Telma Alves Veloso

Exploration and Application of Machine Learning Algorithms to Functional Connectivity Data
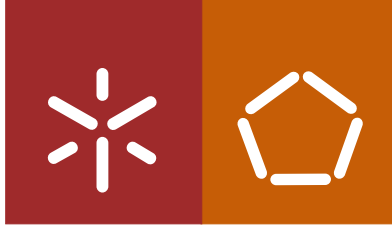
outubro de 2014

Universidade do Minho
Escola de Engenharia

Telma Alves Veloso

Exploration and Application of Machine Learning Algorithms to Functional Connectivity Data

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica
Ramo de Informática Médica

Trabalho efetuado sob a orientação do
Professor Doutor Victor Manuel Rodrigues Alves
e supervisão do
Mestre Paulo César Gonçalves Marques

outubro de 2014

DECLARAÇÃO

Nome: Telma Alves Veloso

Endereço electrónico: telma.veloso@gmail.com

Número do Bilhete de Identidade: 13950233

Título dissertação: Exploration and Application of Machine Learning Algorithms to Functional Connectivity Data

Orientador: Professor Doutor Victor Manuel Rodrigues Alves

Supervisor: Mestre Paulo César Gonçalves Marques

Ano de conclusão: 2014

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica – Ramo de Informática Médica

Universidade do Minho, ____/____/_____

Assinatura: _____

Acknowledgments

First of all, I would like to thank professor Victor Alves for the opportunity to develop this work and to Paulo Marques for all the guidance and time dedicated to the development of this work. Furthermore, I would like to acknowledge Escola de Ciências da Saúde and Instituto de Investigação em Ciências da Vida e Saúde for providing the data used in this work, in particular, the Neuroscience research domain coordinator, professor Nuno Sousa.

I would also like to express thanks to Ricardo Magalhães and Pedro Moreira, and Paulo again, for the good work environment. Furthermore, I appreciate every hour spent with Hugo Gomes and Margarida Gomes, and all the support and relaxed moments they provided me with. In spite of the distance, I can't forget all the e-mails and good old-fashioned postcards from Marta Moreira. They were dearly appreciated. Additionally, I can't express how thankful I am to Cristiano Castro for all the support and patience towards me.

Last, but not least, I would like to thank my family. Without my parents' unconditional support, this work would not be possible. Moreover, I want to thank my brother Tiago, who always cared, even though he didn't understand what I was doing.

Title

Exploration and application of machine learning algorithms to functional connectivity data

Abstract

Methods for the study of the functional connectivity in the brain have seen several developments over the last years, however not yet in a fully realized manner. Machine learning and complex network analysis are two promising techniques that together can help the process of better exploring functional connectivity for future clinical applications.

Machine learning and pattern recognition algorithms are helpful for mining vast amounts of neural data with increasing precision of measures and also for detecting signals from an overwhelming noise component (Lemm, Blankertz, Dickhaus, & Müller, 2011). Complex network analysis, a subset of graph theory, is an approach that allows the quantitative assessment of network properties such as functional segregation, integration, resilience, and centrality (Rubinov & Sporns, 2010). These properties can be fed into classification algorithms as features. This is a new and complex approach that has no standard procedures defined, so the aim of this work is to explore the use of fMRI-derived complex network measures combined with machine learning algorithms in a clinical dataset.

In order to do so, a set of classifiers is implemented on a feature dataset built with brain regional volumes and topological network measures that, in turn, were constructed based on functional connectivity data extracted from a resting-state functional MRI study. The set of classifiers includes the nearest neighbor, support vector machine, linear discriminant analysis and decision tree methods. A set of feature selection methods was also implemented before the classification tasks. Every possible combination of feature selection methods and classifiers was implemented and the performance was evaluated by a cross-validation procedure.

Although the results achieved weren't exceptionally good, the present work generated knowledge on how to implement this recent approach and allowed the conclusion that, for most cases, feature selection improves the performance of the classifier. The results also showed that the decision tree algorithm produces relatively good results without being associated with a feature selection method and that the SVM classifier, together with RFE feature selection method, produced results on the same level as other work done with a similar approach.

Título

Exploração e aplicação de algoritmos de aprendizagem computacional em dados de conectividade funcional

Resumo

Métodos para o estudo de conectividade funcional têm sofrido vários progressos ao longo dos últimos anos, no entanto, as suas potencialidades não estão a ser completamente exploradas. Aprendizagem computacional e análise de redes complexas são duas técnicas promissoras que, em conjunto, são capazes de auxiliar no processo de melhor explorar conectividade funcional para futuras aplicações clínicas. Aprendizagem computacional e reconhecimento de padrões permitem a extração de conhecimento a partir de imensas quantidades de informação neuronal, cada vez com melhor precisão de medidas e são capazes de encontrar sinal de interesse, mesmo na presença de uma grande componente de ruído (Lemm et al., 2011). A análise de redes complexas é uma abordagem que permite a avaliação quantitativa das propriedades de rede (Rubinov & Sporns, 2010). Estas propriedades podem ser usadas em classificação como atributos, o que é considerado uma a abordagem recente e complexa, pelo que não existem ainda procedimentos-padrão definidos. Deste modo, o objetivo deste trabalho é explorar o uso de medidas de redes complexas derivadas de conectividade funcional e combinadas com algoritmos de aprendizagem computacional em dados clínicos.

Para tal, um conjunto de classificadores foi implementado, tendo como atributos volumes de regiões cerebrais e medidas de rede que, por sua vez, foram construídas a partir de dados de conectividade funcional extraídos de um estudo de Ressonância Magnética funcional de repouso. Um conjunto de métodos para a seleção de atributos também foi implementado antes de realizar as tarefas de classificação. Todas as possíveis combinações destes métodos com os classificadores foram testadas e o desempenho foi avaliado através de *cross-validation*.

Apesar dos resultados obtidos não serem excecionalmente bons, o presente trabalho gerou conhecimento sobre a implementação desta nova abordagem e permitiu concluir que, na maioria dos casos, a seleção de características melhora o desempenho do classificador. Os resultados também demonstram que o algoritmo de árvore de decisão produz relativamente bons resultados quando não está associado a um método de seleção de características e que o algoritmo de máquina de suporte vetorial, juntamente com o método de seleção de atributos RFE, deu origem a resultados ao mesmo nível de outro trabalho, realizado com uma abordagem similar.

Index

Index of Figures

Acronyms List

AUC	Area under curve
BOLD	Blood Oxygen Level Dependent
CBF	Cerebral Blood Flow
CNA	Complex Network Analysis
CSF	Cerebrospinal Fluid
CT	Computerized Tomography
DICOM	Default Imaging and Communications on Medicine
DMN	Default Mode Network
EEG	Electroencephalography
EPI	Echo-planar Imaging
FA	Flip Angle
FC	Functional Connectivity
FOV	Field of View
fMRI	Functional Magnetic Resonance Imaging
FSL	FMRIB Software Library
GLM	General Linear Model
GM	Grey Matter
GNB	Gaussian Naive Bayes
GRF	Gaussian Random Field
GUI	Graphical User Interface
ICA	Independent Components Analysis
IT	Information Technology
kNN	k-Nearest Neighbor
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MCI	Mild Cognitive Impairment
MEG	Magnetoencephalography
MI	Medical Informatics
MKL	Multiple Kernel Learning
MNI	Montreal Neurologic Institute
MRI	Magnetic Resonance Imaging
NIfTI	Neuroimaging Informatics Technology Initiative
OCD	Obsessive Compulsive Disorder

PCA	Principal Components Analysis
PET	Positron Emission Tomography
RBF	Radial Basis Function
RFE	Recursive Feature Elimination
ROC	Region of Convergence
ROI	Region of Interest
RSNs	Resting-State Network
SPM	Statistical Parametric Mapping
SVM	Support Vector Machine
TE	Echo Time
TCP	Transductive Conformal Predictor
TR	Repetition Time
WM	White Matter

1 Introduction

Medical Informatics (MI), as a discipline, is considered young and multidisciplinary. The first use of Information Technology (IT) dates back to the 50's and since then it has been rapidly expanding and evolving. Thus, it establishes one of the grounds of medicine and health care and, consequently, is responsible for improving the health of people. MI contributes to a better and more efficient health care as well as encourages groundbreaking research in biomedicine and health-related computer sciences (Haux, 2010). Given its multidisciplinary character, MI is hard to define, nonetheless it follows a possible definition:

Definition | Medical informatics is the science that underlies the academic investigation and practical application of computing and communications technology to healthcare, health education and biomedical research. This broad area of inquiry incorporates the design and optimization of information systems that support clinical practice, public health and research; modeling, organizing, standardizing, processing, analyzing, communicating and searching health and biomedical research data; understanding and optimizing the way in which biomedical data and information systems are used for decision-making; and using communications and computing technology to better educate healthcare providers, researchers and consumers. Tools and techniques developed from health informatics research have become and will remain integral components of the best strategies in biomedical research and the best practices in healthcare delivery and public health management (University of Virginia, n.d.).

Nowadays, MI is part of a hospital's daily life, either in recording and storing patient information on hospital database, monitoring patients' vital signals or in medical imaging diagnosis. Given the fact that this discipline is multidisciplinary, MI can have a role in different areas such as Medical Imaging, Intelligent Systems, Electronic Health Record and Clinical Decision Support Systems. Besides that, this discipline also shows its worth in information processing in order to extract useful knowledge from the massive amount of information that today's technology produces. Picturing medicine without technology allows us to realize how valuable this discipline is (Haux, 2010).

In Portugal, the Associação Portuguesa de Informática Médica was established in the city of Coimbra in November 5th, 1979. This association was created with the intent to promote the use of Informatics in the public health, medical practice and medical investigation domains, as well as the propagation of this discipline (APIM, 2012). Despite their efforts and the increasing number of work and investigation done at academic level, in practice, this discipline still has a lot of space to grow. Portuguese hospitals have information systems implemented in some departments. An information system is an integrated set of components – hardware, software, infrastructure and people – designed to store, process, and analyze data, as well as report results on a regular basis (Zwass, n.d.). Usually, each department has its own information system. Portuguese health institutions are investing in IT systems, such as information systems, however this investment is not coordinated, resulting in advanced small islands that unfortunately are incapable of communicating with each other (Fonseca, 2011).

Still concerning MI and regarding the present work, there's an area that stands out – Medical Imaging. This area provides the methods and techniques that allow visual representation of human organs and tissue. Neuroimaging is the subset of Medical Imaging in charge of the application of such procedures to study brain structure and function. These methods and techniques include Computerized Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), Magnetoencephalography (MEG), and Electroencephalography (EEG).

1.1 Neuroimaging

It has long been an interest of the human being to figure out how the brain works, particularly how it encodes information and interacts with the environment. In order to explore the brain, more and more advanced techniques have arisen for the task, such as functional Magnetic Resonance Imaging (fMRI), EEG and others. EEG is a technique with poor spatial resolution, making it unsuitable for the study of high-level cognitive activities involved with multiple cortices. In its turn, fMRI allows to further explore the brain function as a whole and with a reasonable spatial resolution (Norman, Polyn, Detre, & Haxby, 2006; S. Song, Zhan, Long, Zhang, & Yao, 2011).

fMRI is a non-invasive medical imaging technique used to detect neural activation of different brain regions (Ogawa, Lee, Kay, & Tank, 1990). In the mid-90's, Ogawa discovered that oxygen-poor hemoglobin and oxygen-rich hemoglobin were affected by the magnetic field in different ways. This discovery allowed him to take advantage of this contrast in order to map images of brain activity (Stephanie Watson, 2008), so this technique takes advantage from the fact that a local variation in the oxygen levels happens when a brain region is active. Therefore, fMRI allows us to study patterns of brain activity during the performance of tasks (Josephs, Turner, & Friston, 1997) or at rest (B. Biswal, Yetkin, Haughton, & Hyde, 1995) which in turn can be used to characterize various pathological conditions such as schizophrenia (Cohen, Gruber, & Renshaw, 1996), Alzheimer's disease (Greicius, Srivastava, Reiss, & Menon, 2004), Parkinson's disease (Haslinger et al., 2001) and non-pathological ones, like healthy aging (Hesselmann et al., 2001).

The brain is often seen as a network, built with several regions, anatomically apart but functionally connected, that share information continuously between each other. Using fMRI to study the brain has led to a new concept described below:

Definition | *Functional Connectivity (FC) is the temporal dependence of neuronal activity patterns of anatomically separated brain regions* (M. P. van den Heuvel & Hulshoff Pol, 2010).

When two or more distinct and anatomically apart brain regions show synchronous neuronal activity or react in the same way to a common stimulus or task, it is then possible to claim that there is FC between them. This synchrony is usually measured by dependencies between blood oxygen-level dependent (BOLD) signals obtained for different anatomical regions.

Since the mid-90's, the study of FC has drew the attention of neuroscientists and computer scientists as it opens a new window to the network that is the human brain (Li, Guo, Nie, Li, & Liu, 2009). Once in possession of a large amount of FC data, it becomes essential and relevant to extract useful information and knowledge from it. Different methods have been adopted since then and they range from brain activity mapping, FC analysis through exploratory techniques such as Independent Component Analysis (Comon, 1992) to network analysis via graph theory applied to FC data (J. Wang, Zuo, & He, 2010). A relatively new approach in this area is the use of machine learning, mainly classification algorithms, using measures extracted from neuroimaging data. Such methodologies can be valuable in relation to the development of diagnostic support

tools and possible forecast of the occurrence of several diseases (Castro, Gómez-Verdejo, Martínez-Ramón, Kiehl, & Calhoun, 2013; Dosenbach, Nardos, Cohen, Fair, & Al., 2010; Long et al., 2012; Nouretdinov et al., 2011; Wee et al., 2012). Classification through machine learning allows the extraction of useful knowledge from large amounts of data in a more automatic way, hence speeding up the process and reducing the human error fraction.

