in Table 2.11. The SMO classifier achieves the best accuracy on this subject using all voxels, however the classification accuracy with 350 voxels is very close to the best accuracy. The use of the voxel selection algorithm improves the

performance of all classifiers except SMO classifier. One can also see performance of the Logistic classifier is significantly improved. From Tables 2.5 and 2.11 we can see that the classification accuracies obtained using binary and one vs all overlaps are similar.

Table 2.12: Classification performance for Subject 6 using one vs all overlaps

Number of selected features	BayesNet	SMO	Logistic	IBK (K=3)	Part	J48
2	44.08	44.01	45.04	41.39	38.22	40.63
5	47.80	49.38	52.41	47.25	41.53	45.45
10	50.34	55.10	**59.50**	54.68	**47.73**	45.25
20	51.38	59.85	**62.05**	58.33	**47.04**	**46.97**
50	**54.13**	68.73	**67.91**	**63.09**	**49.17**	44.70
100	**54.41**	76.58	**62.81**	**64.74**	**48.07**	**46.28**
150	**53.93**	78.37	**62.40**	**64.46**	**48.48**	43.53
250	**54.27**	82.51	**62.95**	**66.60**	**48.48**	**46.35**
349	53.65	82.58	59.02	59.64	44.77	45.52

Results for Subject 6 using one vs all overlaps are reported in Table 2.12. Although the SMO classifier achieves the best accuracy on this subject using all voxels, this accuracy with 250 voxels is very close to the best accuracy. The use of the voxel selection algorithm improves the performance of all classifiers except SMO classifier. Results from Tables 2.6 and 2.12 show that the classification accuracies obtained using binary and one vs all overlaps are similar.

Figure 2.4 shows results of one vs all performance for all subjects in the Haxby data. Comparing results from Figures 2.3 and 2.4 we can see that there is no any significant difference in classification accuracy obtained by all classifiers using both the binary and one vs all overlaps.
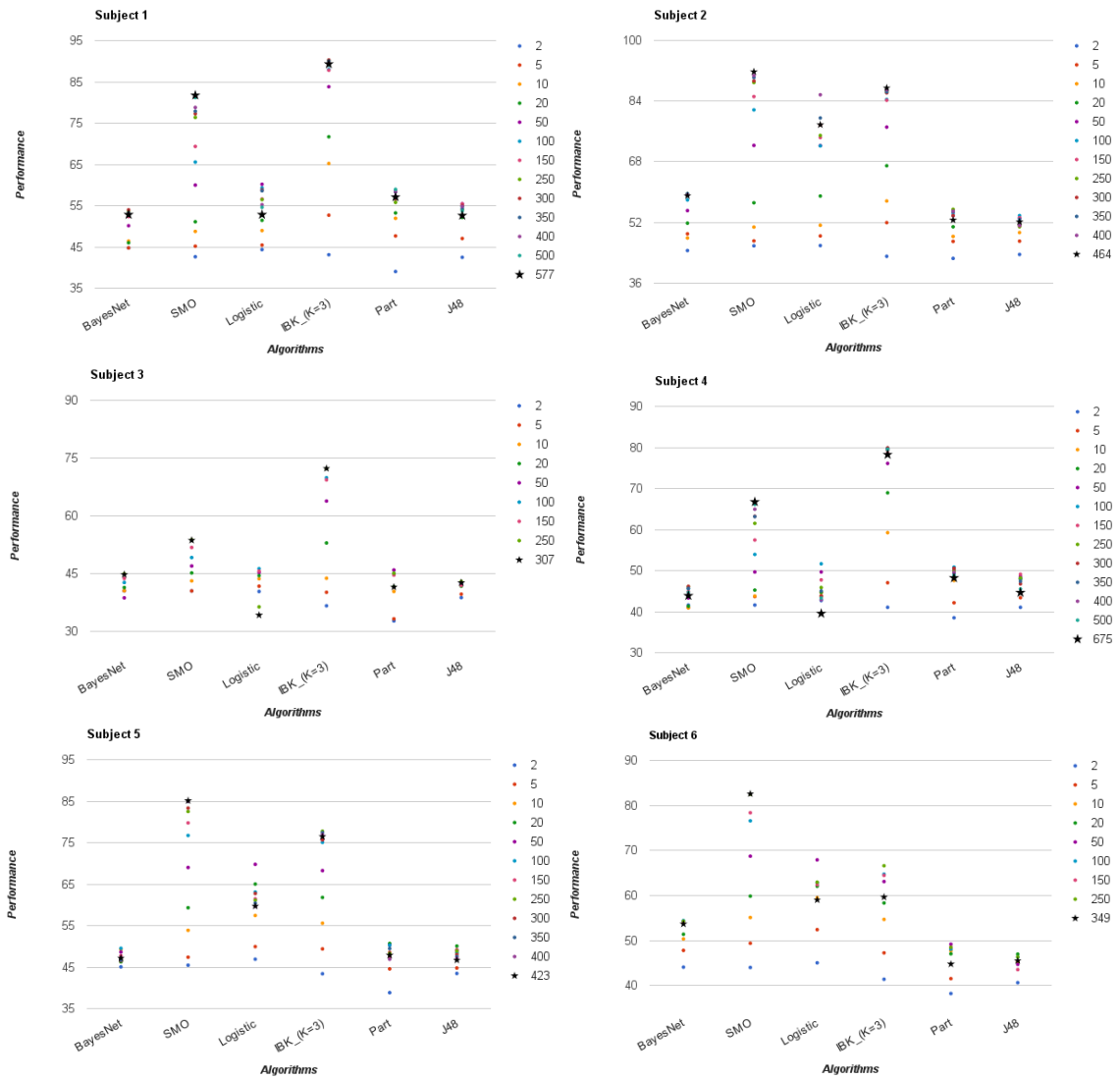
Figure 2.4: Classification performance for Haxby data set using one vs all overlaps. Horizontal axis indicates the accuracy of classification algorithms and vertical axis shows the classifiers that are used. In the right hand side of the graphs number of selected features are shown.

2.1.6 Results for Science 2008 data set

Results for all subjects of Science 2008 data set are presented in Tables 2.13-2.21 (N/A in these tables means that a classifier does not produce any result). Since results for binary and one vs all overlaps do not differ significantly here we present results only for one vs all overlaps. We replace BayesNet and Logistic classifiers by FLR and FURIA classifiers because the former classifiers can not produce any result on this data set using all voxels due to the memory requirements. Results presented in Tables 2.13-2.21 show that the performance of the voxel selection algorithm is very similar throughout of all nine subjects. These results clearly demonstrate that the proposed algorithm is highly efficient in finding the subset of informative voxels in this data set. One can see that the classification accuracy of all classifiers using all 21344 voxels is very low. The algorithm is able to find subsets of 10 or 50 voxels which provide better accuracy for all classifiers. Moreover, this algorithm is able to reduce the number of voxels from 70 to 100 times and at the same time to significantly improve the performance of all classifiers.

Table 2.13: Classification performance for Subject 1 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	13.33	40.00	30.00	30.00	30.00	28.33
50	48.33	46.67	28.33	26.67	35.00	23.33
100	75.00	63.33	33.33	28.33	48.33	13.33
150	75.00	61.67	50.00	21.67	46.67	15.00
200	78.33	68.33	40.00	21.67	43.33	16.67
250	80.00	68.33	43.33	23.33	41.67	10.00
300	76.67	66.67	41.67	21.67	43.33	16.67
350	80.00	68.33	38.33	21.67	45.00	15.00
400	85.00	75.00	40.00	21.67	45.00	16.67
450	81.67	70.00	40.00	20.00	45.00	16.67
500	81.67	68.33	40.00	16.67	48.33	13.33
21764	20.00	28.33	11.67	11.67	3.33	8.33

Table 2.14: Classification performance for Subject 2 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	15.00	18.33	16.67	20.00	28.33	10.00
50	40.00	33.33	25.00	21.67	43.33	16.67
100	53.33	48.33	16.67	20.00	38.33	15.00
150	58.33	51.67	23.33	26.67	21.67	21.67
200	53.33	35.00	20.00	25.00	40.00	21.67
250	55.00	46.67	16.67	26.67	41.67	20.00
300	51.67	45.00	18.33	30.00	41.67	18.33
350	56.67	45.00	20.00	20.00	43.33	25.00
400	56.67	43.33	18.33	18.33	41.67	21.67
450	55.00	55.00	16.67	15.00	41.67	20.00
500	55.00	45.00	16.67	15.00	41.67	20.00
21253	8.33	11.67	13.33	6.67	1.67	6.67

Table 2.15: Classification performance for Subject 3 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	18.33	30.00	26.67	28.33	36.67	26.67
50	55.00	45.00	23.33	28.33	41.67	35.00
100	60.00	46.67	23.33	30.00	48.33	18.33
150	61.67	46.67	18.33	28.33	40.00	20.00
200	70.00	46.67	21.67	20.00	38.33	20.00
250	70.00	51.67	20.00	26.67	40.00	23.33
300	71.67	46.67	35.00	31.67	38.33	20.00
350	71.67	45.00	23.33	38.33	33.33	16.67
400	71.67	50.00	33.33	41.67	35.00	18.33
450	70.00	53.33	33.33	40.00	31.67	20.00
500	71.67	51.67	33.33	36.67	30.00	21.67
20651	20.00	13.33	11.67	5.00	3.33	8.33

Table 2.16: Classification performance for Subject 4 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	31.67	35.00	26.67	30.00	38.33	18.33
50	78.33	53.33	26.67	25.00	70.00	30.00
100	75.00	53.33	25.00	30.00	60.00	20.00
150	81.67	65.00	35.00	26.67	53.33	20.00
200	86.67	68.33	30.00	28.33	56.67	21.67
250	88.33	76.67	25.00	26.67	55.00	35.00
300	90.00	71.67	25.00	25.00	56.67	35.00
350	90.00	73.33	26.67	28.33	55.00	30.00
400	85.00	73.33	25.00	25.00	56.67	21.67
450	88.33	76.67	25.00	25.00	53.33	26.67
500	86.67	78.33	21.67	23.33	55.00	20.00
20395	23.33	23.33	13.33	11.67	N/A	15.00

Table 2.17: Classification performance for Subject 5 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	10.00	16.67	23.33	25.00	21.67	10.00
50	43.33	36.67	23.33	26.67	28.33	10.00
100	46.67	35.00	28.33	21.67	33.33	13.33
150	55.00	31.67	28.33	21.67	33.33	10.00
200	53.33	33.33	31.67	21.67	33.33	13.33
250	53.33	31.67	30.00	18.33	33.33	15.00
300	51.67	30.00	33.33	21.67	35.00	18.33
350	46.67	33.33	31.67	21.67	30.00	13.33
400	43.33	36.67	30.00	23.33	28.33	15.00
450	45.00	35.00	30.00	25.00	23.33	11.67
500	46.67	31.67	30.00	23.33	25.00	10.00
20601	15.00	10.00	5.00	11.67	5.00	11.67

Table 2.18: Classification performance for Subject 6 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	0.00	10.00	3.33	20.00	16.67	11.67
50	13.33	11.67	10.00	16.67	11.67	10.00
100	20.00	6.67	6.67	10.00	13.33	8.33
150	18.33	8.33	10.00	8.33	11.67	6.67
200	15.00	8.33	8.33	8.33	10.00	6.67
250	16.67	5.00	6.67	6.67	10.00	5.00
300	16.67	6.67	10.00	8.33	10.00	6.67
350	15.00	6.67	3.33	6.67	10.00	6.67
400	15.00	6.67	3.33	6.67	8.33	6.67
450	15.00	10.00	3.33	6.67	10.00	5.00
500	15.00	8.33	3.33	11.67	10.00	6.67
19919	11.67	3.33	6.67	1.67	6.67	5.00

Table 2.19: Classification performance for Subject 7 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	20.00	35.00	20.00	13.33	35.00	21.67
50	46.67	30.00	25.00	38.33	45.00	18.33
100	48.33	30.00	30.00	28.33	38.33	21.67
150	48.33	46.67	28.33	26.67	35.00	11.67
200	56.67	31.67	25.00	31.67	33.33	8.33
250	61.67	31.67	26.67	18.33	36.67	13.33
300	60.00	35.00	25.00	26.67	36.67	16.67
350	58.33	36.67	21.67	21.67	33.33	11.67
400	51.67	33.33	18.33	23.33	30.00	10.00
450	51.67	36.67	20.00	18.33	25.00	8.33
500	56.67	33.33	16.67	20.00	25.00	11.67
19750	11.67	10.00	13.33	8.33	5.00	10.00

Table 2.20: Classification performance for Subject 8 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	10.00	25.00	23.33	26.67	35.00	30.00
50	43.33	30.00	15.00	15.00	46.67	13.33
100	58.33	36.67	13.33	21.67	43.33	20.00
150	58.33	35.00	20.00	21.67	43.33	20.00
200	55.00	28.33	20.00	16.67	36.67	18.33
250	53.33	35.00	16.67	8.33	36.67	16.67
300	51.67	26.67	20.00	18.33	38.33	18.33
350	45.00	20.00	21.67	18.33	33.33	13.33
400	43.33	23.33	20.00	18.33	31.67	11.67
450	43.33	23.33	23.33	20.00	35.00	11.67
500	46.67	25.00	20.00	18.33	25.00	10.00
20082	11.67	5.00	5.00	8.33	1.67	10.00

Table 2.21: Classification performance for Subject 9 using one vs all overlaps

Number of selected features	SMO	IBK(K=3)	J48	PART	FLR	FURIA
10	6.67	20.00	18.33	16.67	36.67	18.33
50	65.00	33.33	8.33	16.67	36.67	18.33
100	66.67	48.33	15.00	21.67	36.67	11.67
150	71.67	51.67	13.33	15.00	35.00	16.67
200	68.33	51.67	16.67	16.67	30.00	16.67
250	75.00	45.00	16.67	15.00	28.33	18.33
300	73.33	43.33	16.67	10.00	25.00	21.67
350	73.33	43.33	20.00	8.33	26.67	18.33
400	71.67	40.00	18.33	15.00	28.33	15.00
450	73.33	40.00	21.67	16.67	23.33	10.00
500	71.67	46.67	18.33	25.00	23.33	10.00
21344	13.33	11.67	8.33	8.33	5.00	5.00

The performances of four classifiers for Science 2008 data are illustrated in the Figures 2.5 and 2.6. Horizontal axis indicates the accuracy of classification algorithms and vertical axis shows the classifiers that are used. In the right hand side of the graphs number of selected features are shown. One can see that the One vs all voxel selection algorithm is able to find subsets of 10 or 50 voxels which provide better accuracy for all classifiers in comparison with all voxels.
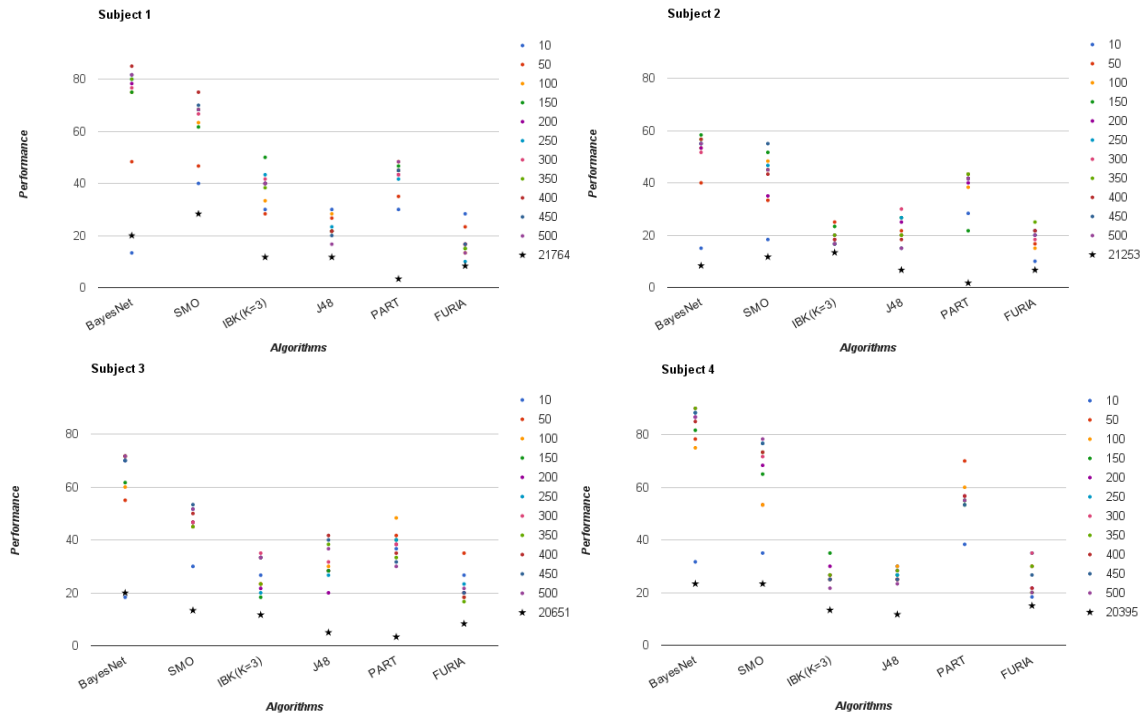

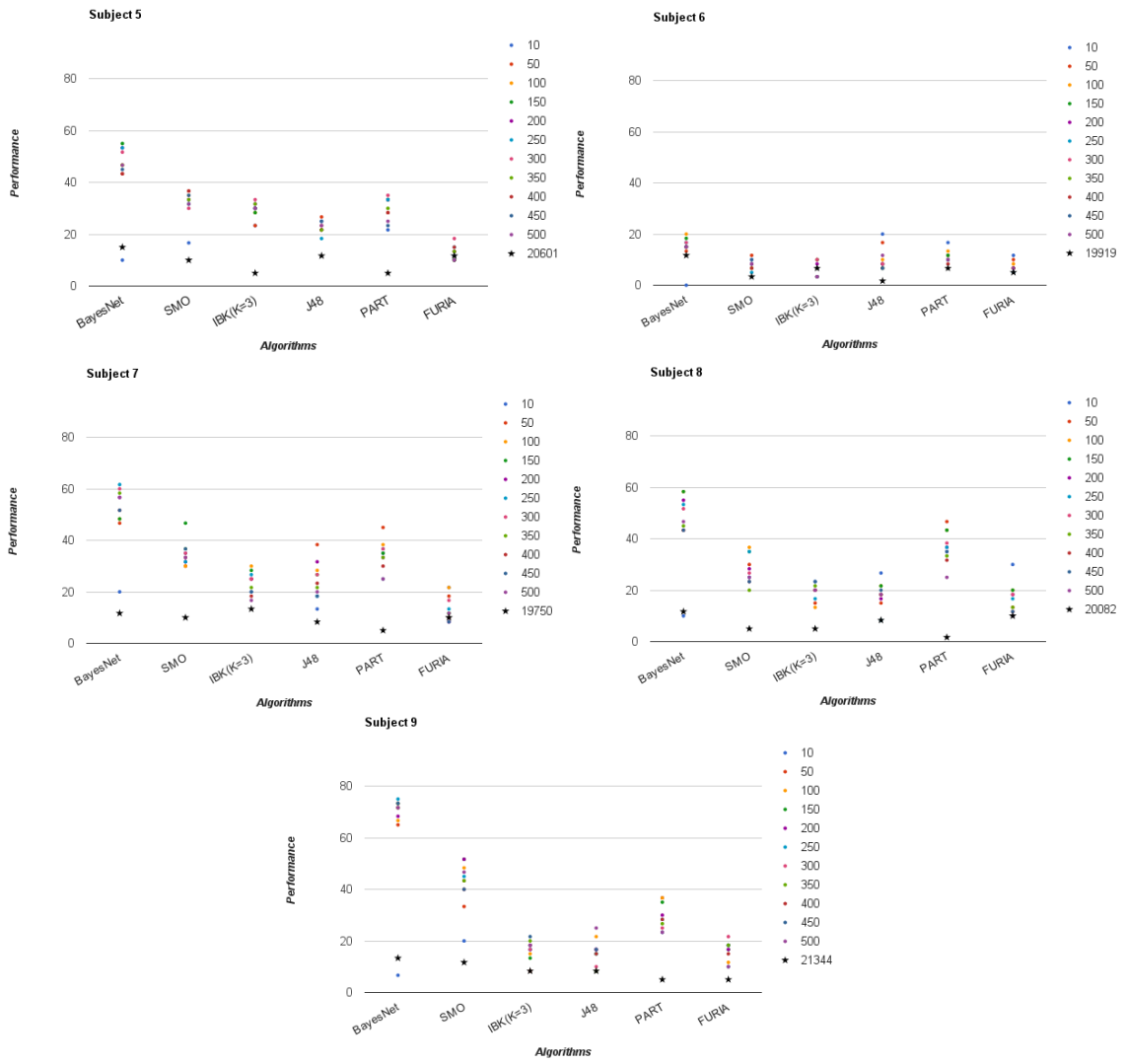Figure 2.5: Classification performance for Science 2008 data set using one vs all overlaps.

58

Figure 2.6: Classification performance for Science 2008 data set using one vs all overlaps (cont.).

2.2 Feature selection algorithm based on Catastrophe model

In this section we introduce a new feature selection algorithm to remove the irrelevant or redundant features in medium size or large scale regression data sets. In this algorithm the importance of a feature is based on it's fitting to the Catastrophe model. Breast Cancer, Parkinson Telemonitoring data and Slice locality data sets are used to evaluate the model. Akaike information criterion value is used for ranking the features in the data set and the proposed algorithm is compared with well-known feature selection algorithm RELIEF. Since our algorithm is based on the approaches from Catastrophe theory and Akaike information criterion, we start with very brief description of them.

2.2.1 Cusp Catastrophe

In this subsection we give a brief description of cusp model. Consider the following dynamical system:

$$\frac{\partial y}{\partial t} = -\frac{\partial V(y; c)}{\partial t}, y \in R^k, c \in R^p,$$ (2.2)

where V is the potential function, $y(t)$ represents the system's state variable(s), c shows one or multiple (control) parameter(s) whose value(s) determine the specific structure of the system. If y is at a point where

$$\frac{\partial V(y; c)}{\partial t} = 0$$ (2.3)

the system is in equilibrium. The function $V(y; c)$ acquires a minimum with respect to y at a non-equilibrium point. Equilibrium points that correspond to minima of $V(y; c)$ are stable equilibrium points because the system will return to such a point after a small perturbation to the system's state. The equilibrium points that correspond to maxima of $V(y; c)$ are unstable equilibrium points because a perturbation of the system's state will cause the system to move away from the equilibrium point towards a stable equilibrium point. Equilibrium points that correspond neither to maxima nor to minima of $V(y; c)$, at which the Hessian matrix $(\partial^2 V(y)/\partial y_i \partial y_j)$ has eigenvalues equal to zero, are called degenerate equilibrium points. When the control variables of the system are changed. System can give rise to unexpected bifurcations in its equilibrium states at these point when the control variables of the system are changed [65, 159, 212].

Cusp model that is the simplest form of Catastrophe and can be formulated as follows:

$$-V(y; \alpha, \beta) = \alpha y + \frac{1}{2}\beta y^2 - \frac{1}{4}y^4, \qquad (2.4)$$

where V is the canonical form of the potential function for the Cusp model and it's equilibrium points is a function of the control parameters α and β (see Figure 2.7). The control parameters are the solution to the equation

$$\alpha + \beta y - y^3 = 0. \qquad (2.5)$$

This equation has one solution if $\delta = 27\alpha - 4\beta^3$ that is greater than zero, and has three solution if $\delta < 0$ [39, 65].
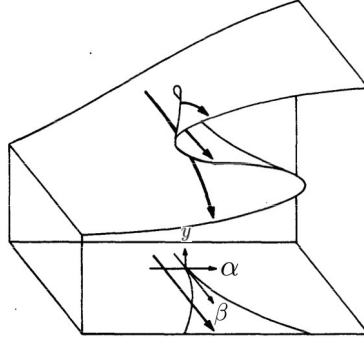

Figure 2.7: Cusp surface [40]

2.2.2 Akaike information criterion

Akaike information criterion (AIC) is a model quality measure for a given data [5, 32]. For a model AIC measure can be defined as follow [29, 158]:

$$AIC = -2log\mathrm{L}(\widehat{\theta}) + 2k, \qquad (2.6)$$

where $\mathrm{L}(\widehat{\theta})$ is the maximized likelihood function and k is the number of free parameters in the model. The smaller value of AIC shows that data is better fit to model. In the proposed algorithm, we used the reverse value of AIC for ranking the features in our data.

2.2.3 The feature selection algorithm

In the Catastrophe theory small change in certain parameters of a system can cause equilibria to appear or disappear [178, 212]. We used this characteristics of the Catastrophe model to find the features that are more affective in regression analysis. In the proposed algorithm the features that better change the dynamic of outcome feature or features are considered as informative features. Assume that we are given a data set A with N features that z is outcome feature. The algorithm takes each feature i from the data set and consider it as bifurcation variable in the Cusp Catastrophe model. If this variable affects the dynamic of the system (outcome feature), it is informative feature. The AIC value of the Cusp model is computed for each feature for ranking. The ranking of a feature i can be formulated as follows:

$$AIC_i = AIC(-V(y; \alpha, i)), \qquad (2.7)$$

where V is the potential function for the Cusp model (see Equation 2.4), AIC_i is the AIC value of the Cusp model for the feature i as bifurcation value (β) and α is the asymmetric value in the Cusp model. Figure 2.8 shows the preparing the input parameters for Cusp model where the the outcome feature is considered as state variable and the features i and the last features are considered as bifurcation and asymmetric values, respectively. The state variable and control values can be computed as follows [65]:

$$y[t] = w[0] + w[1] * Y[t, 1] + ... + w[p] * Y[t, p], \qquad (2.8)$$

$$\alpha[t] = a[0] + a[1] * X[t, 1] + ... + a[p] * X[t, p], \qquad (2.9)$$

$$\beta[t] = b[0] + b[1] * X[t, 1] + ... + b[p] * X[t, p], \qquad (2.10)$$

where $X[t, p]$'s are independent and $Y[t, p]$'s are dependent features in the data set. The vectors $a[j]$'s, $b[j]$'s and $w[j]$'s are estimated by means of maximum likelihood. The rank of each feature i in the data set can be calculated as follows:

$$rank_i \leftarrow \frac{1}{AIC_i}. \qquad (2.11)$$

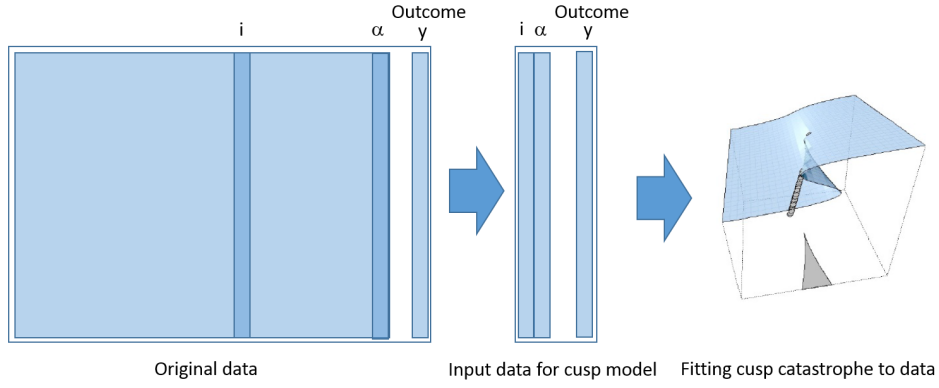More details about the algorithm is shown in the Algorithm 1.

Figure 2.8: Preparing input features for Cusp Catastrophe model

---

Algorithm 1 Feature selection algorithm based on the Cusp Catastrophe model and AIC ranking

---

Step 1: (Initialization) $N \leftarrow$ Number of features , $N_F \leftarrow$ Number of informative features , $\alpha \leftarrow feature_N$, $i \leftarrow 1$ and α is asymmetric variable

Step 2: Let $\beta \leftarrow feature_i$ be bifurcation value in the Cusp model

Step 3: (Fitting the Cusp model using α and β) Let AIC_i be the Akaike information criterion value of the fitting Cusp model using parameters α and β

Step 4: (Ranking the feature) $rank_i \leftarrow \frac{1}{AIC_i}$ is the rank of feature i in the dataset

Step 5: if $\frac{1}{AIC_i} \leq t$ then $feature_i$ is not informative and eliminate it, $i \leftarrow i+1$ and go to 6

Step 6: (Stopping criterion) if $i > N_F$ stop. Otherwise go to Step 2

Step 7: (Retraining informative features) Return N_F informative features.

Here N is the number of all feature in the data set and N_F ($N_F < N$) is the number of informative features. For all features i of the data set their rank in the data set is computed ($rank_i$). The set of informative features with N_F features is the outcome of the algorithm.

2.2.4 RELIEF feature selection algorithm

Next, we give a brief description of the RELIEF algorithm. More detailed description can be found in [97, 100, 154]. For a given data set with m samples, and threshold of relevancy τ ($0 \leq \tau \leq 1$), it detects those features which are statistically relevant to the target concept ($Y = f(X)$). Differences of feature values between two instances X and Y are defined by the following function $diff$ [98].

$$diff(x_k, y_k) = (x_k - y_k)/nu_k, \tag{2.12}$$

63

where nu_k is a normalization unit to normalize the values of $diff$ into the interval $[0, l]$. RELIEF picks a sample composed of m triplets of an instance X, it's same-class instance ($nearHit$) and closest different-class instance ($nearMiss$). RELIEF uses the p-dimensional Euclidean distance for selecting $nearHit$ and $nearMiss$. In every routine the feature weight W vector is updated as follows:

$$W_i = W_{i-1} - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2. \tag{2.13}$$

Then the average feature weight vector relevance is determined for every sample triple. Finally, it chooses the features whose average weight is above the given threshold τ.

2.2.5 Experimental results

The effectiveness of the proposed algorithm is verified using three different data sets: Parkinson's Telemonitoring, Breast Cancer and Slice locality from UCI machine learning repository [28]. Numerical experiments have been carried out on a PC with Processor Intel(R) Core(TM) i5-3470S CPU 2.90 GHz and 8 GB RAM running under Windows 7.

In numerical experiments we apply the proposed algorithm to find a ranking sequence of features in data sets. Then we apply different regression analysis algorithms from WEKA to compute regression error with subsets of features. The following regression analysis algorithms from WEKA are used in numerical experiments:

- Linear regression: Linear regression finds the best curve to fit the data by computing the relationship between a scalar dependent variable y and one or more explanatory variables denoted X. It appplies least squares, which minimizes the sum of the distance from the line for each of points. The actual observations, y_i, may be slightly off the population line because of variability in the population. The equation is $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where ϵ_i is the deviation from the population line which is called the residual [19, 122].

- K nearest neighbors regressor: The algorithm computes the mean of the function values of its K-nearest neighbours [101].

- M5Rulles: It generates rules for numeric prediction by separate-and-conquer and at each iteration builds a model tree using M5 and makes the "best" leaf into a rule [79, 148, 203]

- REPTree: Reptree is a fast tree learner that uses reduced error pruning [209].

2.2.6 Results for Breast cancer data set

Breast Cancer Wisconsin (Prognostic) Data Set contains 30 features with 569 samples. Each record represents follow-up data for one breast cancer case [107, 170] . Table 2.22 presents the error of analysing the data using for regression analysis algorithms. The second row shows the number of features before and after feature selection. Results from this table demonstrate that features selected by the proposed algorithm allow us to reduce the mean absolute error (MAE) regression. MAE is calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |f_i - y_i|, \tag{2.14}$$

where n is the number of observation, f_i is the predicted and y_i is the true values. Although this data set is not noisy the proposed algorithm is able to significantly reduce the number of features without deteriorating the regression error. Regression errors with the subsets of features which are better than that of for all features are presented in bold font.

Table 2.22: Performance of regression analysis algorithms for breast cancer data set

	Original data	After feature selection					
Number of features	30	25	20	15	10	6	5
Linear Regression	0.003	**0.003**	**0.003**	**0.003**	**0.003**	**0.003**	0.004
IBK	0.008	**0.008**	**0.008**	**0.007**	**0.007**	**0.006**	**0.007**
M5P	0.003	**0.003**	**0.003**	**0.003**	**0.003**	**0.003**	0.004
M5Rules	0.003	**0.003**	**0.003**	**0.003**	**0.003**	**0.003**	0.004

2.2.7 Results for Slice locality data set

Slice locality data set consists of 384 features extracted from 53500 CT images. The CT images are from 74 different patients (43 male, 31 female). The class variable of this data set is location of the CT slice on the axial axis of the human body [64]. This data set is available on UCI Machine Learning Repository.

Results for 10 subjects of Slice locality data set are presented in Tables 2.23-2.26. In these tables regression error obtained by regression algorithms are given. The second line in all tables contains number of features of original data and after feature selection. Table 2.23 presents results for all

subjects using IBK algorithm. One can see that the IBK algorithm achieved the better accuracy for all subjects data set except subject number 10 using 380 features. Table 2.24 presents results for all subjects using Logistic regression algorithm. The use of the proposed algorithm allows to improve performance of Logistic regression using 250 features for Subject 1 and 150 features for Subjects 2 and 3. The best performance for Subject 5 achieved using 100 features. Results are almost the same for other Subjects.

Tables 2.25 and 2.26 show results for all patients using M5P and M5Rules algorithms, respectively. Results for these two algorithms are very similar and one can see that the proposed algorithm can improve the accuracy of regression algorithms.

Table 2.23: IBK algorithm performance for 10 subjects from Slice locality data

	Original data	After feature selection						
Number of features	385	380	350	300	250	200	150	100
Patient1	0.059	**0.059**	**0.059**	0.060	0.061	0.065	0.063	0.083
Patient2	0.080	**0.080**	0.081	0.081	0.082	0.083	0.085	0.103
Patient3	0.076	**0.076**	**0.076**	**0.075**	**0.076**	0.077	0.086	0.115
Patient4	0.060	**0.060**	**0.060**	0.061	0.062	0.063	0.066	0.081
Patient5	0.078	**0.078**	**0.078**	0.079	0.080	0.088	0.086	0.090
Patient6	0.349	**0.349**	**0.349**	**0.349**	**0.336**	**0.346**	0.456	0.466
Patient7	0.081	**0.081**	**0.081**	**0.081**	**0.081**	0.087	0.091	0.099
Patient8	0.087	**0.087**	**0.087**	**0.087**	**0.086**	**0.086**	0.093	0.099
Patient9	0.364	**0.364**	0.370	0.370	0.364	0.380	0.494	0.516
Patient10	0.098	**0.098**	0.100	0.104	0.103	0.105	0.110	0.139

Table 2.24: Logistic regression algorithm performance for 10 subjects from Slice locality data

	Original data	After feature selection						
Number of features	385	380	350	300	250	200	150	100
Patient1	0.354	0.392	**0.250**	0.267	0.284	0.326	0.411	0.570
Patient2	0.496	**0.435**	**0.398**	**0.367**	**0.332**	**0.309**	**0.376**	0.621
Patient3	0.258	**0.256**	0.266	**0.228**	**0.226**	**0.226**	**0.247**	0.361
Patient4	0.282	**0.294**	0.305	**0.281**	0.294	**0.269**	0.373	0.476
Patient5	0.928	1.742	2.413	**0.512**	**0.440**	**0.469**	**0.572**	**0.529**
Patient6	0.435	0.439	0.456	0.440	0.456	0.572	2.232	1.514
Patient7	0.515	**0.500**	**0.460**	**0.426**	**0.420**	**0.414**	**0.443**	0.756
Patient8	1.306	**1.272**	**1.275**	**1.275**	1.449	**1.234**	1.457	2.025
Patient9	0.549	**0.539**	0.567	0.532	**0.497**	0.860	1.857	7.839
Patient10	0.570	**0.565**	**0.513**	**0.522**	**0.508**	**0.492**	**0.506**	0.681

Table 2.25: M5P algorithm performance for 10 subjects from Slice localization data

	Original data	After feature selection						
Number of features	385	380	350	300	250	200	150	100
Patient1	0.299	**0.299**	0.301	**0.297**	**0.294**	**0.293**	**0.298**	0.338
Patient2	0.455	**0.455**	**0.440**	**0.443**	**0.441**	0.471	**0.451**	**0.452**
Patient3	0.352	**0.352**	**0.352**	**0.349**	0.358	**0.343**	**0.342**	**0.337**
Patient4	0.341	0.347	0.348	0.350	**0.339**	**0.310**	**0.319**	**0.325**
Patient5	0.458	**0.458**	**0.427**	**0.404**	**0.395**	**0.375**	**0.385**	**0.396**
Patient6	1.334	**1.297**	**1.289**	**1.326**	1.357	**1.136**	**1.229**	**1.291**
Patient7	0.472	**0.467**	**0.472**	**0.472**	**0.469**	0.476	0.475	0.490
Patient8	0.782	0.797	0.801	0.801	**0.720**	**0.744**	**0.728**	**0.728**
Patient9	1.214	**1.214**	**1.175**	**1.189**	**1.152**	**1.020**	1.683	1.754
Patient10	0.561	**0.546**	**0.542**	**0.513**	**0.513**	**0.519**	**0.509**	**0.519**

Table 2.26: M5Rules algorithm performance for 10 subjects for Slice localization data

|	Original data	After feature selection						
Number of features	385	380	350	300	250	200	150	100
Patient1	0.331	0.319	**0.313**	0.368	0.370	**0.322**	**0.272**	2.217
Patient2	0.455	**0.455**	**0.360**	**0.339**	**0.347**	0.557	**0.445**	0.490
Patient3	0.508	**0.508**	**0.508**	**0.477**	**0.432**	**0.413**	**0.388**	**0.420**
Patient4	0.328	**0.307**	**0.311**	**0.328**	0.333	0.294	**0.309**	**0.317**
Patient5	0.481	**0.479**	**0.410**	0.507	0.508	**0.458**	0.492	**0.412**
Patient6	1.562	**1.320**	**1.231**	**1.313**	**1.338**	**1.030**	**1.480**	**1.242**
Patient7	0.783	**0.783**	0.784	**0.783**	**0.559**	**0.500**	**0.412**	**0.611**
Patient8	0.686	**0.687**	0.696	0.696	0.853	0.822	0.755	2.506
Patient9	1.476	**1.476**	**1.220**	**1.249**	**1.162**	**1.260**	**0.968**	1.952
Patient10	0.815	**0.693**	**0.727**	**0.714**	**0.688**	-	1.926	0.586
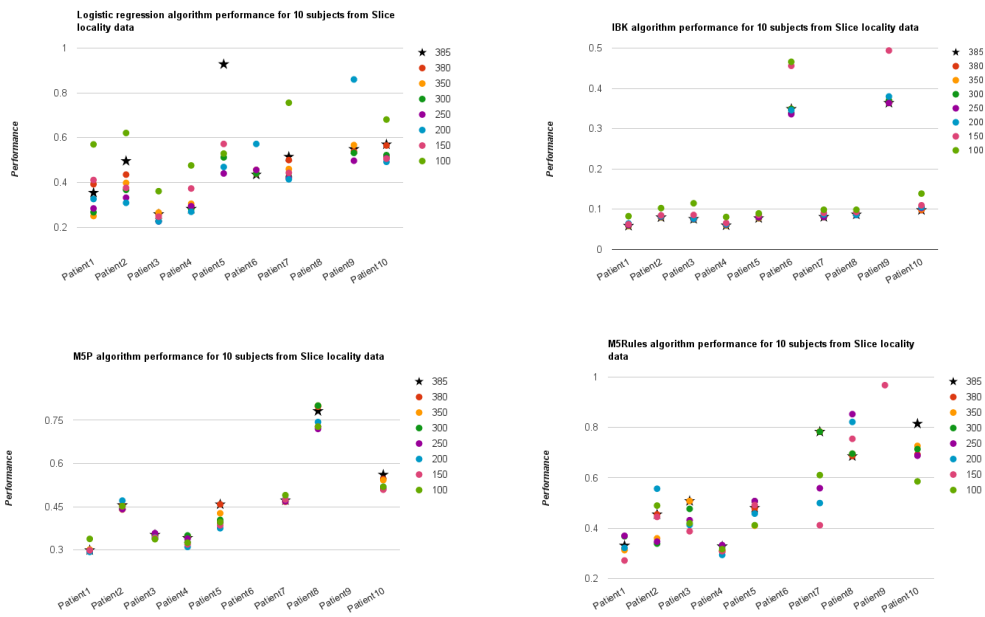
Figure 2.9: Classification algorithms performance for 10 subjects of Slice localization data

Results for Slice locality data set is presented in Figure 2.9. In the right hand side of the graphs number of selected features are shown (star is related to the number of features in the original data set). The use of the proposed algorithm allows to improve performance of algorithms. The IBK and Logistic regression algorithms achieved the better accuracy for all subjects data set except subject number 10. Results for M5P and M5Rules algorithms are very similar and one can see that the proposed algorithm can improve the accuracy of regression algorithms.

2.2.8   Results for Parkinsons Telemonitoring data set

In this section we present the results for Parkinsons Telemonitoring data set. This data set composed of a range of biomedical voice measurements from 42 people with early-stage Parkinson's disease. Here we analyzed 15 subjects from this data set. Results for subjects of Parkinsons Telemonitoring data set are presented in Tables 2.27-2.30. This is illustration of a number of features in original data and after feature selection. Number of features in original data is 18.

Table 2.27 shows the results for the error of the data using IBK regressor algorithm. The use of very small subset of features can provide better performance for almost all subjects. Table 2.28 presents the results for Logistic regression algorithm. The proposed algorithm can reduce the error for more than 70% of cases. The situation is almost the same for the M5P algorithm 2.29, but M5Rulles algorithm provides better performance and the accuracy is increased for all subjects except Subjects 14 and 15.

Table 2.27: IBK algorithm performance for Parkinson's disease data

	Original data	After feature selection							
Number of features	**18**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**
Subject1	0.037	0.038	**0.037**	0.038	0.038	0.040	0.044	0.041	0.042
Subject2	0.039	**0.037**	0.039	0.038	0.039	0.040	**0.036**	0.040	0.042
Subject3	0.030	**0.027**	0.027	0.027	0.027	0.026	0.029	0.027	0.027
Subject4	0.039	**0.034**	0.034	0.035	0.035	0.036	0.034	0.035	0.037
Subject5	0.037	**0.033**	0.032	0.032	0.030	0.029	0.029	0.030	0.031
Subject6	0.034	0.037	**0.035**	0.033	0.033	0.033	0.031	0.031	0.031
Subject7	0.040	**0.033**	0.033	0.034	0.030	0.034	0.035	0.036	0.035
Subject8	0.032	**0.031**	0.033	0.034	**0.032**	0.033	0.036	0.036	0.036
Subject9	0.041	**0.038**	0.038	0.039	0.039	0.037	0.036	0.036	0.039
Subject10	0.044	0.037	0.039	0.039	0.044	0.046	0.044	0.042	0.040
Subject11	0.022	**0.022**	0.022	0.021	0.021	0.021	0.020	0.021	0.023
Subject12	0.030	**0.024**	0.024	0.028	0.029	0.030	0.032	0.032	**0.030**
Subject13	0.040	0.042	0.044	0.042	0.047	**0.039**	0.051	0.049	0.049
Subject14	0.032	**0.030**	0.030	0.031	0.031	0.031	0.031	0.033	0.033
Subject15	0.032	**0.031**	0.030	0.030	0.032	0.031	0.030	0.032	0.032

Table 2.28: Linear regression algorithm performance for Parkinson's disease data

Number of features	Original data	After feature selection							
	18	11	10	9	8	7	6	5	4
Subject1	0.030	**0.028**	**0.028**	**0.028**	**0.028**	**0.028**	**0.028**	**0.029**	**0.029**
Subject2	0.028	**0.028**	0.030	0.031	0.030	0.030	0.030	0.030	0.030
Subject3	0.018	0.019	**0.018**	0.020	0.020	0.020	0.022	0.021	0.021
Subject4	0.029	**0.028**	**0.027**	**0.027**	**0.027**	**0.027**	**0.027**	**0.027**	**0.028**
Subject5	0.024	0.025	0.025	0.025	0.025	0.026	0.026	0.026	0.027
Subject6	0.024	0.025	**0.024**	0.025	0.025	0.025	0.025	0.025	0.025
Subject7	0.024	**0.024**	**0.024**	**0.023**	**0.023**	**0.024**	**0.024**	**0.024**	0.025
Subject8	0.027	0.031	0.035	0.034	0.034	0.034	0.033	0.031	0.034
Subject9	0.029	**0.029**	0.030	0.030	0.030	0.030	0.037	0.037	0.038
Subject10	0.033	**0.033**	**0.033**	**0.033**	**0.032**	**0.032**	**0.032**	**0.032**	0.034
Subject11	0.017	**0.017**	**0.016**	**0.016**	**0.016**	**0.016**	**0.016**	**0.017**	**0.017**
Subject12	0.019	**0.018**	**0.017**	0.021	0.021	0.021	0.020	0.020	0.021
Subject13	0.031	**0.030**	**0.031**	**0.030**	0.032	0.033	0.033	0.033	0.035
Subject14	0.024	**0.020**	**0.019**	**0.019**	**0.019**	**0.020**	**0.020**	**0.020**	0.027
Subject15	0.019	0.020	**0.018**	**0.018**	**0.018**	**0.018**	0.021	0.021	0.022

Table 2.29: M5P algorithm performance for Parkinson's disease data

Number of features	Original data	After feature selection							
	18	11	10	9	8	7	6	5	4
Subject1	0.030	**0.028**	**0.028**	**0.028**	**0.028**	**0.028**	**0.029**	**0.029**	**0.029**
Subject2	0.027	0.028	0.029	0.030	0.030	0.030	0.030	0.030	0.030
Subject3	0.018	0.019	**0.018**	0.020	0.020	0.020	0.022	0.021	0.021
Subject4	0.028	**0.027**	**0.027**	**0.027**	**0.025**	**0.027**	**0.027**	**0.027**	**0.028**
Subject5	0.024	**0.023**	**0.023**	**0.023**	**0.023**	**0.022**	**0.022**	**0.023**	**0.024**
Subject6	0.025	**0.025**	**0.024**	**0.025**	**0.025**	**0.025**	**0.025**	**0.025**	**0.025**
Subject7	0.024	**0.024**	**0.024**	**0.023**	**0.023**	**0.024**	**0.023**	**0.024**	0.025
Subject8	0.024	0.026	0.029	0.029	0.030	0.030	0.029	0.030	0.030
Subject9	0.029	**0.029**	**0.028**	**0.028**	**0.029**	**0.029**	0.031	0.031	0.031
Subject10	0.034	**0.034**	**0.034**	**0.034**	**0.032**	**0.032**	**0.032**	**0.032**	**0.032**
Subject11	0.017	**0.017**	**0.016**	**0.016**	**0.016**	**0.016**	**0.016**	**0.017**	**0.017**
Subject12	0.020	**0.018**	**0.017**	0.021	0.021	0.021	**0.020**	**0.020**	0.021
Subject13	0.033	**0.031**	**0.031**	**0.031**	0.032	0.033	0.033	0.033	0.035
Subject14	0.019	**0.019**	**0.019**	**0.019**	**0.019**	0.021	0.020	0.020	0.023
Subject15	0.019	0.021	**0.019**	**0.019**	**0.019**	**0.019**	0.023	0.023	0.022

Table 2.30: M5Rules algorithm performance for Parkinson's disease data

| Number of features	Original data	After feature selection							
	18	11	10	9	8	7	6	5	4
Subject1	0.030	**0.028**	**0.029**	**0.029**	**0.028**	**0.028**	**0.029**	**0.029**	**0.029**
Subject2	0.027	0.029	0.029	0.031	0.030	0.030	0.030	0.030	0.030
Subject3	0.019	**0.019**	**0.018**	0.020	0.020	0.021	0.022	0.021	0.021
Subject4	0.029	**0.028**	**0.028**	**0.027**	**0.026**	**0.028**	0.030	0.030	**0.028**
Subject5	0.025	**0.023**	**0.023**	**0.023**	**0.023**	**0.022**	**0.022**	**0.023**	**0.024**
Subject6	0.024	0.025	**0.024**	0.025	0.025	0.025	0.025	0.025	0.025
Subject7	0.024	**0.024**	**0.024**	**0.023**	**0.023**	0.025	**0.024**	**0.024**	0.025
Subject8	0.029	0.038	0.043	0.043	0.031	0.031	0.044	0.045	0.031
Subject9	0.031	**0.030**	**0.029**	**0.029**	**0.030**	**0.031**	0.032	0.032	**0.031**
Subject10	0.035	**0.035**	**0.035**	**0.034**	**0.033**	**0.033**	**0.033**	**0.033**	**0.033**
Subject11	0.017	**0.017**	**0.016**	**0.016**	**0.016**	**0.016**	**0.016**	**0.017**	**0.017**
Subject12	0.019	**0.018**	**0.017**	0.021	0.021	0.021	0.020	0.020	0.021
Subject13	0.033	**0.031**	**0.031**	**0.031**	0.032	0.033	0.033	0.033	0.035
Subject14	0.019	0.020	0.020	**0.019**	**0.019**	0.021	0.020	0.020	0.023
Subject15	0.020	0.021	**0.020**	**0.020**	**0.020**	**0.020**	0.023	0.024	0.022
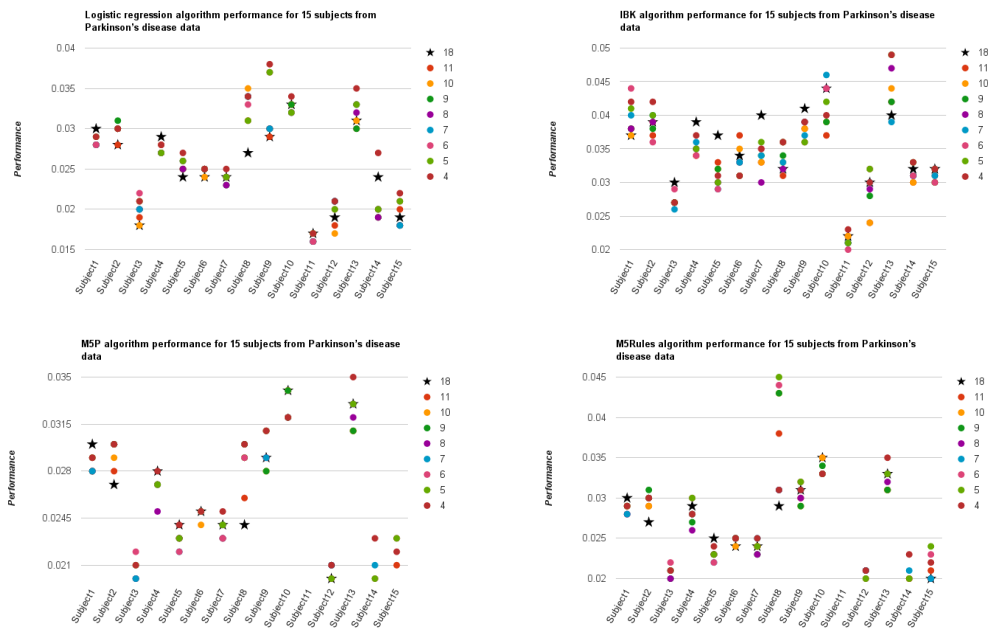
Figure 2.10: Regression analysis algorithms performance for Parkinson's disease data using all features and subset of features obtained by the cusp catastrophe feature selection algorithm

Figure 2.10 illustrates application of different regression analysis algorithms for Parkinson's disease data set. Figure 2.10 indicates that the cusp model reduced the error of algorithms for almost all subjects from Parkinson's disease data set.

Figures 2.11 show the Equilibrium surface (3 dimensional) and control surface (2 dimensional) of fitting the most irrelevant (left) and the most significant features in different data sets using the Cusp Catastrophe model. The informative features have more affect to the system and put the system closer to the bifurcation situation.

Tables 2.31- 2.36 show ranking of the features using the proposed and RELIEF algorithms. The ranking values are not exactly the same, but the for almost all cases the informative features' levels are similar in both ranking results. For example, for the first subject the informative features of 3, 14, 4 and 6 are in the top of the table in both algorithms and less-significant features 2 and 17 are in the bottom.

Table 2.31: Ranking of the features using the proposed and RELIEF algorithms for subject 1 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
3	0.003144	14	0.030901
14	0.003096	3	0.014302
4	0.003052	6	0.014158
6	0.002947	4	0.011554
15	0.002923	5	0.009576
5	0.002732	7	0.009572
7	0.002731	15	0.006487
9	0.002685	12	0.004949
12	0.002586	16	0.004764
16	0.002569	9	0.004034
8	0.002565	11	0.001722
11	0.002564	13	0.0016
10	0.0025	10	0.001595
13	0.0025	2	0.000525
17	0.002358	8	-0.00004
2	0.002351	1	-0.00254
1	0.002351	17	-0.00378

Table 2.32: Ranking of the features using the proposed and RELIEF algorithms for subject 2 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
16	0.003336	14	0.0195
6	0.003299	6	0.00943
14	0.003241	16	0.00907
9	0.003212	12	0.00784
12	0.0032	9	0.00706
4	0.003173	3	0.00582
3	0.003167	4	0.00309
15	0.003154	15	0.003
8	0.003134	11	0.0026
13	0.003069	13	0.00244
10	0.003069	10	0.00244
11	0.003057	7	0.00243
7	0.00304	5	0.00242
5	0.00304	8	0.0023
2	0.002727	2	0.00133
1	0.002727	1	0.00107
17	0.002718	17	-0.00237

Table 2.33: Ranking of the features using the proposed and RELIEF algorithms for subject 3 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
15	0.003585	15	0.024669
6	0.003473	14	0.018446
3	0.003261	6	0.016579
4	0.003031	3	0.013203
14	0.002994	4	0.010286
7	0.002946	5	0.007498
9	0.002946	7	0.00748
5	0.002945	11	0.005778
12	0.002937	12	0.003904
8	0.002934	9	0.00329
11	0.002933	1	0.003219
10	0.00287	8	0.002655
13	0.002869	10	0.002304
16	0.002627	13	0.002297
1	0.002595	17	0.002161
2	0.002589	2	0.000729
17	0.002565	16	-0.00093
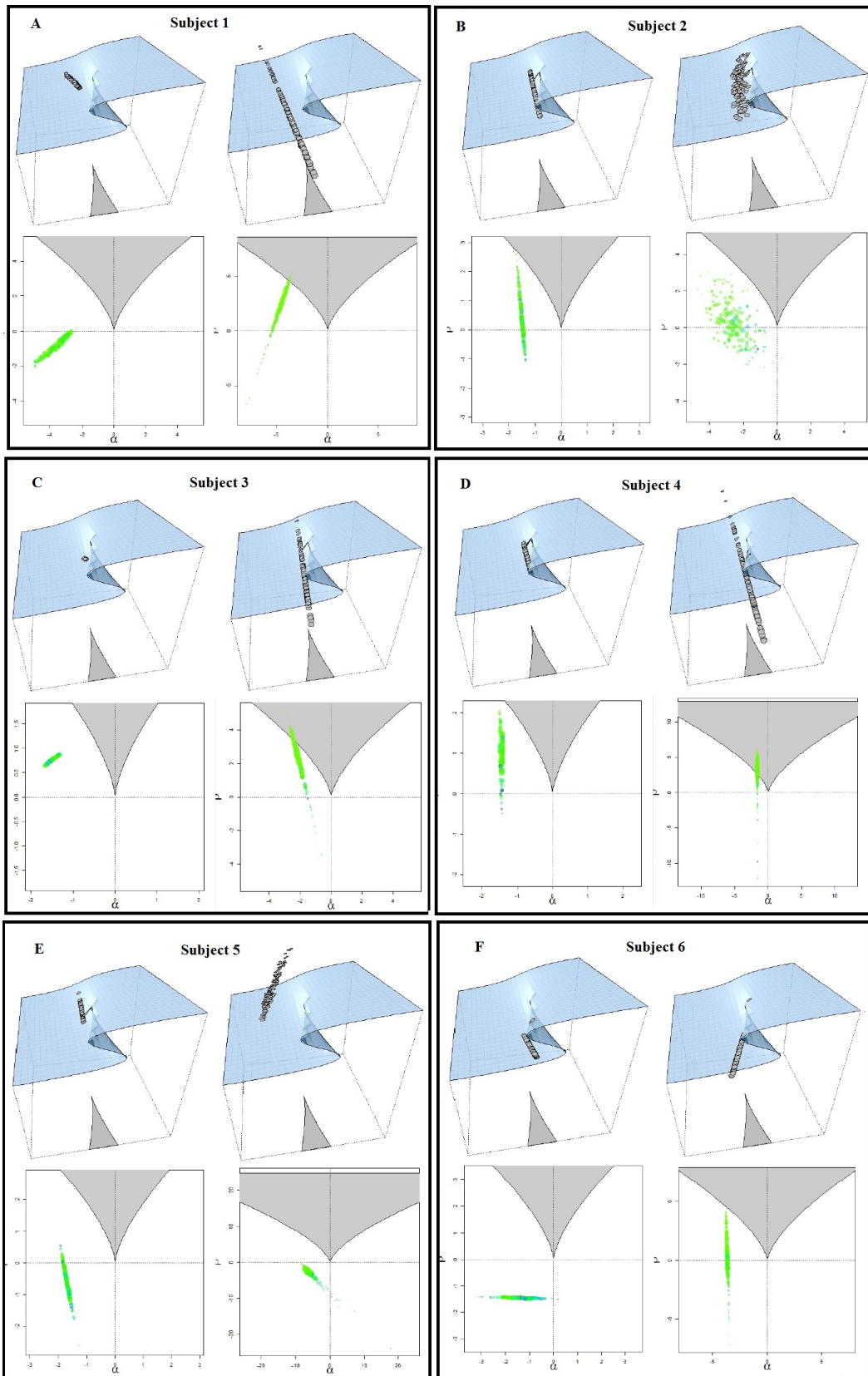
Figure 2.11: Cusp plot: The most least informative features (left) and the most least informative features (right) base on proposed algorithm for subject 1 to subject 6

Table 2.34: Ranking of the features using the proposed and RELIEF algorithms for subject 4 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
3	0.004621	6	0.02566
4	0.00456	3	0.02124
6	0.004473	17	0.01921
5	0.003827	4	0.01823
7	0.003826	14	0.01734
14	0.003417	5	0.01714
15	0.003254	7	0.01711
9	0.002984	2	0.00843
8	0.002968	15	0.00774
13	0.002935	13	0.00711
10	0.002935	10	0.00711
12	0.00293	11	0.00695
11	0.00291	12	0.00676
17	0.002904	8	0.00671
16	0.002793	9	0.00613
1	0.002771	1	0.00519
2	0.00277	16	0.00168

Table 2.35: Ranking of the features using the proposed and RELIEF algorithms for subject 5 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
14	0.003896	14	0.02979
3	0.003671	6	0.02661
4	0.003533	4	0.02327
6	0.003529	3	0.01819
7	0.003189	7	0.01289
5	0.003185	5	0.01287
15	0.003059	9	0.01101
16	0.00253	15	0.0101
9	0.00248	12	0.00659
12	0.002401	11	0.00414
8	0.002372	10	0.00354
11	0.002363	13	0.00354
10	0.002343	8	0.00331
13	0.002343	16	0.00278
2	0.002339	2	0.00244
1	0.002324	17	0.00116
17	0.002314	1	-0.00406

Table 2.36: Ranking of the features using the proposed and RELIEF algorithms for subject 6 from Parkinsons disease data

Feature selection algorithm based on the Cusp model		RELIEF algorithms	
Attribute ID	Rank	Attribute ID	Rank
15	0.003297	14	0.014851
4	0.003173	6	0.014336
3	0.003093	15	0.014142
6	0.003076	17	0.01388
14	0.003062	4	0.012454
7	0.002854	3	0.010648
5	0.002854	7	0.008541
9	0.002691	5	0.008525
12	0.002649	2	0.003976
8	0.002644	12	0.002502
11	0.002619	1	0.001973
16	0.002597	11	0.001794
10	0.002565	9	-6.8E-05
13	0.002565	8	-0.0006
1	0.002418	13	-0.00153
2	0.00234	10	-0.00153
17	0.002315	16	-0.00244

Tables 2.37-2.44 show the mean absolute error and root mean square error for Regression analysis before and after feature selection for 15 subjects. We separated the results of different algorithms from each other. Tables 2.37 and 2.38 shows the results of Linear regression algorithm. The accuracy of analyzing all subjects except subject 2, 9 and 14 using the proposed algorithm compared with original data is improved. The RELIEF algorithm has improvement for almost all subjects, but our algorithm has better performance than RELIEF algorithm.

Tables 2.39-2.40 are the related results for K-nearest neighbors algorithm and they show that both algorithms have better accuracy only for 60% of subjects and the same situation happened for M5Rulles (see the tables 2.41-2.42) and REPTree (2.43-2.44) algorithms, but for some subjects the RELIEF algorithm has better performance.

Table 2.37: Mean absolute error of Linear regression algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

	MAE of Linear Regression		
Subject	Original data	After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithms
1	0.0295	0.0291	0.0282
2	0.0276	0.028	0.028
3	0.0183	0.0183	0.0182
4	0.0292	0.029	0.0292
5	0.0235	0.0235	0.0235
6	0.0239	0.0239	0.024
7	0.0243	0.0242	0.0244
8	0.0266	0.0266	0.028
9	0.0286	0.0288	0.0288
10	0.0333	0.0333	0.0333
11	0.0169	0.0167	0.017
12	0.0193	0.0187	0.0194
13	0.0305	0.0297	0.0315
14	0.019	0.0193	0.0188
15	0.0266	0.0261	0.0266

Table 2.38: Root mean square error of Linear regression algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

| Subject | Original data | RMSE of Linear Regression | |
		After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0386	0.0381	0.0384
2	0.0372	0.0377	0.0377
3	0.0249	0.0249	0.0248
4	0.042	0.0418	0.042
5	0.0336	0.0336	0.0336
6	0.0325	0.0325	0.0322
7	0.0338	0.0338	0.0335
8	0.0401	0.0401	0.0424
9	0.0376	0.0377	0.0375
10	0.0472	0.0472	0.0461
11	0.0239	0.0237	0.024
12	0.025	0.0245	0.0256
13	0.0404	0.0392	0.0425
14	0.0248	0.0253	0.0246
15	0.0322	0.0317	0.0319

Table 2.39: Mean absolute error of IBK algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

| Subject | Original data | MAE of IBK | |
		After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.037	0.038	0.042
2	0.0389	0.0411	0.0411
3	0.0304	0.0297	0.0311
4	0.0394	0.0387	0.0372
5	0.0369	0.0344	0.0356
6	0.034	0.0334	0.0355
7	0.0404	0.0385	0.0389
8	0.0321	0.032	0.032
9	0.0405	0.0399	0.0399
10	0.0439	0.044	0.0433
11	0.0218	0.0231	0.0224
12	0.0297	0.0295	0.0308
13	0.0402	0.0411	0.0402
14	0.0317	0.03	0.0307
15	0.0338	0.0352	0.0335

Table 2.40: Root mean square error of IBK algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

	RMSE of IBK		
Subject	Original data	After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0493	0.0506	0.0548
2	0.0526	0.0537	0.0537
3	0.0379	0.0379	0.0401
4	0.0569	0.0565	0.0567
5	0.0499	0.047	0.0477
6	0.0453	0.0447	0.0457
7	0.0527	0.0504	0.0507
8	0.0462	0.0458	0.0466
9	0.0531	0.0528	0.0536
10	0.056	0.0562	0.0538
11	0.0285	0.0316	0.029
12	0.0385	0.0399	0.0402
13	0.0533	0.0539	0.0532
14	0.038	0.0359	0.0378
15	0.0426	0.0435	0.0427

Table 2.41: Mean absolute error of M5Rules algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

	MAE of M5Rules		
Subject	Original data	After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0299	0.0299	0.0292
2	0.0273	0.0285	0.0276
3	0.0188	0.0181	0.0203
4	0.0291	0.0291	0.0278
5	0.0246	0.0248	0.0233
6	0.0241	0.024	0.024
7	0.0237	0.0233	0.0235
8	0.0286	0.0275	0.0262
9	0.0306	0.0319	0.0306
10	0.0349	0.0349	0.034
11	0.0167	0.0169	0.0175
12	0.019	0.0188	0.0197
13	0.0333	0.0313	0.032
14	0.0196	0.021	0.0209
15	0.0246	0.0249	0.0256

Table 2.42: Root mean square error of M5Rules algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

	RMSE of M5Rules		
Subject	Original data	After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0393	0.0393	0.0393
2	0.0366	0.0377	0.0366
3	0.0259	0.0252	0.0284
4	0.0423	0.0423	0.0415
5	0.0343	0.0345	0.0318
6	0.0327	0.0327	0.0327
7	0.033	0.0324	0.0328
8	0.0457	0.0444	0.0388
9	0.0403	0.0424	0.0401
10	0.0488	0.0488	0.0478
11	0.0219	0.0225	0.0235
12	0.0244	0.0244	0.026
13	0.044	0.041	0.0424
14	0.0258	0.0286	0.0275
15	0.0305	0.0305	0.0314

Table 2.43: Mean absolute error of REPTree algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

	MAE of REPTree		
Subject	Original data	After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0357	0.0357	0.0353
2	0.0344	0.0347	0.0347
3	0.0223	0.0228	0.0226
4	0.0312	0.0308	0.0304
5	0.0272	0.0273	0.0276
6	0.0278	0.028	0.0278
7	0.0273	0.0276	0.0276
8	0.03	0.0311	0.03
9	0.0387	0.0381	0.0387
10	0.0358	0.0358	0.0349
11	0.0183	0.018	0.0184
12	0.0261	0.0267	0.0261
13	0.043	0.043	0.0428
14	0.0263	0.0263	0.0263
15	0.0288	0.0289	0.0293

Table 2.44: Root mean square error of REPTree algorithm after feature selection using the proposed and RELIEF algorithms for Slice locality data set

| Subject | Original data | RMSE of REPTree | |
		After Feature selection using feature selection algorithm based on the Cusp model	After feature selection using RELIEF algorithm
1	0.0458	0.0458	0.0453
2	0.0449	0.0448	0.0448
3	0.0284	0.0288	0.0288
4	0.0449	0.0446	0.0437
5	0.0363	0.0363	0.0367
6	0.0371	0.038	0.0371
7	0.0379	0.0387	0.0383
8	0.0506	0.0547	0.0506
9	0.0519	0.0513	0.052
10	0.0458	0.0458	0.0454
11	0.0251	0.025	0.0252
12	0.0336	0.0356	0.0335
13	0.0538	0.0538	0.0538
14	0.0339	0.0338	0.0339
15	0.0362	0.0363	0.0366

Figure 2.12 provides a comparison between proposed algorithm and the well known RELIEF algorithm for Slice locality data set. Mean absolute error and root mean square error of four classifiers of original data and after feature selection are shown in the Figures. Graphs show that the proposed algorithm improved the accuracy of classification algorithms for almost all subjects using different classifiers.
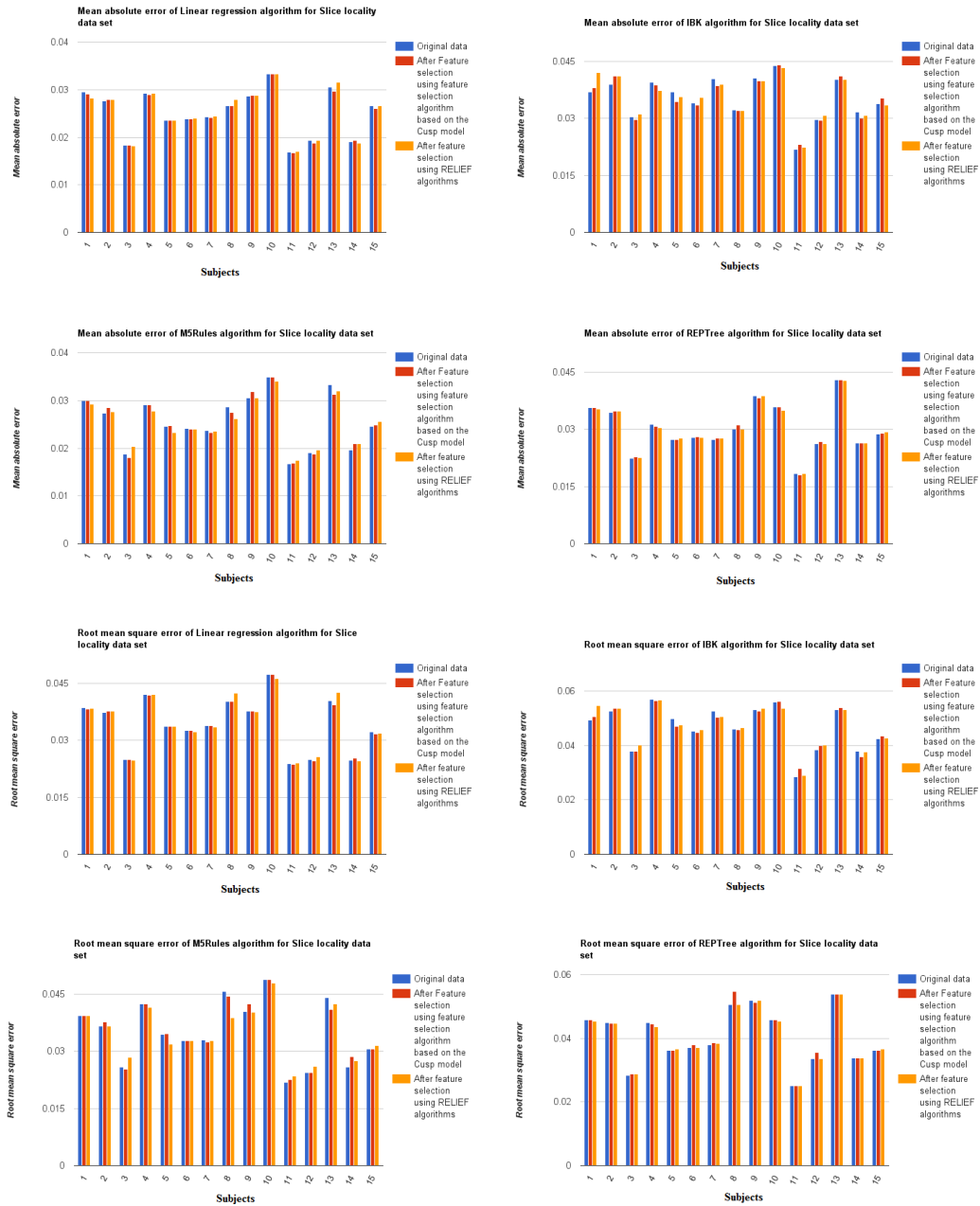
Figure 2.12: Mean square error and root mean square error of classifiers after feature selection using the proposed and RELIEF algorithms for Slice locality data set

2.3   Conclusions

In this chapter we introduced feature selection algorithms for brain data sets. First algorithm selects most significant voxels (features) in fMRI data sets using hyperrectangles. This algorithm uses the n-dimensional rectangles to approximate each class for the given subset of voxels and computes overlaps between classes. Voxels (features) or groups of voxels providing smallest overlaps are identified as the most informative voxels. This algorithm is applied to find a group of most informative voxels in Haxby and Science 2008 data set. Various classifiers from WEKA are used to evaluate classification error for each subset of voxels. Results show that the proposed algorithm allows one significantly decrease the number of voxels and improve the classification accuracy of most of classifiers. These results demonstrate that the proposed algorithm is efficient in finding subset of informative voxels in fMRI data sets.

Second algorithm removes the irrelevant or redundant features of a regression data sets. This algorithm selects significant features based on their fitting to the Catastrophe model and the features that better change the dynamics of the outcome feature or features are considered as informative features. The Akaike information criterion value of the Cusp model is computed for ranking of each feature. We applied this algorithm to three different data sets: Parkinson's Telemonitoring, Breast Cancer and Slice locality from UCI machine learning repository. Results show that the proposed algorithm is efficient in finding significant subset of features in a data set.

Chapter 3

Spike discharge prediction based on Neuro-fuzzy system

This chapter presents the development and evaluation of different versions of adaptive neuro-fuzzy model for prediction of spike discharge. We aim to predict the spike discharge variation using first spike latency and frequency-following interval. The use of animals like cat and rat because of similarity of their brain with human brain is common method for studying spike discharge [61, 145]. For study of spike discharge, we analyzed the Cerebral Cortex data of the Cat [187]. For analysis of the Cat data we applied the following algorithms: Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Wang and Mendel (WM), Dynamic evolving neural-fuzzy inference system (DENFIS), Hybrid neural Fuzzy Inference System (HyFIS), genetic for lateral tuning and rule selection of linguistic fuzzy system (GFS.LT.RS) and subtractive clustering and fuzzy c-means (SBC). Among all these algorithms ANFIS and genetic for lateral tuning and rule selection of linguistic fuzzy system models have better performance.

3.1   Introduction

Recording action potentials (spikes) from the neural cells makes it possible to investigate their health, stability and sensitivity[88]. Different characteristics of electrical activity of neurones can be considered in the study of neural coding. One important concept in this area is spike discharge that is a type of transient waveforms present in the brain activity and include a high correlation with seizure

occurrence [171].

Studies on movement indole illustrated that this process is related to the neuronal discharge [60, 90]. For example, study on activity of arm-related neurons and their relationship between premotor cortical cell activity and direction of arm movement shows that the cells activity vary in orderly fashion with the direction of movement [34]. Also, detection of spike discharge in the electroencephalogram is an important way of diagnosis of the disease [171]. Different algorithms like neural networks, logistic regression and neuro-fuzzy model can be applied for detection of epileptic seizure [135, 171, 173]. There are many similarity between human and animal brain's neural coding and many studies used animal modeling for investigation the spike discharge ([36, 109, 145, 163]). Johnsen et al. [88] analyzed twenty-six pairs of units recorded from twenty-four retinal ganglion cells in the isolated goldfish retina and examined the cross-correlation histogram for the maintained discharge of each pair of cells. Their results showed that it is unlikely that differences in latency could be attributed to unequal effectiveness of the stimuli for the two units. Batuev et al. [23] investigated the postsynaptic response of motor cortex neurons of the cat in response to the stimulation of different modalities and showed that it responds with a wide range of peripheral inputs. The electrical changes in the cerebral cortex can correspond with the electric changes in muscle and nerve [3]. The studies of the functional organisation of the motor cortex show that this cortical area is composed of modules consisting of columnar aggregates of neurones related to different aspects of the same movement [94]. The current-flow and current-source-density analysis of the direct cortical response in the somatosensory cortex of rats show that the activation and magnitude of direct cortical response depends on stimulus strength and frequency [72].

In this chapter the variation of spike discharge as a function of first spike latency and frequency-following interval is analyzed. First spike latency is the time delay between stimulus onset and first action potential [63]. Neuro-fuzzy model is a combination of artificial neural network (ANN) and fuzzy logic approaches. It is a powerful tool for dealing with uncertainty, and widely used for analyzing electrical activity of neurons. It is widely used for analyzing of the electrical activity of the neurones ([67, 80, 143, 172]). The ANFIS method was successfully applied for EEG signals with high accuracy of the results obtained [67]. A feature extraction method through the time-series prediction based on ANFIS model for brain computer interface applications has been proposed by Hsu [80]. In this model ANFISs is used for prediction of time-series for the left and right motor imagery

classification, respectively. It is shown that neuro-fuzzy is an accurate model diagnosing epilepsy [172].

Different versions of neuro-fuzzy model have been used to find the model with higher accuracy. In all models the spike discharge is considered as an output of the model, while first spike latency and spike frequency are considered as inputs. Using neuro-fuzzy model as a predictor of spike discharge, we are able to use insufficient crisp inputs to make accurate decision about spike discharge. We used first spike latency and frequency-following interval in input layer of the neuro-fuzzy system and output was the spike discharge. The structure of this chapter is as follows.

First we discuss about spike discharge, latency and frequency. Section 3.2 provides a brief description of ANFIS, WM, DENFIS, HyFIS and SBC algorithms and section 3.3 presents performance of different neuro-fuzzy algorithms for analysis of cat data.

3.2 Neuro-fuzzy model

Neuro-fuzzy model is a combination of artificial neural networks and fuzzy logic and it uses capabilities of both models. It applies a neural networks structure and at the same time uses *if-then* rules in fuzzy systems. It uses prior knowledge to compute membership function and different learning algorithms of neural networks, including the back-propagation algorithm [180].

The different types of neuro-fuzzy systems used in this chapter are as follow:

- Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

- Wang and Mendel (WM)

- Dynamic evolving neural-fuzzy inference system (DENFIS)

- Hybrid neural Fuzzy Inference System (HyFIS) [96]

- genetic for lateral tuning and rule selection of linguistic fuzzy system (GFS.LT.RS) [7]

- subtractive clustering and fuzzy c-means (SBC) [37, 210]

Here we provide a short description each of them.

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) model is a well-known neuro-fuzzy system that implements a Sugeno fuzzy system and uses a t-norm and differentiable membership function [86, 177]. For a system with two rules we can build the following neuro-fuzzy structure.