1.2 Machine Learning

The fundamental goal of a learner is to generalize from its experience. In this context, generalization is the ability to give good and accurate results on unseen new data, based on the experience acquired along the learning process.

Definition | Machine Learning is a branch of computer science that, allied with statistics, is focused on the development and study of algorithms that can learn from data instead of simply following programmed instructions. Or, as Arthur Samuel said in 1959, *machine learning is a field of study that gives computers the ability to learn without being explicitly programmed* (Samuel, 1959).

A classifier, given a training set, is able to learn the association between an examples' set of attributes (or features) and its corresponding label. This way, it's possible to perceive the classifier as a function that, for a given example, returns a prediction of its corresponding label. There are many classifiers, however, the ones frequently used in fMRI data are the Nearest Neighbor (kNN), Logistic Regression (LR), Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB) and Fisher's Linear Discriminant Analysis (LDA) (Pereira, Mitchell, & Botvinick, 2009).

In particular, the classifier used in (Long et al., 2012) to tell apart between patients with early Parkinson's disease and control subjects was obtained from the SVM method applied to structural and functional images of the participants. This classification method was also used in (Wee et al., 2012) to identify patients with Mild Cognitive Impairment (MCI) and in (Dosenbach et al., 2010) to predict the maturity of the brain from fMRI images. Less frequently used methods include the Multiple Kernel Learning (MKL) described in (Castro et al., 2013) to distinguish between individuals with schizophrenia and control subjects and also the Transductive Conformal Predictor

(TCP) algorithm that can calculate the degree of confidence of the prediction given and is described in (Nouretdinov et al., 2011).

Machine learning and pattern recognition are becoming more and more frequent between the chosen techniques to perform fMRI analysis. These techniques allow to detect subtle, non-strictly localized effects that may be invisible to conventional analysis with univariate statistics (Haynes & Rees, n.d.; Martino et al., 2008; Norman et al., 2006). This technique enables the investigator to take into account the pattern of activity from the whole brain. This activity, in its turn, is measured in several points at the same time (i.e. simultaneously). This technique also allows exploring the inherent multivariate nature of fMRI data.

Multivariate pattern analysis is the application of machine learning techniques to fMRI data and typically involves four steps: (Martino et al., 2008)

1. The selection of voxels (features);
2. The representation of brain activity as points in a multidimensional space;
3. The training of the classifier, with a subset of examples, to define the decision boundary;
4. The performance evaluation of the model created.

1.3 Graph Theory

Today's techniques, whether in the Neuroscience context or in other scientific fields, can provide very large datasets. In the Neuroimaging field, these datasets usually contain anatomical and/or functional connections patterns. In the attempt to characterize these datasets, a new approach has been developed over the last 15 years. This multidisciplinary approach is known as Complex Network Analysis (CNA) and has its origin in the mathematical study of networks, known as Graph Theory (Berge & Doig, 1962). However, this analysis deals with real-life networks above all, which are large and complex (not uniformly random nor ordered) (Rubinov & Sporns, 2010). This method describes complex systems' relevant properties by quantifying topologies of their respective network representations.

Definition | Graph Theory is a branch of mathematics that deals with the formal description and analysis of graphs. A graph (see Figure 1) is defined simply as a set of nodes (vertices) linked by connections (edges), and may be directed or undirected. When describing a real-world system, a graph provides an abstract representation of the system's elements and their interactions (Bullmore & Sporns, 2009).
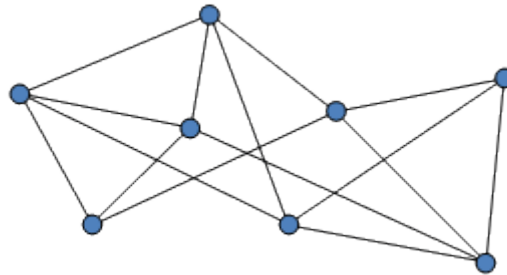
Figure 1 - Graphic representation of a graph. The blue circles are the nodes and the lines are the edges that connect the nodes, making a network. [Adapted from (Boyd, n.d.)]

In the context of brain networks, the analogy between the brain and a graph is more than useful, it's intuitive. Since the brain consists of several different areas that are believed to be connected between each other, one good way to depict a brain is by representing it as a complex network, i.e., a graph (Sporns & Zwi, 2004). The areas of the brain can be seen as nodes and the connections between those areas as the edges of the graph.

A graph can be either weighted or unweighted. In the first case, the edges are attributed a value – weight – representing their importance/significance to the network. A more important or prominent connection should have a higher value. In the opposite case, the unweighted graph has no value assigned to each edge. It can be seen as a weighted graph, only every edge has the same value. There are several measures one can obtain in order to characterize a graph and they can be local – referring to a specific node – or global – referring to the network as a whole.

As said, a network is made by nodes and links. In large-scale brain networks, nodes represent brain regions and links can represent anatomical, functional or effective connections. So, in Neuroscience's context we can find three types of connectivity (Rubinov & Sporns, 2010):

Structural Connectivity: consists of white matter tracts between pairs of brain regions (Greicius, Supekar, Menon, & Dougherty, 2009);

Functional Connectivity: consists of temporal correlations of activity between pairs of brain regions, which may be anatomically apart (M. P. van den Heuvel & Hulshoff Pol, 2010);

Effective Connectivity: represents direct or indirect influences that one region causes on another (K. J. Friston, Harrison, & Penny, 2003).

The characterization of structural connectivity and FC as a network is increasing (Bullmore & Sporns, 2009). This is a result of a combination of reasons. The method previously mentioned, CNA, quantifies brain networks in a reliable way and using only a small number of neurobiological meaningful and easy to compute measures (Hagmann et al., 2008; Sporns & Zwi, 2004). The definition of anatomical and functional connections on the same brain map helps to better investigate relationships between structural connectivity and FC (Honey & Sporns, 2009). Also, by comparing structural or functional network topologies between different subject populations it's possible to detect connectivity abnormalities in neurological and psychiatric disorders (Leistedt et al., 2009; C J Stam, Jones, Nolte, Breakspear, & Scheltens, 2007; Cornelis J Stam & Reijneveld, 2007; L. Wang et al., 2009).

1.4 Problem

Graph theory and machine learning, individually, are not recent fields of investigation, however the combination of both is, especially when applied to FC data. Given the fact that this combination is recent, there isn't much knowledge available on how to better take advantage of them together. The "know-how" is still very limited.

The use of machine learning in neuroimaging data is already a step ahead of statistical inference, because it allows to identify which features are relevant to distinguish two groups instead of just knowing that a significant difference exists between the groups. The arrangement of these three fields of investigation - machine learning, graph theory and FC - is a complex process that has a lot of potential.

The more knowledge there is about this combination, the easier is to build user-friendly software with graphical and/or command line interfaces, which are more accessible to researchers and allows them to focus on their research questions (Hanke et al., 2009). Besides investigation, these tools can also be relevant for diagnosis purpose, in a way that they might help detecting

neurological and psychiatric disorders at an early stage, thus bringing the application of functional neuroimaging to clinical settings.

1.5 Goals

The present work is mainly aimed to explore the use of machine learning algorithms applied to FC data in order to explore their suitability to neuroimaging studies. To this end, a selection of classification algorithms is proposed, as well as the implementation of such selection to the obtained FC data. Furthermore, the FC data will be used to create complex networks. Topological measures of these networks together with the FC data and brain regional volumes will be used as attributes to perform the classification tasks. Finally, a study of the accuracy of these algorithms on the distinction between a group of healthy subjects and subjects with a properly diagnosed neurological pathology, in this case, Obsessive Control Disorder (OCD), will be performed.

One of the issues of FC analysis is the consequential large amount of data. And since data is definitely not the same as knowledge, the need to "dig" for useful knowledge urges. Classification algorithms are somewhat similar to data mining and its use in neuroimaging is increasing, although there isn't much knowledge on how to implement them. Thus the goal of the present work is to investigate the best way to apply machine learning algorithms to FC data derived from graph theory analysis. There are many classifiers that can be used, and for each one of them, there are several parameters that can be adjusted in order to obtain a better performance. Knowing this, it's impossible to state that one classifier is better than the rest. There is no absolute truth about this matter, because it depends on the context of the classification problem, the data's nature and, of course, the combination of the parameters. Hence the purpose of this work is to explore the best combination of parameters and which classifiers to implement in the context of neuroimaging, particularly, FC data.

Other approaches could have been the focus of the present work. Instead of machine learning, for instance, statistic inference could have been used. This approach consists in the process of drawing conclusions about populations or scientific truths from data that is subject to random variation (George Casella & Berger, 2001). Still, since machine learning is somewhat

relatively new to the field of neuroscience and appears to be promising in its results, it was, in fact, the path chosen for the present work.

1.6 Structure of the document

This document can be roughly divided in two parts, the first being of introductory and theoretical background while the second comprises the procedures performed throughout the development of this work and the discussion of the obtained results.

Besides the present chapter, chapters 2 and 3 constitute the first part of the document. Chapter 2 introduces the neuroimaging technique that allows the data used, which is fMRI. This technique was discovered by (Ogawa et al., 1990) and is the mother of a relatively new concept that is FC. This technique also allows the investigation of brain activity patterns that can be analyzed through a series of methods, including CNA. This method models the brain as a complex network, where the nodes are brain regions and the edges are connections, functional or anatomical, between the brain areas.

Chapter 3 presents the concept of Machine Learning, which allows to generalize the structure of a certain dataset to unseen data points, given that they are of the same nature as the dataset. In this chapter, the classification algorithms used in the development of the present work are described in detail, as well as some applications of these algorithms to FC data.

Regarding the second part of the document, chapter 4 describes every step and procedure performed throughout the development of this work, starting with the fMRI images and finishing with the attributes and classifiers. This chapter also explains how the correlation matrices were achieved, as well as the complex networks and how the dataset was built.

Chapter 5 presents the results obtained for each combination of classifier and feature selection algorithm, as well as the results for each classifier without feature selection. Here, values of accuracy, sensitivity and specificity are presented for each classifier implemented in the development of this work. Comparisons and contextualization with previous results in the neuroimaging field are also presented.

Finally, chapter 6 outlines the conclusions drawn from the results obtained. At this point, is possible to infer which classifier produces the best results, as well as the influence feature

selection methods have on the results. Moreover, this chapter presents the main limitations of the work and proposes some improvements that can be made in future works in the field.

2 Functional Connectivity

Medical technology has benefited from evolution, so much that it is now possible for imaging scans to create 3D models of organs and tissues of the human body. Moreover, the use of IT in the medical environment has facilitated the recording, analysis and modeling of systems consisting of several interacting elements. These interactions happen in a structured way, resulting in complex and organized patterns that also portray an intrinsic connectivity network. The brain can be seen as one of these systems, constituted by nerve cells that interact with each other and form networks.

Being the brain a target of curiosity and interest as well as an object of study, methods to investigate intrinsic connectivity in this organ have been developed. One of this methods is fMRI, derived from the more conventional MRI, which is a recent technique of imaging that can help diagnose brain diseases, but also investigate human's mental processes. Both fMRI and MRI are based on the same technology, regarding some differences.

2.1 Magnetic Resonance Imaging

MRI is a noninvasive technique that uses a strong magnetic field and radiofrequency waves to create detailed images of the body. The first time it was used in humans was in July 3, 1977 and it took almost 5 hours to produce one image (Gould & Edmonds, 2010).

This technique takes advantage of the hydrogen atom – abundant in the human body - and its magnetic properties. Initially, the hydrogen atoms are spinning and become aligned when under the influence of a magnetic field. Afterwards, the atoms are hit by a specific radio-frequency pulse, causing them to absorb energy and change their spinning direction, i.e., they are no longer aligned. When the pulse is ceased, the atoms return to the initial alignment, releasing the energy previously absorbed which is picked up by the machine. The signal is then converted into an image by means of the Fourier Transform.

Usually, a superconducting magnet is used. It consists in many coils of wire passed by a current of electricity, thus creating a magnetic field (Gould & Edmonds, 2010). The most common MRI machines yield a 3 Tesla magnetic field, which is equivalent to 50 000x the Earth's magnetic field. In Tallahassee, vertical widebore magnets are capable of performing MRI with 21.1 Tesla

("National High Magnetic Field Laboratory," 2014). However, this is not yet used in humans. Stronger magnet fields produce better images but also more noise, making the acquisition much more uncomfortable for the patient, so the typical magnetic field used is of 3 Tesla.

2.2 Functional Magnetic Resonance Imaging

In its turn, fMRI focus on blood flow and detects oxygen levels in the brain, thus providing an indirect measure of neural activity.

Neural activity is an aerobic process, i.e. consumes oxygen O_2. Given this fact, it's safe to say that the cerebral metabolic rate of O_2 consumption and neural activity go side by side, so it's possible to measure the latter indirectly by means of the former. A healthy brain arterial blood is saturated with oxygen and when this blood arrives to a brain area with increased neural activity and, therefore, with increased O_2 local consumption, the oxygen gradient across the vessels in that area intensifies. This means that the O_2 molecules are abandoning the hemoglobin, i.e., increasing the concentration of deoxyhemoglobin. A higher concentration of deoxyhemoglobin denotes a shorter decay time T_2^*, because it provokes faster de-phasing of excited spins, which, in turn, implies a smaller BOLD signal measured at the echo time. However, studies from the beginning of neural activity research with fMRI showed the exact opposite result – the BOLD signal increases with neural activity, hence the deoxyhemoglobin concentration decreases (Ogawa et al., 1992). This divergence can be explained by a phenomenon called neurovascular coupling. Given the O_2 demand, parallel with increased neural activity, the human body reacts by increasing the cerebral blood flow (CBF), i.e., the vessels dilate locally to enlarge the volume of oxygen-rich blood reaching the area with increased neural activity (see Figure 2). The result of the coupling mechanism overlaps the O_2 consumption, disguising the increase of deoxyhemoglobin concentration and producing a stronger BOLD signal.
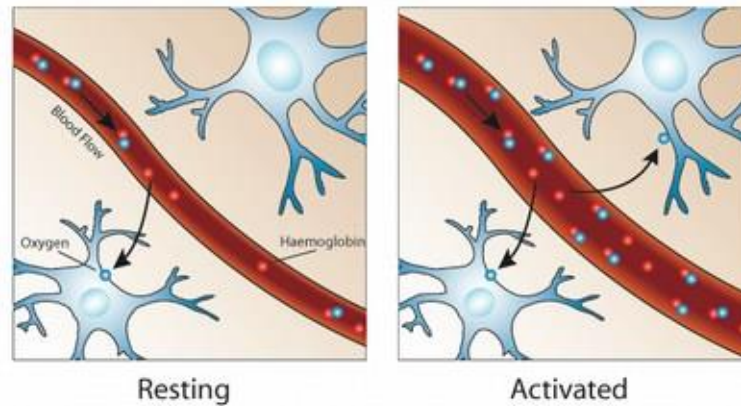
Figure 2 - Representation of the organism's response to an activated brain region. As one can see in the second picture, when an area is activated, the blood vessel is dilated, increasing the blood flow in that specific area. [Adapted from (Devlin, 2007)]

2.2.1 BOLD

The imaging method most used when dealing with brain function is the BOLD contrast. This method is based on MRI images made sensitive to changes in the state of oxygenation of the hemoglobin. The BOLD signal results from the fact that oxyhemoglobin is diamagnetic and deoxyhemoglobin is paramagnetic. This means that oxyhemoglobin has no unpaired electrons in its outer layer and is not affected by the magnetic field. However, deoxyhemoglobin is affected by the magnetic field because it has at least one unpaired electron (Helmenstine, 2014).

Following, Eq. 1 describes the BOLD signal (Murphy, Birn, & Bandettini, 2013):

$$S = M_0 \exp\left(-\frac{TE}{T_2^*}\right) \qquad \text{Eq.1}$$

Where:

$S =$ BOLD signal strength

$M_0 =$ Initial magnetization

$TE =$ Echo time (at which the image is acquired)

$T_2^* =$ Decay time

The initial magnetization (M_0) depends on the number of excited spins in a voxel and the changes in the decay time (T_2^*) are the basis for the BOLD signal. T_2^* is the inverse of the relaxation rate (R_2^*) of the magnetization caused by local susceptibility-induced magnetic field gradients. TE

is chosen to maximize the BOLD contrast and, for a 3 Tesla magnetic field, is usually 30 ms (Murphy et al., 2013).

The brain volume consists in several voxels – a volume unit that can be portrayed as a pixel with a third dimension. For each voxel, the BOLD signal is measured over a period of time, resulting in a timeseries for each voxel. In Figure 3, a schematic representation of the BOLD signal is presented.
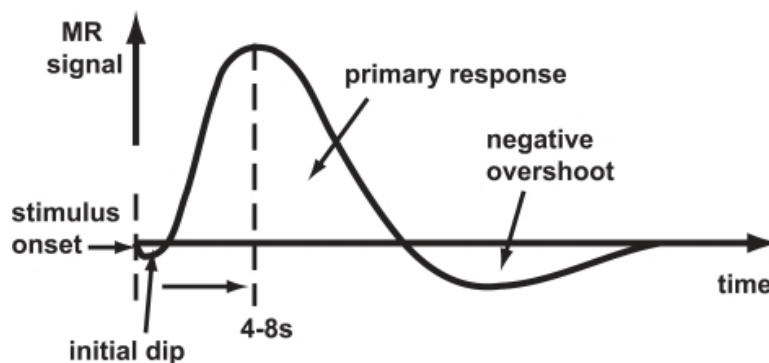
Figure 3 – Schematic representation of the BOLD hemodynamic response function (HRF). The first vertical dashed line is the stimulus beginning. The response initially takes the form a dip that is only seen at high magnetic field strengths. Afterwards, the response hits a peak, about 4 to 8 seconds after the stimulus and then a negative overshot that dips below the baseline. [Adapted from (Kornak, Hall, & Haggard, 2011)*]*

2.2.2 Task-related fMRI

Task-related fMRI is the most common of fMRI acquisition and is used to detect neural activity in response to a determined event, by means of the BOLD signal. When performing a task-related fMRI, the patient must perform an a task, whether talk, answer questions via response pads, move a finger or simply visualize images (Josephs et al., 1997). There is a wide range of fMRI study designs available for neuroscientists who investigate cognition and other brain processes and different acquisition schemes can be used, such as block design or event-related (Amaro & Barker, 2006).

The strategy of a task-related fMRI experiment is to perform some sort of interference in a system (the brain, in this case) and observe the modulation of the system response. In other words, the experimenter executes an action and observes the reaction. Usually the interference caused or the action made is a cognitive task or a stimulus presented to the subject under study

and the reaction is measured by the BOLD signal that shows the hemodynamic response to that specific stimulus. The set of structured, temporal organized and ordered tasks or stimulus is called a paradigm.

In early days, fMRI studies consisted of stimuli presented in sequence within blocked conditions. The reason for this is the historical influence of PET, because this technique studies had investigated changes in blood flow measured over time periods of up to 1 min, while the subjects had to maintain their cognitive engagement (Amaro & Barker, 2006). Over the years, fMRI has evolved to implement several stimulus presentation schemes, such as block and event-related designs, as well as a combination of both (see Figure 4). Minimum degree of complexity is advised when building a paradigm. Also, the experimenter must formulate a hypothesis, preferably with neuroanatomical background.
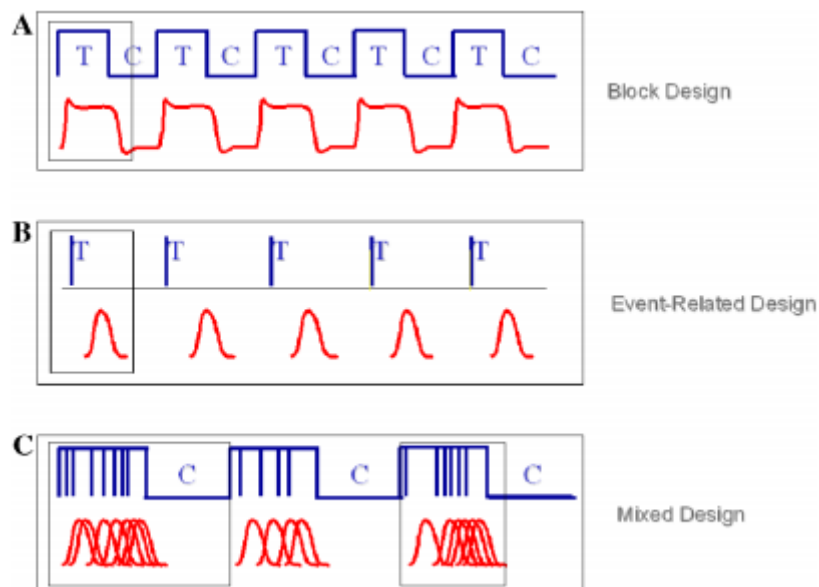
Figure 4 – (A) Block design: stimulus of the same condition are presented subsequently; (B) event-related design: each stimulus' hemodynamic response is detected, and can be analyzed in detail; (C) mixed design: combination of block and event-related. [Adapted from (Amaro & Barker, 2006)*]*

Regarding block design, this type of paradigm is based on maintaining cognitive engagement in a task by presenting stimuli sequentially within a condition, alternating this with other moments (epochs) when a different condition is presented. When two conditions are used, the design is called an "AB block" and two epochs of each condition form a "cycle". The BOLD

response is composed from individual hemodynamic responses from each stimulus and is generally higher in magnitude.

In its turn, event-related design is based on the observation that changes in hemodynamics are rapid and occur within seconds after a neuronal event. These events can be the result of presentation of a stimulus, delay period and response. In this case, stimuli are not presented in a set sequence; the presentation is randomized and the time between stimuli can vary. This paradigm design has emerged to better exploit fMRI, which has good temporal resolution and is sensitive to transient signal changes of brief neuronal events (Buckner, 1998), allowing the temporal characterization of the hemodynamic response. For each stimulus, the matching response is collected by the fMRI scanner.

As shown in Figure 4, both block and event-related designs can be used together as a mixed design. In this case, a combination of events closely presented, intermixed with control condition, provides the technical needs for event-related analysis as well as cognitive state information.

2.3 Resting-State fMRI and FC

Resting-state fMRI aims to investigate the brain's functional connections between different regions of the brain when it is "at rest". In this case the patient is asked to close his/her eyes and try to think about nothing, instead of performing a task in a task-related fMRI. Resting-state fMRI aims to study the spontaneous synchronous neural activity that happens in the human brain when the subject isn't particularly focused, i.e., not performing a task, not thinking in something specific.

Although the mechanisms behind this neural activity are yet to be known, many resting-state networks have been discovered. The most known is the default mode network (DMN) – see Figure 5. DMN was first discovered using an imaging technique called PET (Raichle et al., 2001), however, fMRI is the preferred tool to further investigate this and other resting-state networks. Its better spatial resolution, compared to EEG and MEG, allows the localization and separation of the various resting-state networks simultaneously (Murphy et al., 2013).
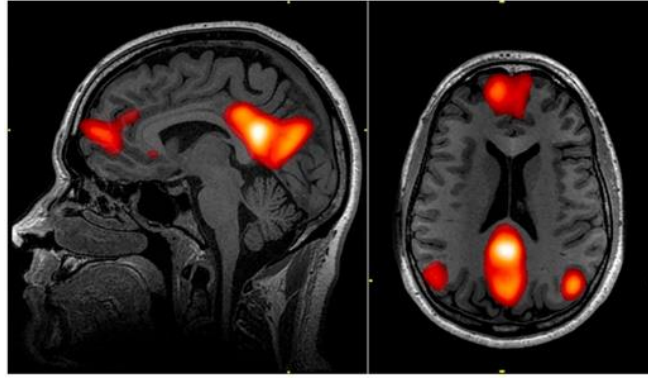
Figure 5 –The Default Mode Network includes areas in the medial pre-frontal cortex, precuneus, and bilateral parietal cortex. [Adapted from (Graner, Oakes, French, & Riedy, 2013)*]*

Compared with task-related fMRI, resting-state fMRI has a downside: since all the voxel's timeseries are acquired at the same time, the resulting measure of FC can be a spurious result, because one or both timeseries of the brain region in cause could be affected by any non-neural activity at that time. Everything that might affect the FC and is not neural activity related is considered noise and, whenever possible, should be omitted or reduced. This includes the cardiac frequency, as well as the patient breathing movements. This is why the patients are asked to stay as still as possible during the acquisition period.

2.3.1 Functional Connectivity

FC has its foundation in the assumption that temporal similarity between BOLD signals of brain regions denotes that they are connected, thus forming a functional network (M. P. van den Heuvel & Hulshoff Pol, 2010). In order to find these temporal similarities, the BOLD signals from all the brain regions must be correlated against each other to find which areas are active. If two given areas are highly correlated, it means they are active at the same time (or during the same task, in task-related fMRI) and are, presumably, functionally connected.

FC is believed to be vital in complex cognitive processes by allowing a continuous integration of information. Thus, regarding the study and investigation of the human brain structure and organization, the analysis of FC shows to be very important. In neuroimaging, FC depicts the level of communication between anatomically apart brain regions, by describing the neuronal activity patterns of those regions.