For given two inputs x_0 and y_0 and corresponding linguistic labels A_i and B_i, each neuron in the first layer of neuro-fuzzy model transmit crisp signal to the next layer (Algorithm 2).

Algorithm 2 ANFIS model

This algorithm has two main stages, the forward and backward steps. The forward step has five layers as follows:

- First layer maps the crisp inputs using bell-shaped membership function as follows:

$$A_i(u) = exp\left[-\frac{1}{2}\left(\frac{u - a_{i1}}{b_{i1}}\right)^2\right]$$

and

$$B_i(u) = exp\left[-\frac{1}{2}\left(\frac{u - a_{i2}}{b_{i2}}\right)^2\right]$$

where $\{a_{i1}, a_{i2}, b_{i1}, b_{i2}\}$ is the parameters set.

- Second layer is responsible for fuzzification and each neuron in this layer determines the fuzzy degree received crisp input.

$$\alpha_1 = A_1(x_0) \times B_1(y_0) = A_1(x_0) \wedge B_1(y_0)$$

and

$$\alpha_2 = A_2(x_0) \times B_2(y_0) = A_2(x_0) \wedge B_2(y_0)$$

- Neurones in the third layer are correspond to fuzzy rules and receive inputs from fuzzification neurons in the second layer. The outputs of layer 3 are as follow:

$$\beta_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}$$

and

$$\beta_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2}$$

- Layer 4 or output membership layer combine all its inputs by using the fuzzy operation union

$$\beta_1 z_1 = \beta_1(a_1 x_0 + b_1 y_0)$$

and

$$\beta_2 z_2 = \beta_2(a_2 x_0 + b_2 y_0)$$

- The last layer is responsible for Defuzzification.

$$o = \beta_1 z_1 + \beta_2 z_2$$

In the backward process the errors are propagated backward and the parameters are updated by gradient descent technique.

Wang and Mendel (WM) model is another type of neuro-fuzzy system that developed by Wang and Mendel [202] that has high performance for regression tasks. First it divides input and outputs

into fuzzy region and assigns a membership function to each regions. Then finds a rule for each pair of input data. In the next step a degree is assigned to each rule. After assigning degrees, they are combined. The final rule is obtained after deleting redundant rules. Algorithm 3 provides more details about WM algorithm.

Algorithm 3 Wang and Mendel (WM)
***
Division numerical input and output data spaces into fuzzy regions
Generate fuzzy IF-THEN rules covering the training data
Determining a degree for each rule
Eliminating redundant rules and obtaining a final rule base

Dynamic evolving neural-fuzzy inference system (DENFIS) is another fuzzy inference systems that developed by Kasabov et al. [91]. Output of the system is based on m-most activated fuzzy rules and evolving clustering method is applied to determine the cluster center (Algorithm 4).

Algorithm 4 Dynamic evolving neural-fuzzy inference system (DENFIS) model
***
Choose cluster center from training data
Determine the cluster centers using the evolving clustering method
partition the input space and to find optimal parameters on the consequent part
Update the parameters on the consequent part

Hybrid neural Fuzzy Inference System (HyFIS) has two general steps for learning [96]. In the first step the Wang and Mendel is used for knowledge acquisition. In the second step the input vector is propagated forward in the network and parameter updating is performed using backpropagating the error using a gradient descending approach [181].

Algorithm 5 Hybrid neural Fuzzy Inference System (HyFIS)
***
Uses the techniques of Wang and Mendel to acquire the knowledge
Use gradient descent-based to learn parameters of the structure

---

GFS.LT.RS: GFS.LT.RS is proposed by R. Alcala et al. [7] that performs an evolutionary lateral tuning of membership functions in constructing FRBS model to obtain higher accurate linguistic models (algorithm 6).

Algorithm 6 genetic for lateral tuning and rule selection of linguistic fuzzy system (GFS.LT.RS)
***
Uses the Wang and Mendel to to construct the population
Evaluate the chromosome using Mean square error
Minimize the number of rules

Subtractive clustering and fuzzy c-means (SBC) [37, 210] is checking each data point's distance from all other data points to find the cluster centers. More details about SBC algorithm is provided in the Algorithm 7

Algorithm 7 Subtractive clustering and fuzzy c-means (SBC)

   Use subtractive clustering method to obtain the cluster centeres (generating the rules)
   Choose the highest potential as the cluster centere
   Update the potential of each data point
   Optimise the cluster centers using fuzzy c-means

3.3   computational results

To verify the effectiveness of the neuro-fuzzy algorithms we carried out a number of numerical experiments with cortex of the somatosensory/motor system of the Cat data set on a PC with Processor Intel(R) Core(TM) i5-3470S CPU 2.90 GHz and 8 GB RAM running under Windows XP. The cortex of the somatosensory/motor system of the Cat data is publicly available from [187]. This data is based on recording neurones of extracellularly in postcruciate cerebral cortex of cats. It is neuronal responsiveness of each of the four paws to strong cortical surface stimulation to understand facilitatory and inhibitory modulation of wide-field neurons by small-field neurones. Two groups of data from the Cerebral Cortex of the Cat data sets are considered for evaluation of the algorithms: Contralateral Forepaw (CF) Cortex (Chloralose) and Contralateral Hindpaw (CH) Cortex (Chloralose). The Contralateral Forepaw (CF) Cortex (Chloralose) is based on the measurements of 4,272 neurons, but Contralateral Hindpaw (CH) Cortex (Chloralose) contains data of 991 neurons. Various versions of neuro-fuzzy algorithms from R package are used to evaluate the algorithms' error for each data. The R Project for Statistical Computing is an environment for statistical computing and graphics that contains a comprehensive libraries of machine learning and statistical analysis applications that is available on [195].

3.3.1   Results of Forepaw (CF) Cortex (Chloralose) analysis

The Ipsilateral and Contralateral data from Forepaw Cortex data are considered for analysis. The results of application of neuro-fuzzy algorithm to the Ipsilateral Forepaw Cortex data are presented on the Figures 3.3.1-3.3.1. The first spike latency and ipsilateral forepaw frequency following interval (msec) are used as inputs, while mean spikes per discharge is used as output of the model. Each figure

contains the actual spikes per discharge value that is computed using neuro-fuzzy algorithm. Also, some statistics about the analysis is illustrated in each figure. The results show that the smallest Root Mean Square Error (RMSE) is obtained by HYFIS algorithm (RMSE=1.34) and the biggest RMSE is obtained by WM model (RMSE=2.72).
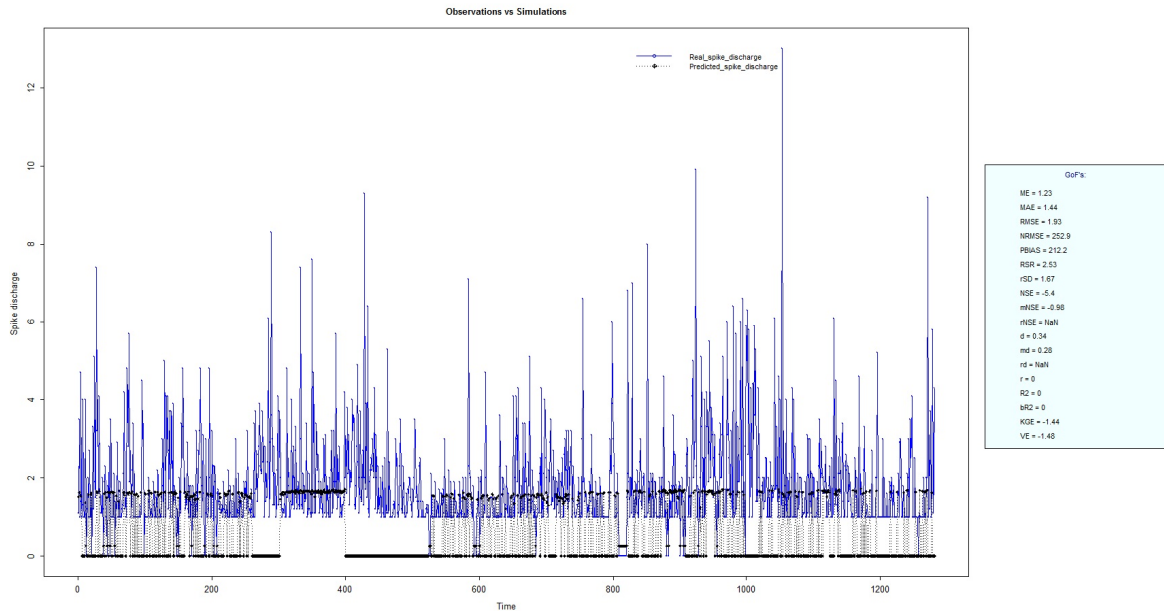
Figure 3.1: Spike Discharge prediction for cat Ipsilateral Forepaw Cortex using ANFIS algorithm

Figures 3.3.1-3.3.1 present the results of application for neuro-fuzzy algorithm for the Contralateral Forepaw Cortex data. Again the 1st spike latency and ipsilateral forepaw frequency following interval (msec) are used as inputs and mean spikes per discharge is used as output of the model. The results demonstrate that the smallest Root Mean Square Error (RMSE) is obtained using HYFIS algorithm (RMSE=0.93) and the biggest RMSE is obtained by WM model (RMSE=4.27).

3.3.2 Results of Hindpaw Cortex (Chloralose) data analysis

The Hindpaw Cortex is divided into two parts: the Contralateral Forepaw Cortex and Ipsilateral Hindpaw Cortex. Then different neuro-fuzzy algorithms have been applied to them. Figures 3.3.1-3.3.1 present the results of application of neuro-fuzzy algorithm for the Contralateral Forepaw Cortex data. The best RMSE is obtained using GFS LT RS (RMSE=2.06), the smallest Root Mean Square Error (RMSE) is obtained using HYFIS algorithm (RMSE=0.93), and the biggest RMSE is btained by WM model (RMSE=4.27).
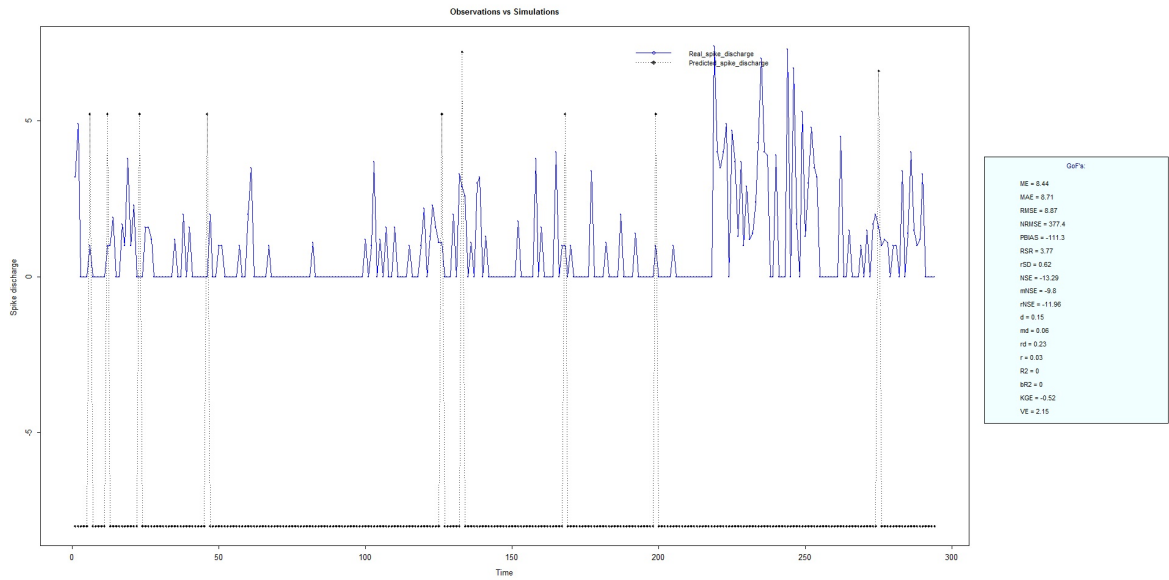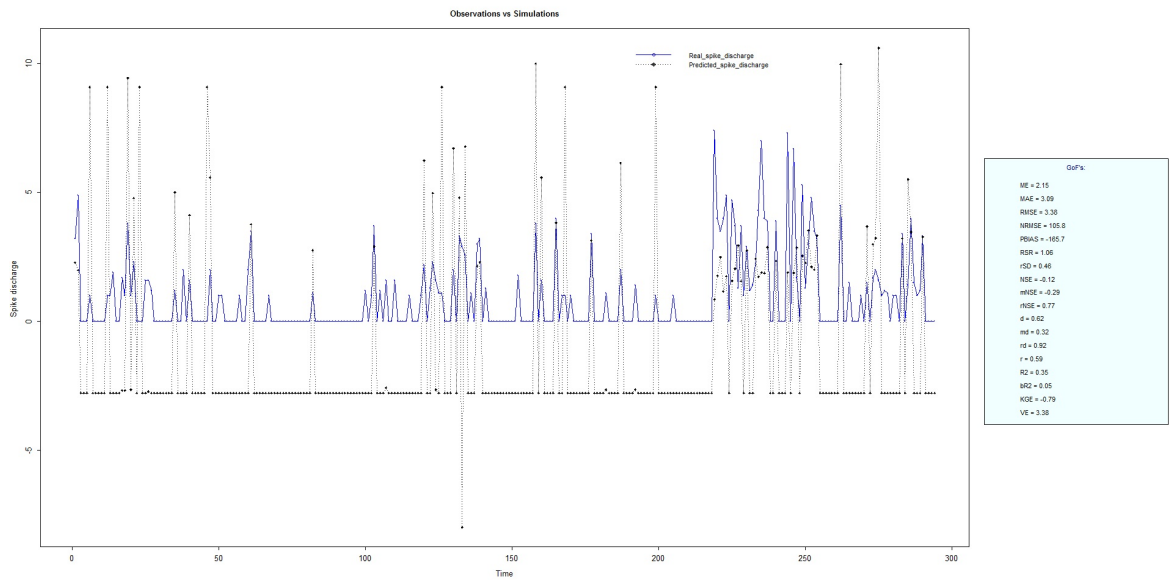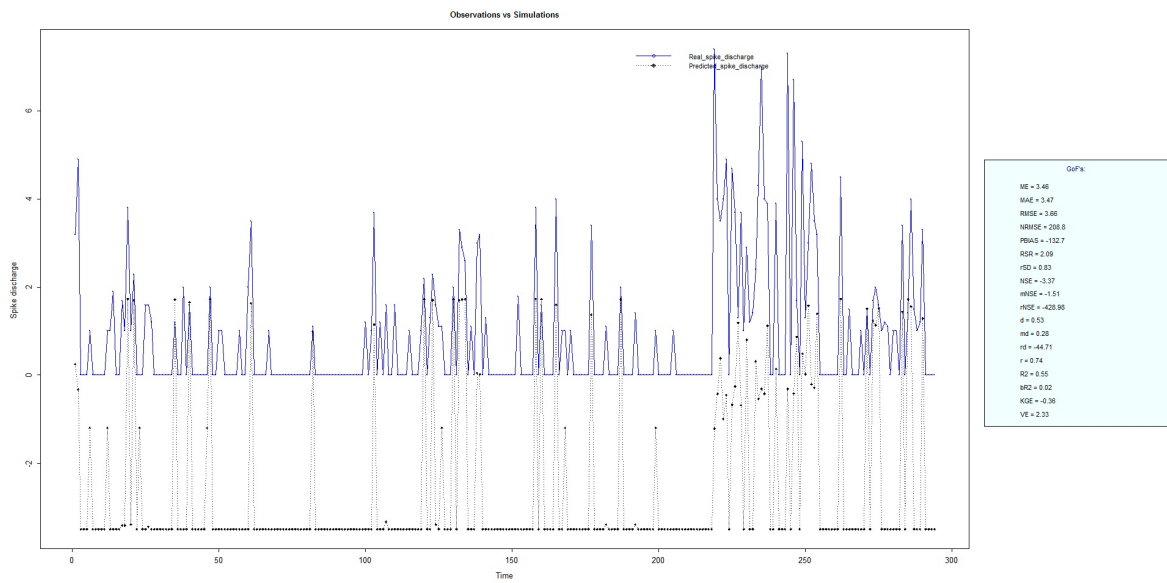
90
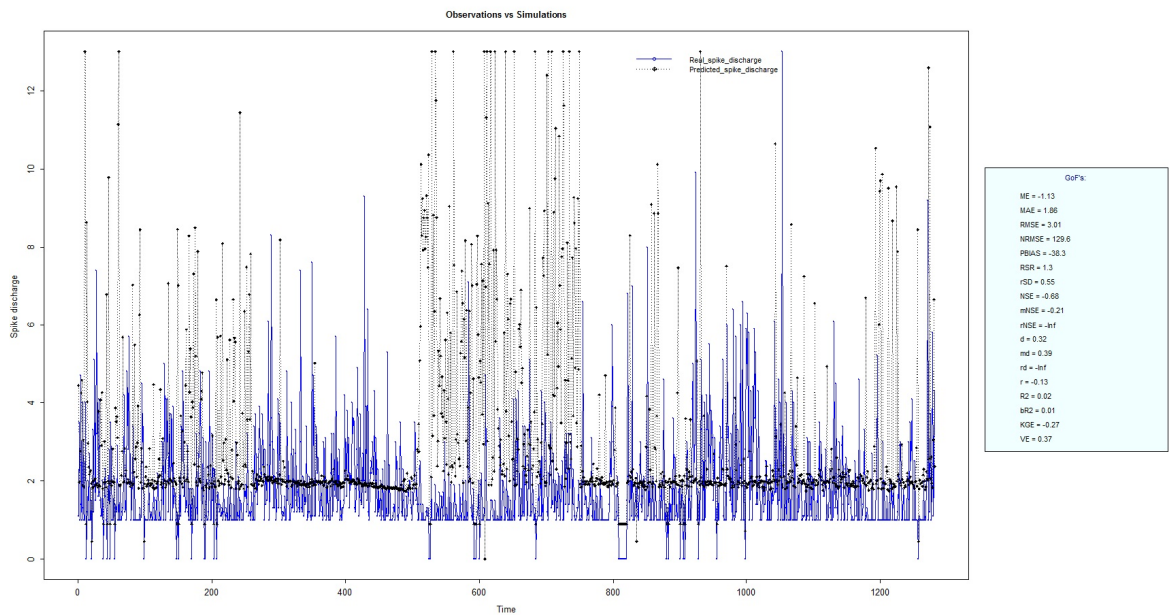
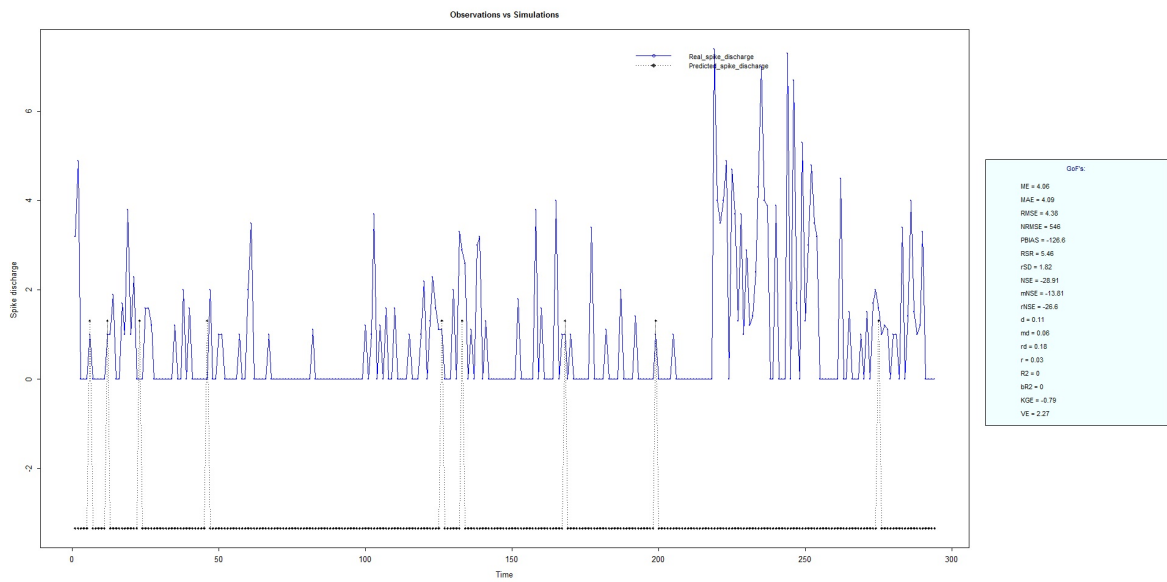Figure 3.2: Spike Discharge prediction for cat Ipsilateral Forepaw Cortex using Denfis algorithm
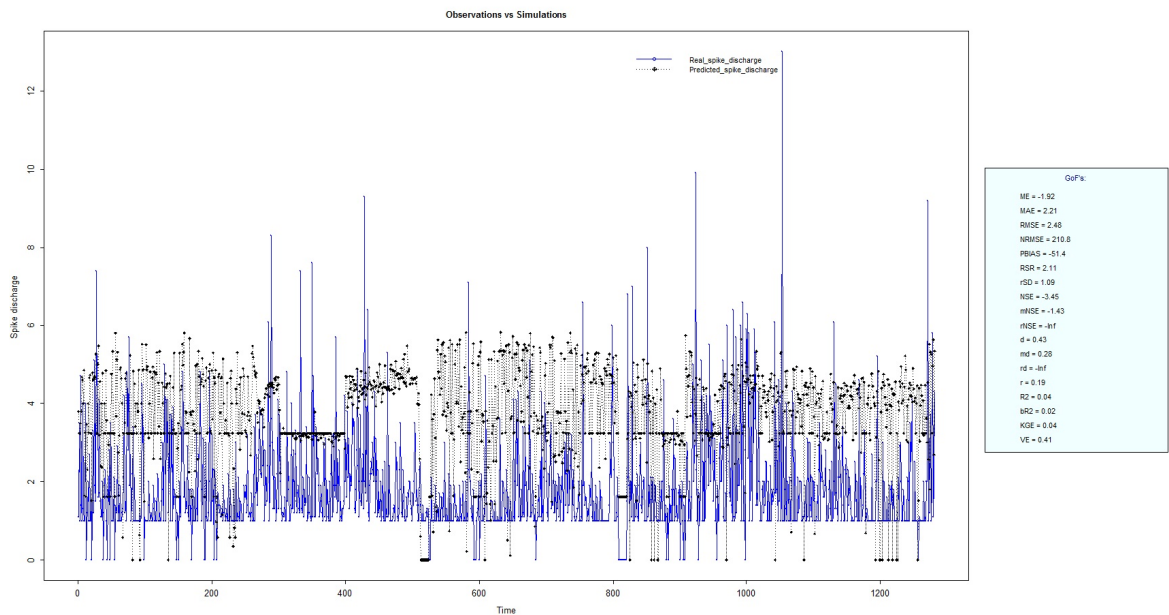
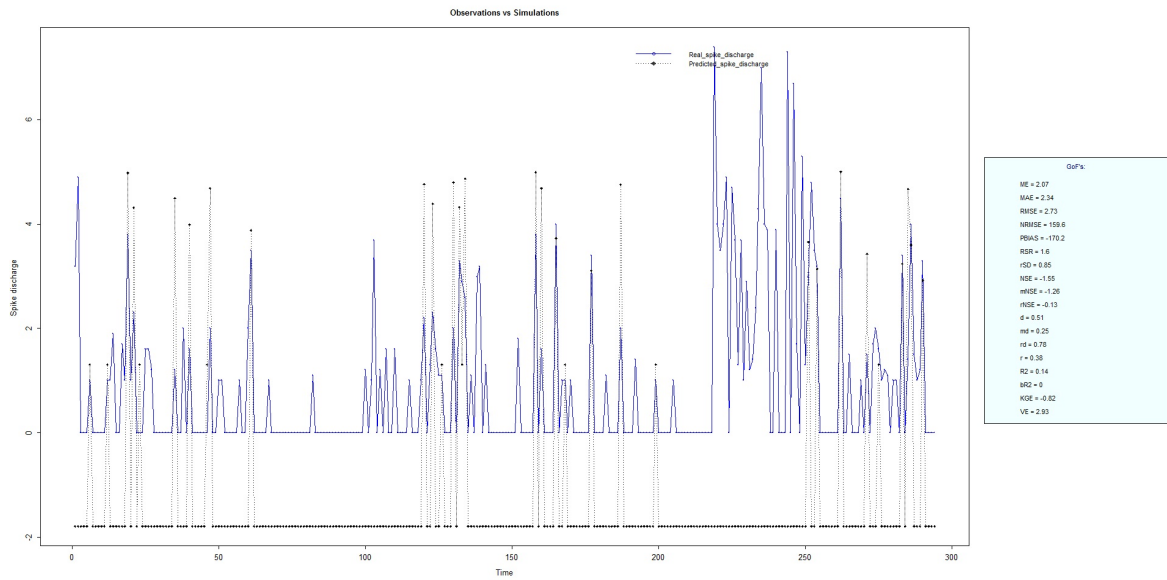Figure 3.3: Spike Discharge prediction for cat Ipsilateral Forepaw Cortex using GFS LT RS algorithm
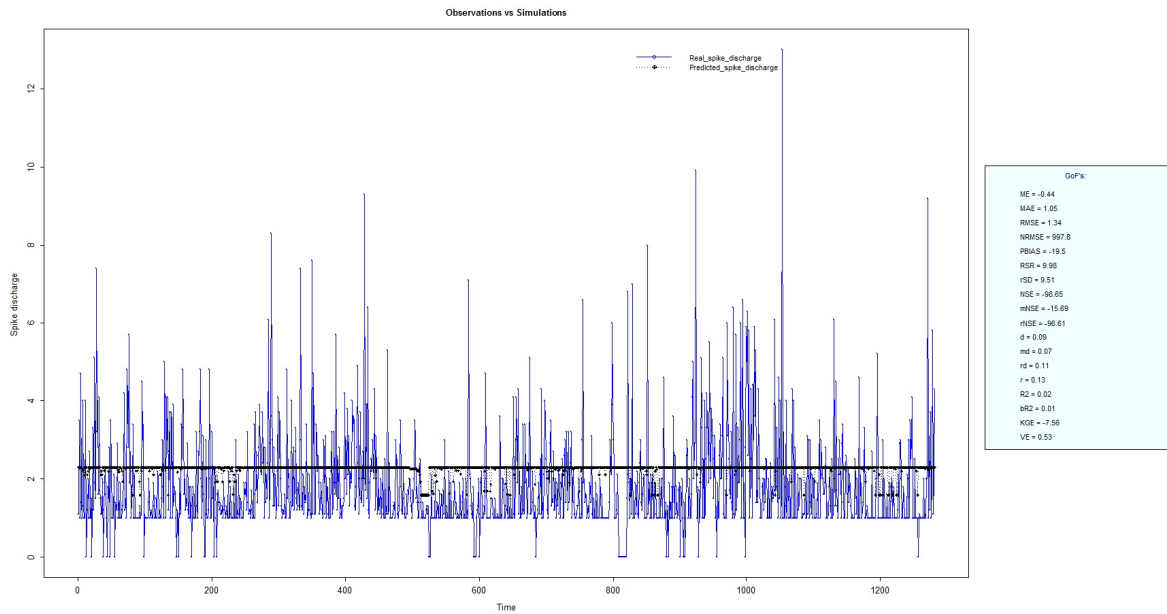
Figure 3.4: Spike Discharge prediction for cat Ipsilateral Forepaw Cortex using HYFIS algorithm
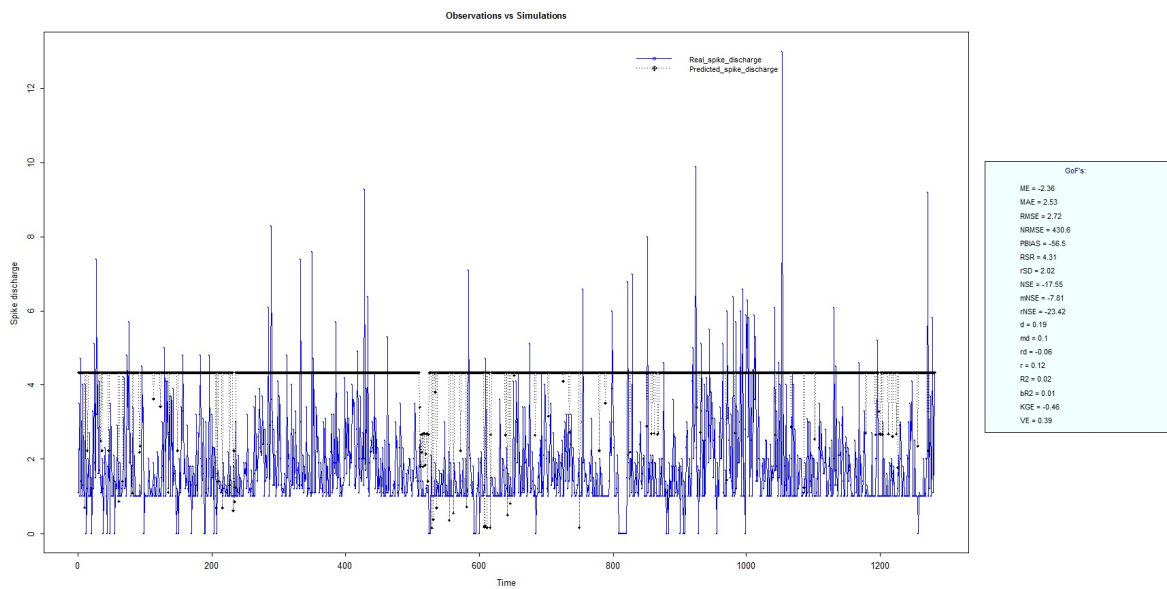
Figure 3.5: Spike Discharge prediction for cat Ipsilateral Forepaw Cortex using WM algorithm
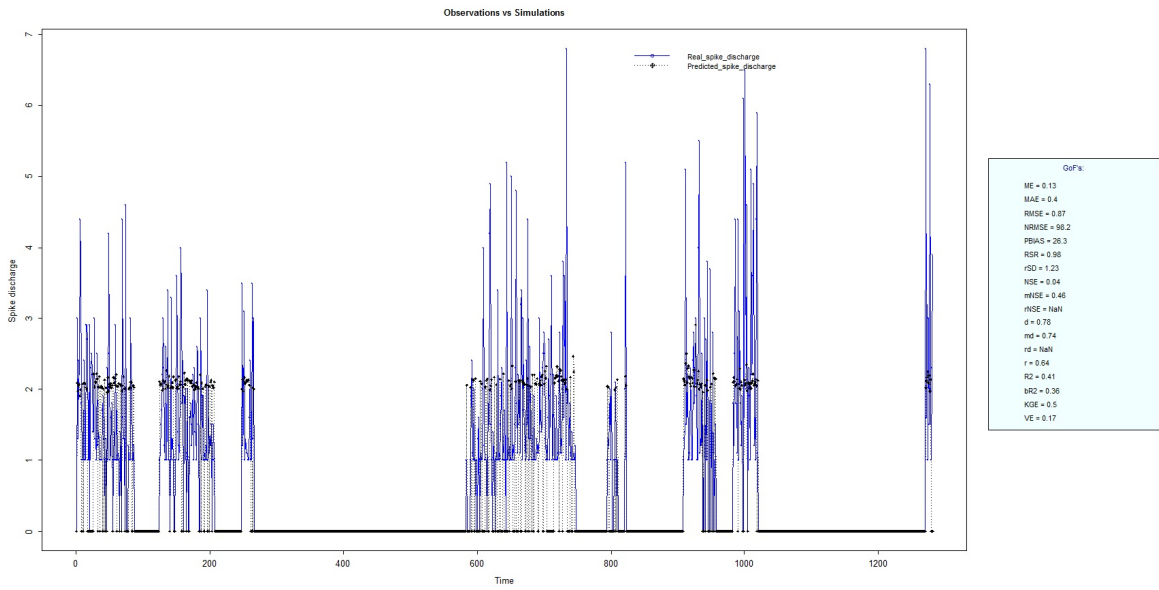
Figure 3.6: Spike Discharge prediction for cat Contralateral Forepaw Cortex using ANFIS algorithm
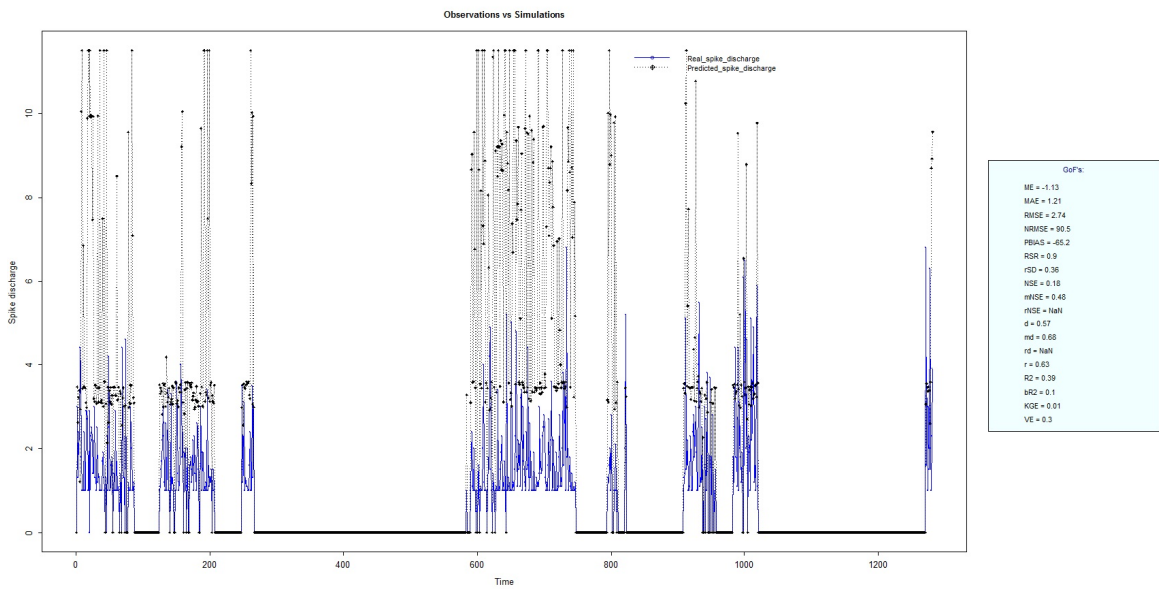
Figure 3.7: Spike Discharge prediction for cat Contralateral Forepaw Cortex using Denfis algorithm
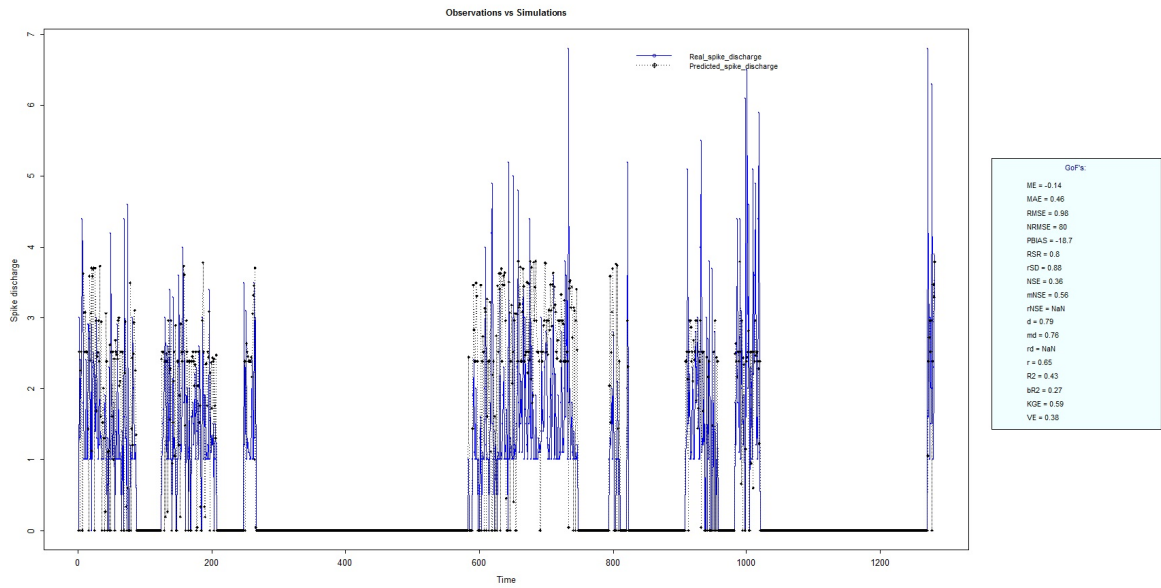
Figure 3.8: Spike Discharge prediction for cat Contralateral Forepaw Cortex using GFS LT RS algorithm
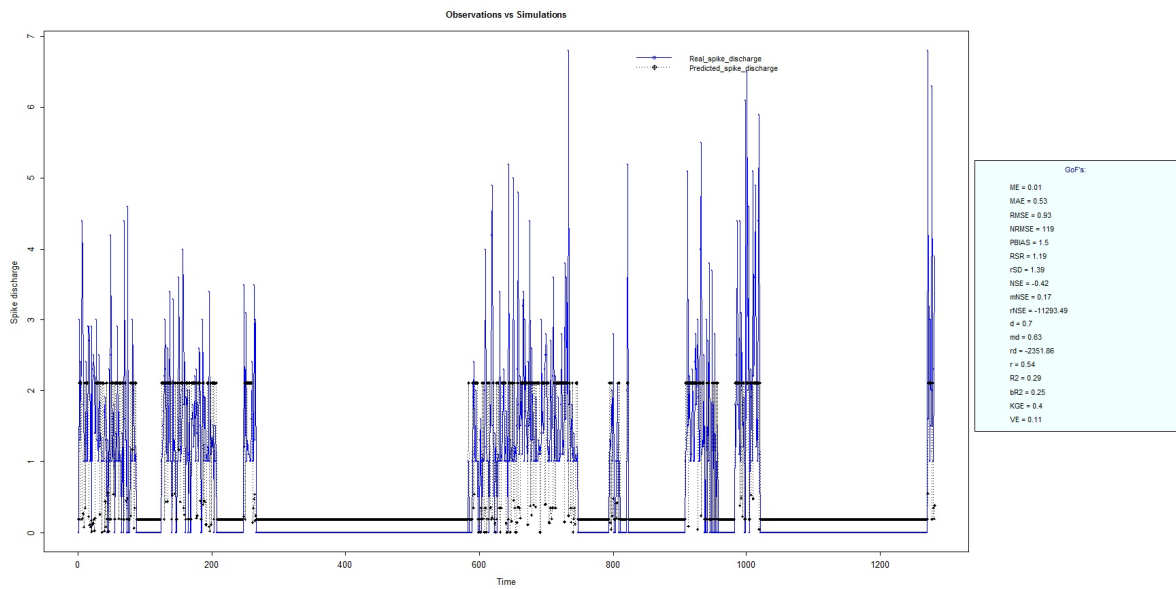
Figure 3.9: Spike Discharge prediction for cat Contralateral Forepaw Cortex using HYFIS algorithm