2.3.2 Resting State Networks

Several resting-state networks (RSNs) have been identified over time. RSNs is the designation of a set of brain regions identified in FC studies as being anatomically apart but functionally connected when the brain is at rest. These sets of brain regions can also be designated as low frequency networks, because these networks mostly consist of low frequency spectrums. These networks can be seen in Figure 6, where a schematic representation of the brain areas active in each RSN is shown. In spite of different test subjects, different methods and even different imaging techniques, the results converge to the same resting-state networks (B. B. Biswal, 2012; J. Damoiseaux, 2006; M. P. van den Heuvel & Hulshoff Pol, 2010).
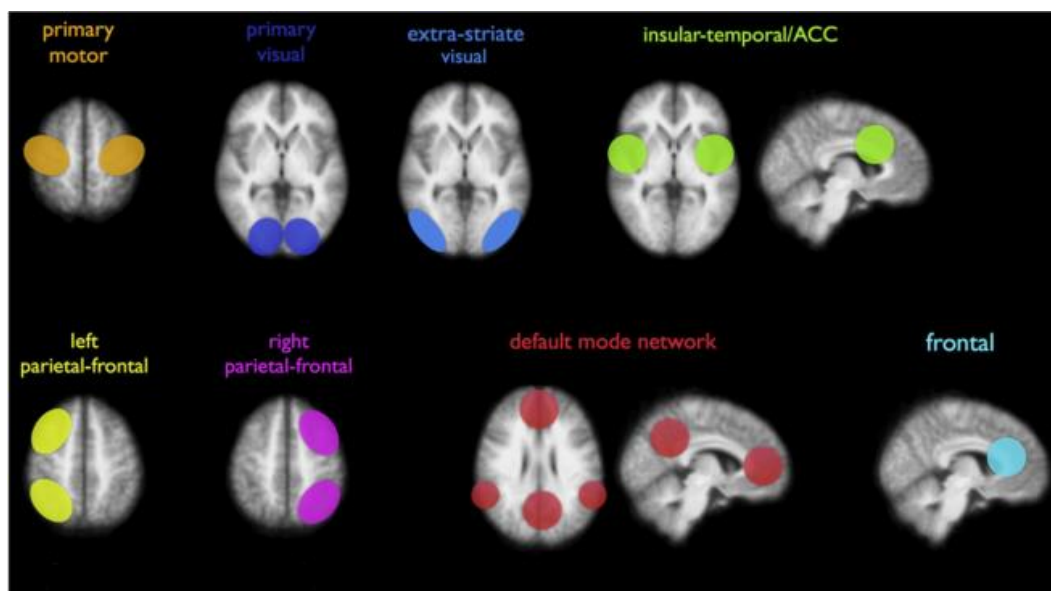
Figure 6 - Resting-state networks. Adapted from (M. P. van den Heuvel & Hulshoff Pol, 2010)*.*

These networks are alleged to be the reflection of intrinsic demands of energy from brain cells - neurons - which go off together with a common functional purpose (Cole, Smith, & Beckmann, 2010; Saini, 2004). It is possible to reliably detect and to reproduce results of the detected RSNs both at individual and group level across a range of analysis techniques (J. Damoiseaux, 2006; Greicius et al., 2004; J. Wang et al., 2010). Moreover, a determined set of co-activating functional systems is consistently found across subjects (C. F. Beckmann, DeLuca, Devlin, & Smith, 2005; J. Damoiseaux, 2006), stages of cognitive development (Cole et al., 2010; Horsch et al., 2007) and degrees of consciousness (Cole et al., 2010; Greicius et al., 2008). Thus

enforcing the assumption of RSNs as core functional networks that support perceptual and cognitive processes in the human brain.

2.4 Methods for the investigation of FC

Several methods to analyze fMRI data are described in literature (Li et al., 2009). In fMRI studies, a method called General Linear Model (GLM) is often used (K. Friston & Holmes, 1995). Concerning FC, other approaches are available and include Principal Component Analysis (PCA) (K. Friston & Frith, 1993), Independent Component Analysis (ICA) (C. F. Beckmann et al., 2005; Calhoun, Adali, Pearlson, & Pekar, 2001), clustering (Cordes, Haughton, & Carew, 2002; M. van den Heuvel, Mandl, & Hulshoff Pol, 2008) and seed methods (Cordes et al., 2002; M. Song et al., 2008). These approaches can be divided in model-dependent methods and model-free (also known as data-driven) methods (M. P. van den Heuvel & Hulshoff Pol, 2010). A flow chart representing an organized view of current methods is visible in Figure 7.
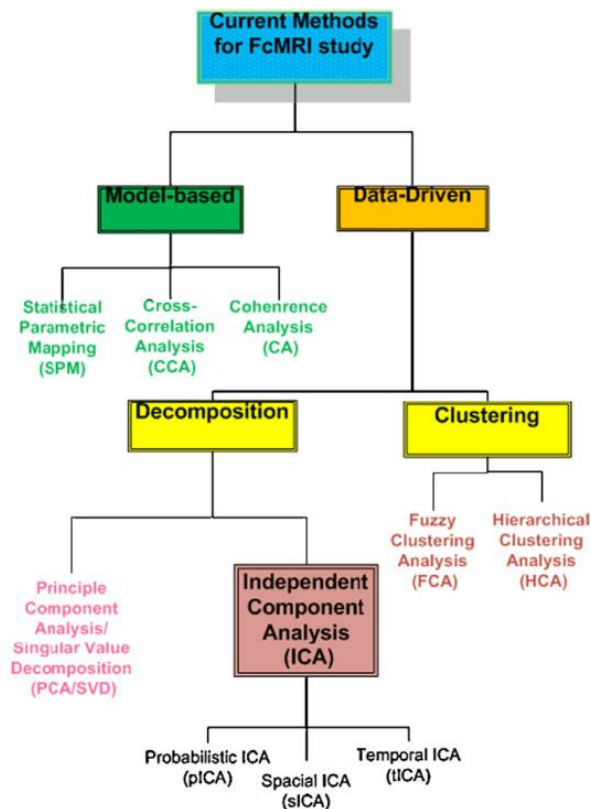
Figure 7 - Methods developed for functional connectivity MRI study. [Adapted from (Li et al., 2009)]

STATISTICAL PARAMETRIC MAPPING

This is a model-based method used to find activation patterns consequential from cognitive tasks. Over the years, statistical parametric mapping (SPM) has come to refer to the conjoint use of the GLM and Gaussian Random Field (GRF) theory to make typical inferences about spatially extended data through statistical parametric maps (Li et al., 2009). In few words, SPM uses GLM to estimate the parameters that could explain the data and uses GRF to resolve the multiple comparison problems in making statistically powerful inferences (K. Friston & Holmes, 1995). This method was initially used for FC detection in a resting-state data set by (Greicius, Krasnow, Reiss, & Menon, 2003). All brain voxels go through a scaling and filtering process and are then averaged in a certain seed, which is considered as a covariate on interest in the first-level SPM analysis. Afterwards, the contrast images corresponding to this regressor are determined individually for each subject and enter into a second-level random effect analysis, so to determine the brain areas that show significant FC across all subjects belonging to the dataset. SPM is an extremely univariate approach since a statistic (e.g. t-value) is calculated for every voxel, using the general linear model (Fumiko Hoeft, 2008).

SEED-BASED CORRELATION ANALYSIS

This model-based analysis starts with the *a priori* selection of a voxel, cluster or atlas region from which the timeseries data is extracted; this is called a seed. This selection can be done based on literature or functional activation maps from a localizer experiment. Afterwards, the data is used as a regressor in a linear correlation analysis or in a GLM analysis so to calculate whole-brain, voxel-wise FC maps of correlation with the seed region. This process is considered univariate because each voxel's data is regressed against the model apart from the other voxels (Cole et al., 2010; Jo, Saad, Simmons, Milbury, & Cox, 2010). In other words, the correlation between the seed's timeseries and every other voxels' timeseries is calculated. The correlation coefficient r can be calculated by the Eq. 2.

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \qquad \text{Eq.2}$$

In the previous expression, X is the timeseries from the voxel under the scope and Y is the timeseries from the seed voxel. In its turn, $\overline{X}$ and $\overline{Y}$ are averaged timeseries, respectively.

The areas with bigger correlation index are the ones activated at the same time as the seed and are hence functionally connected. This results in a FC map that depicts the functional connections of the predefined brain region (seed) and provides information about which areas are connected to the seed and to what extent. Given that this method is considerably straightforward and simple, some researchers prefer it. However, the information retrieved by the FC map is limited to the seed selected, which makes it more difficult to perform an analysis on the whole-brain scale, because it requires to perform every step for each voxel (M. P. van den Heuvel & Hulshoff Pol, 2010). Another downside concerns the noise that might be present in the seed voxel or region that can influence the timeseries and thus confound the results.

PRINCIPAL COMPONENTS ANALYSIS

This is a technique widely used for data analysis and its fundamental goal is to represent the fMRI timeseries by way of a combination of orthogonal contributors. Each contributor consists in a temporal pattern – a principal component – multiplied with a spatial pattern, called an Eigen map. Next, Eq.3 shows how a timeseries X can be represented with principal components.

$$X = \sum_{i=1}^{p} S_i U_i V_i^T \qquad \text{Eq.3}$$

In Eq.3, S_i denotes the singular value of X, U_i denotes the ith principal component and V_i represents the corresponding eigen map. P represents the number of components. Typically, the vectors which contribute less to the data variance are excluded, resulting in refined signal data with most of the signal preserved. The generated eigen maps depicts the connectivity between different brain regions and the bigger the absolute eigen value is, the bigger is the correlation between those brain regions (Li et al., 2009).

This FC technique has been applied to some studies (Baumgartner et al., 2000) but its application has some constrains regarding FC because it fails to identify activations at lower contrast-to-noise ratios when other sources of signal are present, such as physiological noise. In addition, there isn't a consensus about the optimum number of components. In spite of these disadvantages, this method is good at detecting the extensive regions of correlated voxels. Still, this

method can be used as a preprocessing step to achieve dimensionality reduction since it reduces second-order dependency between each component.

COMPLEX NETWORK ANALYSIS

Complex network systems and neuroscience are two study fields that meet to create this new perspective of FC analysis (Bullmore & Sporns, 2009; Onias et al., 2013). In this kind of analysis, commonly, several brain regions are defined as nodes of a network and connections (edges) between regions are modeled as the FC between each pair of regions.

The first step of CNA comprises the selection of the nodes. This can be done by one of three ways, being image voxels, segregation based on brain functional division and anatomical brain divisions. The latter is the most common approach and uses brain atlases to segment the brain in a number of regions that will be the nodes. On the other hand, the links constitute measures of functional or effective connectivity between pairs of nodes. These values have their origin in the timeseries computed from the average hemodynamic response at each node. These timeseries will be correlated against each other for all the nodes, yielding a $N \times N$ symmetric correlation matrix. This matrix can be used to create a network straight way, however it is commonly thresholded in order to obtain a binary matrix or to obtain the x percentage of strongest connections (being x a value determined by the experimenter). When all matrices are thresholded at the same value for all subjects, it is possible to calculate the desired metrics. Every subject will have a specific value, hence, group averages may be computed and a statistical comparison between different groups can be performed.

This approach has been used by numerous studies that have found complex network changes related to different conditions in basic neuroscience (Schröter et al., 2012), neurology (C J Stam et al., 2007) and psychiatry (Liu et al., 2008).

INDEPENDENT COMPONENT ANALYSIS

Unlike the model-based methods, data-driven methods allow the exploration of whole-brain FC patterns without having to previously define a region of interest or seed. ICA, used in several RSNs studies (C. Beckmann, 2005; Calhoun et al., 2001), is one of these methods.

Starting with a signal that is a combination of several sources mixed together by an unknown process, this method tries to find the unknown signal sources. Particularly, in FC analysis, ICA tries to find a combination of sources that might explain RSNs' patterns, based on the

assumption that the signal sources are statistically independent from non-Gaussian distributions. Iteratively, this method finds the components by maximizing the degree of independence amongst them. ICA analysis can be spatial, when the components are independent in the space domain, or they can be temporal, when the components are independent in the time domain (Sui, Adali, Pearlson, & Calhoun, 2009).

2.5 Other Applications

MRI is not only notable for the flexibility of the image contrast between tissues that it permits but also for the range of anatomical and physiological studies that can be undertaken with this technology. Despite its relative maturity, MRI technology is still very dynamic and new applications are being developed and adopted into clinical practice at an impressive rate.

In cancer patients, MRI is still an essential method to establish the staging of cancer and preferred over CT for specific sites of the body including the uterus and bladder, prostate, ovaries, and head and neck cancer. Contrast-enhanced MRI can reduce the number of biopsies in women with abnormal mammograms; and in difficult cases it can reveal residual cancer and help in treatment planning.

Regarding Magnetic Resonance Angiography (MRA), the increased speed of newer MRI systems, together with improved resolution and software processing methods, allows impressive angiographic imaging results through all of the human body. Given the high quality of MRA images, physicians are able to make clinical decisions based on these images. This way, it's possible to avoid invasive procedures that would have return the same diagnosis and also to perform a better planning of the treatment procedures.

On the subject of cardiac MRI, recent studies have demonstrated that cardiac magnetic resonance (CMR) can produce images of myocardial perfusion that compare positively with those obtained using positron emission tomography and single photon emission computed tomography (Ibrahim et al., 2002).

As for the central nervous system, the use of perfusion and diffusion imaging is becoming a clinical standard in addition to regular MR imaging to help guide the early treatment of stroke. Furthermore, improvements have been observed in MR spectroscopy for the characterization of brain tumors and the follow-up treatment.

Software for the coregistration of fMRI data with images from other radiological modalities continues to be developed. This technology is particularly important for stereo-tactical surgical and radio-surgical treatment planning. (A. Maidment, Seibert, & Flynn, n.d.).

3 Machine Learning

Machine learning can be found in a numerous amount of applications, such as spam filtering, optical character recognition, search engines and computer vision (Cortes & Vapnik, 1995; Wernick, Yang, Brankov, Yourganov, & Strother, 2010).

The learning algorithm or classifier must learn a number of parameters called *features* from a set of examples. After the learning process, the classifier is a model that represents the relationship between the features and the class (or label) for that specific set of examples. Later, the classifier can be used to verify if the features used contain information about the class of an example. This relationship is tested by applying the classifier in another set. *The training and testing examples are independently drawn from an "example distribution"; when judging a classifier on a test set, we are obtaining an estimate of its performance on* **any** *test set from the* **same** *distribution. Quote from* (Pereira et al., 2009)*.*

These sets are organized in a matrix where each row is an example and, since an example is a row of features, each column is a feature. The testing set has one column missing – the label. The features can have different weight values (w) amongst them, meaning that some features can be more important than others to the result.

A classifier can be seen as a function that receives an example and returns a class label prediction for that example. The various types of classifiers are distinguished by the specific kind of function they learn (Pereira et al., 2009).

Usually, the data from a classification problem is divided in two sets – the train and test set – and both these sets are constituted by examples. In their turn, these examples are an instance of the problem's population and each one of them has a group of attributes or features. These features are common for all examples, i.e. all the examples have the same features. However the value of the feature varies amongst the examples. In a simple, yet effective, explanation, please consider the mock data set depicted in Table 1.

Table 1 – Mock data set that embodies a typical representation of data concerning classification problems

Car	Brand	Color	Year	Condition
Car_1	Opel	Grey	2010	Good
Car_2	Nissan	Red	2003	Bad
Car_3	BMW	Blue	2003	Average

This dataset has 3 examples, identified in the first column. The second, third and fourth columns – Brand, Color, and Year – are the features that characterize the objects. All three objects have the same features, however, not all have the same values. The last column is the attribute that depends from the other three and is considered the target value, i.e., the label. Regarding the test set, the last column is omitted, because that is the value intended to predict by the learner. All 5 columns are considered attributes, however they aren't all the same. The first and last columns are special attributes because they have a role. The former uniquely identifies each one of the examples and the latter is the target value, i.e., the value that identifies the examples in any way and must be predicted for new examples (Akthat & Hahne, 2012).

The more examples used in the training set, the better the classifier will be. When using voxels in the whole brain, it usually results in a few tens of examples and, at least, a few hundreds of features. A possible outcome of this is a phenomenon called *over-fitting*, which is the likelihood of finding a function that will produce good results in the training set, however it's not guaranteed that it will do well in the test set (Pereira et al., 2009). It's very important to avoid this phenomenon in machine learning because it means the trained classifier, i.e. the obtained model, is not appropriate to use in new data. This happens usually when the number of training examples is low and so the learner may adjust to a specific subset of features that have no causal relation to the target function. Despite the increasing performance in the training examples, the model will deliver bad performance results on unseen (new) data sets.

To avoid this problem, from among the functions with good results, one must chose a simple one. A simple function is one where the prediction depends on a linear combination of the features that can reduce or increase the influence of each one of them. When working with linear classifiers, one can be sure that each feature only affects the prediction with its weight and it doesn't interact with other features, thus giving us the measure of its influence on that prediction. Learning a linear classifier is equivalent to learning a line that separates points in the two classes

as well as possible (Pereira et al., 2009). A representation of a linear classifier is shown in Figure 8.
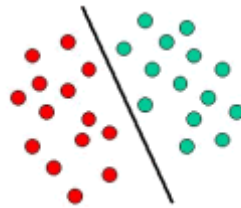
Figure 8 - Representation of a linear classifier that separates objects from two different classes, red and green.

3.1 Supervised vs. Unsupervised

Usually, machine learning tasks can be divided in two categories - supervised and unsupervised learning.

Concerning supervised learning, the algorithm receives as input a set of data examples as well as their respective output (i.e., label), and the goal is to learn the rule that explains the relationship between the examples and their label. This way, when new examples are supplied to the algorithm, it can predict the label, based on the learned rule.

As for unsupervised learning, the algorithm receives only the data examples and no labels. In this case, trying to find natural groups (or clusters) in the input data is the goal. This can also be used to detect patterns in the input data examples.

3.2 Types of Algorithms

Regarding the classifiers that learn a classification function, these fall into two categories, *discriminative* or *generative* models. The first ones' goal is to learn how to predict straight from the training set of examples. Usually, this requires learning a prediction function with a given parametric form by setting its parameters. In the second case, a statistical model of the join probability, $p(x, y)$, of the inputs x and the label y is learned. This model is capable of generating an example that belongs to a given class. The distribution of feature values on the example is

modelled like $p(x|y = A)$ and $p(x|y = B)$, where A and B are two distinct classes. Afterwards, this is inverted via Bayes Rule to classify and the prediction result is the class with the largest probability (Ng & Jordan, 2002).

General consensus shows that discriminative classifiers are almost always preferred over the generative ones. One of the reasons for this is that "*one should solve the [classification] problem directly and never solve a more general problem as an intermediate step [such as modeling $p(x|y)$]*" as Vapnik said (Vapnik, 1998).

3.2.1 Algorithms

Numerous learning algorithms are available nowadays, including neural networks, SVM and as well as simpler ones, such as the kNN, amongst others. Each learning algorithm is unique and requires a specific set of parameters in order to perform. Hence, the performance results may vary with the algorithm used, apart from the dataset used.

From the vast range of learning algorithm existing, some discriminative approaches for machine learning algorithms are described ahead. In these descriptions, the definition of the parameters is included, as well as some theory about each algorithm.

NEAREST NEIGHBOR

kNN is one of the simplest classification algorithms and it doesn't need to learn a classification function (Pereira et al., 2009), i.e., there is no model to fit. The basis behind this classifier is that close objects are more likely to belong to the same category.

Examples from the test set are classified according to the k closest examples - the neighbors - from the training set. Being n the number of features, all objects are set in the n-dimensional space. A new, unlabeled object is also set in the same n-dimension space and each object in the surroundings adds a vote to one of the possible label classifications. The new object is classified according to the most similar and close objects. This similarity can be measured by the lowest Euclidian distance, for instance. However, there are several ways to measure this distance. When dealing with binary classification problems, i.e. when there are two possible labels to assign, it's better to choose an odd k, to avoid tied votes.

Euclidean distance – the Euclidean distance between two points is the length of the line segment that links them. If $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_n)$ are two points in a n-dimension space, then the distance d between them is calculated by Eq.4.

$$d(A, B) = d(B, A) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2} \qquad \text{Eq.4}$$

Manhattan distance – the distance between two points in a grid based on a strictly horizontal and/or vertical path (i.e., the diagonals aren't used as path) is called the Manhattan distance when is the result of the sum of the horizontal and vertical components, whereas the diagonal distance might be calculated by means of the Pythagorean Theorem.

Chebychev Distance – this metric is defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension. The Chebychev distance between two vectors (or points) is shown in Eq.5

$$d = \max_i(|a_i - b_i|) \qquad \text{Eq.5}$$

SUPPORT VECTOR MACHINE

SVM is a method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates objects of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables (Cortes & Vapnik, 1995).

As mentioned before, a linear classifier uses a straight line to separate objects in their respective groups, however, most cases aren't this simple and require a more complex structure in order to optimally separate the objects in their class.

SVM is considered a hyperplane classifier because it uses lines (or planes) to distinguish objects from different classes. The basic idea behind SVM algorithms is the use of a kernel (mathematical function) to rearrange the original objects. After this process called mapping, the objects can then be linearly separated. This way, it is not necessary to "draw" a more complex structure, because a straight line is now able to separate the objects in their respective classes.

There are 4 kernel functions that can be used – Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid – and their expression are represented in Eq.6.

$$K(X_i, X_j) = \begin{cases} X_i \cdot X_j & Linear \\ (\gamma X_i \cdot X_j + C)^d & Polynomial \\ exp\left(-\gamma|X_i - X_j|^2\right) & RBF \\ tanh(\gamma X_i \cdot X_j + C) & Sigmoid \end{cases} \text{ where } K(X_i, X_j) = \phi(X_i) \cdot \phi(X_j) \text{ Eq.6}$$

The kernel function represents a dot product of input data points mapped into the higher dimensional features space by transformation ϕ. The parameters γ and C are defined by the user and d from the polynomial function is the degree.

In order to construct an optimal hyperplane, SVM runs an iterative training algorithm aimed to minimize an error function. The form of this error function distinguishes SVM models between one of four groups. Two are used for classification tasks, such as C-SVM classification; Nu-SVM classification; and the other two for regression tasks, being Epsilon-SVM regression; Nu-SVM regression.

The first type presented above, C-SVM classification, is described by Eq.7.

$$\frac{1}{2}w^T + C\sum_{i=1}^{N}\xi_i \text{ where } y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \dots, N \qquad \text{Eq.7}$$

The parameter C is the capacity constant, w is the vector of coefficients, b is a constant and ξ_i represents parameters for handling nonseparable data (inputs). The index i labels the N training cases. Note that $y \in \pm 1$ represents the class labels and ξ represents the independent variables. The kernel ϕ is used to transform the input data to the feature space. It should also be noted that the larger the C, the more the error is penalized. Thus, C should be chosen carefully to avoid over-fitting.

The second type, nu-SVM classification, minimizes the function in Eq.8.

$$\frac{1}{2}w^T w - v\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i \quad \text{where} \quad \begin{cases} y_i(w^T\phi(x_i) + b) \geq \rho - \xi_i \\ \xi_i \geq 0 \\ i = 1, ..., N \\ \rho \geq 0 \end{cases} \qquad \text{Eq.8}$$

Before using a dataset as input to a SVM learner, it's necessary to ensure the data is pre-processed for such learner. SVM requires that each example is represented as a vector of real numbers, hence, any nominal attributes must be converted to numerical data. It's considered a good practice to normalize the dataset. This prevents attributes with bigger ranges from dominating over the ones with smaller ranges. It also helps to avoid calculation problems during the algorithm process (Hsu, Chang, & Lin, 2003). A possible manner to perform this is by the transformation of the data into z-scores.

The best first choice for the kernel function is the RBF, because it maps the samples nonlinearly into a higher dimensional space. Unlike the linear kernel, it can handle cases where the relation between the class labels and the attributes in nonlinear. Besides, RBF has less numerical difficulties (Hsu et al., 2003).

DECISION TREE

This learner assumes the form of an upside-down tree: the root is at the top and the leaves grow downwards. One of the principal advantages of the representation of data in this model is its simple and easy comprehension.

This classifier, although simple to represent, has several parameters defining it (Akthat & Hahne, 2012).

Criterion is used to decide if a node is declared a leaf or if a sub-tree is created underneath it, i.e., if the attribute is split. It also determines the number of branches that a sub-tree is allowed to have. This criterion can take form of one of the following values:

- Information gain: the attribute with minimum entropy is split. This method tends to select attributes with a large number of values

- Gain ratio: adjusts the information gain for each attribute in a way that enables the extent and uniformity of the attribute values

- Gini index: measures the impurity of a dataset. By splitting a determined attribute, the average gini index of the resulting subsets decreases

- Accuracy: selects an attribute to split so the accuracy of the entire tree is maximized.

Minimal size for split is the number of examples a node must have in order for to create a new branch. If a new branch isn't created then it means that node becomes a leaf, ie, the end of its respective ramification. When a new example arrives to a leaf, a label is assign to it, because it reaches the end of the algorithm.

Minimal leaf size sets the minimum number of examples classified by a leaf in a tree. If there are a big number of leaves with only a few examples, this value should be increased.

Minimal gain determines if a node is split. In order to be split, a node's gain must be higher than this parameter. If the value set is too high, the harder it is to split a node, resulting in a tree with a single node.

Maximal depth restricts the size of the decision tree, which means that when the tree depth is equal to this value, the tree generation process will be stopped. In order to run the algorithm without this bound, the user should set the value to -1 (in RapidMiner software)

Confidence is the parameter that specifies the confidence level used for the pessimistic error calculation of pruning. Pruning is a technique to reduce the size of decision trees, that, in order to do that, removes sections that have low influence in the classification process. This technique reduces the final classifier's complexity and improves the predictive accuracy by removing sections that may be based on noise or erroneous data.

Number of prepruning alternatives is the number of alternative nodes where splitting will be tried after the prepruning prevented a certain node from splitting. Prepruning is a parallel process running along with the tree generation process and prevents splitting from happening to a node when the discriminative power of the entire tree doesn't benefit from it.

Usually, the decision tree algorithm runs together with the prepruning and pruning processes, however the user can disable one or both this processes.

FISHER'S LINEAR DISCRIMINANT ANALYSIS

This method's goal is to find a linear combination of features that together are able to separate, in the best way, two or more classes of examples. The resulting combination is then used as a linear classifier.

This method resembles PCA in the fact that both search for linear combinations of variables that explain the data better, i.e., it tries to find naturally occurring groups. LDA is also closely related to analysis of variance ANOVA and regression analysis, in the way that they also try to express one dependent variable as a linear combination of other features. The difference between these two methods and LDA is the fact that the dependent variable is numeric in the former ones while in the latter method is a categorical variable (i.e., the class label) (Akthat & Hahne, 2012).

This method can be used to serve one of two purposes. It can be used to assess the adequacy of classification or it can be used to assign objects to one of a certain number of known groups.

ARTIFICIAL NEURAL NETWORK

An artificial neural network, usually mentioned as simply neural network, is a simplified mathematical model of the human central nervous system. It consists in an interconnected structure of computational elements, frequently called neurons or nodes. The connections between these nodes are sometimes referred to as synapses. During the learning process, this algorithm adjusts the weight of the connections between nodes to achieve the desired state for the classifier. Usually, this algorithm is used to model complex relationships between inputs and outputs or to find patterns in data (Akthat & Hahne, 2012).