94
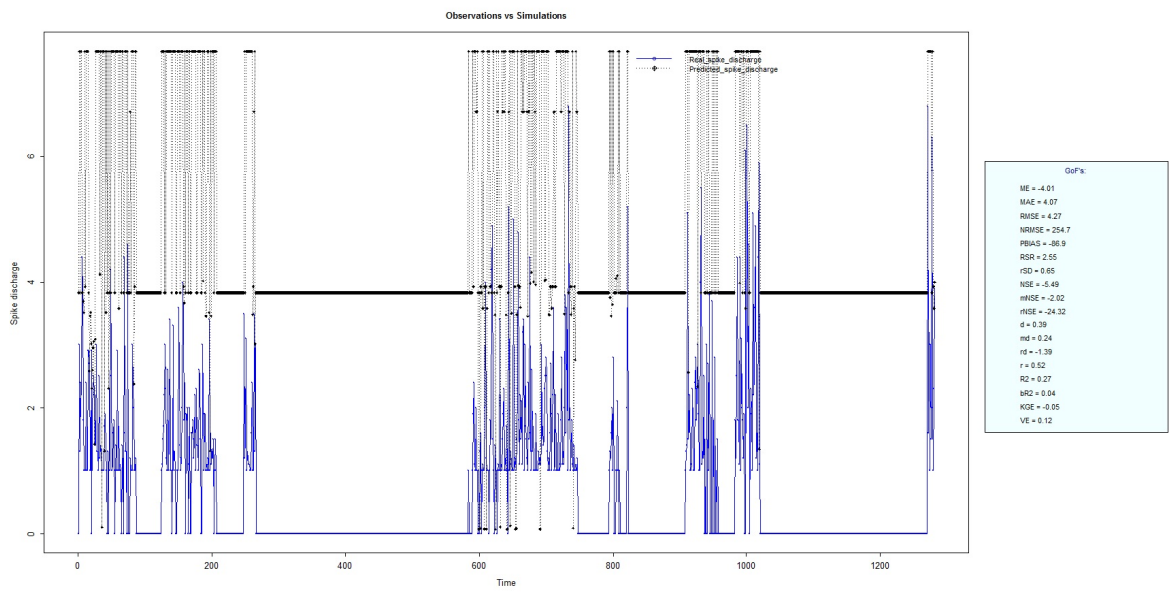
Figure 3.10: Spike Discharge prediction for cat Contralateral Forepaw Cortex using WM algorithm
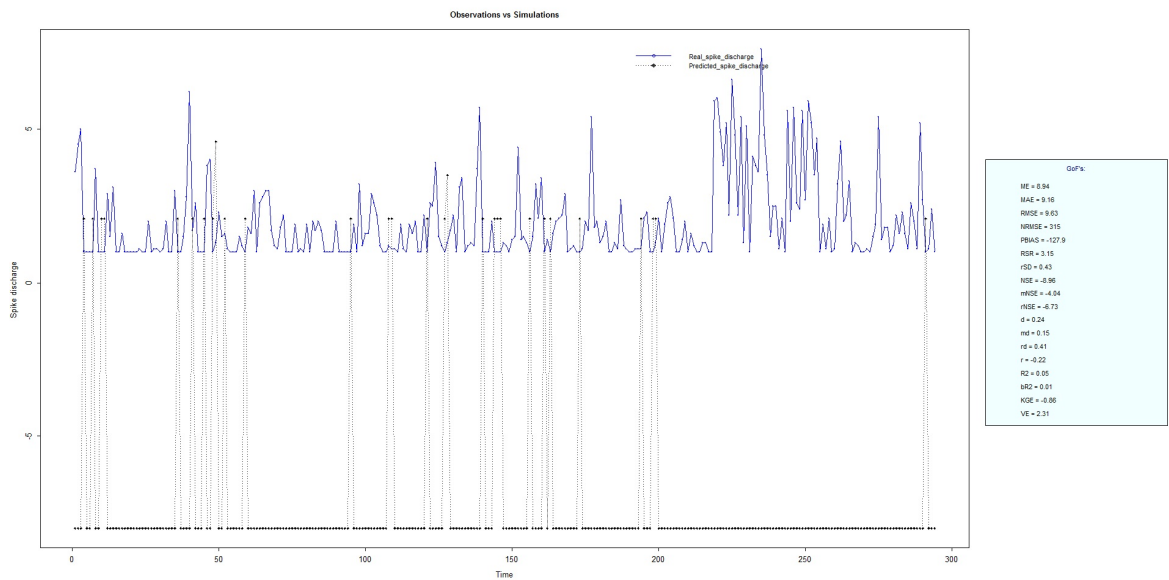
Figure 3.11: Spike Discharge prediction for cat Contralateral Hindpaw Cortex using ANFIS algorithm
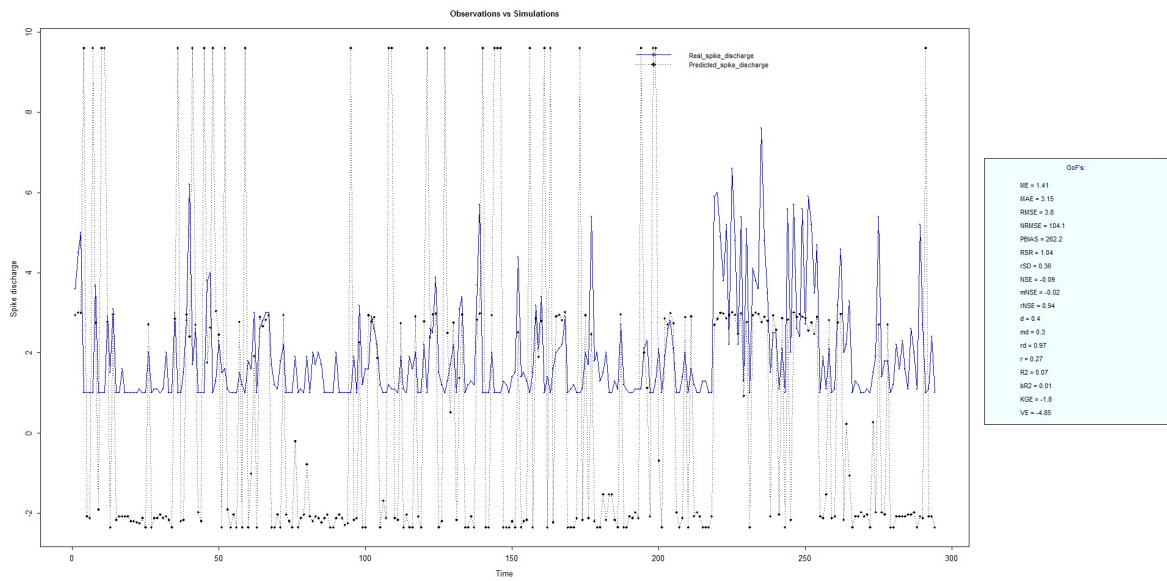
Figure 3.12: Spike Discharge prediction for cat Contralateral Hindpaw Cortex using Denfis algorithm
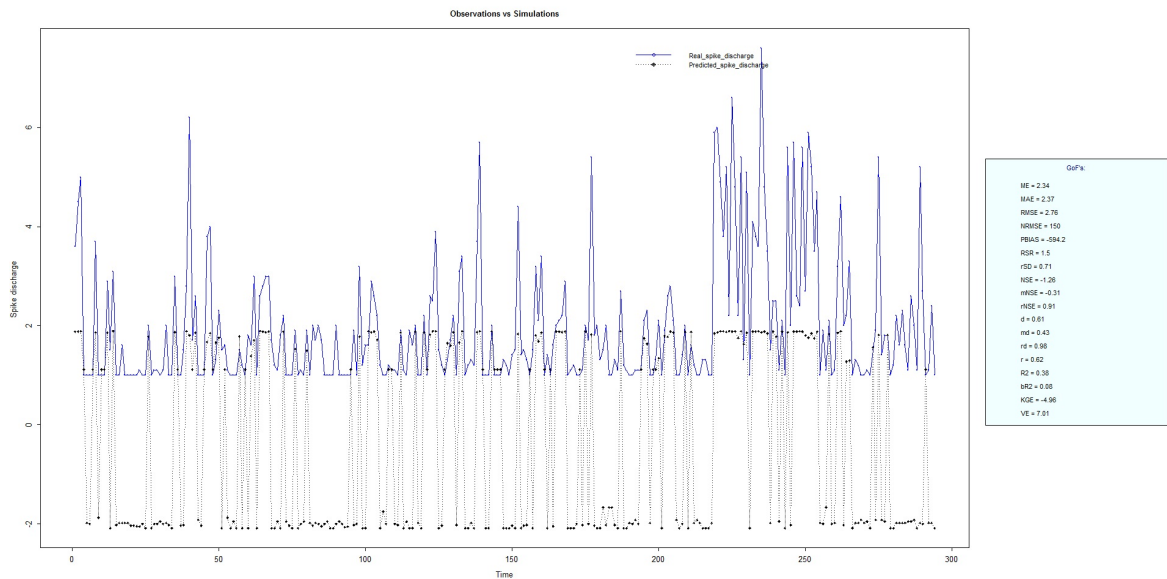
Figure 3.13: Spike Discharge prediction for cat Contralateral Hindpaw Cortex using SBC algorithm
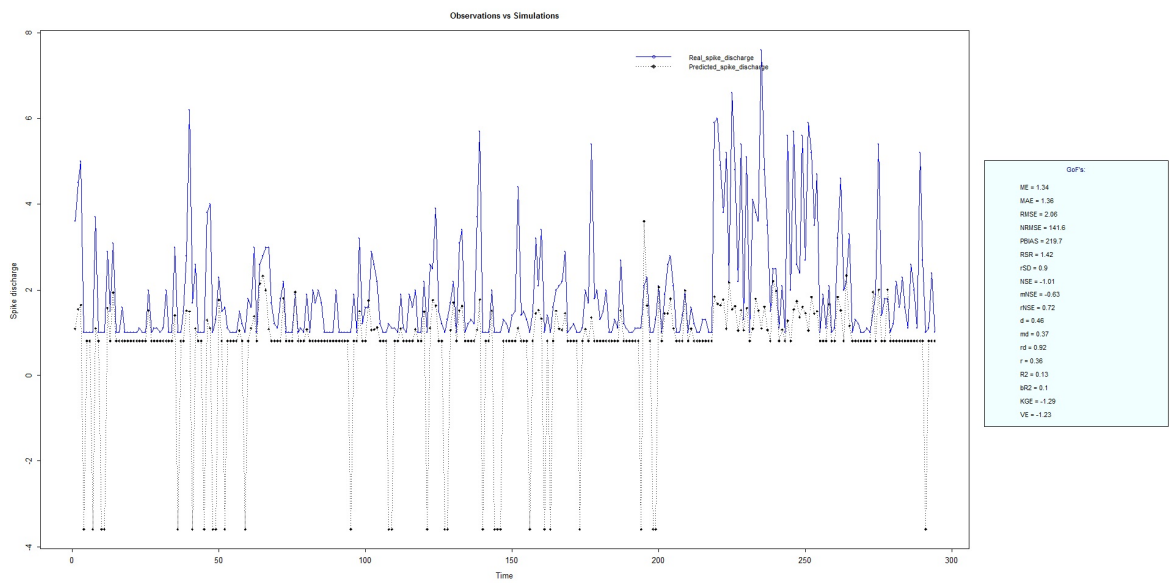
Figure 3.14: Spike Discharge prediction for cat Contralateral Hindpaw Cortex using GFS LT RS algorithm
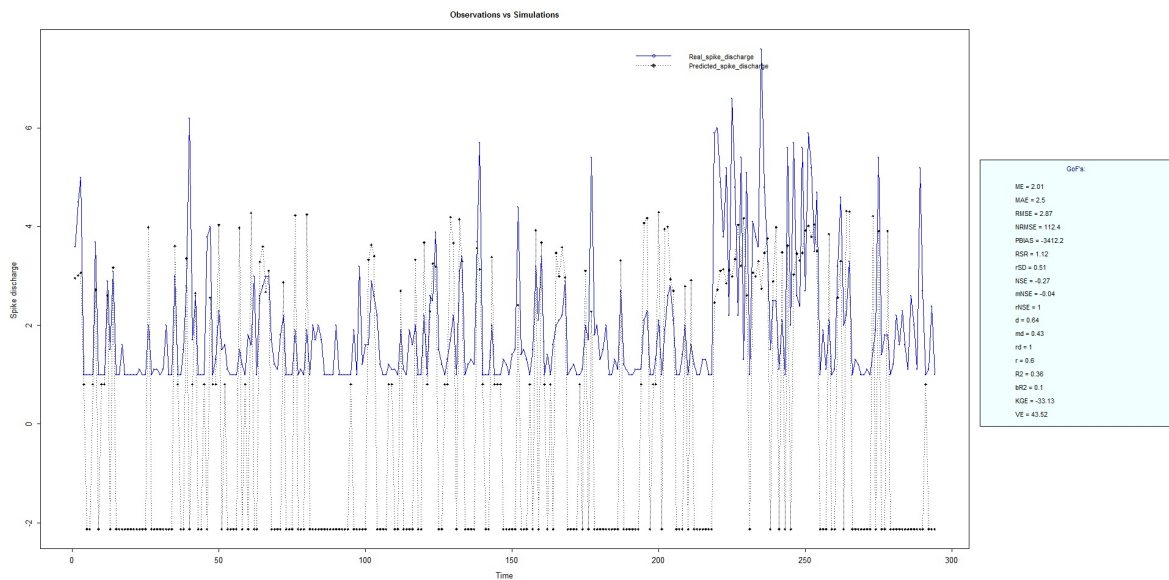
Figure 3.15: Spike Discharge prediction for cat Contralateral Hindpaw Cortex using WM algorithm

Results of application of the algorithms to the Ipsilateral Hindpaw Cortex data are presented in figures 3.3.2-3.3.2. The WM algorithm provides better accuracy compared with other algorithm (RMSE=2.73)

Figure 3.16: Spike Discharge prediction for cat Ipsilateral Hindpaw Cortex using ANFIS algorithm

Figure 3.17: Spike Discharge prediction for cat Ipsilateral Hindpaw Cortex using Denfis algorithm

Figure 3.18: Spike Discharge prediction for cat Ipsilateral Hindpaw Cortex using SBC algorithm

Figure 3.19: Spike Discharge prediction for cat Ipsilateral Hindpaw Cortex using GFS LT RS algorithm

Figure 3.20: Spike Discharge prediction for cat Ipsilateral Hindpaw Cortex using WM algorithm

3.4 Conclusion

In this chapter we presented the development and evaluation of different versions of adaptive neuro-fuzzy model including Adaptive Neuro-Fuzzy Inference Systems, Wang and Mendel, Dynamic evolving neural-fuzzy inference system, Hybrid neural Fuzzy Inference System, genetic for lateral tuning and rule selection of linguistic fuzzy system and subtractive clustering and fuzzy c-means algorithms for prediction of Spike discharge. Results reveal that Spike discharge can be predicted using the neuro-fuzzy model where first spike latency and frequency-following interval are the inputs and spike discharge is the output of the model.

Chapter 4

Modularity detection in the brain

In this chapter we discuss the concept of modularity in the brain and the algorithms for discovery of meaningful modules or clusters in the brain. First we introduce basic concepts of complex networks that can reveal a lot of information about different functions of the brain. Then we describe the application of data mining algorithm to "resting-state brain data". Analysis of these data sets can be considered as a case study.

4.1 Modularity in the brain and complex networks

Recent studies in the area of neuroimaging showed that connectivity in the brain network can be described by small-world properties, because in the brain networks, nodes are highly clustered and average minimal path connecting pair of nodes is small compared to the size of the network [168]. Below we provide a brief review of complex networks' methods for analyzing and finding communities in fMRI data sets. We first focus on communities in the brain networks and how to find them; we, then present our method for finding informative voxels in state dependent fMRI data.

4.2 Complex networks

By definition complex system is a composition of interconnected elements in such a way, that global property of the system is not clear from characteristics of its individual parts, but these parts altogether show some global or emergent properties. Figure 4.1 shows the structure of the internet that, an example complex system which can be analysed with a complex networks approach. Many

complex systems can be properly described by complex networks whose nodes represent individuals or organisations, and links mimic the interactions among them [136]. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution [16]. Figure 5.2 shows the popularity of an actor using number of links (the probability that an actor has k links) has power-low tail for large k, following $p(k) \sim k^{-\lambda_{actor}}$ distribution where typically $\lambda_{actor} = 2.3 \pm 0.1$. Figure 4.2 shows the The distribution function of connectivities for various large networks : (A) actor collaboration network with 212,250 vertices, (B) WWW network with 325,729 vertices and (C) power grid data network with 4941 vertices [16].
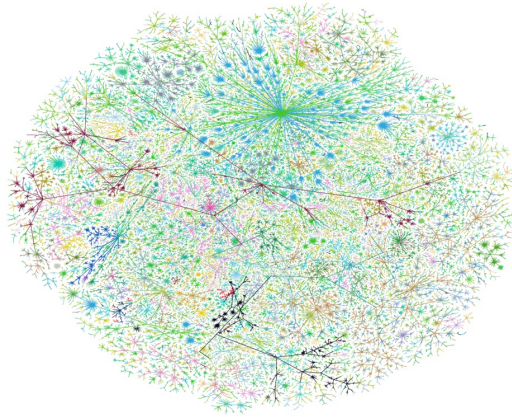
Figure 4.1: Structure of Internet [123]
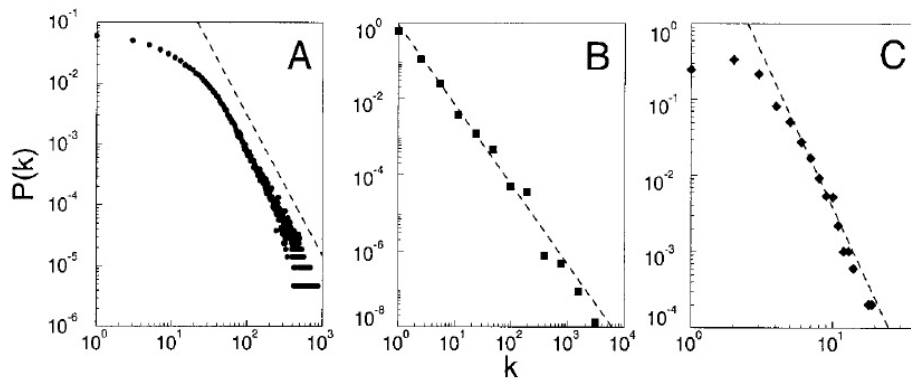
Figure 4.2: The distribution function of connectivities for various large networks [16].

4.3 Identifying modularity and community structure in complex networks

Modularity optimization is an NP-complete problem, and many algorithms for this have exponential time complexity. Therefore, it is not efficient to run this function on larger graphs, but graphs with up to fifty vertices should be fine, and graphs with a couple of hundred vertices might be possible to consider. Newman [124] developed an algorithm to find optimal community in very large network, that can be found via greedy optimization of modularity. Community of a given network with n vertices can be found as follow:

$$Q = 1/2 \sum_{i,j} (A_{ij} - \frac{k_i k_j}{M})(s_i s_j + 1) \tag{4.1}$$

Where Q is modularity, A_{ij} is adjacency matrix, $\frac{k_i k_j}{M}$ is the probability of an edge between two vertices is proportional to their degrees and $(s_i s_j + 1)$ is vertices that are in the same community.
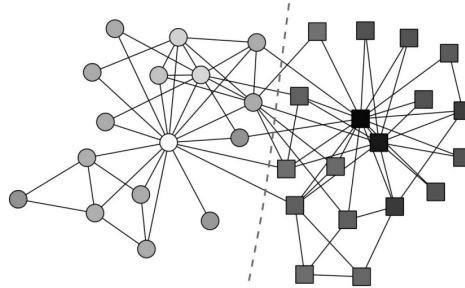
Figure 4.3: Splitting a social network (karate club) into two groups using division algorithm [124]

Ravasz et al. [152] developed an algorithm to calculate topological overlap as follow:

$$O(i,j) = \frac{J(i,j)}{min(k_i, k_j)} \tag{4.2}$$

Where O is topological overlap for each pair of nodes, i and j, $J(i,j)$ denotes the number of nodes to which both i and j are linked; k_i, k_j is the degree of i and j, respectively.

Hollme et al. [78] introduced an algorithm for decomposing biochemical networks into subnetworks based on the global network structure. For an undirected graph the betweenness centrality of reaction nodes is defined as:

$$C_B(r) = \frac{1}{k_{in}(r)} \sum \frac{\sigma_r(s,t)}{\sigma(s,t)} \tag{4.3}$$

Where $C_B(r)$ is centrality of reaction nodes ($r \in R$), $\sigma_r(s,t)$ is the number of shortest paths between s and t that passes through r, $\sigma(s,t)$ is the total number of shortest paths between s and t, $k_{in}(r)$ is the in-degree of node r.

As mentioned before, modularity in large networks reveals significant information about structure of network [152]. We can deduce from modularity of the brain networks that these modules could be correspond to different functions [165]. Recent applications based on graph theoretical analysis in the neuroscience about consistent description of the brain functional network community structure have been addressed [10, 20, 115]. One of effective algorithm for analyzing brain data addressed in [43], that cuts the graph into a pre-specified number of clusters. Here minimized cutting cost can be computed as follow:

$$cut(A,B) = \sum_{v_i \in A, v_i \in B} W_{ij} \tag{4.4}$$

Where K is pre-specified number of clusters, i and j are indices for two clusters A and B and W_{ij} is the non-negative weight of two voxels, v_i and v_j that shows the similarity between the voxel. The normalized cut that minimizes the similarity between clusters and maximizes the similarity within clusters, can be found as follow:

$$Ncut(A,B) = \frac{cut(A,B)}{\sum_{v_i \in A, v_n \in V} w_{in}} + \frac{cut(A,B)}{\sum_{v_j \in B, v_n \in V} w_{jn}} \tag{4.5}$$

4.4 Modularity in brain networks

As mentioned before, modularity in large networks reveals significant information about the structure of the network. These modules can correspond to different functions. For example hierarchical modularity in human brain functional networks with more than 1800 regional nodes analyzed using graph theoretical tools in [115] in order to identify nested modular structure at several hierarchical levels. The results obtained showed that human brain functional networks have hierarchical modular organization with a fair degree of similarity between subjects. The largest five modules at the high-

est level of the hierarchy were medial occipital, lateral occipital, central, parieto-frontal and fronto-temporal systems; occipital modules demonstrated less sub-modular organization than modules comprising regions of multimodal association cortex 1.5. The hub nodes play a key role in inter-modular connectivity. Another study [9] addressed the decreasing of modularity in functional brain networks in schizophrenia, with proportionally more inter-modular edges and fewer intra-modular edges. Figure 4.5 shows the group difference in modularity for two subjects in each clinical sample. Black edges represent intra-modular connections, between brain regions in the same functional community. Red edges represent inter-modular connections, between brain regions in different functional communities. On average there are more inter-modular connections and less intra-modular connections in the networks of patients with childhood-onset schizophrenia compared to healthy participants [9].
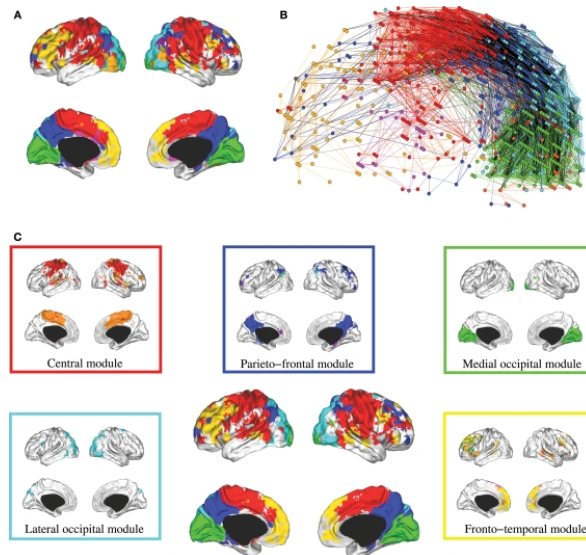
Figure 4.4: Sub-modular decomposition of the five largest modules (shown centrally) illustrates that the medial occipital module has no major submodules whereas the fronto-temporal modules has many sub-modules [115]

4.5 Oxford data

Oxford data from 1000 Functional Connectomes Project used in this project to study the modularity in the human brain that contain the data of 22 healthy brains [192]. Figure 4.6 shows the structure of the input data of the subject 02248 after preprocessing and dimension reduction which contains 1008 brain regions with 166 time points brain activity. Last three columns are the 3 dimensional coordinates for each region.
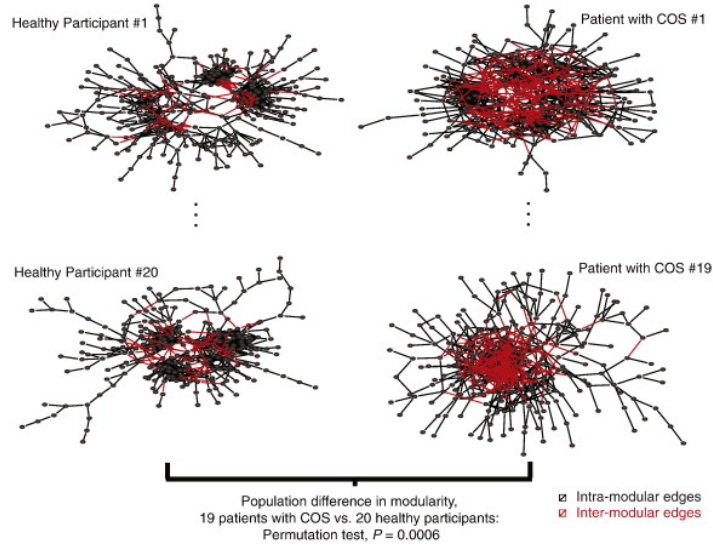
Figure 4.5: The group difference in modularity for two subjects in each clinical sample [9]

4.6 Modified global k-means algorithm for minimum sum-of-squares clustering problems

The modified global k-means clustering algorithm is another method for analysis of large data sets. This algorithm computes clusters incrementally and computes as many clusters as a data set contains, with respect to a given tolerance. In this section we present the brief description of the modified global k-means algorithm. Let the tolerance $\epsilon > 0$ be given. The modified global k-means algorithm:

Step 1. (Initialization). Select a tolerance $\varepsilon > 0$. Compute the center $x^1 \in \mathbb{R}^n$ of the set A. Let f^1 be the corresponding value of the objective function. Set $k = 1$.

Step 2. (Computation of the next cluster center). Set $k = k + 1$. Let $x^1, \ldots, x^{k-1}$ be the cluster centers for $(k-1)$-partition problem. Find a starting point $\bar{y} \in \mathbb{R}^n$ for the k-th cluster center (details of finding starting point can be found in [14]).

Step 3. (Refinement of all cluster centers). Select $(x^1, \ldots, x^{k-1}, \bar{y})$ as a new starting point, apply k-means algorithm to solve k-partition problem. Let $y^1, \ldots, y^k$ be a solution to this problem and f^k be the corresponding value of the objective function.

Step 4. (Stopping criterion). If

$$\frac{f^{k-1} - f^k}{f^1} < \varepsilon \qquad (4.6)$$

107

Figure 4.6: Oxford data: 22 healthy brains

then stop, otherwise set $x^i = y^i$, $i = 1, \ldots, k$ and go to Step 2.

It is clear that $f^k \geq 0$ for all $k \geq 1$ and the sequence $\{f^k\}$ is decreasing, that is,

$$f^{k+1} \leq f^k \forall k \geq 1.$$

This means that the stopping criterion in Step 4 will be satisfied after finite many iterations. Thus Algorithm 4.6 computes as many clusters as the data set A contains with respect to the tolerance $\varepsilon > 0$.

The choice of the tolerance $\varepsilon > 0$ is crucial for Algorithm 4.6. Large values of ε can result in the appearance of large clusters whereas small values can produce artificial clusters. The recommended values for ε are $\varepsilon \in [0.01, 0.1]$.

4.6.1 Computational results

To verify the effectiveness of the proposed algorithms we carried out a number of numerical experiments with Oxford time series data ([192]) set with 1008 nodes and 166 brain activities for each node. The clusters are computed based different tolerances. More than %80 of clusters are stable. For example cluster number 47 is completely stable, also after some iteration cluster number 54 becomes stable, but cluster number 50 is not stable. There are some stable clusters that we cant consider them as informative community, because they are only noise with a few elements. Our future

work is to find communities in that are spatiallycontiguous, because these provide more information about structure of the brain and are easier to interpret.
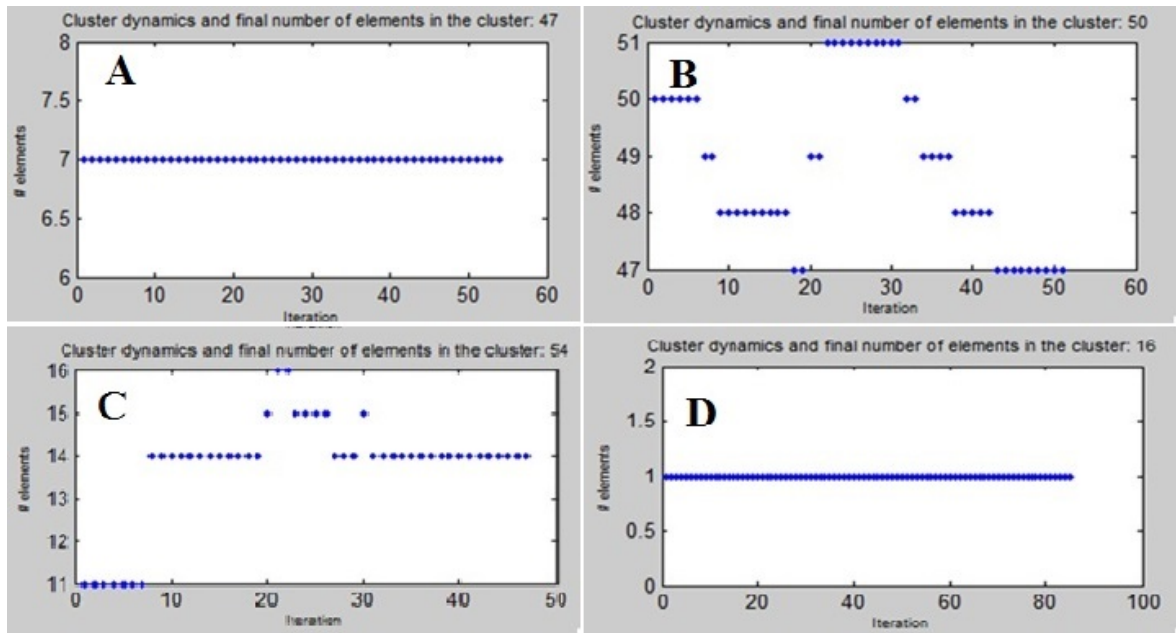

Figure 4.7: Stability in clusters using Modified global k-means algorithm for subject 02248 from Oxford data. Horizontal axis shows the number of data point in the cluster, vertical axis is number of iteration and Tolerance=0.01. (A) a cluster that is stable in all iteration, (B) unstable cluster. (C) a cluster that becomes stable after some iteration and (D) a cluster that can be considered as a noise because it has only one data point. Horizontal axis shows the number of data point in the cluster, vertical axis is number of iteration and Tolerance=0.01

Figures 6-6 from appendix demonstrate the three dimensional coordinates of data points in each cluster for subject 02248. They are the computed clusters in the last iteration using 0.01 for tolerance. Many modules or clusters have symmetric shapes that are related to left and right hemispheres of the brain. Figure 4.8 shows some of the symmetric clusters (clusters 1, 2, 13, 21, 26 and 30) for subject 02248.

As mentioned before the Modified Global K-means algorithm find a new cluster center in each iteration and calculates new cluster. It means that some of the old clusters get smaller. The clusters that keep the shape in different iteration can be called stable clusters. Figures 6-6 from appendix show the stability of clusters based on 0.0001 for tolerance. In each figure the vertical axis shows the number of data points in the cluster, the horizontal axis is the iteration and the cluster number is above the graph. It is clear that many clusters are stable and can be related to different functions of the brain. Figure 4.9 shows cluster stability of 8 clusters based on 0.0001 for tolerance for 10 subjects. We can
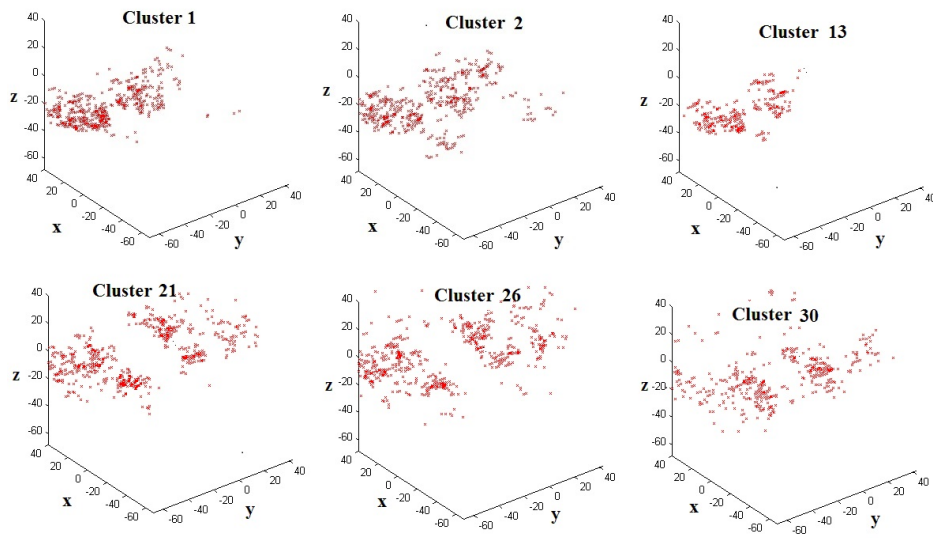
Figure 4.8: Cluster 3d 02248

see that there only a few unstable clusters in each group.

4.7 Conclusion

In this chapter we discussed the concept of community and modularity in the brain. We provided a brief description about some of commonly used modularity detection algorithms, then we introduced the algorithm that we developed and evaluated for resting-state Oxford data. The results of analyzing of Oxford data shows that more than 80% of obtained clusters are stable.
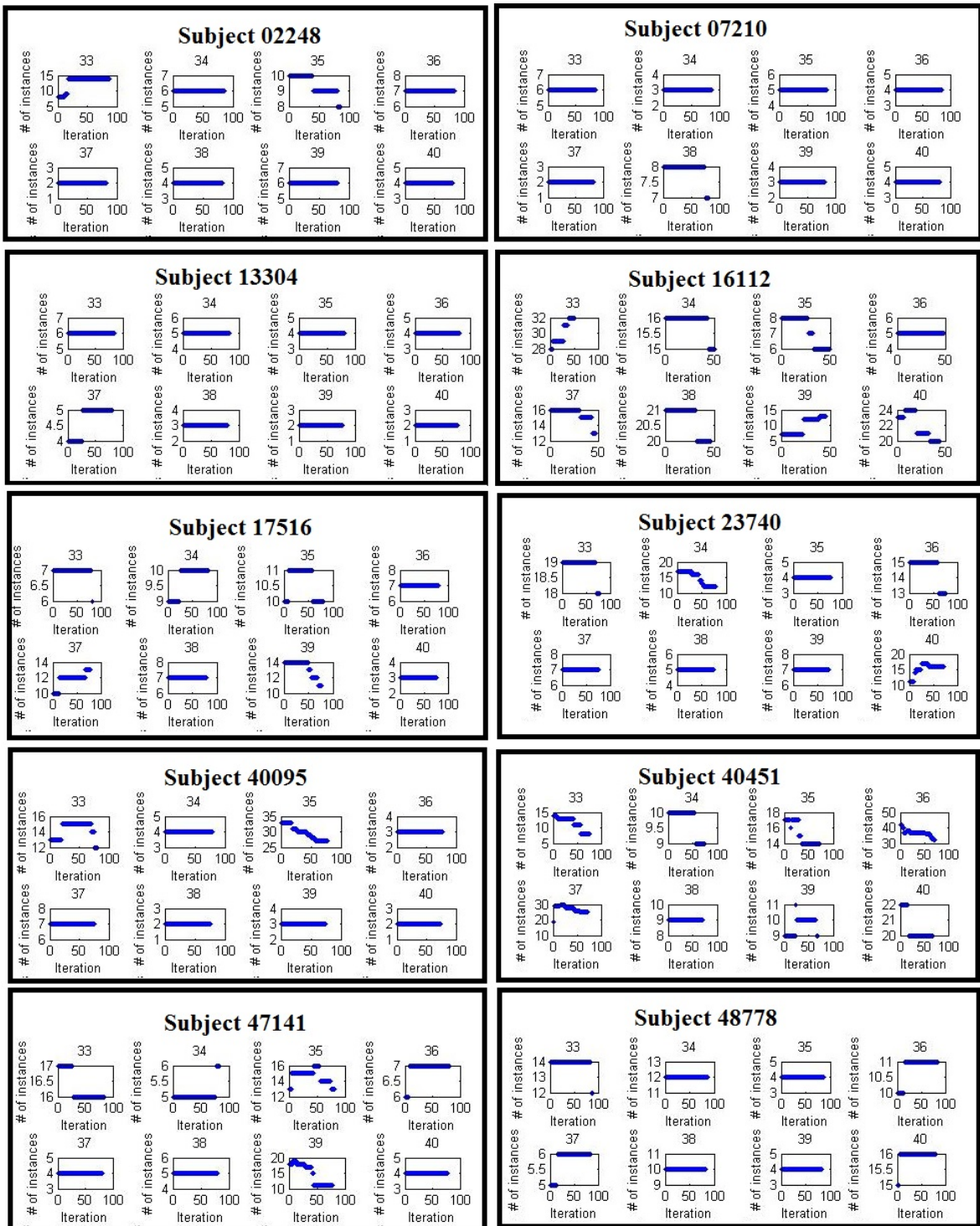
Figure 4.9: cluster stability of clusters 33 to 40 based on 0.0001 for tolerance for 10 subjects

Chapter 5

Functional connectivity differences between men and women in healthy brains and schizophrenia

Schizophrenia is a complex and multifactorial psychiatric disorder, increasingly associated with abnormalities of functional brain connectivity, primarily detected through non-invasive brain imaging methods such as functional magnetic resonance imaging (MRI). Previous studies of schizophrenia have largely converged on abnormal functional connectivity patterns in schizophrenia involving frontal, temporal and limbic regions. However few studies have considered gender-specific influences on these effects, despite clear evidence that gender modulates the clinical presentation and progression of schizophrenia, and is consequently likely to influence the underlying patterns of anatomical and functional dysconnectivity underpinning the disorder. Here we used whole-brain functional MRI and data driven analyses to study gender-specific differences in functional connectivity in 48 men and women with and without schizophrenia. Our results reproduced some of the previously observed major connectivity alterations in the disorder, but importantly additionally detected gender-specific differences in dysconnectivity involving the left superior temporal gyrus, a brain region associated with basic speech perception. These results provide new evidence for gender-specific abnormalities of schizophrenia, and their implications in distinguishing between the male and female phenotypes of the disorder.