The connections between the nodes can be of different arrangements. A node can be connected to other nodes, forming a divergent connection; or several nodes can be connected to one single node, thus forming a convergent connections. Moreover, nodes can form chains or cyclic structures. Several learning algorithms can be used to train the neural network, back propagation being one of them. This algorithm is divided in two steps, propagation and weight update, which are repeated until the performance of the network reaches a determined value. Then, the output values are compared to the correct answers in order to determine the error. Afterwards, this error is fed back into the network, so the algorithm can use this information to adjust the weights of each

connection and to reduce the error percentage. This process is repeated for a determined number of training cycles, usually making the network converge to a state where the error is considered small.

3.3 Large Datasets

Many types of data analysis and classification problems have become significantly harder because of the increasing dimensionality of data. This is called the *curse of dimensionality* (Warren Powel, 2008) in the literature. Regarding supervised classification, the available training set might be too small (i.e., very few examples) which prevents the creation of a reliable model. In spite of the small training set, the number of features can be very high and some might not be relevant to the model concerned.

As a consequence of this phenomenon, classification accuracy might suffer, not to mention high computational cost and memory usage from the learning algorithms (Janecek, Gansterer, Demel, & Ecker, 2008). In order to avoid these problems, feature selection and dimensionality reduction methods have been developed. Some of this method are described ahead.

3.3.1 Feature Selection

The two basic and significant steps to the process of fMRI analysis aimed at classification are feature selection and feature based classification (S. Song et al., 2011).

Feature selection is a task performed to test if all the attributes belonging to a certain dataset are effectively relevant or to see if it's possible to obtain a better model by omitting some of the original attributes ("Data Mining and Rapid Miner," 2010). Nowadays, several methods are available to perform feature selection. These methods can be divided in two categories: filters and wrappers. Methods based on statistical tests are considered filters. On the other hand, wrappers use a learning algorithm to search and report optimal feature subsets (Saeys, Inza, & Larrañaga, 2007).

FORWARD SELECTION

This is considered a wrapper approach and an iterative method. It starts with no attributes and, in each iteration, temporarily adds each one of the attributes unused (in the first iteration, adds all of them). Afterwards, for each of the attributes added, it calculates the performance (of a classifier) and permanently adds the attribute that results in the highest increase of performance. The following iteration starts with the modified selection of attributes. The user can set a number of speculative rounds to define how many times the stopping criterion should be ignored. If the performance increases during a speculative round, the selection continues. However, if the performance never increases during the speculative rounds, all attributes added will be removed.

A downside of this approach is the fact that it tends to overfit to the learning scheme used in this feature selection method. Although it increases the performance of the learner used, it doesn't retrieve the best, generally relevant subset of features (Schowe, 2011)

A related approach is the Backwards Elimination, which follows the same logic but applied in an opposite way. Instead of adding attributes in each iteration, the Backward Elimination, as the name suggests, temporarily removes all attributes and calculates the performance of a classifier. Afterwards, definitely removes the attribute with the least decrease of performance.

RECURSIVE FEATURE ELIMINATION

Recursive Feature Elimination (RFE) is a method that trains an SVM and repeatedly discards the features with the lowest weights and retrains. Instead of discarding all features with small influence at once or discarding all but the largest, this method works in an iterative way (Schowe, 2011):

1. A linear SVM is trained on all remaining features, yielding β (the feature's importance);

2. The fraction r or fixed number c of features with smallest $|\beta_i|$ is discarded

3. If only k features are left, finish; else go to 1.

For example, a determined feature would have been discarded if only the importance (beta) from the first run had been considered. But since the importance is recalculated in each iteration, and maybe some features that shared some information with this specific feature were removed in the first run, this particular feature receives a higher ranking.

PRINCIPAL COMPONENTS ANALYSIS

First of all, PCA is not a feature selection method, in a way that it doesn't select a subset of features from the original one. However this method is able to reduce the number of features, by grouping in the same component attributes or features that explain the variance in the dataset. In order to do this, this method uses a mathematical procedure, orthogonal transformation, to convert a set of observations of possibly correlated attributes into a set of values of uncorrelated attributes called principal components. This transformation assures that the first principal component's variance is as high as possible and that each succeeding component has the highest variance possible given the constraint that it must be orthogonal to the preceding components, i.e., they must be uncorrelated (Akthat & Hahne, 2012).

The PCA algorithm can be used to linearly transform the data while both reducing the dimensionality and preserve most of the explained variance at the same time.

FILTERS

Another approach to feature selection is the use of filters. These methods use a measure to score a feature subset and only a determined number of features, with a determined score, is selected. This measure, typically, is chosen so to be fast to compute, without losing the meaning and usefulness of the dataset. Mutual information and Pearson product-moment correlation coefficient are two measures that are most frequently used in this type of methods for feature selection (Estévez & Tesmer, 2009).

Some filters, instead of a feature subset, deliver a feature ranking, allowing to choose the top most important features. In spite of being faster and less memory consuming as wrapper methods, the result of a filter method is not adjusted to a specific type of classification model.

3.4 Performance Evaluation

Several criteria can be used to evaluate the performance of classification algorithms and, usually, different measures evaluate different characteristics of the classifier. This is a matter of on-going research (Costa, Postal, Lorena, & Ad, 2007).

The conventional measure of classifier performance is accuracy, which is the probability of a correct answer. Still, other measures are available, such as the Receiver Operating Characteristic (ROC) curve. The area under this curve (AUC) can be a measure of classifier performance since it represents the probability that a randomly chosen positive example is correctly rated with greater doubt than a randomly chosen negative example. Other methods include F-score, average precision, precision/recall break-even point, square entropy and cross-entropy and can be divided in three groups, threshold metrics, ordering/rank metrics, and probability metrics (Caruana & Niculescu-Mizil, 2006).

Inside threshold metrics, it is possible to find accuracy and F-score. The performance of a classifier, when evaluated by these methods, is defined whether the prediction is above or below the threshold; it does not matter how close it is to the threshold.

The ordering/rank metrics depend on the order of the cases and not the actual predicted values. These metrics measure how well the positive cases are ordered before negative cases and can be seen as a summary of model performance across all possible thresholds. This category includes AUC, average precision and precision/recall break-even point.

As for the probabilistic measures, these interpret the predicted value of each case as the conditional probability of that case being in the positive class. The square entropy and cross entropy fall inside this group of metrics.

3.4.1 Cross-Validation

As previously mentioned, the more examples used in the training of the algorithm, the better the classifier will be, however, in order to obtain a good estimate of the classifier, one should not use just a few examples for testing. Meanwhile, it's not possible to train and test on the same data example, because the goal is to obtain an estimation of the true accuracy of the classifier. To solve this problem, an iterative procedure called Cross-Validation (or n-fold cross-validation) can be implemented.

The dataset is divided in n subsets of equal size, afterwards one subset is used as test set and the other $n-1$ subsets are used to train the classifier in order to obtain a model. There is one particular application of this procedure that allows to take the most advantage of the dataset. This is called *leave-one-out* cross-validation and happens when the number of subsets n is the same as the number of examples belonging to the dataset. Considering a dataset with m examples,

in each iteration, one example will be used as the new unseen example (with the label omitted) to test the classifier, while the remaining $m - 1$ examples are used to train the classifier and obtain. This process has as many iterations as examples, in this case, m iterations.

This is also used to estimate the statistical performance of a classifier, which means, it's used to estimate how accurately a classification model will perform in practice (Akthat & Hahne, 2012). Since cross-validation allows the learner to train with more examples, this procedure also avoids the over-fitting problem.

3.5 Software

RapidMiner is an environment for Machine Learning and Data Mining processes. It follows a modular operator concept that allows the design of complex nested operator chains and it also allows for the data handling to be transparent to the operators. RapidMiner introduces concepts of transparent data handling and process modeling in a way that facilitates the configuration for end users. This tool also permits the user to create plugins ("Data Mining and Rapid Miner," 2010).

This software contains a large range of operators. These operators are grouped by the tasks they perform. Some operators deal with data, whether to import or export, as well as to access the repository. The user can import his or her data to the RapidMiner program and use an operator to access it every time it's necessary. Other operators are in charge of data transformation, such as sorting, filtering and conversion. The most important feature is the modelling. Operators under this category are the ones that allow the classification tasks. Moreover, some operators can evaluate performance and significance.

A group of operators, set in an ordered, structured and logical manner, is called a process. The RapidMiner user can build a process by simply dragging and dropping operators in the process window. Each operator has defined input and output ports and, for instance, an operator output can be another operator input. The user can link these two ports in order to make the data flow throughout the process. Besides the process view, the user can see the final output of the process in the results view. These results can be saved into the repository to latter be used in another process or they can be exported. This software also lets the user install plugins, for

instance, within the present work, a Feature Elimination Extension was installed in order to perform the RFE-SVM algorithm.

3.6 Applications to neuroimaging

Classifier-based analysis was first used to investigate neural representations of faces and objects in ventral temporal cortex and showed that the representations of different object categories are spatially distributed and overlapping. Plus, it also revealed that they have a similarity structure related to stimulus properties and semantic relationships (Haxby et al. 2001; Hanson et al. 2004; O'Toole et al. 2005).

In 2012, (Long et al., 2012) developed a non-invasive technology intended to be used in the diagnosis of early Parkinson's disease. In this work, the authors used resting-state fMRI and structural brain images. Features were extracted for functional and structural images. Regarding the functional ones, characteristics were extracted at three different levels, including amplitude of low-frequency fluctuations, regional homogeneity and regional FC strength. As for the structural images, volumes of white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) structures were also extracted. In order to reduce the number of features, a two sample t-test was performed and only the features with significant differences between the two groups, controls and patients, were selected. Later, SVM algorithm was implemented as a classifier for early Parkinson's disease patients and control subjects. To evaluate the performance of this classifier, the leave-one-out cross-validation method was used.

Nouretdinov and colleagues (Nouretdinov et al., 2011) proposed a general probabilistic classification method to produce measures of confidence for MRI data, by applying transductive conformal predictor to MRI images. This predictor was applied to both structural and functional MRI data in patients and healthy controls to investigate diagnostic and prognostic prediction of depression. The motivation for this work is based on the crucial requirement of predictions' quality for clinical applications of machine learning, also known as the confidence of the classification output. This method benefits the determination of the risk of error associated to machine learning application in the clinical environment.

A different approach was developed by Chong-Yaw Wee and colleagues (Wee et al., 2012), using multi-spectrum FC networks to identify MCI patients. The goal is to accurately

distinguish MCI patients from subjects with normal aging and it uses multi-spectrum networks to characterize BOLD signal changes. The timeseries from several defined regions of interest (ROIs) were averaged and subsequently band-pass filtered. Afterwards, these timeseries were decomposed into 5 frequency sub-bands and five connectivity networks were built, one from each frequency sub-band. Clustering coefficient of each ROI against every other ROI was used as feature and leave-one-out cross-validation was used to evaluate the SVM classifier performance. This classification framework allows the early detection of functional brain abnormalities and positively contributes to the treatment of potential Alzheimer's disease patients.

Usually, the typical fMRI study generates more features than examples, which is not the best scenario when dealing with machine learning. Frequently, a large dataset includes features that do not contribute to the classification task, because they carry no useful information. On the other side, the bigger the number of examples, the better model the classifier will be. But to avoid this problems, several mechanism are accessible, including feature selection methods and cross-validation procedures.

4 Materials and Methods

As mentioned before, the main purpose of this work is to produce knowledge on how to combine machine learning and complex network analysis and to see how well learning algorithms perform with network topological measures as features. In order to solve the problem – lack of knowledge – a set of learning algorithms was used with FC data to see which combination of parameters and feature selection methods produced better results.

Typical neuroscience/neuroimaging studies focus on characterizing and describing differences between groups of patients with a certain condition and healthy controls. Hence, the approach described above was applied to FC data from an fMRI study with OCD patients and healthy subjects as controls.

With the intention of being applied to classification, the dataset must have, inherent to itself, the target value. In other words, the dataset must contain useful information to answer the classification question "to which of this classes does the example belongs to". In order to obtain useful information from raw neuroimaging data, several processing procedures need to be applied. For the present work, images from all subjects needed to be preprocessed in order to increase the quality of the data. Then, for the network analysis, the functional networks of the subjects needed to be created and network metrics calculated. Afterwards several datasets were built according to the feature selection method applied. Next, classification algorithms were implemented and the performance was measured for each combination of classifier and feature selection method.

4.1 Sample and Image Acquisitions

In this study, a total of 62 individuals participated, 19 males and 43 females. All the participants were right-handed and had no history of neurological or comorbidity disorders. Out of the 62 participants, 24 were patients that had been diagnosed with Obsessive Compulsive Disorder and the remaining participants were controls. OCD patients were evaluated regarding duration of illness, obsession and compulsion's types, as well as medication taken. To all participants, the aim of the study was explained and they participated as volunteers and gave written informed consent.