5.1 Introduction

Functional connectivity refers to analyzing the temporal relationship of different regions of the brain [11] and its correlation between different brain regions activities. The following different methods can be used for this analysis: Seed-to-Voxel, ROI-to-ROI and Voxel-to-Voxel [207]. Studies of brain connectivity based on two types of experiments, such as resting state and task-related data however resting-state fMRI is more robust, reliable, and can elucidates the functional organization of the human brain [207]. Resting state functional connectivity network provides an essential architecture of the human brain's functions [207]. Functional magnetic resonance imaging is widely used technology for study of connectivity in the brain and it can be used for either of healthy and diseased brains. For example, decreasing fMRI blood-oxygen-level dependent (BOLD) signal activity in two areas shown on the figure 5.1-A. Figure 5.1-B shows the time series of BOLD signal for these areas.
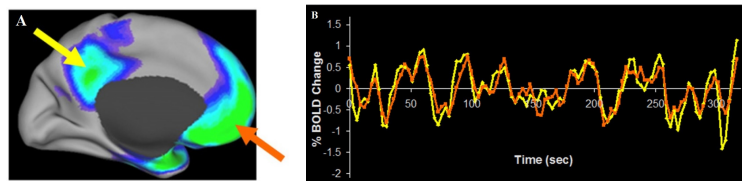

Figure 5.1: Similarity of FMRI BOLD signal activity in two areas in the resting state[149]

Generally, FMRI data is noisy and contains many irrelevant data points in time and space and before analysing it we need to use different preprocessing techniques like noise reduction or we can use dimensional reduction methods to get meaningful features of this data. Different methods like statistical summaries (t-test or correlation test), isolating region of interest (ROI) and dimension reduction methods such as principle component analysis (PCA) or independent component analysis (ICA) can be used for dimension reduction of fMRI data. These methods are well-known for extracting common signals across the brain. Component-based noise correction method (CompCor) [24] is an efficient strategy for noise reduction of fMRI data and its application to fMRI data is addressed in [11, 84]. For efficiency, only after preprocessing fMRI data, different machine learning and complex networks algorithms can be used for analyzing it.

Brain networks can be described by small-world properties, because the nodes of the network are highly clustered and an average minimal path connecting pair of nodes is small compared to the network's size [31, 168].Complex networks method is an efficient method for analysing the functional

connectivity in the brain, that is measured using different graph measures like average path length and small-worldness. Usually, the original brain data is de-composing to a brain networks in such a way, that functional connectivity measures are computed and based on these measures graph structure is created. Complex network approach used in [198] for analyzing bilateral inferior/superior frontal cortex and temporal pole regions and their capacity to communicate with other regions of the brain.

Abnormality in the connectivity in schizophrenia is an interesting topic in the neuroscience area. For example, Fornito et al. [52] showed that there are connectivity deficits between frontal cortex and posterior regions and occurred irrespective of task context in the first-episode schizophrenia. Another research [51] addressed the abnormality in resting-state functional connectivity in bipolar patients.

In terms of connectivity significant difference between bipolar patients and healthy subjects was observed between the Medial prefrontal cortex and the right dorsolateral prefrontal cortex. In [11] functional-connectivity measures used for independent components to classify schizophrenia patients and healthy controls during resting-state. Modularity and local connectivity impairments of brain functional networks in childhood-onset schizophrenia is addressed in [10]. They used Complex network analysis to show the alteration of inter- versus intra-modular connections between network nodes. They illustrated that there are more inter-modular connections and less intra-modular connections in the networks of patients with childhood-onset schizophrenia compared to healthy participants. Figure 5.1 hows the decreasing of modularity (edges within clusters) and increased global efficiency (edges between modules) in schizophrenic patients relative to controls.
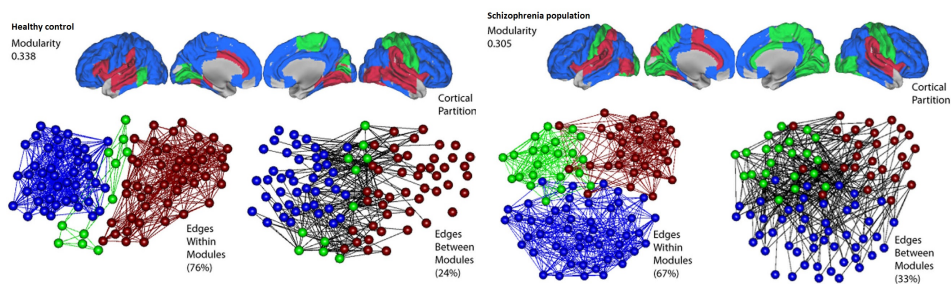

Figure 5.2: Difference of modularity between health control (left) and schizophrenic population (right) [10]

Recent studies addressed alterations of functional connectivity of prefrontal cortex in schizophrenia [41, 54, 106]. In [41] resting state functional connectivity magnetic resonance imaging used to identify within prefrontal cortex (PFC) dysconnectivity in schizophrenia and showed that the right dorsolateral prefrontal cortex (DLPFC) and left inferior frontal junction (IFJ) has reduction within-
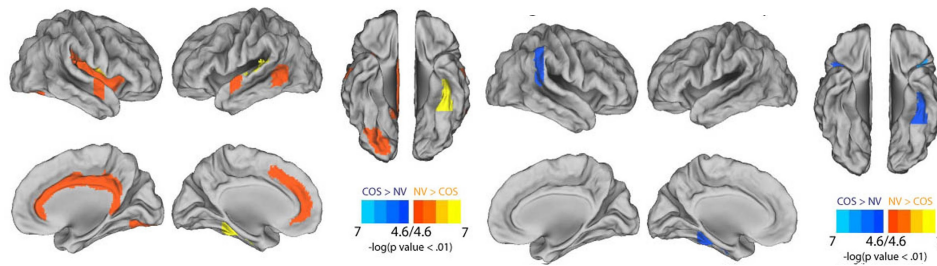
Figure 5.3: Difference in clustering (left) and efficiency (right) between schizophrenic patients and healthy control [10]

PFC connectivity for patients compared with control subjects (figure 5.1). Also, The dorsolateral prefrontal cortex is underconnected with prefrontal cortex and overconnected with posterior cortex in schizophrenia.
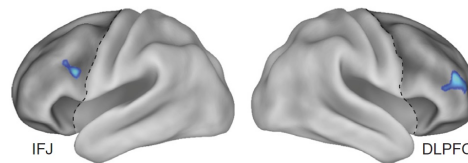

Figure 5.4: Within prefrontal cortex (PFC) dysconnectivity in schizophrenia [41]

Middle temporal gyrus and inferior temporal gyrus have abnormal connectivity in schizophrenia patient [76, 129]. Also, the superior temporal gyrus is strongly implicated in the pathophysiology of schizophrenia [46]. Figure 5.5 shows the relative volumes of the middle temporal gyrus and inferior temporal gyrus in chronic schizophrenia patients. Compared with control healthy, the schizophrenia patients show gray matter volume reductions in the left middle temporal gyrus and bilateral inferior temporal gyrus [41].

As mentioned before, analyzing the statistical dependencies between regional time series could be based on source of interest. The difference between connectivities can help us to discriminate the patient group and healthy controls [11]. In this chapter, we focused on connectivity difference in male and females schizophrenia patients. First, we considered the connectivity difference between patients and healthy control subjects, then we applied ROI-to-ROI analysing (analysing the connectivity between seed region and all other ROIs) and complex networks methods for patients' data to see the difference between genders in patients data.
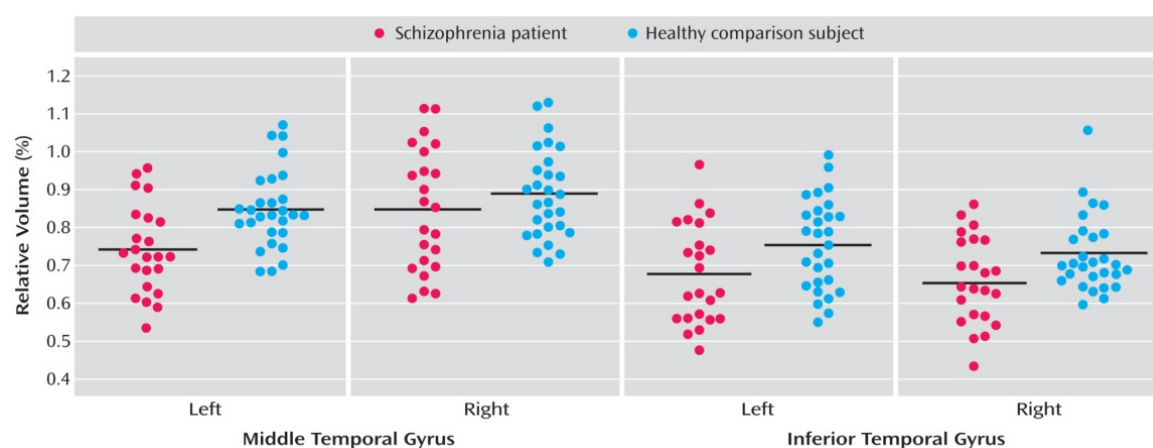
Figure 5.5: Relative Volumes of the Middle Temporal Gyrus and Inferior Temporal Gyrus by Hemisphere in Male Patients With Chronic Schizophrenia (Horizontal lines represent the mean relative volume) [129].

5.2 Gender difference in schizophrenia

In this section we discuss about the functional connectivity difference of schizophrenia in male and female patients. Although, functional connectivity of many areas in the brain is similar for different genders, but there are some distinction for some areas in some experiments. This mental disorder is common in women, but the age of onset is early in women[44, 141]. There is reduction in left hippocampus formation in male [199] and the men have smaller ratio of grey matter in the caudate, hippocampus, and temporal gyrus (are involved functions with language, thinking and memory) [111, 199]. Figure 5.6 demonstrated the regional changes in brain gray matter in patients with schizophrenia.

Schmithorst et al. [160] showed that association between intelligence and the functional connectivity linking Broca's area to auditory processing area, including Wernicke's areas and the right posterior superior temporal gyrus in boys is greater than girl, but for girls the association is greater the linking left posterior superior temporal gyrus to Wernicke's areas bilaterally. Also, investigating functional connectivity measuring the gamma phase synchrony in schizophrenia shows that global functional connectivity (lower gamma phase synchrony) declines in chronic schizophrenia subjects compared with healthy subjects and this reduction is most apparent in female patients [164].

Compared with patients with healthy comparison subjects gray matter volume over time in the left superior temporal gyrus has significant decreases [92]. A study on handedness as a differen-
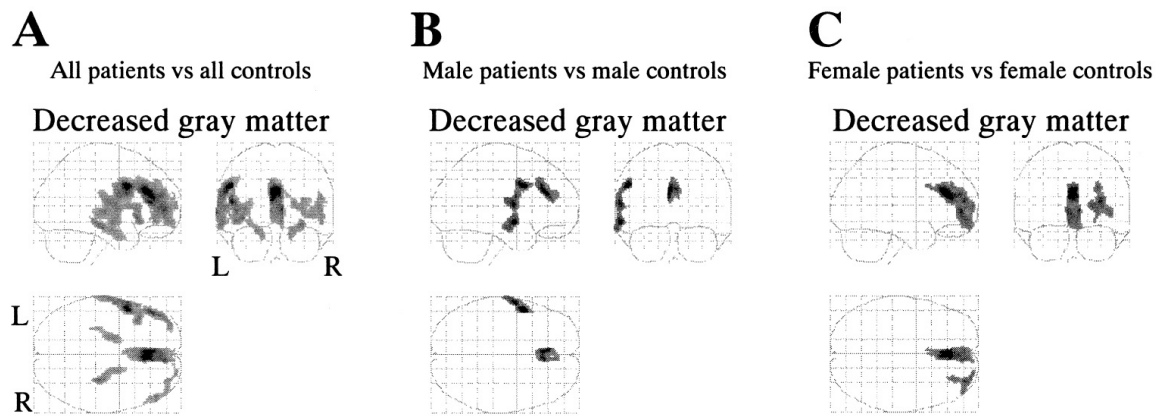
Figure 5.6: Regional changes in brain gray matter in patients with schizophrenia [175]

tiating variable [77] shows that gray matter volumes in the superior temporal gyrus in left-handed schizophrenic men is smaller. More evidences regarding abnormality in between different genders and left and right hemisphere of superior temporal gyrus in schizophrenia patients' can be found [38, 114, 150, 151, 174, 200].

The current study has two steps. First we analyzed the difference in functional connectivity between schizophrenia patients and healthy controls. Then we evaluated the difference between male and female dysfunctionality has been evaluated. Our results in the first step confirmed the previous works about abnormality in the prefrontal cortex in the SP's. We used the CONN-fMRI Functional connectivity toolbox that is a toolbox for functional connectivity analyses of fMRI data [207] that publicly available on [193]. It provides different techniques of analyzing of functional connectivity of resting state and task-related data.

5.3 Material and method

5.3.1 COBRE data

We included 48 subjects from Centre for Biomedical Research Excellence COBRE data set; 24 healthy controls and 24 schizophrenia patients (12 men and 12 women in each group) in the study. In the original data set (see table 5.3.1) there are more male subjects than females and to allow the same effects of the both genders in the results we used the above arrangement. COBRE data contains raw anatomical and functional MRI data from patients with Schizophrenia and healthy controls. Some

papers that by the COBRE group published based on this data [33, 112, 169]. This data set is available

on [184] and contains raw anatomical and functional MR data from patients with Schizophrenia and

healthy controls, ranging in age from 18 to 65 years old. Resting fMRI, anatomical MRI, phenotypic

data for every participant including: gender, age, handedness and diagnostic information are released.

	N	Age(SD)	%Female	%Right-Handed
Schizophrenia	72	38.16(13.89)	0.19	0.83
Patients	74	35.82(11.58)	0.31	0.96

Table 5.1: COBRE data [11]

5.4 Setting up the data and the related parameters

In this section we discuss about setting up the structural and functional data, defining region of

interests and setting up the parameters.

5.4.1 Structural and functional preprocessing of the data

The CONN and SPM toolboxes are used for spatially preprocessing (Realignment, coregistering,

normalization, ...) of structural and functional data. Figures 5.4.1 and 5.4.1 show the structural and

functional data for each subject. The functional volumes are coregistered with the region of interest

and structural volumes (figure 5.4.1).
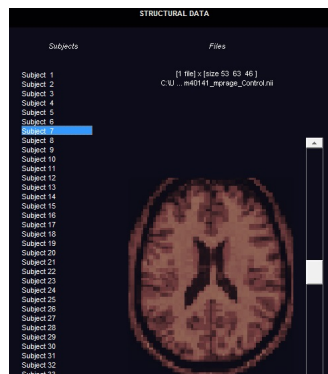
Figure 5.7: Structural data setup

5.4.2 Regions of interest

Regions of interest and all the Brodmann areas defined from Talairach daemon assigned to all

subjects. By segmentation of structural image for each subject, grey matter, white matter and cere-
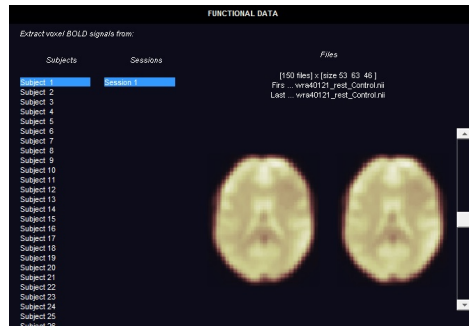
Figure 5.8: Setting up of the functional data. The window shows the first (left) and last (right) scan for each subject
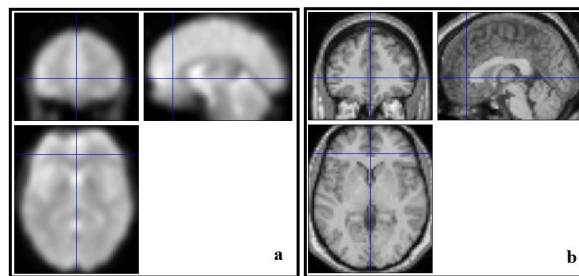
Figure 5.9: Functional data (a) and the related structural data (b)

brospinal fluid (CSF) masks generated. Here, the time series of interest are the number of PCA components.
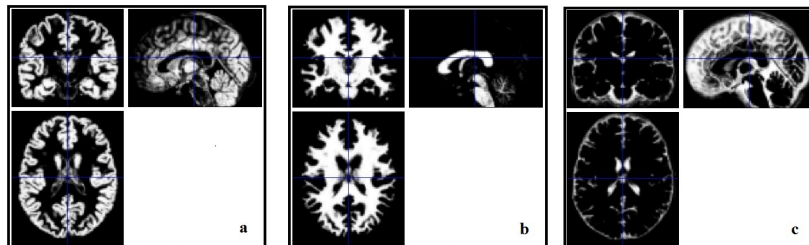
Figure 5.10: Regions of interest for subject 40121 from COBRE data. Grey matter (a), White matter (b) and Cerebrospinal fluid (c) masks.

5.4.3 First and second level of covariates

In this step the realignment parameters in BOLD model is defined (first level covariate), then in the second level covariate the group level regressor is performed. We categorized the data input data into 4 groups; Control females, Control males, Patient females and Patient male. After defining the experiment data, the functional data is imported, then the structural data is segmented to define the grey matter, white matter, cerebrospinal fluid region of interest. By performing PCA on within region

of interests the ROIs time series is extracted.

5.5 Denoising of the data

Before analysing the data we need to explore and remove the confounds. Using CONN toolbox different source of possible confounds like cerebrospinal fluid and white matter signal and within-subject covariate (realignment parameters) are considered. Figure 5.11 shows the white matter and CSF and the related signals for subject 40121 from COBRE data. We chose the 5 dimension that is the number of temporal components are being used. Similarly, the number of dimension for white matter was 5 and the derivative order was 0.
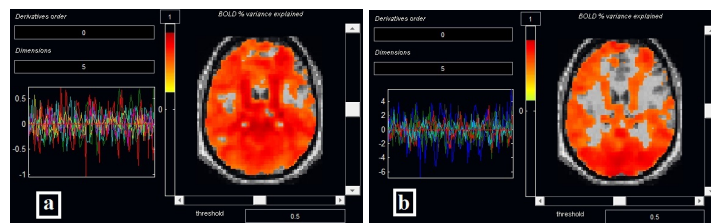
Figure 5.11: Signals and the total variance explained in white matter (a) and CSF (b) for subject 40121

The histogram plot 5.12 display r value before and after confound removal and the band-pass filter is set to [0.008 0.09].
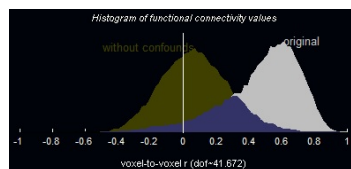
Figure 5.12: R value before and after confound removal

The CONN and SPM toolboxes used for spatially preprocessing (Realignment, coregistering, normalization) of structural and functional data of functional data is done using SPM toolbox. The functional volumes are coregistered with the region of interest and structural volumes. ROI-to-ROI correlational analysis were carried out by the CONN toolbox [207] that is publicly available on [193]and SPM8 (Statistical Parametric Mapping). The preprocessing of the functional images considered of band-pass filtering of 0.008 - 0.09 Hz, motion correction, registration to structural images and spatial normalization to the MNI template. Then to reduce the physiological noise source, a Component

Based Noise Correction Method (CompCor) has been used [24]. CompCor can be used for the reduction of noise in both blood oxygenation level dependent and perfusion-based functional magnetic resonance imaging data. False discovery rate correction is used to multiple hypothesis testing. Number of PCA components to be extracted for each ROI is set to one. It means that the time-series of interest is defined as the average BOLD activation within the ROI voxels, but it's possible to define it as the principal eigenvariates of the time-series within the ROI voxels. Regions of interest and all the Brodmann areas defined from Talairach daemon assigned to all subjects. By segmentation of structural image for each subject, grey matter, white matter and cerebrospinal fluid (CSF) masks were generated.

5.5.1 Calculating functional connectivity measures

Functional connectivity analysis was done using CONN toolbox to examine sex difference in schizophrenia. Bivariate correlation is used as a functional connectivity measures between two areas. General linear model (GLM) [113] used for comparison of connectivity results between genders.

5.5.2 ROI based analysis

As discussed before, temporal gyrus is altered in schizophrenia and it's abnormality is different for men and women [44, 111, 141]. We investigated the hypothesis of connectivity differece in schizophrenia and used ROI analysing for this area. Two-sample t-test analyses computed via SPM8 [58] to compare the connectivity results of patients vs. controls and male patient vs. women patient to compare the connectivity across two group. Connectivity values (Fisher-transformed correlation coefficients) between the seed and the identified ROI was extracted from the connectivity map.

Different source of interest can be defined for analyzing. Figure 5.13 shows the results connectivity analyzing based left primary somatosensory cortex and left associative visual cortex for subject 40121 from COBRE data that voxels with correlation coefficient 0.5 are colored.

Different measures can be used for the analyzing the connectivity that is listed in the table 5.2. x and y are two BOLD time series vectors and X and Y are matrices created by concatenating horizontally one or several x and y vectors. The brackets ([]) show the operation of zeroing all the nondiagonal elements in a matrix. Here we used bivariate correlation [207].
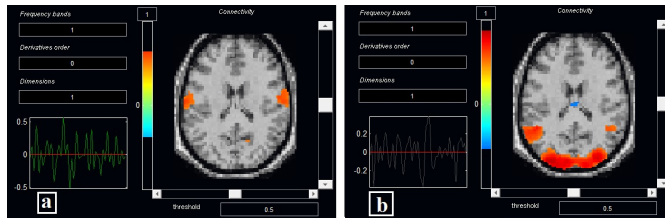
Figure 5.13: ROI-based connectivity of left primary somatosensory cortex (a) and left associative visual cortex (b) for subject 40121

Table 5.2: Functional connectivity measures

Bivariate regression	$b = (x^t.x)^{-1}.(x^t.y)$
Bivariate correlation	$r = (x^t.x)^{1/2}.b.(y^t.y)^{-1/2}$
Multivariate regression	$B = (X^t.X)^{-1}.(X^t.Y)$
Semipartail correlation	$R = [(X^t.X)^{-1}]^{-1/2}.B.[Y^t.Y]^{-1/2}$

5.5.3 First level voxel-based analysis

For a subject or condition it is possible to perform voxel-to-voxel analyzing that applies matrix of voxel-to-voxel connectivity values and there is no need for priori region of interest or seed analysis. In this method we can investigate whole brain connectivity. The voxel based analyzing can be based on connectivity pattern (Principal Component Analysis) between a voxel and the rest of the brain (*MVPA*). Another voxel based measeure is avialable in CONN toolboxes is *Indexes* that calculates the the average local connectivity between each voxel and its neighbors (Integrated Local Correlation) [47] or instead of average, the spatial asymmetry of the local connectivity can be used (Radial Correlation Contrast) [62]. Also, instead of local connectivity, global connectivity pattern between a voxel and the rest of the brain can be used (e.g. Radial Similarity Contrast) [95]. More details about measuring the Index can be found in [207].

5.6 Second level analysis

In the second level analysis step the between-subject contrast can be consider (e.g. to compared different groups like male vs. femals to see main effects in the connectivity within each group). In the ROI-to-ROI analyses, the first-level connectivity-measure matrix is used and the results can be thresholded at the desired p-value threshold. In this step by graph theoretical analyzing method provides the network measures like efficiency, centrality, and cost/degree to test the between-subject contrast.

5.7   Results: Healthy women vs. healthy men

In this section we discuss about some of our results about functional connectivity difference between men and women. Two groups of healthy subjects (12 men and 12 women) from COBRE data is used for this experience. We used all similar setting that we discussed before. Using ROI-based analyzing, we found significant differences in Insular Cortex, Temporal Gyrus,Cingulate Cortex, Opercularis and Premotor cortex.

Table 5.3: Difference between healthy women and healthy men

1	(L)_ Dorsal anterior Cingulate Cortex
2	(L)_ Dorsal Posterior Cingulate Cortex
3	(R)_ Dorsal Posterior Cingulate Cortex
4	(L)_ IFC pars opercularis
5	(L)_ Ventral Posterior Cingulate Cortex
6	(R)_ Dorsal Posterior Cingulate Cortex
7	(L)_ Ventral Posterior Cingulate Cortex
8	(R)_ Ventral Posterior Cingulate Cortex
9	L_Insular_Cortex
10	L_Middle_Temporal_Gyrus
11	PCC
12	rsREL_Left Anterior Sup Temp Gyrus

Table 5.4: Difference between all healthy subjects and all patients

1	(R). Primary Somatosensory Cortex
2	(R). Middle Temporal Gyrus
3	(L). Superior Temporal Gyrus
4	(R). Superior Temporal Gyrus
5	(L). Ventral Anterior Cingulate Cortex
6	(L). Posterior Entorhinal Cortex
7	(R). Posterior Entorhinal Cortex
8	(L). Anterior Entorhinal Cortex
9	(L). Parahippocampal cortex
10	(R). Fusiform gyrus
11	(L). Temporopolar Area
12	(R). Temporopolar Area
13	(L). Primary Auditory Cortex
14	(R). Primary Auditory Cortex
15	(L). Subcentral Area
16	(R). IFC pars opercularis
17	(L). Premotor Cortex
18	(L). Dorsolateral Prefrontal Cortex
19	(R). Dorsolateral Prefrontal Cortex
20	rsREL.Right Posterior Sup Temp Gyrus (60,-30,24)
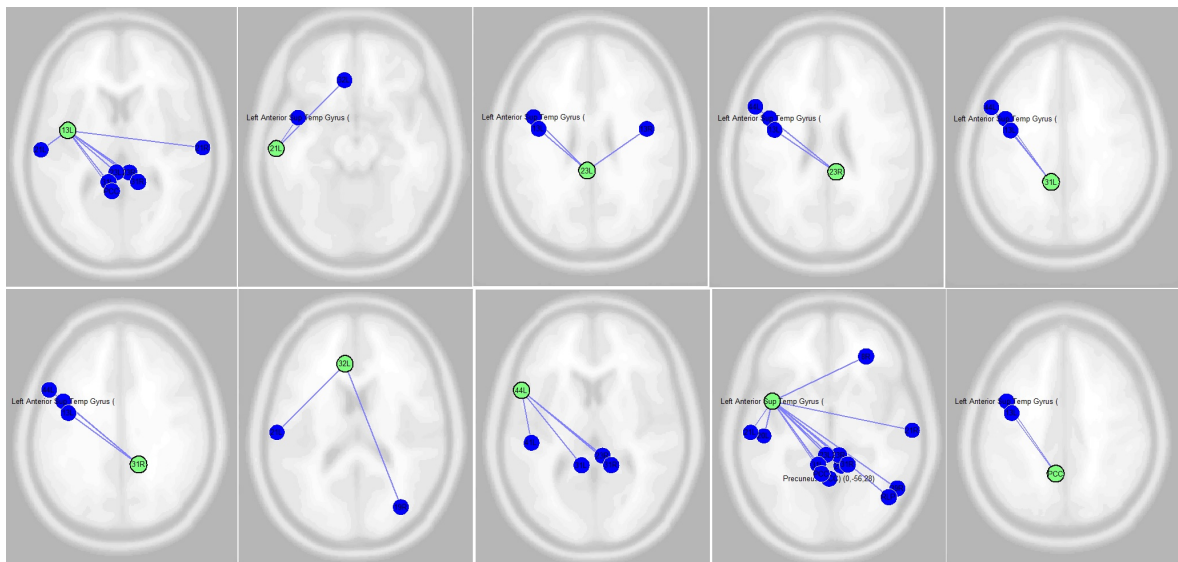
Figure 5.14: Healthy women vs. healthy men. L Insular Cortex, L Middle Temporal Gyrus, L Ventral Posterior Cingulate Cortex, R Ventral Posterior Cingulate Cortex, L Dorsal Posterior Cingulate Cortex, R Dorsal Posterior Cingulate Cortex, L Dorsal anterior Cingulate Cortex, L IFC pars opercularis, Left Anterior Sup Temp Gyrus,PCC. ROI-to-ROI connections from connectivity matrix (99 × 99 ROIs), p-FDR=0.05
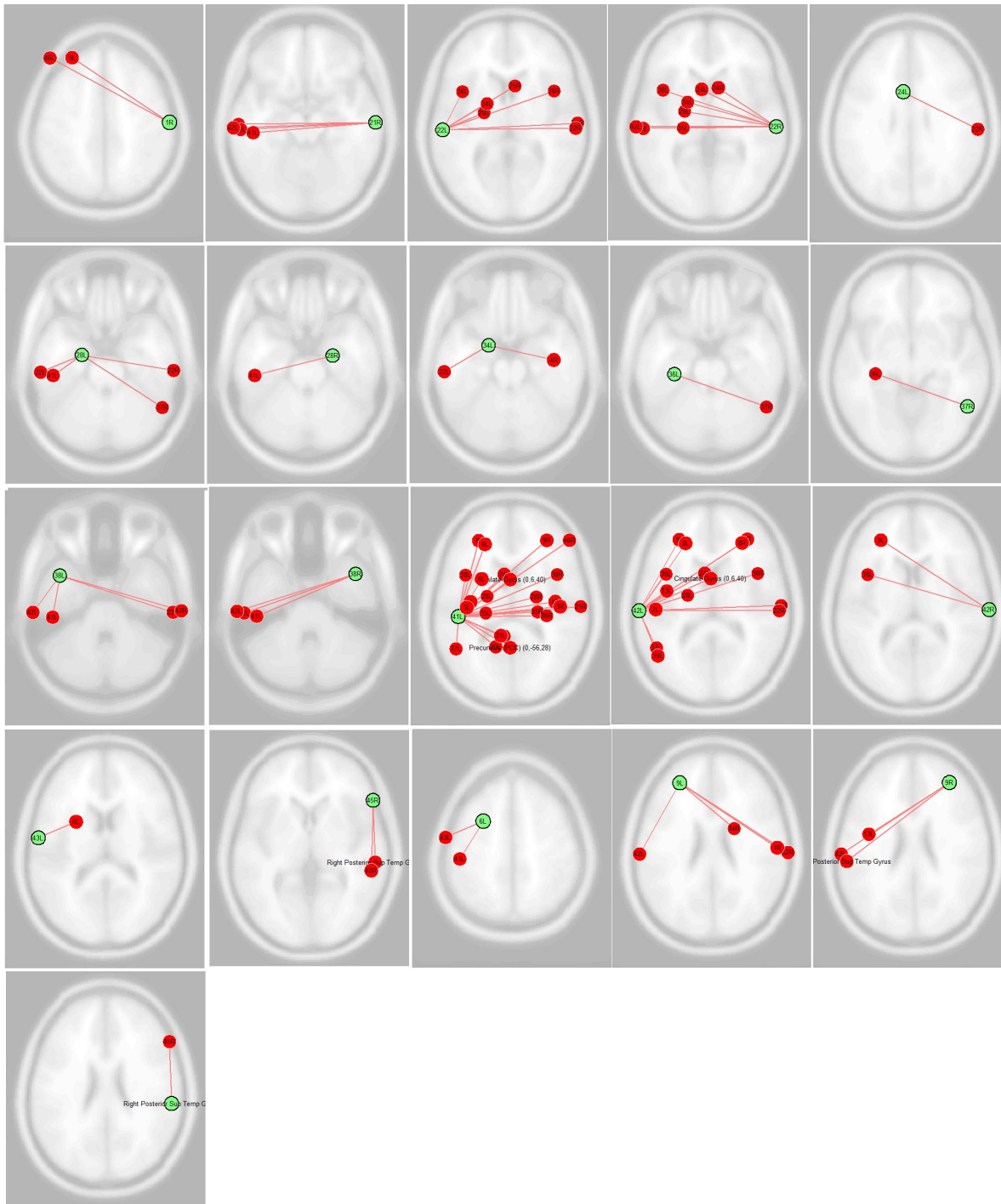
Figure 5.15: All healthy vs. all patients. (R). Primary Somatosensory Cortex,(R). Middle Temporal Gyrus, (L). Superior Temporal Gyrus, (R). Superior Temporal Gyrus, (L). Ventral Anterior Cingulate Cortex, (L). Posterior Entorhinal Cortex, ,(R). Posterior Entorhinal Cortex, (L). Anterior Entorhinal Cortex (L). Parahippocampal cortex, (R). Fusiform gyrus,, (L). Temporopolar Area (R). Temporopolar Area, (L). Primary Auditory Cortex,, (L). Primary Auditory Cortex,, (R). Primary Auditory Cortex (L). Subcentral Area, (R). IFC pars opercularis, (L). Premotor Cortex, (L). Dorsolateral Prefrontal Cortex, (R). Dorsolateral Prefrontal Cortex,, Right Posterior Sup Temp Gyrus (60,-30,24) . ROI-to-ROI connections from connectivity matrix (99 × 99 ROIs), p-FDR=0.05

125

5.8 Connectivity difference based on Network-Based Statistics

The ROI-to-ROI functional connectivity networks properties for the selected between-subjects the graph network's measures are calculated as follow:

1. Network nodes are limited the network analyzed to that defined by a subset of ROIs.

2. The Network edges defined for the connectivity threshold above which two ROIs are considered connected. It can be defined based on correlation scores, z-scores, or cost values. Here cost=0.15 selected for thresholding the nodes.

3. For each ROI the corresponding measure effect size (global efficiency, local effiency, or cost), as well as T- values, uncorrected p-values, and FDR-corrected p- values for the specified second-level analysis. Here for the thresholding I selected 0.05 for uncorrected p-value.

Here we demonstrate a brief description about the graph theoretical measurements that are used in this step. More description about the measurements can be found in [2, 31, 104].

- **Cost**: Cost is the proportion of connected neighbors for each node n in the graph G. It can be defined as follow [207]:

$$C_n(G) = \frac{1}{\|G\| - 1} . \|G_n\| \tag{5.1}$$

where $C_n(G)$ is cost of node n in the graph G in ROI-level and $\|G\|$ represents the number of nodes in graph G.

- **Global efficiency**: It is the average inverse shortest path distance from node n to all other nodes in the graph [207]. For each node n in the graph G the global efficiency $E_n^{global}(G)$ can be calculated as:

$$E_n^{global}(G) = \frac{1}{\|G\| - 1} . \sum_{m \neq n \in G} d_{nm}^{-1}(G) \tag{5.2}$$

where $d_{nm}(G)$ is the shortest path distance between nodes n and m in graph G.

- **Local efficiency**: Is the average global efficiency across all nodes in the local subgraph of node n [207] and can be defined as follow:

$$E_n^{local}(G) = E_n^{global}(G_n) \tag{5.3}$$

126

where $E_n^{local}(G)$ is the local efficiency of node n in the graph G.

- **Betweenness centrality**: For each node n betweenness centrality is the proportion of all shortest-paths in the network containing it [30]. It can be presented as:

$$C_G(n) = \sum_{s \neq n \neq t \in V} \frac{\sigma_{st}(n)}{\sigma_{st}} \qquad (5.4)$$

where $C_G(n)$ is the betweenness centrality of node n in the graph G, V is the set of vertices, $\sigma_{st}(n)$ is the number of those paths that pass through n and σ_{st} is total number of shortest paths from node s to node t.

- **Average path length**: It is the average shortest-path distance from each node n to all other nodes in the graph.

- **Clustering coefficient**: Is the proportion of connected nodes across all nodes neighboring node n. For a node n in a graph G the clustering coefficient C_n can be defined as follow:

$$C_n = \frac{2l_n}{k_n(k_n - 1)} \qquad (5.5)$$

where k_n is the number of neighbors of node n and l_n is the number of connected pairs between all neighbors of node n [18, 204].

- **Degree**: Degree is the number of connected neighbors for each node n in the graph G.