The subjects were scanned on clinical approved 1.5 T Siemens Magnetom Avanto TIM-system MRI scanner at Hospital de Braga. Concerning the structural analysis, a T1 3D MPRAGE (magnetization prepared rapid gradient echo) scan was performed with the following parameters: 176 sagital slices, repetition-time (TR) = 2730, echo-time (TE) = 3.48 ms, slice thickness = 1 mm, slice gap = 0 mm, voxel size = 1x1 mm², field-of-view (FoV) = 256×256 mm, flip angle (FA) = 7°. Functional images were collected axially using an echo-planar imaging (EPI) sequence sensitive to BOLD contrast. The acquisition parameters were: 30 slices, TR = 2000 ms, TE = 30 ms, slice thickness= 3.5 mm, slice gap = 0.48 mm, voxel size = 3.5 x 3.5 mm², FoV = 1344×1344 mm, FA = 90° and 180 volumes. During the resting state scan, the subjects were instructed to remain still, with their eyes closed, but still awake, to remain as motionless as possible and think of nothing in particular. None of the participants fell asleep during the acquisition. Foam pads were used in both head sides in order to reduce head motion during the acquisition.

4.2 Preprocessing

The first stage of the preprocessing work flow was entirely done with the BrainCAT tool (Marques, Soares, Alves, & Sousa, 2013). This software performs an automated and standard multimodal analysis of combined fMRI/Diffusion Tensor Imaging (DTI) data, resorting to open source tools. The standard pipeline for fMRI images is implemented by this tool and includes the following steps:

1.   **Format Conversion**: the first step is to convert the DICOM images obtained from the MRI to the NIfTI (Neuroimaging Informatics Technology Initiative) format. This is accomplished with the open source tool *dcm2nii* ("DICOM to NIfTI conversion," n.d.)*.

2.   **Initial Volumes Removal**: after that, the initial volumes of the acquisition are removed. In this work, the first 5 volumes (out of 180) were removed, in order to assure the magnetic field was stable.

3.   **Motion Correction**: this guarantees that all anatomical key-places stay in the same site across all the acquisition's volumes through a rigid alignment. This step also outputs two graphic representations of the translational and rotational

movement and subjects that surpassed the limits of 2 mm in translation and $2°$ in rotation were considered to present too much movement and were excluded.

4. **Slice Timing Correction**: this corrects timing differences between slices, since they are acquired in different points in time and the fMRI analysis assumes the complete volume as a one-time acquisition.

5. **Skull Stripping**: this step removes the non-brain structures.

6. **Normalization:** is performed to put all subject's data in the same standard space (in this case, MNI 152), so the acquisitions are transformed, space-wise, to force anatomical key-points to be in the same position across all the subjects.

7. **Smoothing:** this step helps reduce some noise and correct some problems introduced by the previous step (normalization). However, in the present work, this step was not performed because the images acquired were already smoothed.

8. **Band-pass Filtering**: the images were filtered between the frequencies of 0.01 Hz and 0.08 Hz because RSNs are related with low frequency BOLD fluctuations and high frequency ones are, generally, associated with cardiac beating and respiratory movements.

This preprocessing workflow is the recommended for analysis of RSNs using ICA. However, for network analysis some other preprocessing steps are recommended. As so, it was necessary to take some steps back and perform an additional step to remove possible confounding influence from WM and cerebrospinal fluid CSF. In order to do so, the FMRIB Software Library (FSL) was used. This software is a library with tools to analyze fMRI, MRI and DTI brain images. Most of the tools provided can be run both from the command line and in GUI (i.e., Graphical User Interfaces) ("FMRIB Software Library," 2014).

To remove WM and CSF signal confounds, it was decided to regress out the signals coming from these sources, by means of a GLM. First, brain segmentation was performed with FSL's *fast* command to obtain WM and CSF. Afterwards, these masks were eroded to guarantee that grey matter signals would not be included in the nuisance regressors. Afterwards, the WM and CSF masks were used with the FSL's *fslmeants* command to obtain the average timeseries for each signal. These timeseries were then fed to the GLM as regressors, using FSL's *fsl_glm* command and the residuals of the model were used as data of interest.

After this step, the preprocessing pipeline was resumed at step number 6 and normalization was performed, as well as the band-pass filtering.

In order to remove the noise coming from WM and CSF signals, other approaches were investigated. One of those approaches consisted in Independent Components labeling for artifact removal (Tohka, Foerde, Aron, & Tom, 2008). Although promising, this approach was abandoned because it required the user to specifically tell the classifier which components are considered noise before the training step. To do that, the user must visually identify the noise components, which is hard for the unexperienced user. Besides, some components can be very ambiguous, making the choice very uncertain and user dependent, thus not fully exploratory. Other reason for the rejection of this approach was the lack of data. After using data to show the classifier noise components examples and using data to train the classifier, very few was left to actually test the classifier.

4.3 Correlation Matrices

For this work, 160 brain areas were chosen as ROIs to be the nodes of the network. A sphere was created for each one of these ROIs, and each sphere was centered in the center of gravity of each ROI. In other words, each ROI had a determined shape and size, usually irregular, so the center of gravity for each one of these ROIs was calculated. A sphere was then created for each one of the 160 ROIs and it was centered at the same point as the center of gravity of the corresponding ROI. These spheres are an approximation of the ROIs. Then, FLS's *fslmeants* command was used to obtain the average timeseries for all the 160 ROIS. All the voxels inside a sphere were averaged together to obtain the average timeseries of all 160 ROIs. And, with this step, the preprocessing is complete. This pre-processing procedure was necessary to perform before building the correlation matrices in order to assure their quality.

Given that part of this work's goal is to build a complex network having the brain as model, i.e., construct a graph where the nodes are brain regions and the edges are the connections between the brain regions, it is mandatory to obtain the edges between the previously defined 160 nodes. In order to do so, it is necessary to correlate each one of the 160 timeseries acquired with

each other and build a correlation matrix for each subject. This is also called an adjacency matrix. The result is 62 160x160 correlation matrices, by the end of this step.

This step was performed using Python libraries, such as Nitime ("Nitime: Time-series Analysis for Neuroscience," n.d.), as resource. For each subject, the text file with the ROIs average timeseries was used as input to obtain the correlation matrix. By means of the *CorrelationAnalyzer* method, the correlation matrix was assembled through the calculation of the correlation coefficient between every pair-wise combination of timeseries contained in the file. This matrix is of dimensions 160x160 and is symmetric, i.e., the diagonal is 1, since it represents the correlation of one brain region with itself and the top triangle has the same exact information as the bottom one.

Afterwards, since some areas were negatively correlated and network metrics are computed only with positive values, the matrix was converted to its absolute version in order to make all correlation values positive. Now, this matrix is the starting point to build a graph and subsequently calculate topological network metrics. Network metrics are known to vary according to network density (J. S. Damoiseaux & Greicius, 2009; Liang et al., 2012; Rubinov & Sporns, 2010; van Wijk, Stam, & Daffertshofer, 2010). Consequently, in order to compare these metrics across subjects, it is mandatory that the networks have the same density. So, in order to accomplish this, all matrices were thresholded at the values of 5, 10, 15, 20 and 25% of strongest correlations. In other words, having a matrix thresholded at 5% means that only the 5% strongest correlations remain, while the others are cancelled. This means that, for each one of the 62 subjects, 5 matrices were created.

These thresholded matrices were created by means of a Python function created within this work, based on an existing Matlab function (MathWorks, n.d.). Given the threshold desired, this function calculates how many connections correspond to that value. Next, it orders all connections in a decreasing order and selects the number of connections that belong above the threshold by annulling the ones remaining.

An example of these matrices can be visualized in Figure 9. To simplify, brain regions names were shortened, but the correspondence to the full brain areas can be seen in the Appendices section, at the end of the document.

Figure 9 – Example of a correlation matrix acquired via Nitime Python package. The scale ranges from blue to dark-red, being blue a strong negative correlation and dark-red a strong positive correlation. As one can see in the zoomed-in circle, the LPUT area is less correlated to LPT than it is to LCAU.

These matrices showed that some subjects had very strong correlations (the dark red color of the image). After careful examination it was observed that these high correlations were due to the presence of a common signal to all regions, known to be the global brain signal. In order to try to correct these matrices, another step back was taken and another regressor was added to the regression of the WM and CSF signals. This regressor represents the average timeseries of the whole brain over the acquisition. This means that Global Signal Regression was performed. This is a controversial process. Some people claim that it should never be performed because it removes

important part of the neural signal, which is the focus in study and it also introduces, undoubtedly, negative activation measures in standard fMRI analysis (Murphy, Birn, Handwerker, Jones, & Bandettini, 2009). On the other hand, some defend that this procedure enhances 'true' negative relationships existing between cognitive control RSNs (Weissenbacher et al., 2009).

4.4 Network Metrics

Once the matrices were complete, everything was set to create graphs from them. Although different structures, both a correlation matrix and a graph represent the same information and so, they are equivalent. To create the graphs, a method called *mkgraph* was used from the Nitime library. This method requires an adjacency matrix (correlation matrix) as an input and returns a NetworkX graph object. NetworkX is another Python library that allows the creation, manipulation and study of the structure, dynamics and functions of complex networks ("NetworkX - High productivity software for complex networks," 2014). This library was used to calculate the network metrics that would be used as features, with exception to the network measure Global Efficiency, which was determined by a Python function made within this work.

From the large number of metrics that can be extracted from a network, only a few were chosen for the present work. Amongst these metrics, some are calculated locally, in the scope of a node while other are calculated globally, considering the whole network. The criterion for metric selection was based on the metrics that are more commonly reported in neuroimaging studies involving network analyses.

Regarding the global ones, the following metrics were obtained (Bullmore & Sporns, 2012; Rubinov & Sporns, 2010):

- Global Efficiency: is calculated as the average inverse of the shortest path length; regarding brain networks, this measure usually represents the overall capacity for parallel information transfer and integrated processing;

- Clustering Coefficient: represents the occurrence of clustered connectivity around individual nodes;

- Transitivity: is a normalized variant of the clustering coefficient that avoids being disproportionately influenced by nodes with a low degree;

- Assortativity: measures the similarity of connections in the graph with respect to the node degree.

As for the local ones, these were the ones extracted:

- Degree: the number of edges connected to a node. If a node has a high degree, i.e., if a lot of edges "visit" him this might mean that it's an important node;

- Betweenness Centrality: is an indicator of a node's centrality in the network; a node with high betweenness centrality has a large influence in the transportation of items through the network, assuming that the shortest path is preferred to transport items;

- Triangles: is the number of triangles that a node belongs to.

As mentioned above, network measures can either be local or global. Local measures of individual network elements, such as nodes and links, usually depict how these elements are integrated in the network. As for global measures, these ones provide an overall description of the network. However, network measures can also be grouped by the characteristics they measure. This way, measures of functional integration estimate how easily brain regions communicate and share information with distributed areas and are based on shortest path lengths measures, such as global efficiency. In its turn, measures of segregation quantify the existence of densely interconnected groups of brain regions, which suggests an organization of statistical dependencies pointing to a segregated neural processing, when these clusters are found in functional networks. One such metric is the clustering coefficient based on a vertex neighborhood triangle count. Measures of centrality are based on node degree or on the length and number of shortest paths between nodes. A high degree node interacts with many other nodes in the network and might be import in the information transportation. As for measures of resilience, they can quantify the vulnerability of the network. The assortativity coefficient, one of those measures, represents the correlation between the degrees of all nodes on two opposite ends of a link (Rubinov & Sporns, 2010).

4.5 Dataset

Segmentation of structural images was already done, and as part of the present work, this segmentation was revised to make sure they presented sufficient quality. In order to do so, FreeSurfer recommended reconstruction (Moreau, 2014) was performed to make sure the WM and pial surfaces were correctly defined, as well as the segmentation of anatomical brain regions for all subjects.

Volumes for different brain regions were calculated. In order to do that, segmentation from two distinct brain atlases were used: the subcortical segmentation of FreeSurfer and cortical segmentation according to the Destrieux Atlas (Destrieux, Fischl, Dale, & Halgren, 2010). It was decided to include volumes of brain regions as features because it is known that some structural changes may occur in OCD patients (Pujol et al., 2004; Radua & Mataix-Cols, 2009).

The previously mentioned metric values, together with brain region volumes and demographic values resulted in the feature dataset used for this work. This dataset contained 2531 attributes. From the local measures, 2400 attributes (160 nodes x 5 densities x 3 network metrics); from the global measures, 20 attributes (4 network measures x 5 densities); 109 from brain regions volumes and 2 demographic values (sex and age).

This dataset also includes two special attributes, the label, which identifies each one of the 62 subjects as control (C) or as OCD patient (O) and the identification of the subjects, ID. Every attribute is numerical, with exception to the ID, which is polynomial, and the label, that is binomial.

4.6 Feature Selection

For the present work, the feature selection methods used were described earlier in the document, particularly in Chapter 3, namely, the RFE, forward selection and PCA. Apart from PCA, which is a dimensionality reduction method, the methods used are considered wrapper approaches, because they use a learning algorithm in order to find the optimal subset of features. In the present work, this choice was taken because, although filters are suitable to large datasets, they have not been proved as effective as wrappers (Abraham, Simha, & Iyengar, 2007).

These algorithms were implemented via RapidMiner, which allows the user to take advantage of structured operators, ready to use, that only require the parameters input.

Whenever the PCA operator was used for the dimensionality reduction, it was decided to keep the variance of 0.999 because the aim was to keep the components that explained the most of the variance. As a result, 61 components were used as features.

Regarding the RFE operator, the parameter defining the number of attributes to select was set to 30, because it's thought to be a reasonable number of attributes to keep, considering the original number of attributes and the classification process. The parameter of the cost was set to zero so the software would estimate the best value.