5.9 Conclusion

In this chapter abnormality of connectivity and it's difference between male and female schizophrenia patients investigated. Our results revealed abnormal temporal gyrus, cingulate cortex, entorhinal cortex and auditory cortex are involved in schizophrenia and support former hypothesis. Also. some of these abnormalities are not identical for men and women. To do this resting state functional data and structural data of 48 healthy controls and schizophrenia patients data and we chose identical numbers of subjects for each genders in each group. The ROI-to-ROI analysing and Network-Based Statistic are done to see the abnormal connectivity in schizophrenia patients. We demonstrated that alteration in connectivity of left superior temporal gyrus is difference between genders.
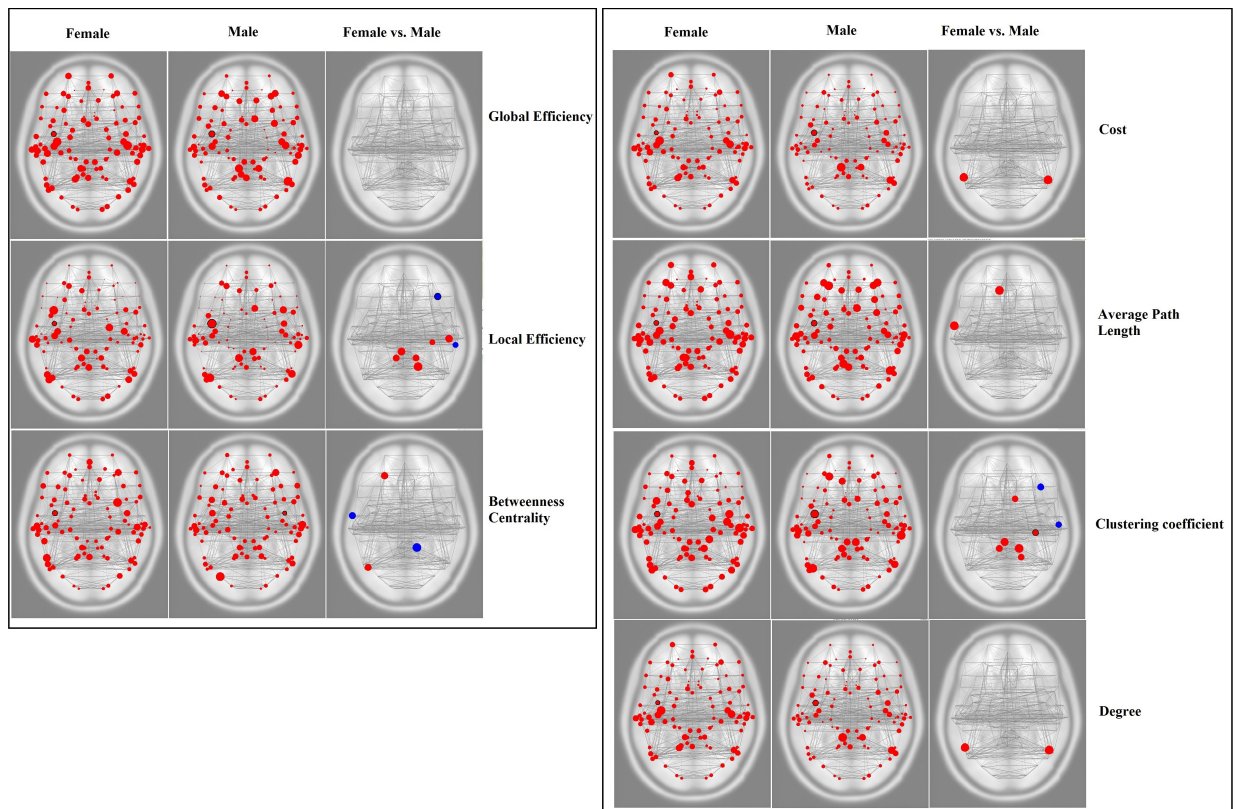
Figure 5.16: Graph-theory measures for the Healthy Women vs. Healthy Men
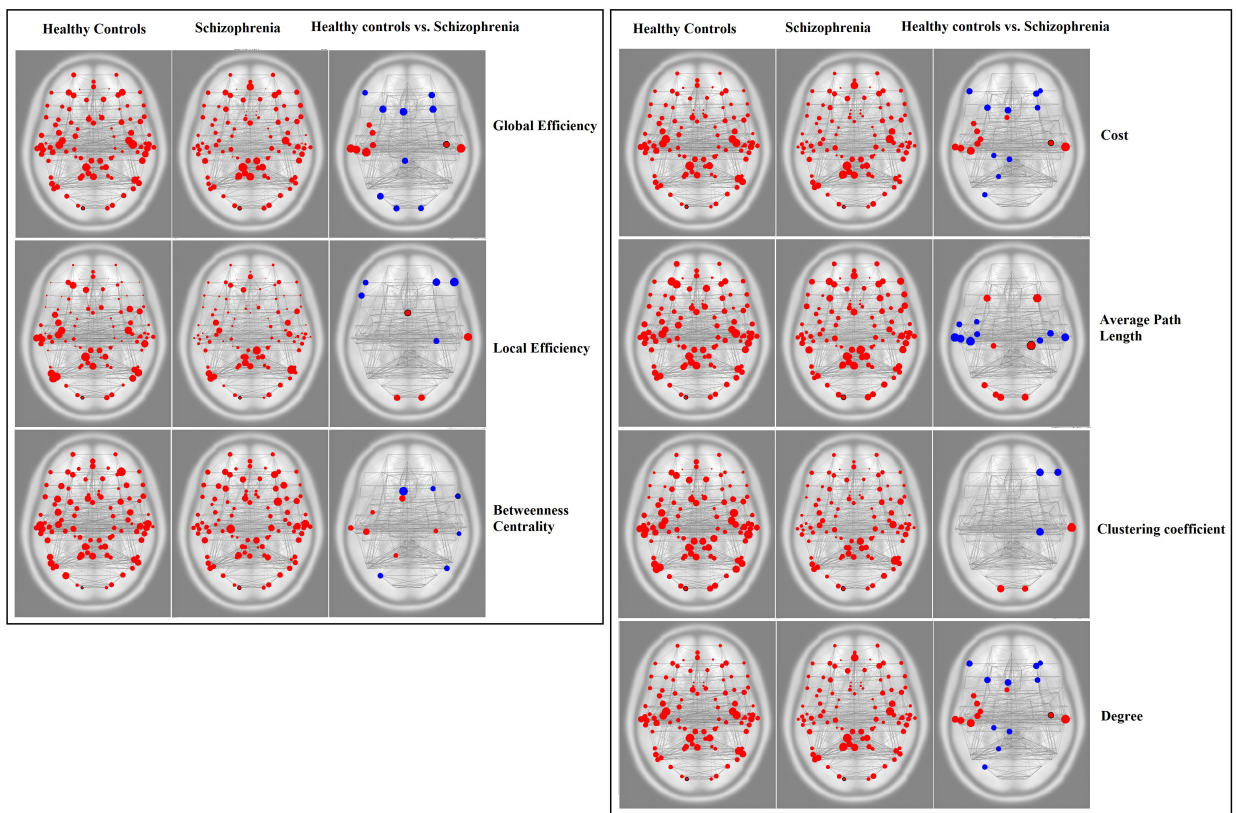
Figure 5.17: Graph-theory measures for all healthy subjects vs. all patients

Chapter 6

Conclusion and future works

In this thesis we considered two main questions in neuroscience; representation of information by neural activity and the structure of the healthy brain. To deal with these problems we looked at several data-sets based acquired using a variety of methods, mainly functional Magnetic Resonance Imaging, and analyzed using a variety of algorithms.

In Chapter 1 we introduced that data and algorithms that are more common in the neuroscience. Although, there are many data sets that we can use for testing the accuracy of prediction and decoding cognitive states and functional connectivity analysis, we focused on a small sample of relevant data sets. For example Science 2008 and Haxby data sets are two famous data sets that we used to test the classification algorithms. Also, we used a new data set called COBRE data that contains heathy brains and the brains with schizophrenia data to analyze the gender differences in schizophrenia.

In Chapter 2 we introduced applications of new feature selection algorithms to brain data sets. Theses new algorithms are overlapping feature selection algorithm and feature selection based on Catastrophe model. Haxby, Science 2008, breast cancer and Parkinson's data sets used for evaluation of these algorithms.

In Chapter 3 we showed the development and evaluation of Neuro-fuzzy model for prediction of Spike discharge. Our results showed that spike discharge can be modeled accurately using the Neuro-fuzzy approach and it can be predicted using first spike latency and frequency-following interval. We showed that ANFIS and genetic for lateral tuning and rule selection of linguistic fuzzy system model have better performance compared with other versions of Neuro-fuzzy models.

In Chapter 4 we discussed application of clustering algorithms and graph theoretical analysis

algorithms to fMRI data for finding modules in the brain.

In Chapter 5 we analyzed a schizophrenia data set and connectivity difference between healthy control brains and schizophrenia. Also, in this chapter we compared male and female patients in schizophrenia. We showed that there is significant difference in some areas like Left Superior Temporal Gyrus between men and women patients.

The following can be considered future work.

- With the development of new technologies for collecting brain data new models and algorithms are necessary to analyze them. Such algorithms should be efficient for solving feature selection, supervised and unsupervised classification problems in high dimensional data sets.

- New mathematical approaches are necessary to model processes in brain. Such approaches may include graph theory, optimization and data mining.

- Applications of these mathematical algorithms and approaches to higher quality data sets of patients with diseases such as schizophrenia will allow to reveal the underlying processes giving rise to psychiatric symptoms.

In conclusion, we demonstrate a successful application of computational techniques to complex brain data sets. The use of these techniques will become increasingly important in the future as new technologies will generate larger and more complex data which could only be analyzed in automated ways. The development of these computational techniques should be combined with development of imaging methods to match the future increase of complexity of the acquired data.

Appendix

Three dimensional location of data points of clusters: subject 02248 and tholerance=0.01 (from chapter 5)

Cluster elements in cluster: 49   Cluster elements in cluster: 50
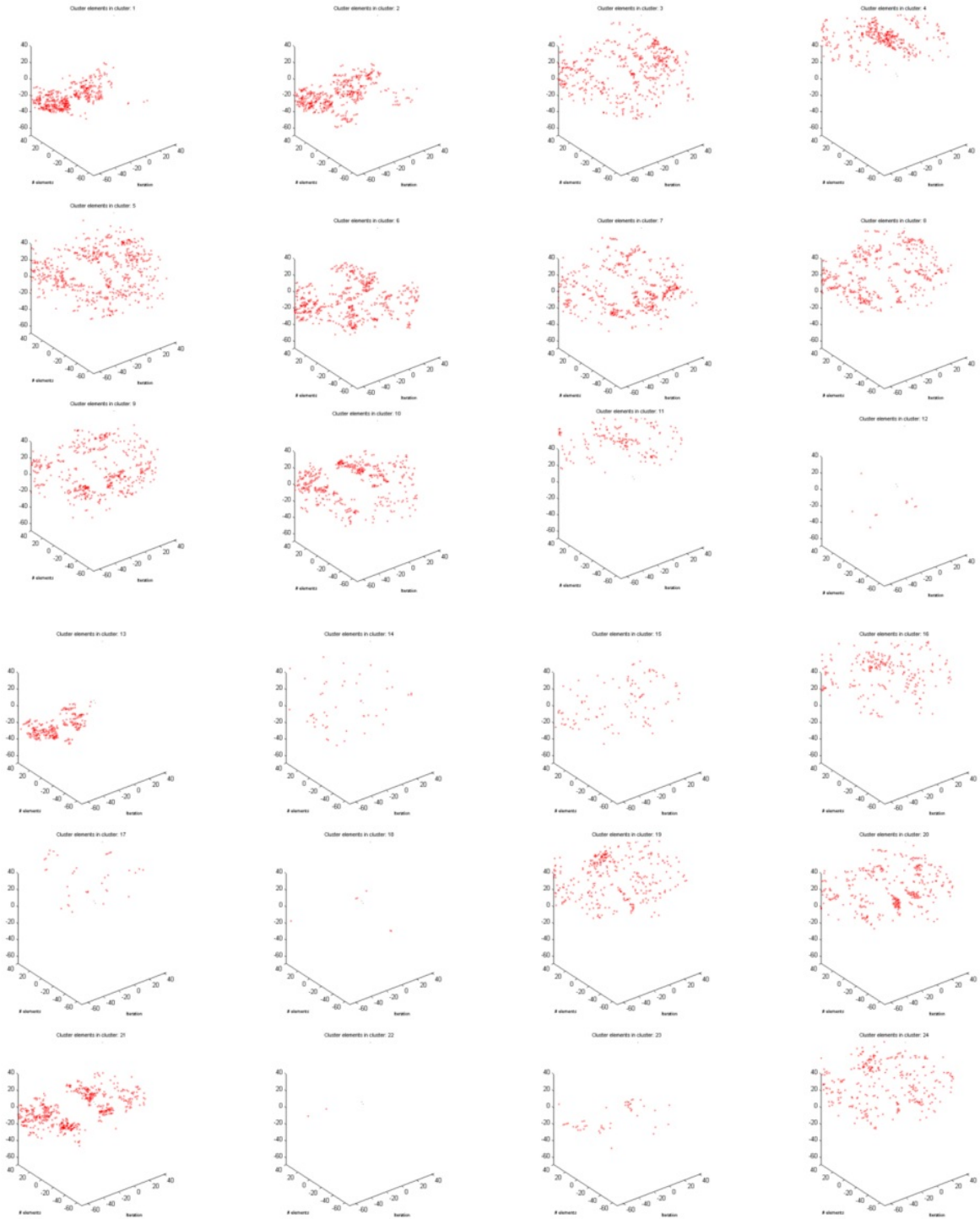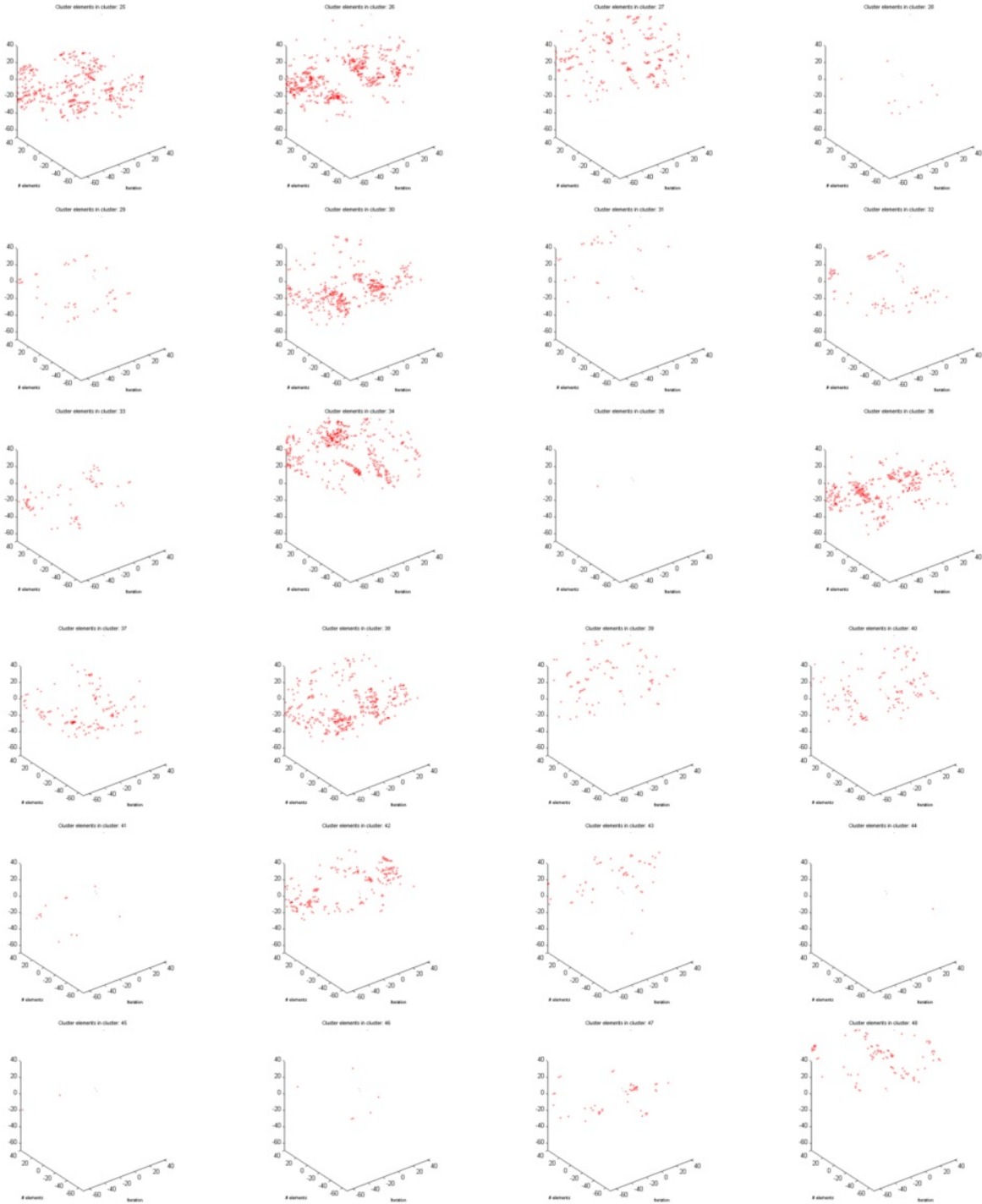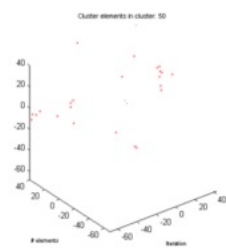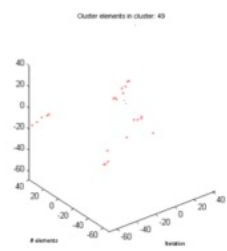
Cluster stability in different iteration: subject 02248 and tolerance 0.0001 (from Chapter 5)
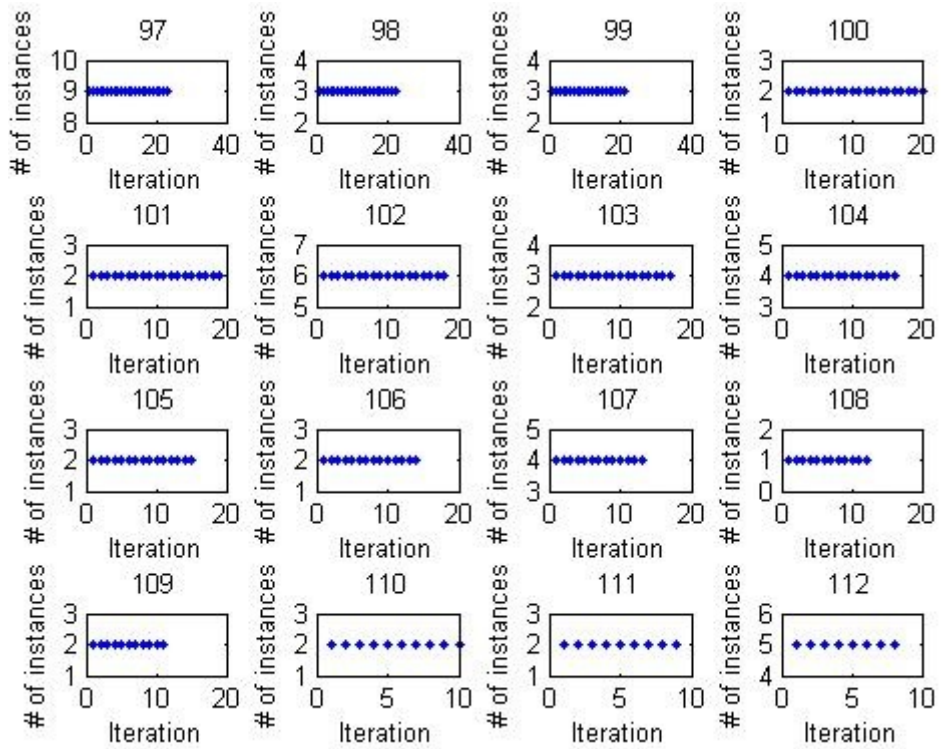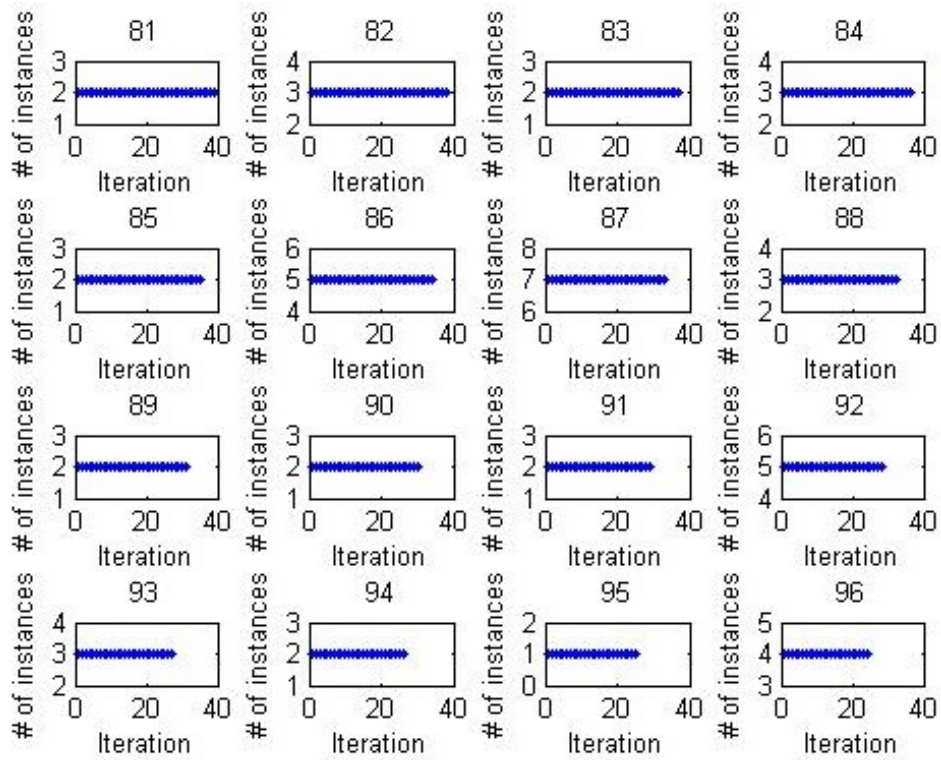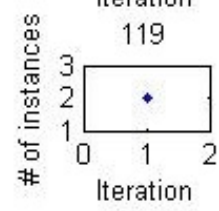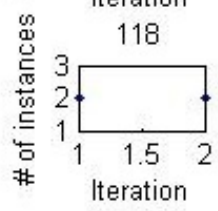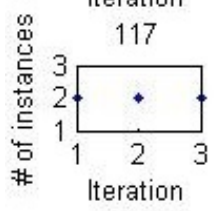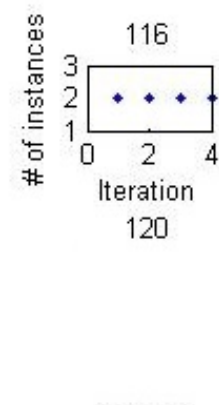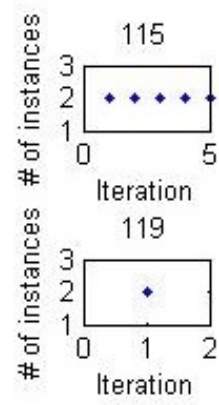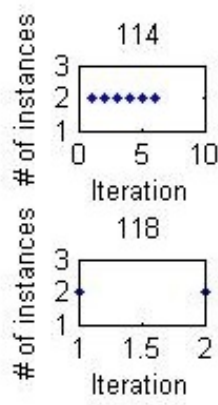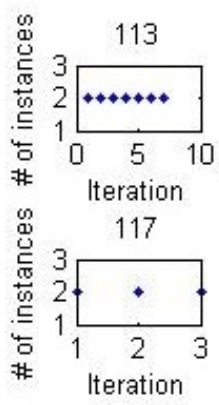
This part is related to Chapter 6

Results: Healthy women vs. healthy men (from Chapter 6)

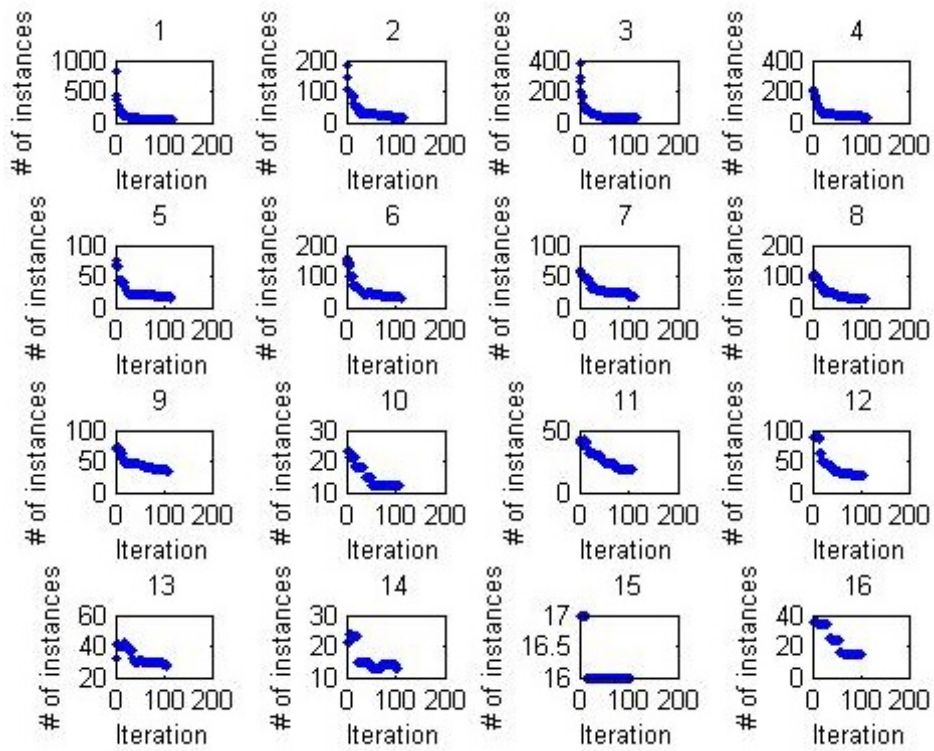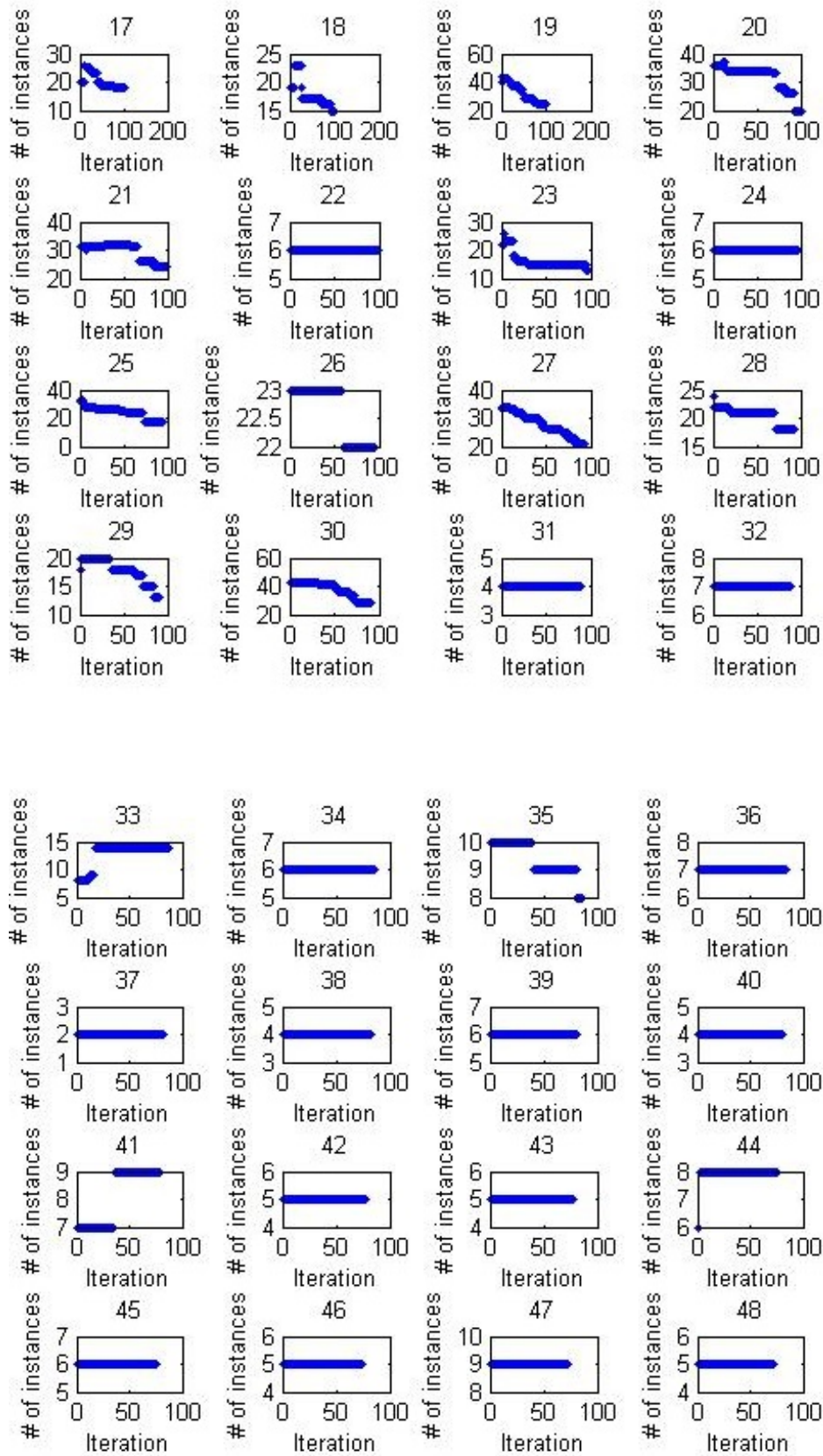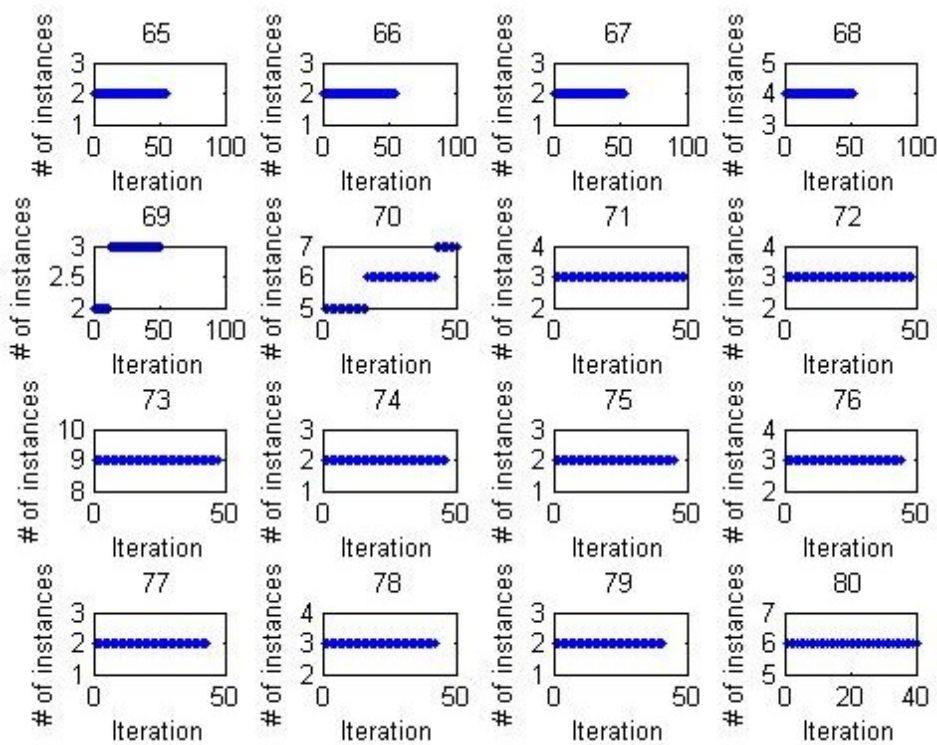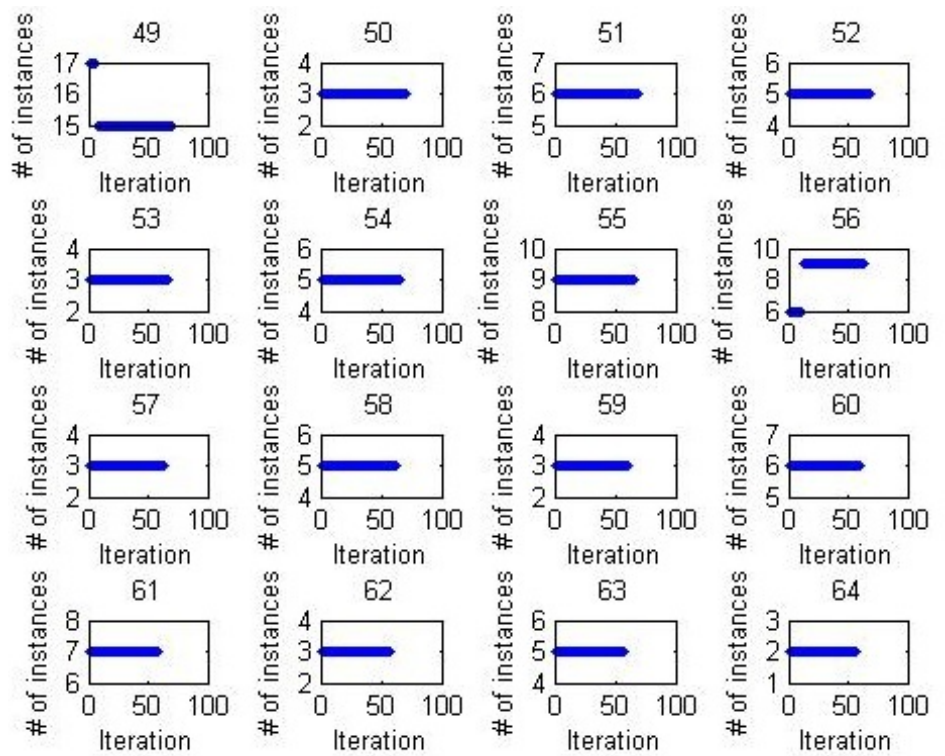| 001 BA.1 (L). Primary Somatosensory Cortex	051 BA.37 (L). Fusiform gyrus
002 BA.1 (R). Primary Somatosensory Cortex	052 BA.37 (R). Fusiform gyrus
003 BA.10 (L). Anterior Prefrontal Cortex	053 BA.38 (L). Temporopolar Area
004 BA.10 (R). Anterior Prefrontal Cortex	054 BA.38 (R). Temporopolar Area
005 BA.11 (L). Orbitofrontal Cortex	055 BA.39 (L). Angular gyrus
006 BA.11 (R). Orbitofrontal Cortex	056 BA.39 (R). Angular gyrus
007 BA.13 (L). Insular Cortex	057 BA.4 (L). Primary Motor Cortex
008 BA.13 (R). Insular Cortex	058 BA.4 (R). Primary Motor Cortex
009 BA.17 (L). Primary Visual Cortex	059 BA.40 (L). Supramarginal Gyrus
010 BA.17 (R). Primary Visual Cortex	060 BA.40 (R). Supramarginal Gyrus
011 BA.18 (L). Secondary Visual Cortex	061 BA.41 (L). Primary Auditory Cortex
012 BA.18 (R). Secondary Visual Cortex	062 BA.41 (R). Primary Auditory Cortex
013 BA.19 (L). Associative Visual Cortex	063 BA.42 (L). Primary Auditory Cortex
014 BA.19 (R). Associative Visual Cortex	064 BA.42 (R). Primary Auditory Cortex
015 BA.2 (L). Primary Somatosensory Cortex	065 BA.43 (L). Subcentral Area
016 BA.2 (R). Primary Somatosensory Cortex	066 BA.43 (R). Subcentral Area
017 BA.20 (L). Inferior Temporal Gyrus	067 BA.44 (L). IFC pars opercularis
018 BA.20 (R). Inferior Temporal Gyrus	068 BA.44 (R). IFC pars opercularis
019 BA.21 (L). Middle Temporal Gyrus	069 BA.45 (L). IFC pars triangularis
020 BA.21 (R). Middle Temporal Gyrus	070 BA.45 (R). IFC pars triangularis
021 BA.22 (L). Superior Temporal Gyrus	071 BA.46 (L). Dorsolateral Prefrontal Cortex
022 BA.22 (R). Superior Temporal Gyrus	072 BA.46 (R). Dorsolateral Prefrontal Cortex
023 BA.23 (L). Ventral Posterior Cingulate Cortex	073 BA.47 (L). Inferior Prefrontal Gyrus
024 BA.23 (R). Ventral Posterior Cingulate Cortex	074 BA.47 (R). Inferior Prefrontal Gyrus
025 BA.24 (L). Ventral Anterior Cingulate Cortex	075 BA.5 (L). Somatosensory Association Cortex
026 BA.24 (R). Ventral Anterior Cingulate Cortex	076 BA.5 (R). Somatosensory Association Cortex
027 BA.25 (L). Subgenual cortex	077 BA.6 (L). Premotor Cortex
028 BA.25 (R). Subgenual cortex	078 BA.6 (R). Premotor Cortex
029 BA.27 (L). Piriform Cortex	079 BA.7 (L). Somatosensory Association Cortex
030 BA.27 (R). Piriform Cortex	080 BA.7 (R). Somatosensory Association Cortex
031 BA.28 (L). Posterior Entorhinal Cortex	081 BA.8 (L). Dorsal Frontal Cortex
032 BA.28 (R). Posterior Entorhinal Cortex	082 BA.8 (R). Dorsal Frontal Cortex
033 BA.29 (L). Retrosplenial Cingulate Cortex	083 BA.9 (L). Dorsolateral Prefrontal Cortex
034 BA.29 (R). Retrosplenial Cingulate Cortex	084 BA.9 (R). Dorsolateral Prefrontal Cortex
035 BA.3 (L). Primary Somatosensory Cortex	085 rsREL.Precuneus (PCC) (0,-56,28)
036 BA.3 (R). Primary Somatosensory Cortex	086 rsREL.Right Inferior Parietal Lobe (RLP) (48,-60,38)
037 BA.30 (L). Cingulate Cortex	087 rsREL.Left Inferior Parietal Lobe (LLP) (-42,-68,38)
038 BA.30 (R). Cingulate Cortex	088 rsREL.Med Prefrontal Cortex (MPFC) (0,54,-8)
039 BA.31 (L). Dorsal Posterior Cingulate Cortex	089 rsREL.Right Posterior Sup Temp Gyrus (60,-30,24)
040 BA.31 (R). Dorsal Posterior Cingulate Cortex	090 rsREL.Right Anterior Sup Temp Gyrus (54,8,-2)
041 BA.32 (L). Dorsal anterior Cingulate Cortex	091 rsREL.Cingulate Gyrus (0,6,40)
042 BA.32 (R). Dorsal anterior Cingulate Cortex	092 rsREL.Right Superior Frontal Gyrus (30,22,52)
043 BA.33 (L). Anterior Cingulate Cortex	093 rsREL.Left Superior Frontal Gyrus (-28,22,52)
044 BA.33 (R). Anterior Cingulate Cortex	094 rsREL.Left Posterior Sup Temp Gyrus (-60,-30,20)
045 BA.34 (L). Anterior Entorhinal Cortex	095 rsREL.Left Anterior Sup Temp Gyrus (-44,4,-4)
046 BA.34 (R). Anterior Entorhinal Cortex	096 LLP
047 BA.35 (L). Perirhinal cortex	097 MPFC
048 BA.35 (R). Perirhinal cortex	098 PCC
049 BA.36 (L). Parahippocampal cortex	099 RLP
050 BA.36 (R). Parahippocampal cortex	

Table 6.2: (L) Dorsal anterior Cingulate Cortex

Targets	beta	T(22)	p-unc	p-FDR
BA.21 (L). Middle Temporal Gyrus	**-0.3**	**-4.72**	**0.000103**	**0.010197**
BA.19 (R). Associative Visual Cortex	**-0.29**	**-3.93**	**0.000714**	**0.035348**
BA.22 (L). Superior Temporal Gyrus	-0.25	-3.54	0.001833	0.06049
BA.22 (R). Superior Temporal Gyrus	-0.2	-3.11	0.005058	0.125192
BA.38 (L). Temporopolar Area	-0.2	-2.8	0.010476	0.204602
BA.19 (L). Associative Visual Cortex	-0.27	-2.72	0.0124	0.204602
Grey Matter	-0.23	-2.6	0.016325	0.23088
BA.38 (R). Temporopolar Area	-0.16	-2.26	0.033727	0.36936
BA.23 (R). Ventral Posterior Ci*rtex	-0.18	-2.25	0.034752	0.36936
BA.18 (R). Secondary Visual Cortex	-0.14	-2.22	0.037309	0.36936
BA.18 (L). Secondary Visual Cortex	-0.16	-2.17	0.041325	0.371923
BA.1 (L). Primary Somatosensory*rtex	0.18	2.09	0.048413	0.386337
BA.21 (R). Middle Temporal Gyrus	-0.18	-2.07	0.050731	0.386337
rsREL.Precuneus (PCC) (0,-56,28)	-0.18	-1.91	0.068704	0.444616
BA.17 (R). Primary Visual Cortex	-0.12	-1.86	0.075581	0.444616
BA.47 (L). Inferior Prefrontal Gyrus	-0.15	-1.85	0.078384	0.444616
BA.39 (L). Angular gyrus	-0.16	-1.84	0.079083	0.444616
BA.44 (L). IFC pars opercularis	-0.15	-1.81	0.083965	0.444616
BA.46 (R). Dorsolateral Prefron*rtex	-0.17	-1.74	0.095712	0.444616
BA.29 (L). Retrosplenial Cingul*rtex	0.14	1.74	0.095863	0.444616
BA.37 (L). Fusiform gyrus	-0.17	-1.69	0.105284	0.444616
BA.43 (L). Subcentral Area	-0.12	-1.69	0.106	0.444616
BA.34 (L). Anterior Entorhinal *rtex	-0.17	-1.68	0.107394	0.444616
BA.31 (L). Dorsal Posterior Cin*rtex	-0.14	-1.68	0.107786	0.444616
rsREL.Cingulate Gyrus (0,6,40)	0.16	1.58	0.128131	0.471705
BA.43 (R). Subcentral Area	-0.12	-1.55	0.135724	0.471705
BA.45 (R). IFC pars triangularis	-0.12	-1.54	0.137024	0.471705
BA.45 (L). IFC pars triangularis	-0.14	-1.54	0.138838	0.471705
BA.4 (L). Primary Motor Cortex	-0.12	-1.53	0.13963	0.471705
BA.3 (R). Primary Somatosensory*rtex	-0.11	-1.5	0.147288	0.471705
rsREL.Right Superior Frontal Gy*,52)	-0.12	-1.5	0.147963	0.471705
BA.9 (R). Dorsolateral Prefront*rtex	-0.14	-1.48	0.15247	0.471705
BA.34 (R). Anterior Entorhinal *rtex	-0.11	-1.46	0.157377	0.47213
BA.35 (R). Perirhinal cortex	-0.16	-1.43	0.167678	0.482841
BA.33 (L). Anterior Cingulate Cortex	-0.15	-1.42	0.170701	0.482841
PCC	-0.1	-1.33	0.195923	0.508877
BA.39 (R). Angular gyrus	-0.13	-1.33	0.196002	0.508877
BA.4 (R). Primary Motor Cortex	-0.1	-1.33	0.196866	0.508877
BA.30 (R). Cingulate Cortex	-0.11	-1.3	0.2067	0.508877
BA.33 (R). Anterior Cingulate Cortex	-0.13	-1.29	0.210507	0.508877
BA.13 (L). Insular Cortex	-0.08	-1.29	0.211501	0.508877
LLP	-0.11	-1.27	0.215887	0.508877
RLP	-0.11	-1.21	0.23847	0.54242
BA.28 (R). Posterior Entorhinal*rtex	-0.13	-1.19	0.246811	0.54242
BA.35 (L). Perirhinal cortex	-0.13	-1.18	0.250614	0.54242
rsREL.Left Inferior Parietal Lo*,38)	-0.11	-1.16	0.257097	0.54242

Table 6.3: (L) Dorsal Posterior Cingulate Cortex

Targets	beta	T(22)	p-unc	p-FDR
BA.13 (L). Insular Cortex	**-0.26**	**-3.93**	**0.000713**	**0.034421**
BA.44 (L). IFC pars opercularis	**-0.26**	**-3.88**	**0.000814**	**0.034421**
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.31**	**-3.77**	**0.001043**	**0.034421**
BA.44 (R). IFC pars opercularis	-0.28	-3	0.006614	0.163687
rsREL.Right Anterior Sup Temp G*,-2)	-0.24	-2.73	0.012338	0.244295
BA.13 (R). Insular Cortex	-0.22	-2.47	0.021929	0.361829
BA.4 (R). Primary Motor Cortex	-0.18	-2.35	0.027996	0.376402
BA.29 (L). Retrosplenial Cingul*rtex	0.26	2.23	0.036632	0.376402
rsREL.Left Posterior Sup Temp G*,20)	-0.17	-2.16	0.042039	0.376402
BA.3 (R). Primary Somatosensory*rtex	-0.17	-2.11	0.04659	0.376402
rsREL.Right Posterior Sup Temp *,24)	-0.2	-2.08	0.048985	0.376402
BA.42 (L). Primary Auditory Cortex	-0.2	-2.07	0.050854	0.376402
BA.4 (L). Primary Motor Cortex	-0.18	-2.05	0.052755	0.376402
BA.22 (R). Superior Temporal Gyrus	-0.22	-2.04	0.053229	0.376402
BA.45 (R). IFC pars triangularis	-0.18	-2	0.058426	0.384568
BA.40 (R). Supramarginal Gyrus	-0.22	-1.97	0.062152	0.384568
BA.3 (L). Primary Somatosensory*rtex	-0.16	-1.92	0.068364	0.398122
BA.19 (R). Associative Visual Cortex	-0.15	-1.81	0.083212	0.457668
BA.6 (R). Premotor Cortex	-0.2	-1.79	0.088034	0.458702
BA.22 (L). Superior Temporal Gyrus	-0.19	-1.7	0.103441	0.46577
BA.24 (L). Ventral Anterior Cin*rtex	-0.15	-1.69	0.104371	0.46577
BA.32 (L). Dorsal anterior Cing*rtex	-0.14	-1.68	0.107786	0.46577
BA.42 (R). Primary Auditory Cortex	-0.14	-1.65	0.113105	0.46577
BA.20 (R). Inferior Temporal Gyrus	0.14	1.63	0.11707	0.46577
BA.10 (L). Anterior Prefrontal *rtex	0.11	1.61	0.122104	0.46577
BA.33 (L). Anterior Cingulate Cortex	-0.14	-1.57	0.131136	0.46577
BA.41 (R). Primary Auditory Cortex	-0.14	-1.57	0.131639	0.46577
BA.11 (L). Orbitofrontal Cortex	0.13	1.56	0.132282	0.46577
BA.28 (L). Posterior Entorhinal*rtex	0.16	1.54	0.137587	0.46577
BA.9 (R). Dorsolateral Prefront*rtex	-0.12	-1.5	0.148491	0.46577
BA.24 (R). Ventral Anterior Cin*rtex	-0.14	-1.49	0.149131	0.46577
BA.43 (R). Subcentral Area	-0.13	-1.47	0.156813	0.46577
BA.41 (L). Primary Auditory Cortex	-0.15	-1.46	0.157495	0.46577
Grey Matter	-0.15	-1.45	0.162095	0.46577
BA.11 (R). Orbitofrontal Cortex	0.1	1.43	0.167858	0.46577
BA.45 (L). IFC pars triangularis	-0.12	-1.42	0.169981	0.46577
BA.2 (R). Primary Somatosensory*rtex	-0.15	-1.4	0.174076	0.46577
rsREL.Precuneus (PCC) (0,-56,28)	0.14	1.37	0.184976	0.48191
BA.25 (R). Subgenual cortex	0.1	1.33	0.196192	0.4869
BA.20 (L). Inferior Temporal Gyrus	0.11	1.32	0.198883	0.4869
BA.36 (R). Parahippocampal cortex	0.13	1.32	0.201876	0.4869
BA.30 (L). Cingulate Cortex	0.11	1.3	0.206564	0.4869
BA.35 (L). Perirhinal cortex	0.12	1.27	0.216826	0.495951
BA.32 (R). Dorsal anterior Cing*rtex	-0.12	-1.26	0.220423	0.495951
BA.18 (R). Secondary Visual Cortex	-0.1	-1.22	0.234053	0.497394

Table 6.4: (L) IFC pars opercularis

Targets		beta	T(22)	p-unc	p-FDR
BA.31 (R).	**Dorsal Posterior Cin*rtex**	**-0.31**	**-4.44**	**0.000207**	**0.020471**
BA.23 (R).	**Ventral Posterior Ci*rtex**	**-0.27**	**-3.99**	**0.000621**	**0.026848**
BA.31 (L).	**Dorsal Posterior Cin*rtex**	**-0.26**	**-3.88**	**0.000814**	**0.026848**
BA.41 (L).	**Primary Auditory Cortex**	**-0.27**	**-3.64**	**0.001449**	**0.035865**
BA.43 (L).	Subcentral Area	-0.23	-2.93	0.00776	0.150163
BA.29 (R).	Retrosplenial Cingul*rtex	-0.2	-2.78	0.01084	0.150163
BA.23 (L).	Ventral Posterior Ci*rtex	-0.19	-2.78	0.011029	0.150163
PCC		-0.22	-2.73	0.012134	0.150163
BA.41 (R).	Primary Auditory Cortex	-0.18	-2.14	0.044071	0.386097
BA.25 (L).	Subgenual cortex	-0.16	-2.13	0.044766	0.386097
BA.29 (L).	Retrosplenial Cingul*rtex	-0.15	-2.1	0.047136	0.386097
BA.9 (R).	Dorsolateral Prefront*rtex	-0.16	-2.1	0.047791	0.386097
BA.34 (R).	Anterior Entorhinal *rtex	-0.18	-2.07	0.0507	0.386097
Grey Matte	r	-0.16	-1.99	0.05859	0.408127
BA.34 (L).	Anterior Entorhinal *rtex	-0.17	-1.96	0.062791	0.408127
BA.32 (R).	Dorsal anterior Cing*rtex	-0.16	-1.93	0.06596	0.408127
RLP		-0.19	-1.83	0.080409	0.434934
BA.32 (L).	Dorsal anterior Cing*rtex	-0.15	-1.81	0.083965	0.434934
rsREL.Prec	uneus (PCC) (0,-56,28)	-0.19	-1.8	0.084986	0.434934
BA.30 (L).	Cingulate Cortex	-0.13	-1.78	0.089504	0.434934
BA.43 (R).	Subcentral Area	-0.15	-1.74	0.096393	0.434934
BA.30 (R).	Cingulate Cortex	-0.11	-1.74	0.096652	0.434934
BA.22 (L).	Superior Temporal Gyrus	-0.17	-1.67	0.108432	0.445316
BA.39 (R).	Angular gyrus	-0.18	-1.67	0.110012	0.445316
BA.24 (R).	Ventral Anterior Cin*rtex	-0.11	-1.65	0.112453	0.445316
rsREL.Righ	t Inferior Parietal L*,38)	-0.17	-1.6	0.123232	0.469228
BA.25 (R).	Subgenual cortex	-0.14	-1.57	0.129818	0.471752
BA.8 (R).	Dorsal Frontal Cortex	-0.12	-1.56	0.133425	0.471752
BA.28 (R).	Posterior Entorhinal*rtex	-0.15	-1.54	0.138264	0.472004
BA.21 (L).	Middle Temporal Gyrus	-0.18	-1.45	0.160477	0.514094
BA.27 (L).	Piriform Cortex	0.08	1.43	0.165738	0.514094
BA.11 (L).	Orbitofrontal Cortex	-0.1	-1.43	0.166172	0.514094
BA.28 (L).	Posterior Entorhinal*rtex	-0.14	-1.39	0.178041	0.534123
MPFC		-0.16	-1.37	0.185311	0.539582
BA.20 (L).	Inferior Temporal Gyrus	-0.15	-1.33	0.198028	0.542579
BA.27 (R).	Piriform Cortex	0.08	1.32	0.200555	0.542579
BA.20 (R).	Inferior Temporal Gyrus	-0.15	-1.31	0.202782	0.542579
BA.35 (L).	Perirhinal cortex	-0.12	-1.25	0.225521	0.570675
rsREL.Righ	t Superior Frontal Gy*,52)	-0.11	-1.23	0.230417	0.570675
BA.42 (R).	Primary Auditory Cortex	-0.1	-1.23	0.232907	0.570675
BA.36 (L).	Parahippocampal cortex	-0.13	-1.22	0.23634	0.570675
BA.11 (R).	Orbitofrontal Cortex	-0.06	-1.12	0.27461	0.632436
BA.7 (R).	Somatosensory Associa*rtex	-0.1	-1.09	0.285811	0.632436
BA.21 (R).	Middle Temporal Gyrus	-0.13	-1.09	0.286013	0.632436
BA.9 (L).	Dorsolateral Prefront*rtex	-0.08	-1.08	0.290138	0.632436

Table 6.5: (L) Ventral Posterior Cingulate Cortex

Targets	beta	T(22)	p-unc	p-FDR
BA.13 (L). Insular Cortex	**-0.25**	**-5.22**	**0.000031**	**0.003089**
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.27**	**-3.91**	**0.000745**	**0.033781**
BA.13 (R). Insular Cortex	**-0.25**	**-3.78**	**0.001024**	**0.033781**
BA.22 (R). Superior Temporal Gyrus	-0.26	-3.24	0.00378	0.093551
BA.41 (R). Primary Auditory Cortex	-0.2	-2.79	0.010742	0.156123
BA.44 (L). IFC pars opercularis	-0.19	-2.78	0.011029	0.156123
rsREL.Right Posterior Sup Temp *,24)	-0.22	-2.73	0.012312	0.156123
BA.42 (L). Primary Auditory Cortex	-0.24	-2.68	0.013596	0.156123
rsREL.Right Anterior Sup Temp G*,-2)	-0.23	-2.66	0.014193	0.156123
rsREL.Left Posterior Sup Temp G*,20)	-0.2	-2.55	0.018057	0.178763
BA.41 (L). Primary Auditory Cortex	-0.2	-2.23	0.036173	0.322663
BA.4 (R). Primary Motor Cortex	-0.12	-2.19	0.039111	0.322663
BA.44 (R). IFC pars opercularis	-0.2	-2.09	0.048612	0.370195
BA.22 (L). Superior Temporal Gyrus	-0.17	-2	0.057663	0.40776
BA.42 (R). Primary Auditory Cortex	-0.16	-1.87	0.075105	0.430405
BA.19 (R). Associative Visual Cortex	-0.14	-1.86	0.076729	0.430405
BA.24 (R). Ventral Anterior Cin*rtex	-0.17	-1.85	0.078092	0.430405
BA.20 (R). Inferior Temporal Gyrus	0.13	1.84	0.078777	0.430405
rsREL.Cingulate Gyrus (0,6,40)	-0.18	-1.82	0.082603	0.430405
BA.43 (R). Subcentral Area	-0.14	-1.78	0.088512	0.438136
BA.4 (L). Primary Motor Cortex	-0.13	-1.71	0.101822	0.461053
BA.11 (R). Orbitofrontal Cortex	0.09	1.7	0.102456	0.461053
BA.40 (R). Supramarginal Gyrus	-0.15	-1.67	0.109914	0.473109
BA.40 (L). Supramarginal Gyrus	-0.13	-1.55	0.134396	0.537033
BA.33 (R). Anterior Cingulate Cortex	-0.12	-1.53	0.140616	0.537033
BA.2 (L). Primary Somatosensory*rtex	-0.1	-1.49	0.150775	0.537033
BA.20 (L). Inferior Temporal Gyrus	0.11	1.49	0.151163	0.537033
BA.23 (R). Ventral Posterior Ci*rtex	-0.17	-1.47	0.156063	0.537033
BA.45 (R). IFC pars triangularis	-0.12	-1.46	0.158511	0.537033
BA.33 (L). Anterior Cingulate Cortex	-0.1	-1.43	0.165739	0.537033
BA.24 (L). Ventral Anterior Cin*rtex	-0.12	-1.41	0.172203	0.537033
BA.35 (L). Perirhinal cortex	0.12	1.41	0.173586	0.537033
BA.10 (L). Anterior Prefrontal *rtex	0.12	1.37	0.183268	0.538278
BA.37 (R). Fusiform gyrus	-0.1	-1.35	0.189423	0.538278
BA.3 (L). Primary Somatosensory*rtex	-0.09	-1.35	0.1903	0.538278
BA.6 (R). Premotor Cortex	-0.12	-1.29	0.211015	0.564253
Grey Matter	-0.11	-1.28	0.215384	0.564253
BA.19 (L). Associative Visual Cortex	-0.1	-1.25	0.224284	0.564253
BA.3 (R). Primary Somatosensory*rtex	-0.07	-1.22	0.23374	0.564253
BA.17 (R). Primary Visual Cortex	-0.07	-1.22	0.234301	0.564253
BA.39 (L). Angular gyrus	0.11	1.19	0.247523	0.564253
BA.5 (L). Somatosensory Associa*rtex	-0.1	-1.18	0.249786	0.564253
BA.32 (R). Dorsal anterior Cing*rtex	-0.13	-1.17	0.254047	0.564253
BA.43 (L). Subcentral Area	-0.1	-1.17	0.255671	0.564253
BA.17 (L). Primary Visual Cortex	0.08	1.17	0.256479	0.564253

Table 6.6: (R) Dorsal Posterior Cingulate Cortex

Targets	beta	T(22)	p-unc	p-FDR
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.31**	**-4.53**	**0.000164**	**0.006959**
BA.44 (L). IFC pars opercularis	**-0.31**	**-4.44**	**0.000207**	**0.006959**
BA.13 (L). Insular Cortex	**-0.29**	**-4.43**	**0.000211**	**0.006959**
BA.44 (R). IFC pars opercularis	-0.26	-3.23	0.003843	0.082388
BA.13 (R). Insular Cortex	-0.26	-3.2	0.004161	0.082388
BA.40 (R). Supramarginal Gyrus	-0.25	-2.98	0.006889	0.113667
BA.10 (L). Anterior Prefrontal *rtex	0.21	2.55	0.01811	0.240575
rsREL.Right Anterior Sup Temp G*,-2)	-0.21	-2.52	0.01944	0.240575
BA.45 (R). IFC pars triangularis	-0.2	-2.4	0.025228	0.277511
rsREL.Left Posterior Sup Temp G*,20)	-0.2	-2.27	0.033084	0.277757
BA.6 (R). Premotor Cortex	-0.25	-2.19	0.039017	0.277757
BA.4 (R). Primary Motor Cortex	-0.19	-2.19	0.039051	0.277757
BA.42 (L). Primary Auditory Cortex	-0.21	-2.19	0.039523	0.277757
BA.29 (L). Retrosplenial Cingul*rtex	0.24	2.16	0.042271	0.277757
rsREL.Right Posterior Sup Temp *,24)	-0.18	-2.13	0.044174	0.277757
BA.3 (R). Primary Somatosensory*rtex	-0.19	-2.13	0.04489	0.277757
BA.41 (R). Primary Auditory Cortex	-0.17	-1.99	0.058947	0.343279
BA.43 (R). Subcentral Area	-0.16	-1.9	0.070344	0.37635
BA.24 (R). Ventral Anterior Cin*rtex	-0.18	-1.87	0.074781	0.37635
BA.22 (R). Superior Temporal Gyrus	-0.18	-1.86	0.07603	0.37635
BA.25 (L). Subgenual cortex	0.1	1.7	0.102962	0.469208
BA.42 (R). Primary Auditory Cortex	-0.16	-1.69	0.104268	0.469208
BA.24 (L). Ventral Anterior Cin*rtex	-0.15	-1.62	0.119451	0.472461
BA.2 (R). Primary Somatosensory*rtex	-0.17	-1.61	0.122243	0.472461
BA.45 (L). IFC pars triangularis	-0.14	-1.59	0.125717	0.472461
rsREL.Med Prefrontal Cortex (MP*,-8)	0.13	1.58	0.128873	0.472461
BA.19 (R). Associative Visual Cortex	-0.15	-1.56	0.133966	0.472461
BA.25 (R). Subgenual cortex	0.11	1.51	0.144101	0.472461
BA.41 (L). Primary Auditory Cortex	-0.15	-1.51	0.145123	0.472461
rsREL.Cingulate Gyrus (0,6,40)	-0.19	-1.51	0.145393	0.472461
BA.33 (L). Anterior Cingulate Cortex	-0.11	-1.5	0.148978	0.472461
BA.32 (R). Dorsal anterior Cing*rtex	-0.14	-1.47	0.155048	0.472461
BA.22 (L). Superior Temporal Gyrus	-0.18	-1.46	0.157487	0.472461
BA.47 (R). Inferior Prefrontal Gyrus	-0.11	-1.4	0.175554	0.483844
BA.11 (L). Orbitofrontal Cortex	0.09	1.39	0.178421	0.483844
BA.28 (L). Posterior Entorhinal*rtex	0.1	1.35	0.191489	0.483844
BA.11 (R). Orbitofrontal Cortex	0.08	1.34	0.193867	0.483844
BA.3 (L). Primary Somatosensory*rtex	-0.11	-1.34	0.195401	0.483844
BA.4 (L). Primary Motor Cortex	-0.12	-1.32	0.200801	0.483844
BA.7 (L). Somatosensory Associa*rtex	-0.1	-1.32	0.200942	0.483844
BA.5 (L). Somatosensory Associa*rtex	-0.12	-1.31	0.203548	0.483844
rsREL.Left Superior Frontal Gyr*,52)	0.1	1.3	0.207745	0.483844
Grey Matter	-0.13	-1.29	0.210155	0.483844
BA.43 (L). Subcentral Area	-0.09	-1.24	0.228113	0.513254
BA.9 (R). Dorsolateral Prefront*rtex	-0.08	-1.2	0.243605	0.535103

Table 6.7: (R) Ventral Posterior Cingulate Cortex

Targets	beta	T(22)	p-unc	p-FDR
BA.44 (L). IFC pars opercularis	**-0.27**	**-3.99**	**0.000621**	**0.042188**
BA.13 (L). Insular Cortex	**-0.27**	**-3.86**	**0.000852**	**0.042188**
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.29**	**-3.67**	**0.001334**	**0.044035**
BA.22 (R). Superior Temporal Gyrus	-0.27	-3.29	0.003346	0.082809
rsREL.Right Anterior Sup Temp G*,-2)	-0.29	-3.07	0.005619	0.111262
BA.13 (R). Insular Cortex	-0.27	-2.91	0.008129	0.134122
BA.41 (R). Primary Auditory Cortex	-0.23	-2.77	0.011141	0.149372
rsREL.Left Posterior Sup Temp G*,20)	-0.21	-2.74	0.01207	0.149372
BA.42 (L). Primary Auditory Cortex	-0.25	-2.67	0.013888	0.152766
BA.32 (R). Dorsal anterior Cing*rtex	-0.23	-2.59	0.016619	0.164526
BA.44 (R). IFC pars opercularis	-0.25	-2.38	0.026137	0.211064
rsREL.Right Posterior Sup Temp *,24)	-0.19	-2.38	0.026371	0.211064
BA.24 (R). Ventral Anterior Cin*rtex	-0.21	-2.36	0.027715	0.211064
BA.47 (R). Inferior Prefrontal Gyrus	-0.15	-2.28	0.032751	0.22657
BA.32 (L). Dorsal anterior Cing*rtex	-0.18	-2.25	0.034752	0.22657
BA.24 (L). Ventral Anterior Cin*rtex	-0.19	-2.21	0.038101	0.22657
BA.47 (L). Inferior Prefrontal Gyrus	-0.12	-2.17	0.04106	0.22657
BA.45 (R). IFC pars triangularis	-0.19	-2.17	0.041195	0.22657
BA.41 (L). Primary Auditory Cortex	-0.21	-2.05	0.051999	0.270944
rsREL.Cingulate Gyrus (0,6,40)	-0.21	-1.99	0.058725	0.290689
BA.42 (R). Primary Auditory Cortex	-0.17	-1.92	0.068219	0.321605
BA.43 (R). Subcentral Area	-0.16	-1.83	0.080705	0.363174
BA.1 (L). Primary Somatosensory*rtex	0.1	1.8	0.084841	0.365184
BA.40 (R). Supramarginal Gyrus	-0.13	-1.71	0.101873	0.420226
BA.33 (R). Anterior Cingulate Cortex	-0.14	-1.67	0.10884	0.42659
BA.29 (L). Retrosplenial Cingul*rtex	0.2	1.64	0.115503	0.42659
BA.4 (R). Primary Motor Cortex	-0.09	-1.62	0.120131	0.42659
BA.34 (L). Anterior Entorhinal *rtex	-0.1	-1.61	0.120652	0.42659
rsREL.Left Superior Frontal Gyr*,52)	0.12	1.56	0.132087	0.446173
BA.22 (L). Superior Temporal Gyrus	-0.16	-1.55	0.135204	0.446173
BA.23 (L). Ventral Posterior Ci*rtex	-0.17	-1.47	0.156063	0.494535
BA.10 (L). Anterior Prefrontal *rtex	0.11	1.44	0.1653	0.494535
BA.20 (R). Inferior Temporal Gyrus	0.1	1.42	0.168589	0.494535
BA.20 (L). Inferior Temporal Gyrus	0.09	1.41	0.171098	0.494535
BA.33 (L). Anterior Cingulate Cortex	-0.11	-1.4	0.174836	0.494535
BA.27 (R). Piriform Cortex	-0.12	-1.38	0.182657	0.502306
BA.8 (L). Dorsal Frontal Cortex	0.1	1.33	0.198446	0.530977
BA.37 (L). Fusiform gyrus	-0.09	-1.26	0.219818	0.572683
BA.19 (L). Associative Visual Cortex	-0.11	-1.24	0.226357	0.574598
BA.19 (R). Associative Visual Cortex	-0.1	-1.17	0.25502	0.631173
Grey Matter	-0.1	-1.14	0.265853	0.64128
BA.1 (R). Primary Somatosensory*rtex	0.07	1.13	0.272058	0.64128
BA.31 (R). Dorsal Posterior Cin*rtex	0.11	1.08	0.293709	0.676213
BA.6 (R). Premotor Cortex	-0.11	-1.02	0.318142	0.715821
BA.2 (R). Primary Somatosensory*rtex	-0.08	-0.99	0.33459	0.736098

Table 6.8: L Insular Cortex

Targets	beta	T(22)	p-unc	p-FDR
BA.23 (L). Ventral Posterior Ci*rtex	**-0.25**	**-5.22**	**0.000031**	**0.003089**
BA.31 (R). Dorsal Posterior Cin*rtex	**-0.29**	**-4.43**	**0.000211**	**0.010439**
BA.31 (L). Dorsal Posterior Cin*rtex	**-0.26**	**-3.93**	**0.000713**	**0.018262**
BA.23 (R). Ventral Posterior Ci*rtex	**-0.27**	**-3.86**	**0.000852**	**0.018262**
PCC	**-0.27**	**-3.83**	**0.000922**	**0.018262**
BA.21 (L). Middle Temporal Gyrus	**-0.31**	**-3.43**	**0.00237**	**0.039104**
BA.21 (R). Middle Temporal Gyrus	**-0.34**	**-3.28**	**0.003381**	**0.047822**
rsREL.Precuneus (PCC) (0,-56,28)	-0.2	-2.99	0.006772	0.083807
BA.22 (R). Superior Temporal Gyrus	-0.3	-2.82	0.009896	0.108859
BA.39 (L). Angular gyrus	-0.17	-2.55	0.018238	0.174288
BA.39 (R). Angular gyrus	-0.19	-2.52	0.019365	0.174288
BA.29 (R). Retrosplenial Cingul*rtex	-0.17	-2.4	0.025067	0.206799
RLP	-0.15	-2.28	0.032435	0.247001
BA.41 (R). Primary Auditory Cortex	-0.2	-2.23	0.035945	0.247771
BA.42 (R). Primary Auditory Cortex	-0.18	-2.21	0.037541	0.247771
rsREL.Right Inferior Parietal L*,38)	-0.17	-2.16	0.042203	0.26113
BA.41 (L). Primary Auditory Cortex	-0.19	-2.09	0.048108	0.266088
BA.32 (R). Dorsal anterior Cing*rtex	-0.15	-2.08	0.04928	0.266088
BA.20 (L). Inferior Temporal Gyrus	-0.17	-2.06	0.051067	0.266088
BA.45 (R). IFC pars triangularis	-0.18	-1.94	0.06508	0.322148
BA.22 (L). Superior Temporal Gyrus	-0.19	-1.83	0.080354	0.378812
BA.29 (L). Retrosplenial Cingul*rtex	-0.13	-1.77	0.089796	0.404083
BA.38 (R). Temporopolar Area	-0.18	-1.72	0.099584	0.428643
BA.9 (R). Dorsolateral Prefront*rtex	-0.12	-1.65	0.112809	0.455092
BA.8 (R). Dorsal Frontal Cortex	-0.13	-1.64	0.114922	0.455092
rsREL.Med Prefrontal Cortex (MP*,-8)	-0.12	-1.58	0.129216	0.492016
Grey Matter	-0.16	-1.51	0.145544	0.533663
BA.35 (R). Perirhinal cortex	-0.14	-1.47	0.155813	0.55091
rsREL.Left Superior Frontal Gyr*,52)	0.11	1.42	0.170392	0.571547
LLP	-0.12	-1.41	0.173196	0.571547
BA.24 (R). Ventral Anterior Cin*rtex	-0.11	-1.34	0.193806	0.59628
BA.46 (L). Dorsolateral Prefron*rtex	0.1	1.3	0.206768	0.59628
BA.30 (L). Cingulate Cortex	-0.11	-1.3	0.207072	0.59628
MPFC	-0.09	-1.3	0.208527	0.59628
BA.32 (L). Dorsal anterior Cing*rtex	-0.08	-1.29	0.211501	0.59628
BA.30 (R). Cingulate Cortex	-0.11	-1.27	0.218218	0.59628
BA.43 (L). Subcentral Area	-0.13	-1.24	0.226901	0.59628
BA.20 (R). Inferior Temporal Gyrus	-0.12	-1.23	0.23349	0.59628
rsREL.Left Inferior Parietal Lo*,38)	-0.1	-1.22	0.237139	0.59628
BA.25 (L). Subgenual cortex	-0.12	-1.21	0.240921	0.59628
rsREL.Left Anterior Sup Temp Gy*,-4)	-0.09	-1.18	0.252549	0.606635
BA.2 (R). Primary Somatosensory*rtex	-0.11	-1.16	0.258279	0.606635
BA.1 (R). Primary Somatosensory*rtex	-0.1	-1.15	0.263488	0.606635
BA.47 (R). Inferior Prefrontal Gyrus	-0.1	-1.11	0.277265	0.623847
BA.35 (L). Perirhinal cortex	-0.1	-1.09	0.286622	0.630568

Table 6.9: L Middle Temporal Gyrus

Targets	beta	T(22)	p-unc	p-FDR
BA.32 (L). Dorsal anterior Cing*rtex	**-0.3**	**-4.72**	**0.000103**	**0.010197**
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.42**	**-4.23**	**0.000342**	**0.016947**
BA.13 (L). Insular Cortex	-0.31	-3.43	0.00237	0.078208
rsREL.Right Anterior Sup Temp G*,-2)	-0.3	-2.84	0.00944	0.233628
BA.40 (R). Supramarginal Gyrus	-0.26	-2.7	0.012933	0.248856
BA.42 (L). Primary Auditory Cortex	-0.26	-2.55	0.01841	0.248856
BA.43 (L). Subcentral Area	-0.17	-2.51	0.020046	0.248856
BA.2 (L). Primary Somatosensory*rtex	-0.22	-2.51	0.02011	0.248856
BA.13 (R). Insular Cortex	-0.24	-2.2	0.038801	0.372374
BA.6 (R). Premotor Cortex	-0.21	-2.15	0.042718	0.372374
BA.32 (R). Dorsal anterior Cing*rtex	-0.2	-2.1	0.0479	0.372374
Grey Matter	-0.14	-2.07	0.050849	0.372374
BA.22 (L). Superior Temporal Gyrus	-0.23	-2.06	0.051661	0.372374
BA.44 (R). IFC pars opercularis	-0.24	-2.01	0.056422	0.372374
rsREL.Left Inferior Parietal Lo*,38)	0.28	2.01	0.056822	0.372374
MPFC	-0.19	-1.97	0.062152	0.372374
BA.24 (R). Ventral Anterior Cin*rtex	-0.16	-1.94	0.064875	0.372374
BA.9 (R). Dorsolateral Prefront*rtex	-0.19	-1.92	0.067704	0.372374
BA.17 (R). Primary Visual Cortex	-0.11	-1.87	0.07543	0.372412
rsREL.Right Posterior Sup Temp *,24)	-0.17	-1.83	0.080588	0.372412
BA.47 (R). Inferior Prefrontal Gyrus	-0.19	-1.83	0.080724	0.372412
BA.24 (L). Ventral Anterior Cin*rtex	-0.16	-1.82	0.082758	0.372412
rsREL.Right Inferior Parietal L*,38)	0.18	1.75	0.093918	0.404255
LLP	0.25	1.71	0.1019	0.420336
rsREL.Cingulate Gyrus (0,6,40)	-0.13	-1.67	0.109364	0.432864
BA.4 (L). Primary Motor Cortex	-0.14	-1.65	0.113682	0.432864
BA.41 (L). Primary Auditory Cortex	-0.15	-1.61	0.121796	0.446585
BA.42 (R). Primary Auditory Cortex	-0.14	-1.5	0.14789	0.478978
RLP	0.19	1.49	0.149598	0.478978
BA.43 (R). Subcentral Area	-0.12	-1.49	0.151622	0.478978
BA.44 (L). IFC pars opercularis	-0.18	-1.45	0.160477	0.478978
BA.3 (L). Primary Somatosensory*rtex	-0.12	-1.45	0.161012	0.478978
BA.39 (L). Angular gyrus	0.21	1.44	0.164219	0.478978
BA.22 (R). Superior Temporal Gyrus	-0.15	-1.43	0.165526	0.478978
BA.4 (R). Primary Motor Cortex	-0.13	-1.42	0.169336	0.478978
BA.18 (R). Secondary Visual Cortex	-0.09	-1.39	0.179928	0.494802
rsREL.Med Prefrontal Cortex (MP*,-8)	-0.14	-1.32	0.199528	0.533873
BA.2 (R). Primary Somatosensory*rtex	-0.14	-1.3	0.206671	0.535642
BA.3 (R). Primary Somatosensory*rtex	-0.11	-1.28	0.214971	0.535642
BA.25 (L). Subgenual cortex	-0.13	-1.27	0.216421	0.535642
BA.35 (L). Perirhinal cortex	0.13	1.22	0.234912	0.536403
BA.33 (R). Anterior Cingulate Cortex	0.09	1.22	0.236464	0.536403
BA.18 (L). Secondary Visual Cortex	-0.1	-1.21	0.239343	0.536403
rsREL.Left Posterior Sup Temp G*,20)	-0.1	-1.21	0.239834	0.536403
BA.36 (R). Parahippocampal cortex	0.13	1.2	0.24382	0.536403

Table 6.10: PCC

Targets	beta	T(22)	p-unc	p-FDR
rsREL.Left Anterior Sup Temp Gy*,-4)	**-0.34**	**-3.86**	**0.00085**	**0.045654**
BA.13 (L). Insular Cortex	**-0.27**	**-3.83**	**0.000922**	**0.045654**
rsREL.Right Anterior Sup Temp G*,-2)	-0.28	-3.43	0.002377	0.076463
rsREL.Left Posterior Sup Temp G*,20)	-0.2	-3.3	0.003238	0.076463
rsREL.Right Posterior Sup Temp *,24)	-0.26	-3.23	0.003862	0.076463
BA.9 (R). Dorsolateral Prefront*rtex	-0.23	-3.04	0.006076	0.088446
BA.42 (L). Primary Auditory Cortex	-0.23	-3.02	0.006254	0.088446
BA.2 (L). Primary Somatosensory*rtex	-0.19	-2.94	0.00763	0.094425
BA.44 (R). IFC pars opercularis	-0.27	-2.75	0.01156	0.100456
rsREL.Cingulate Gyrus (0,6,40)	-0.24	-2.73	0.012118	0.100456
BA.44 (L). IFC pars opercularis	-0.22	-2.73	0.012134	0.100456
BA.40 (R). Supramarginal Gyrus	-0.28	-2.73	0.012176	0.100456
BA.13 (R). Insular Cortex	-0.22	-2.63	0.01542	0.117429
BA.6 (R). Premotor Cortex	-0.29	-2.5	0.02017	0.142633
BA.3 (L). Primary Somatosensory*rtex	-0.18	-2.45	0.022702	0.149836
BA.20 (L). Inferior Temporal Gyrus	0.2	2.3	0.031108	0.192482
BA.22 (L). Superior Temporal Gyrus	-0.23	-2.18	0.040608	0.225613
BA.10 (L). Anterior Prefrontal *rtex	0.14	2.16	0.041873	0.225613
BA.2 (R). Primary Somatosensory*rtex	-0.19	-2.13	0.044489	0.225613
BA.20 (R). Inferior Temporal Gyrus	0.2	2.12	0.045578	0.225613
BA.4 (L). Primary Motor Cortex	-0.16	-2.08	0.049508	0.233395
BA.41 (L). Primary Auditory Cortex	-0.17	-1.98	0.060851	0.27383
BA.24 (L). Ventral Anterior Cin*rtex	-0.16	-1.85	0.078339	0.327725
BA.41 (R). Primary Auditory Cortex	-0.17	-1.82	0.082367	0.327725
BA.24 (R). Ventral Anterior Cin*rtex	-0.16	-1.82	0.082759	0.327725
BA.22 (R). Superior Temporal Gyrus	-0.19	-1.75	0.094636	0.350226
BA.42 (R). Primary Auditory Cortex	-0.15	-1.74	0.095516	0.350226
BA.4 (R). Primary Motor Cortex	-0.12	-1.65	0.113931	0.37476
RLP	0.2	1.64	0.115401	0.37476
BA.7 (R). Somatosensory Associa*rtex	-0.15	-1.63	0.116854	0.37476
BA.29 (L). Retrosplenial Cingul*rtex	0.2	1.63	0.117349	0.37476
BA.11 (L). Orbitofrontal Cortex	0.12	1.59	0.125086	0.386986
BA.3 (R). Primary Somatosensory*rtex	-0.12	-1.53	0.140304	0.393334
BA.40 (L). Supramarginal Gyrus	-0.17	-1.52	0.142465	0.393334
Grey Matter	-0.14	-1.5	0.147894	0.393334
BA.19 (R). Associative Visual Cortex	-0.12	-1.5	0.148857	0.393334
LLP	0.18	1.49	0.149532	0.393334
rsREL.Right Inferior Parietal L*,38)	0.17	1.49	0.150977	0.393334
rsREL.Precuneus (PCC) (0,-56,28)	0.19	1.47	0.156741	0.397882
BA.6 (L). Premotor Cortex	-0.13	-1.44	0.162786	0.401401
BA.35 (L). Perirhinal cortex	0.14	1.42	0.168542	0.401401
rsREL.Right Superior Frontal Gy*,52)	-0.16	-1.42	0.170291	0.401401
BA.36 (R). Parahippocampal cortex	0.13	1.36	0.188315	0.433562
BA.32 (L). Dorsal anterior Cing*rtex	-0.1	-1.33	0.195923	0.440827
BA.11 (R). Orbitofrontal Cortex	0.1	1.27	0.216309	0.462038

Table 6.11: rsREL.Left Anterior Sup Temp Gyrus

Targets	beta	T(22)	p-unc	p-FDR
BA.31 (R). Dorsal Posterior Cin*rtex	**-0.31**	**-4.53**	**0.000164**	**0.016271**
BA.21 (L). Middle Temporal Gyrus	**-0.42**	**-4.23**	**0.000342**	**0.016947**
BA.23 (L). Ventral Posterior Ci*rtex	**-0.27**	**-3.91**	**0.000745**	**0.018872**
PCC	**-0.34**	**-3.86**	**0.00085**	**0.018872**
BA.31 (L). Dorsal Posterior Cin*rtex	**-0.31**	**-3.77**	**0.001043**	**0.018872**
rsREL.Precuneus (PCC) (0,-56,28)	**-0.32**	**-3.73**	**0.001158**	**0.018872**
BA.23 (R). Ventral Posterior Ci*rtex	**-0.29**	**-3.67**	**0.001334**	**0.018872**
BA.8 (R). Dorsal Frontal Cortex	**-0.24**	**-3.26**	**0.003564**	**0.040179**
BA.29 (R). Retrosplenial Cingul*rtex	**-0.26**	**-3.22**	**0.003905**	**0.040179**
BA.39 (R). Angular gyrus	**-0.22**	**-3.18**	**0.004359**	**0.040179**
BA.20 (L). Inferior Temporal Gyrus	**-0.29**	**-3.17**	**0.004464**	**0.040179**
BA.21 (R). Middle Temporal Gyrus	**-0.35**	**-3.05**	**0.005922**	**0.048855**
RLP	**-0.24**	**-3**	**0.006544**	**0.049831**
rsREL.Right Inferior Parietal L*,38)	-0.25	-2.96	0.007153	0.050584
BA.41 (L). Primary Auditory Cortex	-0.2	-2.49	0.020596	0.132265
BA.41 (R). Primary Auditory Cortex	-0.23	-2.46	0.022341	0.132265
BA.30 (L). Cingulate Cortex	-0.23	-2.45	0.022712	0.132265
BA.22 (L). Superior Temporal Gyrus	-0.22	-2.39	0.025858	0.142219
BA.42 (R). Primary Auditory Cortex	-0.23	-2.35	0.028066	0.146239
BA.30 (R). Cingulate Cortex	-0.2	-2.32	0.029937	0.148187
BA.29 (L). Retrosplenial Cingul*rtex	-0.19	-2.28	0.032911	0.149356
Grey Matter	-0.22	-2.27	0.033669	0.149356
BA.39 (L). Angular gyrus	-0.21	-2.25	0.034699	0.149356
BA.22 (R). Superior Temporal Gyrus	-0.26	-2.18	0.040166	0.165687
BA.20 (R). Inferior Temporal Gyrus	-0.19	-2.13	0.044855	0.177626
BA.43 (L). Subcentral Area	-0.19	-2.11	0.046755	0.17803
BA.8 (L). Dorsal Frontal Cortex	-0.18	-2.04	0.054043	0.198157
BA.32 (R). Dorsal anterior Cing*rtex	-0.15	-1.93	0.066253	0.23425
BA.13 (R). Insular Cortex	-0.15	-1.78	0.089585	0.305825
BA.35 (L). Perirhinal cortex	-0.18	-1.73	0.097153	0.318309
LLP	-0.17	-1.72	0.099784	0.318309
BA.45 (R). IFC pars triangularis	-0.18	-1.7	0.104168	0.318309
BA.9 (R). Dorsolateral Prefront*rtex	-0.11	-1.69	0.106103	0.318309
MPFC	-0.14	-1.65	0.112821	0.320183
BA.35 (R). Perirhinal cortex	-0.18	-1.65	0.113196	0.320183
rsREL.Med Prefrontal Cortex (MP*,-8)	-0.16	-1.63	0.117319	0.322628
BA.9 (L). Dorsolateral Prefront*rtex	-0.14	-1.58	0.127712	0.336662
rsREL.Left Inferior Parietal Lo*,38)	-0.16	-1.57	0.13174	0.336662
BA.43 (R). Subcentral Area	-0.14	-1.56	0.132624	0.336662
rsREL.Right Superior Frontal Gy*,52)	-0.14	-1.55	0.136495	0.337825
BA.38 (R). Temporopolar Area	-0.19	-1.48	0.153089	0.369655
BA.2 (R). Primary Somatosensory*rtex	-0.1	-1.35	0.189853	0.44751
BA.3 (R). Primary Somatosensory*rtex	-0.12	-1.34	0.19449	0.447779
BA.1 (R). Primary Somatosensory*rtex	-0.11	-1.28	0.215441	0.484742
BA.4 (R). Primary Motor Cortex	-0.1	-1.19	0.245088	0.539194

Bibliography

[1] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, 2014.

[2] Sophie Achard and Ed Bullmore. Efficiency and cost of economical brain functional networks. *PLoS Computational Biology*, 3(2):e17, 2007.

[3] Edgar Douglas Adrian. The spread of activity in the cerebral cortex. *The Journal of Physiology*, 88(2):127–161, 1936.

[4] David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[5] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

[6] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

[7] Rafael Alcala, Jesus Alcala-Fdez, and Francisco Herrera. A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. *Fuzzy Systems, IEEE Transactions on*, 15(4):616–635, 2007.

[8] Andrew L Alexander, Jee Eun Lee, Mariana Lazar, and Aaron S Field. Diffusion tensor imaging of the brain. *Neurotherapeutics*, 4(3):316–329, 2007.

[9] Aaron Alexander-Bloch, Renaud Lambiotte, Ben Roberts, Jay Giedd, Nitin Gogtay, and Ed Bullmore. The discovery of population differences in network community structure: new

methods and applications to brain functional networks in schizophrenia. *Neuroimage*, 59(4): 3889–3900, 2012.

[10] Aaron F Alexander-Bloch, Nitin Gogtay, David Meunier, Rasmus Birn, Liv Clasen, Francois Lalonde, Rhoshel Lenroot, Jay Giedd, and Edward T Bullmore. Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in Systems Neuroscience*, 4, 2010.

[11] Ariana Anderson and Mark S Cohen. Decreased small-world functional network connectivity and clustering across resting state networks in schizophrenia: an fmri classification tutorial. *Frontiers in Human Neuroscience*, 7, 2013.

[12] Ioannis N Athanasiadis, Vassilis G Kaburlasos, Pericles A Mitkas, and Vassilios Petridis. Applying machine learning techniques on air quality data for real-time decision support. In *First international NAISO symposium on information technologies in environmental engineering (ITEE'2003), Gdansk, Poland*. Citeseer, 2003.

[13] Nii O Attoh Okine. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in Engineering Software*, 30(4):291–302, 1999.

[14] Adil M Bagirov. Modified global k-means algorithm for minimum sum-of-squares clustering problems. *Pattern Recognition*, 41(10):3192–3199, 2008.

[15] Adil M. Bagirov, Brent Ferguson, Sasha Ivkovic, G Saunders, and John Yearwood. New algorithms for multi-class cancer diagnosis using tumor gene expression signatures. *Bioinformatics*, 19(14):1800–1807, 2003.

[16] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[17] Albert Laszlo Barabasi and Eric Bonabeau. Scale-free networks. *Scientific American*, 288(5): 50–59, 2003.

[18] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[19] Jesse L Barlow. Numerical aspects of solving linear least squares problems. *Handbook of Statistics*, 9:303–376, 1993.

[20] Kelly Anne Barnes, Alexander L Cohen, Jonathan D Power, Steven M Nelson, Yannic BL Dosenbach, Francis M Miezin, Steven E Petersen, and Bradley L Schlaggar. Identifying basal ganglia divisions in individuals using resting-state functional connectivity mri. *Frontiers in Systems Neuroscience*, 4, 2010.

[21] Peter J Basser, James Mattiello, and Denis LeBihan. Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, 1994.

[22] Danielle Smith Bassett and E D Bullmore. Small-world brain networks. *The Neuroscientist*, 12(6):512–523, 2006.

[23] AS Batuev and AA PIROGOV. Postsynaptic responses of motor cortex neurons of cats to sensory stimijlaation of different modalities. *Acta Neurobiologiae Experimentalis*, 34:317–321, 1974.

[24] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*, 37(1):90–101, 2007.

[25] Pierre Bellec, Pedro Rosa-Neto, Oliver C Lyttelton, Habib Benali, and Alan C Evans. Multi-level bootstrap analysis of stable clusters in resting-state fmri. *Neuroimage*, 51(3):1126, 2010.

[26] Christopher M Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[27] Bharat B Biswal. Resting state fmri: a personal history. *Neuroimage*, 62(2):938–944, 2012.

[28] Catherine Blake and Christopher J Merz. {UCI} repository of machine learning databases. 1998.

[29] Hamparsum Bozdogan. Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44(1):62–91, 2000.

[30] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.

[31] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[32] Kenneth P Burnham and David R Anderson. Multimodel inference understanding aic and bic in model selection. *Sociological Methods & Research*, 33(2):261–304, 2004.

[33] Vince D Calhoun, Jing Sui, Kent Kiehl, Jessica Turner, Elena Allen, and Godfrey Pearlson. Exploring the psychosis functional connectome: aberrant intrinsic networks in schizophrenia and bipolar disorder. *Frontiers in Psychiatry*, 2, 2011.

[34] R Caminiti, PB Johnson, C Galli, S Ferraina, Y Burnod, and A Urbano. Making arm movements within different parts of space: the premotor and motor cortical representation of a coordinate system for reaching to visual targets. *The Journal of Neuroscience*, 11(5):1182–1197, 1991.

[35] Thomas Carlson, Paul R Schrater, and Sheng He. Patterns of activity in the categorical representations of objects. *Cognitive Neuroscience*, 15(5):704–717, 2003.

[36] Pauline Cavelier and Jean-Louis Bossu. Dendritic low-threshold ca2+ channels in rat cerebellar purkinje cells: possible physiological implications. *The Cerebellum*, 2(3):196–205, 2003.

[37] Stephen Chiu. Method and software for extracting fuzzy classification rules by subtractive clustering. In *Fuzzy Information Processing Society, 1996. NAFIPS., 1996 Biennial Conference of the North American*, pages 461–465. IEEE, 1996.

[38] Eva WC Chow, Andrew Ho, Corie Wei, Eduard HJ Voormolen, Adrian P Crawley, and Anne S Bassett. Association of schizophrenia in 22q11. 2 deletion syndrome and gray matter volumetric deficits in the superior temporal gyrus. *American Journal of Psychiatry*, 168(5):522–529, 2011.

[39] Loren Cobb. Estimation theory for the cusp catastrophe model. In *Proceedings of the Section on Survey Research Methods*, pages 772–776, 1980.

[40] Loren Cobb and Bill Watson. Statistical catastrophe theory: An overview. *Mathematical Modelling*, 1(4):311–317, 1980.

[41] Michael W Cole, Alan Anticevic, Grega Repovs, and Deanna Barch. Variable global dysconnectivity and individual differences in schizophrenia. *Biological Psychiatry*, 70(1):43–50, 2011.

[42] Robert W Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29(3):162–173, 1996.

[43] R Cameron Craddock, G Andrew James, Paul E Holtzheimer, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, 2012.

[44] Kathryn R Cullen, Sanjiv Kumra, James Regan, Marcus Westerman, and S Charles Schulz. Atypical antipsychotics for treatment of schizophrenia spectrum disorders. *Psychiatric Times*, 25(3):61–66, 2008.

[45] Federico De Martino, Giancarlo Valente, No l Staeren, John Ashburner, Rainer Goebel, and Elia Formisano. Combining multivariate voxel selection and support vector machines for mapping and classification of fmri spatial patterns. *Neuroimage*, 43(1):44, 2008.

[46] Chao Deng and Xu-Feng Huang. Increased density of gabaa receptors in the superior temporal gyrus in schizophrenia. *Experimental Brain Research*, 168(4):587–590, 2006.

[47] Gopikrishna Deshpande, Stephen LaConte, Scott Peltier, and Xiaoping Hu. Integrated local correlation: a new measure of local coherence in fmri data. *Human Brain Mapping*, 30(1): 13–23, 2009.

[48] G Detre, S Moore Polyn, C Moore, V Natu, B Singer, J Cohen, JV Haxby, and KA Norman. The multi-voxel pattern analysis (mvpa) toolbox. In *Human Brain Mapping*, 2006.

[49] Linda Douw, Marjolein De Groot, Edwin Van Dellen, Jan J Heimans, Hanneke E Ronner, Cornelis J Stam, and Jaap C Reijneveld. Functional connectivityis a sensitive predictor of epilepsy diagnosis after the first seizure. *PLoS One*, 5(5):e10839, 2010.

[50] Turgut Durduran, Regine Choe, W B Baker, and A G Yodh. Diffuse optics for tissue monitoring and tomography. *Reports on Progress in Physics*, 73(7):076701, 2010.

[51] Pauline Favre, Monica Baciu, Cedric Pichat, Thierry Bougerol, and Mircea Polosan. fmri evidence for abnormal resting-state functional connectivity in euthymic bipolar patients. *Journal of Affective Disorders*, 165:182–189, 2014.

[52] Alex Fornito, Jong Yoon, Andrew Zalesky, Edward T Bullmore, and Cameron S Carter. General and specific functional connectivity disturbances in first-episode schizophrenia during cognitive control performance. *Biological Psychiatry*, 70(1):64–72, 2011.

[53] Alex Fornito, Ben J Harrison, Andrew Zalesky, and Jon S Simons. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences of the United States of America*, 109(31):12788–12793, 2012.

[54] Alex Fornito, Andrew Zalesky, and Michael Breakspear. Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage*, 80:426–444, 2013.

[55] Michael D Fox and Michael Greicius. Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*, 4, 2010.

[56] Eibe Frank and Ian H Witten. Generating accurate rule sets without global optimization. 1998.

[57] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[58] Karl J Friston. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36, 2011.

[59] Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994.

[60] Q-G Fu, D Flament, JD Coltz, and TJ Ebner. Relationship of cerebellar purkinje cell simple spike discharge to movement kinematics in the monkey. *Journal of Neurophysiology*, 78(1): 478–491, 1997.

[61] D Giaretta, M Avoli, and P Gloor. Intracellular recordings in pericruciate neurons during spike and wave discharges of feline generalized penicillin epilepsy. *Brain research*, 405(1):68–79, 1987.

[62] Gadi Goelman. Radial correlation contrasta functional connectivity mri contrast to map changes in local neuronal communication. *Neuroimage*, 23(4):1432–1439, 2004.

[63] Tim Gollisch and Markus Meister. Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111, 2008.

[64] Franz Graf, Hans-Peter Kriegel, Matthias Schubert, Sebastian Pölsterl, and Alexander Cavallaro. 2d image registration in ct images using radial image descriptors. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 607–614. Springer, 2011.

[65] Raoul PPP Grasman, Han LJ van der Maas, and E-J Wagenmakers. Fitting the cusp catastrophe in r: A cusp-package primer. *Journal of Statistical Software*, 32(8):1–28, 2009.

[66] Gabriele Gratton and Monica Fabiani. Shedding light on brain function: the event-related optical signal. *Trends in Cognitive Sciences*, 5(8):357–363, 2001.

[67] Inan Güler and Elif Derya Übeyli. Adaptive neuro-fuzzy inference system for classification of eeg signals using wavelet coefficients. *Journal of Neuroscience Methods*, 148(2):113–121, 2005.

[68] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

[69] Matti Hamalainen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413, 1993.

[70] Michael Hanke, Yaroslav O Halchenko, Per B Sederberg, Stephen José Hanson, James V Haxby, and Stefan Pollmann. Pymvpa: A python toolbox for multivariate pattern analysis of fmri data. *Neuroinformatics*, 7(1):37–53, 2009.

[71] S J Hanson, T Matsuka, and J V Haxby. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a. *Neuroimage*, 23(1):156–166, 2004.

[72] GW Harding. The currents that flow in the somatosensory cortex during the direct cortical response. *Experimental Brain Research*, 90(1):29–39, 1992.

[73] GW Harding and AL Towe. Neuron response to direct sensorimotor cortex stimulation in cats: Local and interareal correlation with wide-field modulation. 1995.

[74] T. Hastie, R. Tibshirani, and J. Friedman. *Model assessment and selection*. Springer, 2009.

[75] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

[76] Yoshio Hirayasu, Martha E Shenton, Dean F Salisbury, Chandlee C Dickey, Iris A Fischer, Paola Mazzoni, Tanya Kisler, Hajime Arakaki, Jun Soo Kwon, Jane E Anderson, et al. Lower left temporal lobe mri volumes in patients with first-episode schizophrenia compared with psychotic patients with first-episode affective disorder and normal subjects. *American Journal of Psychiatry*, 155(10):1384–1391, 1998.

[77] Dorothy P Holinger, Martha E Shenton, Cynthia G Wible, Robert Donnino, Ron Kikinis, Ferenc A Jolesz, and Robert W McCarley. Superior temporal gyrus volume abnormalities and thought disorder in left-handed schizophrenic men. *American Journal of Psychiatry*, 156(11): 1730–1735, 1999.

[78] P Holme, M Huss, and H Jeong. Subnetwork hierarchies of biochemical pathways. *Bioinformatics (Oxford, England)*, 19(4):532, 2003.

[79] Geoffrey Holmes, Mark Hall, and Eibe Frank. Generating rule sets from model trees. In *Twelfth Australian Joint Conference on Artificial Intelligence*, pages 1–12. Springer, 1999.

[80] Wei-Yen Hsu. Eeg-based motor imagery classification using neuro-fuzzy prediction and wavelet fractal features. *Journal of Neuroscience Methods*, 189(2):295–302, 2010.

[81] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[82] John R Hughes. Autism: the first firm finding= underconnectivity? *Epilepsy & Behavior*, 11 (1):20–24, 2007.

[83] Jens Hühn and Eyke Hüllermeier. Furia: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3):293–319, 2009.

[84] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430, 2000.

[85] Aapo Hyvarinen, Juha Karhunen, and Erkki Oja. Independent component analysis. *Studies in Informatics and Control*, 11(2):205–207, 2002.

[86] J S R Jang. Anfis: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3):665–685, 1993.

[87] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[88] JAMES A Johnsen and MICHAEL W Levine. Correlation of activity in neighbouring goldfish ganglion cells: relationship between latency and lag. *The Journal of Physiology*, 345(1):439–449, 1983.

[89] Vassilis G Kaburlasos, Ioannis N Athanasiadis, and Pericles A Mitkas. Fuzzy lattice reasoning (flr) classifier and its application for ambient ozone estimation. *International Journal of Approximate Reasoning*, 45(1):152–188, 2007.

[90] JF Kalaska. What parameters of reaching are encoded by discharges of cortical cells. *Motor control: Concepts and Issues*, pages 307–330, 1991.

[91] Nikola K Kasabov and Qun Song. Denfis: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *Fuzzy Systems, IEEE Transactions on*, 10(2): 144–154, 2002.

[92] Kiyoto Kasai, Martha E Shenton, Dean F Salisbury, Yoshio Hirayasu, Chang-Uk Lee, Aleksandra A Ciszewski, Deborah Yurgelun-Todd, Ron Kikinis, Ferenc A Jolesz, and Robert W McCarley. Progressive decrease of left superior temporal gyrus gray matter volume in patients with first-episode schizophrenia. *American Journal of Psychiatry*, 160(1):156–164, 2003.

[93] S. Sathiya Keerthi, Shirish Krishnaj Shevade, Chiranjib Bhattacharyya, and Karuturi Radha Krishna Murthy. Improvements to platt's smo algorithm for svm classifier design. *Neural Computation*, 13(3):637–649, 2001.

[94] Asaf Keller. Intrinsic synaptic organization of the motor cortex. *Cerebral Cortex*, 3(5):430–441, 1993.

[95] Jae-Hun Kim, Jong-Min Lee, Hang Joon Jo, Sook Hui Kim, Jung Hee Lee, Sung Tae Kim, Sang Won Seo, Robert W Cox, Duk L Na, Sun I Kim, et al. Defining functional sma and pre-sma subregions in human mfc using resting state fmri: functional connectivity-based parcellation method. *Neuroimage*, 49(3):2375–2386, 2010.

[96] Jaesoo Kim and N Kasabov. Hyfis: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. *Neural Networks*, 12(9):1301–1319, 1999.

[97] Kenji Kira and Larry A. Rendell. A practical approach to feature selection. In Derek H. Sleeman and Peter Edwards, editors, *Ninth International Workshop on Machine Learning*, pages 249–256. Morgan Kaufmann, 1992.

[98] Kenji Kira and Larry A Rendell. The feature selection problem: Traditional methods and a new algorithm. In *AAAI*, pages 129–134, 1992.

[99] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1):273–324, 1997.

[100] Igor Kononenko. Estimating attributes: Analysis and extensions of relief. In Francesco Bergadano and Luc De Raedt, editors, *European Conference on Machine Learning*, pages 171–182. Springer, 1994.

[101] Oliver Kramer. Unsupervised k-nearest neighbor regression. *arXiv preprint arXiv:1107.3600*, 2011.

[102] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10): 3863, 2006.

[103] Shih-pi Ku, Arthur Gretton, Jakob Macke, and Nikos K Logothetis. Comparison of pattern recognition methods in classifying high-resolution bold signals obtained at high magnetic field in monkeys. *Magnetic resonance imaging*, 26(7):1007–1014, 2008.

[104] Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19):198701, 2001.

[105] Saskia Le Cessie and JC Van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, pages 191–201, 1992.

[106] Meng Liang, Yuan Zhou, Tianzi Jiang, Zhening Liu, Lixia Tian, Haihong Liu, and Yihui Hao. Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging. *Neuroreport*, 17(2):209–213, 2006.

[107] Olvi L Mangasarian, W Nick Street, and William H Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4):570–577, 1995.

[108] J L Marchini and P Lafaye de Micheaux. Analyzefmri: Functions for analysis of fmri datasets stored in the analyze or nifti format. *R Package*, pages 1–1, 2009.

[109] J Mariño, A Canedo, and J Aguilar. Sensorimotor cortical influences on cuneate nucleus rhythmic activity in the anesthetized cat. *Neuroscience*, 95(3):657–673, 1999.

[110] Geoffrey Holmes Bernhard Pfahringer Peter Reutemann Ian H. Witten Mark Hall, Eibe Frank. The weka data mining software: An update. *SIGKDD Explorations*, 11(1), 2009.

[111] L Marsh, GD Pearlson, and SS Richards. Structural brain changes in schizophrenia: Mri replication of a post-mortem study. In *Proceedings of International Congress on Sehizophrenia Research. Tuscon*, pages 21–25, 1991.

[112] Andrew R Mayer, David Ruhl, Flannery Merideth, Josef Ling, Faith M Hanlon, Juan Bustillo, and Jose Cañive. Functional imaging of the hemodynamic sensory gating response in schizophrenia. *Human Brain Mapping*, 34(9):2302–2312, 2013.

[113] Peter McCullagh. Generalized linear models. *European Journal of Operational Research*, 16 (3):285–292, 1984.

[114] Robert K McNamara, Therese Rider, Ronald Jandacek, and Patrick Tso. Abnormal fatty acid pattern in the superior temporal gyrus distinguishes bipolar disorder from major depression and schizophrenia and resembles multiple sclerosis. *Psychiatry Research*, 215(3):560–567, 2014.

[115] David Meunier, Renaud Lambiotte, Alex Fornito, Karen D Ersche, and Edward T Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3, 2009.

[116] T M Mitchell. Chapter 1; generative and discriminative classifiers: Naive bayes and logistic regression. *Machine Learning, Sep*, 21:17, 2006.

[117] T M Mitchell, R Hutchinson, R S Niculescu, F Pereira, X Wang, M Just, and S Newman. Learning to decode cognitive states from brain images. *Machine Learning*, 57(1):145–175, 2004.

[118] T M Mitchell, S V Shinkareva, A Carlson, K M Chang, V L Malave, R A Mason, and M A Just. Predicting human brain activity associated with the meanings of nouns. *Science*, 320 (5880):1191, 2008.

[119] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, 2008.

[120] Janaina Mourao-Miranda, Emanuelle Reynaud, Francis McGlone, Gemma Calvert, and Michael Brammer. The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. *Neuroimage*, 33(4):1055–1065, 2006.

[121] Mary Beth Nebel, Ani Eloyan, Anita D Barber, and Stewart H Mostofsky. Precentral gyrus functional connectivity signatures of autism. *Frontiers in Systems Neuroscience*, 8, 2014.

[122] John Neter, William Wasserman, and Michael H Kutner. *Applied linear regression models*. Irwin Homewood, IL, 1983.

[123] Mark EJ Newman. The structure and function of complex networks. *SIAM Review*, 45(2): 167–256, 2003.

[124] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[125] R.S. Niculescu. *Exploiting parameter domain knowledge for learning in bayesian networks*. PhD thesis, SRI International, 2005.

[126] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[127] K A Norman, S M Polyn, G J Detre, and J V Haxby. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in Cognitive Sciences*, 10(9):424–430, 2006.

[128] S. Ogawa, T.M. Lee, A.S. Nayak, and P. Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, 14 (1):68–78, 1990.

[129] Toshiaki Onitsuka, Martha E Shenton, Dean F Salisbury, Chandlee C Dickey, Kiyoto Kasai, Sarah K Toner, Melissa Frumin, Ron Kikinis, Ferenc A Jolesz, and Robert W McCarley. Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an mri study. *American Journal of Psychiatry*, 161(9):1603–1611, 2004.

[130] Dirk Ostwald, Camillo Porcaro, and Andrew P Bagshaw. An information theoretic approach to eeg–fmri integration of visually evoked responses. *Neuroimage*, 2010.

[131] Alice J O'toole, Fang Jiang, Hervé Abdi, and James V Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17 (4):580–590, 2005.

[132] Alice J O'Toole, Fang Jiang, Hervé Abdi, Nils Pénard, Joseph P Dunlop, and Marc A Parent. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience*, 19(11): 1735–1752, 2007.

[133] M Palatucci and T Mitchell. Classification in very high dimensional problems with handfuls of examples. *Knowledge Discovery in Databases: PKDD 2007*, pages 212–223, 2007.

[134] Mark Palatucci and Andrew Carlson. On the chance accuracies of large collections of classifiers. In *Proceedings of the 25th international conference on Machine learning*, pages 744–751. ACM, 2008.

[135] Clement CC Pang, Adrian RM Upton, Glenn Shine, and Markad V Kamath. A comparison of algorithms for detection of spikes in the electroencephalogram. *Biomedical Engineering, IEEE Transactions on*, 50(4):521–526, 2003.

[136] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200, 2001.

[137] F. Pereira and G. Gordon. The support vector decomposition machine. In *The 23rd international conference on Machine learning*, pages 689–696. ACM, 2006.

[138] F. Pereira, G. Detre, and M. Botvinick. Generating text from functional brain images. *Frontiers in Human Neuroscience*, 5, 2011.

[139] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1 Suppl):S199, 2009.

[140] Luiz Pessoa and Srikanth Padmala. Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cerebral Cortex*, 17(3):691–701, 2007.

[141] Marco M Picchioni and Robin M Murray. Schizophrenia. *BMJ*, 335(7610):91–95, 2007.

[142] John C Platt. 12 fast training of support vector machines using sequential minimal optimization. 1999.

[143] V Pohl and E Fahr. Neuro-fuzzy recognition of k-complexes in sleep eeg signals. In *Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference*, volume 1, pages 789–790. IEEE, 1995.

[144] S.M. Polyn, V.S. Natu, J.D. Cohen, and K.A. Norman. Category-specific cortical activity precedes retrieval during memory search. *Science*, 310(5756):1963, 2005.

[145] DA Prince and D Farrell. Centrencephalic spike-wave discharges following parenteral penicillin injection in cat. In *Neurology*, volume 19, page 309. LIPPINCOTT WILLIAMS & WILKINS 227 EAST WASHINGTON SQ, PHILADELPHIA, PA 19106, 1969.

[146] Martin Pyka, A Balz, A Jansen, A Krug, and Eyke Hüllermeier. A weka interface for fmri data. *Neuroinformatics*, 10(4):409–413, 2012.

[147] John Ross Quinlan. *C4. 5: programs for machine learning*, volume 1. Morgan kaufmann, 1993.

[148] Ross J. Quinlan. Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence*, pages 343–348. World Scientific, 1992.

[149] Marcus E Raichle and Abraham Z Snyder. A default mode of brain function: a brief history of an evolving idea. *Neuroimage*, 37(4):1083–1090, 2007.

[150] RP Rajarethinam, JR DeQuardo, R Nalepa, and R Tandon. Superior temporal gyrus in schizophrenia: a volumetric magnetic resonance imaging study. *Schizophrenia Research*, 41 (2):303–312, 2000.

[151] J Tilak Ratnanather, Clare B Poynton, Dominic V Pisano, Britni Crocker, Elizabeth Postell, Shannon Cebron, Elvan Ceyhan, Nancy A Honeycutt, Pamela B Mahon, and Patrick E Barta. Morphometry of superior temporal gyrus and planum temporale in schizophrenia and psychotic bipolar disorder. *Schizophrenia Research*, 150(2):476–483, 2013.

[152] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[153] E D Reichle, P A A Carpenter, and M A Just. The neural basis of strategy and skill in sentence-picture verification. *Cognitive Psychology*, 40:261–295, 2000.

[154] Marko Robnik-Sikonja and Igor Kononenko. An adaptation of relief for attribute estimation in regression. In Douglas H. Fisher, editor, *Fourteenth International Conference on Machine Learning*, pages 296–304. Morgan Kaufmann, 1997.

[155] Cristina Rosazza and Ludovico Minati. Resting-state brain networks: literature review and clinical applications. *Neurological Sciences*, 32(5):773–785, 2011.

[156] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[157] I Rustandi. Hierarchical gaussian naïve bayes classifier for multiple-subject fmri data. *AIS-TATS*, 2007.

[158] Yosiyuki Sakamoto, Makio Ishiguro, and Genshiro Kitagawa. Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, 1986.

[159] Peter Timothy Saunders. *An introduction to catastrophe theory*. Cambridge University Press, 1980.

[160] Vincent J Schmithorst and Scott K Holland. Sex differences in the development of neuroanatomical functional connectivity underlying intelligence found using bayesian connectivity analysis. *Neuroimage*, 35(1):406–419, 2007.

[161] S V Shinkareva, R A Mason, V L Malave, W Wang, T M Mitchell, and M A Just. Using fmri brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, 3(1):e1394, 2008.

[162] S V Shinkareva, V L Malave, R A Mason, T M Mitchell, and M A Just. Commonality of neural representations of words and pictures. *NeuroImage*, 54(3):2418–2425, 2011.

[163] Evgenia Sitnikova. Thalamo-cortical mechanisms of sleep spindles and spike–wave discharges in rat model of absence epilepsy (a review). *Epilepsy Research*, 89(1):17–26, 2010.

[164] Shameran Slewa-Younan, Evian Gordon, Anthony W Harris, Albert R Haig, Kerri J Brown, Pierre Flor-Henry, and Leanne M Williams. Sex differences in functional connectivity in first-episode and chronic schizophrenia patients. *American Journal of Psychiatry*, 161(9):1595–1602, 2004.

[165] Olaf Sporns. *Networks of the Brain*. MIT press, 2011.

[166] Olaf Sporns and Rolf Kötter. Motifs in brain networks. *PLoS Biology*, 2(11):e369, 2004.

[167] Olaf Sporns and Jonathan D Zwi. The small world of the cerebral cortex. *Neuroinformatics*, 2 (2):145–162, 2004.

[168] Olaf Sporns, Christopher J Honey, and Rolf Kötter. Identification and classification of hubs in brain networks. *PloS one*, 2(10):e1049, 2007.

[169] J M Stephen, B A Coffman, R E Jung, J R Bustillo, C J Aine, and V D Calhoun. Using joint ica to link function and structure using meg and dti in schizophrenia. *NeuroImage*, 83:418–430, 2013.

[170] W Nick Street, Olvi L Mangasarian, and William H Wolberg. An inductive learning approach to prognostic prediction. In *ICML*, pages 522–530. Citeseer, 1995.

[171] Abdulhamit Subasi. Automatic detection of epileptic seizure using dynamic fuzzy neural networks. *Expert Systems with Applications*, 31(2):320–328, 2006.

[172] Abdulhamit Subasi. Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. *Computers in Biology and Medicine*, 37(2):227–244, 2007.

[173] Abdulhamit Subasi and Ergun Erçelebi. Classification of eeg signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine*, 78(2):87–99, 2005.

[174] Jinhua Sun, Jerome J Maller, Lanting Guo, and Paul B Fitzgerald. Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. *Brain Research Reviews*, 61(1):14–32, 2009.

[175] Michio Suzuki, Shigeru Nohara, Hirofumi Hagino, Kenzo Kurokawa, Takashi Yotsutsuji, Yasuhiro Kawasaki, Tsutomu Takahashi, Mie Matsui, Naoto Watanabe, Hikaru Seto, et al. Regional changes in brain gray and white matter in patients with schizophrenia demonstrated with voxel-based analysis of mri. *Schizophrenia Research*, 55(1):41–54, 2002.

[176] Karsten Tabelow, Jonathan D Clayden, Pierre Lafaye de Micheaux, Jörg Polzehl, Volker J Schmid, and Brandon Whitcher. Image analysis and statistical inference in neuroimaging with r. *NeuroImage*, 55(4):1686–1693, 2011.

[177] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *Systems, Man and Cybernetics*, (1):116–132, 1985.

[178] René Thom. Mathematical models of morphogenesis. *Springer*, 1983.

[179] Ryota Tomioka, Kazuyuki Aihara, and Klaus-Robert Müller. Logistic regression for single trial eeg classification. *Advances in Neural Information Processing Systems*, 19:1377–1384, 2007.

[180] Lefteri H Tsoukalas and Robert E Uhrig. *Fuzzy and neural approaches in engineering*. John Wiley & Sons, Inc., 1996.

[181] Sau Wai Tung, Chai Quek, and Cuntai Guan. T2-hyfis-yager: Type 2 hybrid neural fuzzy inference system realizing yager inference. In *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*, pages 80–85. IEEE, 2009.

[182] `http://afni.nimh.nih.gov/afni/download`.

[183] `http://data.pymvpa.org/datasets/haxby2001/`.

[184] `http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html`.

[185] `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/`.

[186] `http://m.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf`.

[187] `http://oto2.wustl.edu/bbears/arnie/catcrtx.htm`.

[188] `http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/`.

[189] `http://www.cs.waikato.ac.nz/ml/weka/`.

[190] `http://www.fil.ion.ucl.ac.uk/spm/`.

[191] `http://www.mathworks.com.au/products/matlab/`.

[192] `http://www.nitrc.org`.

[193] `http://www.nitrc.org/projects/conn`.

[194] `http://www.pymvpa.org/`.

[195] `http://www.r-project.org/`.

[196] `http://www.stat.columbia.edu/~fwood/Teaching/w4315/Spring2010/PCA/slides.pdf`.

[197] Peter E Valk. *Positron emission tomography: basic sciences*. Springer, 2003.

[198] Martijn P van den Heuvel, René CW Mandl, Cornelis J Stam, René S Kahn, and Hilleke E Hulshoff Pol. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *The Journal of Neuroscience*, 30(47):15915–15926, 2010.

[199] M van der Werf, M Hanssen, S Köhler, M Verkaaik, FR Verhey, R van Winkel, J van Os, and J Allardyce. Systematic review and collaborative recalculation of 133 693 incident cases of schizophrenia. *Psychological Medicine*, 44(01):9–16, 2014.

[200] Marie-José van Tol, Lisette van der Meer, Richard Bruggeman, Gemma Modinos, Henderikus Knegtering, and André Aleman. Voxel-based gray and white matter morphometry correlates of hallucinations in schizophrenia: The superior temporal gyrus does not stand alone. *NeuroImage: Clinical*, 4:249–257, 2014.

[201] Anja Wagner, Heiko Mahrholdt, Thomas A Holly, Michael D Elliott, Matthias Regenfus, Michele Parker, Francis J Klocke, Robert O Bonow, Raymond J Kim, and Robert M Judd. Contrast-enhanced mri and routine single photon emission computed tomography (spect) perfusion imaging for detection of subendocardial myocardial infarcts: an imaging study. *The Lancet*, 361(9355):374–379, 2003.

[202] L-X Wang and Jerry M Mendel. Generating fuzzy rules by learning from examples. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(6):1414–1427, 1992.

[203] Y. Wang and I. H. Witten. Induction of model trees for predicting continuous classes. In *Poster papers of the 9th European Conference on Machine Learning*. Springer, 1997.

[204] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998.

[205] Neil A Weiss and Carol A Weiss. *Introductory Statistics*. Pearson Education, 2012.

[206] Brandon Whitcher and Volker J Schmid. dcemris4: a package for medical image analysis. *R Package*, 2010.

[207] Susan Whitfield Gabrieli and Alfonso Nieto Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connectivity*, 2(3):125–141, 2012.

[208] Cohen William et al. Fast effective rule induction. In *Twelfth International Conference on Machine Learning*, pages 115–123, 1995.

[209] Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[210] Ronald R Yager and Dimitar P Filev. Generation of fuzzy rules by mountain clustering. *Journal of Intelligent and Fuzzy Systems*, 2(3):209–219, 1994.

[211] F Zerrin Yetkin, Wade M Mueller, Thomas A Hammeke, George Lee Morris III, and Victor M Haughton. Functional magnetic resonance imaging mapping of the sensorimotor cortex with tactile stimulation. *Neurosurgery*, 36(5):921–925, 1995.

[212] Erik Christopher Zeeman. *Catastrophe theory: Selected papers, 1972–1977*. Addison-Wesley, 1977.