As for forward selection, the number of attributes was also set to 30, so to keep uniformity throughout this work's course. The number of rounds performed after the stop criterion is reached was set to 5. This value is higher than the default value proposed by the software because of the intention to avoiding a locally optimal value. This way, the selection of the best features is better assured. This operator is a nested operator, which means it runs a subprocess. In this subprocess, a classifier is used to evaluate performance of the features. This classification task also requires the use of cross-validation, but, in this case, it was decided not to perform leave-one-out cross validation because it was very time consuming. Instead, 8 fold-validation was performed, meaning that 8 groups were created and the classifier ran 8 times, having one group as test set and the remaining 7 as train sets. For the cross-validation operator of the main classifier, leave-one-out cross-validation was still performed.

4.7 Algorithms and Parameters

The selection of algorithms to implement in the present work consists in two variations of kNN (3NN and 5NN), as well as SVM, LDA and decision tree algorithms. The kNN approaches prevailed because of its simplicity. SVM is a frequent appearance regarding neuroimaging studies allied with FC. As for decision tree, this method was chosen because of its similarity with medical guidelines, which depict a reasoning and decision making flow. As for the LDA, this method was selected because it was convenient, relatively to time and implementation.

Regarding the SVM classification process, from all available operators from RapidMiner software, the choice went to the libSVM implementation (Chang & Lin, 2011). This decision was

taken because it's the most common implementation used in fMRI data (Cox & Savoy, 2003; Fekete et al., 2013; Long et al., 2012; Sato et al., 2013; Weygandt et al., 2012).

Still concerning the SVM classifier process, in order to find the best parameter values of C and γ, a grid search should be performed according to (Hsu et al., 2003). After defining a set of possible values for both parameters, several combinations of those values were tested and the classifier produced best results at $C = 1.000001$ and $\gamma = 0.00000032$.

As for the Nearest Neighbor algorithms, since these operators run considerably fast, every possible distance measure was tested for the dataset in question, without feature selection, to see which one performed the best. In result, 3NN was used with the Euclidean Distance and the 5NN with the Manhattan Distance.

LDA operator had no parameters to define and so it was run as is.

Regarding the decision tree, the procedure adopted was to perform a small scale grid-search. As an outcome, this operator was used with the criterion Accuracy, minimal size for split of 4, minimal leaf size of 3, minimal gain of 0.2 and confidence of 0.25. The maximal depth parameter was set to -1, which implies that no bound was set.

Before the dataset enters the classifiers operators, sometimes it's necessary to perform a preprocessing step. In this case, the operator *Normalize* was used to perform normalization, particularly, z-transformation. This is helpful in cases where the dataset values are of different units and scales. This z-transformation converts data into Normal Distribution with mean of 0 and variance of 1. A z-score z is calculated by the formula described by Eq. 9.

$$z = (X - \mu)/\sigma \qquad \text{Eq. 9}$$

In Eq. 9, X denotes the value, μ denotes the mean and σ denote the variance or standard deviation.

This step was performed due to recommendations from (Akthat & Hahne, 2012) and (Hsu et al., 2003) when dealing with SVM classifiers and methods that use the Euclidean distance, such as kNN. It was additionally performed in the processes that included PCA as feature dimensionality reduction method.

4.8 Performance Measures

Every one of the 5 classifiers was tested with all the feature selection methods selected and without any feature selection approach. To achieve these combinations, 20 processes were constructed. The performance of the classifiers was evaluated by the Performance operator, which delivers a list of performance criteria values, automatically determined to fit the learning task type (Akthat & Hahne, 2012).

An example of a process' result in a confusion matrix, as shown in Figure 10.

accuracy: 70.97% +/- 45.39% (mikro: 70.97%)

	true C	true O	class precision
pred. C	30	10	75.00%
pred. O	8	14	63.64%
class recall	78.95%	58.33%	

Figure 10 – Example of the results outcome by RapidMiner software.

This table provides the accuracy (first row), as well as precision and recall for each class. The recall values for C (control) and for O (OCD patients) are, in fact, the values of specificity and sensitivity, respectively. Sensitivity and specificity are the statistical measures of the performance of a binary classification test.

Sensitivity, also known as the true positive rate, is calculated as the fraction of the number of correctly classified OCD patients over the number of true OCD patients, in this case is equal to 14/(14+10). This value corresponds to the recall value for the class O.

As for specificity, also known as the true negative rate, is the fraction of the number of correctly classified Controls over the total number of controls, in this case is 30/(30+8). This value corresponds to the recall value for the class C.

5 Results

In the scope of the present work, 5 classification algorithms were tested, together with 3 feature selection methods. Besides, every classification algorithm was tested with the complete dataset, i.e., without any feature selection method associated.

In order to evaluate performance, accuracy, sensitivity, specificity, and precision measures were calculated. Accuracy is an overall measure of the classifier's performance, however, all the measures as a whole must be taken into consideration when evaluating the performance of a classifier.

Accuracy is the percentage of correct predictions, for both controls and OCD patients together. Sensitivity is the proportion of correctly classified OCD patients, i.e., is the percentage of OCD patients that are correctly identified as having this disorder. As for specificity, it is the proportion of correctly classified control patients or, in other words, the percentage of healthy people who is correctly identified as not having the disorder. The precision, which, in this case, is presented for the OCD label, is the percentage of actual OCD patients, amongst all OCD predicted cases. Ideally, a classifier is 100% sensitive and 100% specific, however, in theory, any classifier has a minimum error bound. Consequently, the best case scenario is to find a good balance between sensitivity and specificity.

5.1 3NN

Concerning the 3NN classifier, which uses the Euclidean distance to find an examples label, Table 2 shows the obtained results for this classifier. In this case, no difference is seen between the classifier without feature selection and with PCA. Both Forward Selection and RFE improve the algorithm's performance, being Forward Selection the best result, with an accuracy value of 64.52% ± 47.85%. This means that 64.52% of all subjects were correctly classified. The deviation value, 47.85%, is considerably high, but it's justified by the leave-one-out cross-validation step. An accuracy value was calculated at each iteration step, and the final result is the average of the accuracy values calculated in every iteration step. Some iterations might have had higher

accuracy values, while others might have had lower values, resulting in a very wide standard deviation value.

Table 2 – Results of 3NN classifier without feature selection, with PCA, Forward Selection and RFE methods. The performance values calculated include Accuracy, Sensitivity, Specificity and Precision measures

	No Feature Selection	**PCA**	**Forward Selection**	**RFE**
Accuracy	61.29% ± 48.71%	61.29% ± 48.71%	64.52% ± 47.85%	59.68% ± 49.05%
Sensitivity	12.05%	12.50%	54.17%	29.17%
Specificity	92.11%	92.11%	71.05%	78.95%
Precision	50.00%	50.00%	54.14%	46.67%

5.2 5NN

The 5NN algorithm used the Manhattan distance to determine an example's label. Here, the pairing with Forward Selection resulted in the best performance values for this classifier with accuracy of 70.97% ± 45.39%, sensitivity of 62.5% and specificity of 76.32%. The remaining feature selection algorithms, Forward Selection and RFE, did not introduce improvements to the classifier's performance as one can see in Table 3. In fact, the combination of these feature selection methods with 5NN even deteriorated the results, which means that this classifier doesn't benefit from these feature selection methods.

Table 3 - Results of 5NN classifier without feature selection, with PCA, Forward Selection and RFE methods. The performance values calculated include Accuracy, Sensitivity, Specificity and Precision measures

	No Feature Selection	**PCA**	**Forward Selection**	**RFE**
Accuracy	69.35% ± 46.10%	62.90% ± 48.31%	70.97% ± 45.39%	59.68% ± 49.05%
Sensitivity	37.50%	33.33%	62.50%	33.33%
Specificity	89.47%	81.58%	76.32%	76.32%
Precision	69.23%	53.33%	62.50%	47.06%

5.3 SVM

As for the SVM algorithm, PCA dimensionality reduction made no difference when compared to without any feature selection approach as is shown in Table 4. This singularity only happened to the 3NN classifier as well. One possible explanation is the normalization step performed for both these classifiers, however, no evidence of this was found. In its turn, the RFE method significantly improved the results, as well as forward selection, that also improved the performance of SVM relatively to without feature selection. The best results of accuracy being 67.74% ± 46.75%, sensitivity 70.83%, specificity 65.79% and precision 56.67% belong to the combination of SVM and RFE algorithms for classification and feature selection, respectively. Given the fact that RFE uses a SVM classifier to evaluate the performance of the dataset, it can be considered that the resulting dataset, after feature selection, is tuned for this specific classification algorithm, thus resulting in better performance values.

Table 4 - Results of SVM classifier without feature selection, with PCA, Forward Selection and RFE methods. The performance values calculated include Accuracy, Sensitivity, Specificity and Precision measures

	No Feature Selection	PCA	Forward Selection	RFE
Accuracy	51.61% ± 49.97%	51.61% ± 49.97%	56.45% ± 49.58%	67.74% ± 46.75%
Sensitivity	16.67%	16.67%	50.00%	70.83%
Specificity	73.68%	73.68%	60.53%	65.79%
Precision	28.57%	28.57%	44.44%	56.67%

5.4 LDA

Table 5 shows the results of LDA performance. It's possible to see that this classifier doesn't perform well without some feature selection approach because it predicted all examples as controls. This results in a very good measure of specificity, because the examples that are actual controls were all correctly classified, however, none of the OCD patients was rightly predicted, which resulted in an extremely low sensitivity result. In this case, all the feature selection methods used in combination with LDA have improved its performance, being RFE the one with bigger

positive impact. RFE together with LDA lead to an accuracy value of 59.68% ± 49.05%, sensitivity of 41.67% and specificity of 71.05%. The precision of the class O was of 47.62%, meaning that from all the subjects classified as having the condition of OCD, more than half of them weren't, in fact, OCD patients.

Table 5 - Results of LDA classifier without feature selection, with PCA, Forward Selection and RFE methods. The performance values calculated include Accuracy, Sensitivity, Specificity and Precision measures

	No Feature Selection	PCA	Forward Selection	RFE
Accuracy	61.29% ± 48.71%	60.04% ± 12.27%	51.61% ± 49.97%	59.68% ± 49.05%
Sensitivity	0.00%	16.67%	25.00%	41.67%
Specificity	100.00%	86.84%	68.42%	71.05%
Precision	0.00%	44.44%	33.33%	47.62%

5.5 Decision Tree

Regarding the decision tree classifier, this algorithm was the one that did not benefit from the combination with feature selection approaches. None of the feature selection approaches used improved this classifier's results. However, this classifier shows relatively good results when implemented by itself only, with and accuracy of 70.97% ± 45.39%, meaning that more than 2 thirds of the subjects were correctly classified. As for the sensitivity value of 58.33%, one can confirm that more than half the OCD patients was correctly classified and nearly 80% of the controls were identified as such. As is visible in Table 6, PCA was the method with the bigger negative influence, resulting in a 0% sensitivity, which means that no OCD patient was correctly classified. The specificity value of 94.74% means that not all the controls were classified as such, in other words, some controls were wrongly classified as OCD patients.

Table 6 - Results of Decision Tree classifier without feature selection, with PCA, Forward Selection and RFE methods. The performance values calculated include Accuracy, Sensitivity, Specificity and Precision measures

	No Feature Selection	PCA	Forward Selection	RFE
Accuracy	70.97% ± 45.39%	58.06% ± 49.35%	69.35% ± 46.10%	61.29% ± 48.71%
Sensitivity	58.33%	0.00%	45.83%	33.33%
Specificity	78.95%	94.74%	84.21%	78.95%
Precision	63.64%	0.00%	64.71%	50.00%

From an overall perspective, one of the best results are from the decision tree classifier without feature selection with accuracy of 70.97%, sensitivity of 58.33%, and specificity of 78.95% and class O precision of 63.64%. The other considered good result is from the SVM classifier combined with RFE. This combination resulted in an accuracy of 67.74%, sensitivity of 70.83%, specificity of 65.79% and class O precision of 56.67%. When evaluating a classifier's performance, one should take into consideration all the measures presented. Focusing on only one of them may lead to wrong conclusions. For instance, regarding Table 5, if one only takes into consideration the accuracy value, the results of LDA without feature selection appear to be better than the remaining ones, because 61.29% is a higher number than all the other ones achieved for this classifier. Yet, it is not true that this case is better than the others, because it did not classified any OCD patient correctly.

Forward Selection seems to improve the results achieved relatively to the case without feature selection, except for the decision tree algorithm that didn't improve from any feature selection method. PCA use shows to be irrelevant when combined with 3NN and SVM classification algorithms. RFE also improved the results obtain, except for the decision tree and 5NN algorithms. The classifier with the poorest overall results is the LDA one. This classifier showed to be unable to correctly classify any OCD patient when performed without a feature selection method combined.

5.6 Discussion

Other studies have been conducted in order to distinguish OCD patients from healthy controls by means of machine learning. Moreover, studies that combine the use of machine learning and graph theory applied to neuroimaging have also been developed. Some of these studies will be concisely described ahead. Results from these studies will also be compared to the results from the present work.

A study by (Weygandt et al., 2012) performed two analysis for pattern recognition in OCD patients. In the first one, the researchers used a SVM classifier to decode the category of a currently viewed image from fMRI patterns for several pairs of categories. The purpose was to investigate whether fear-inducing, disgust-inducing and neutral stimuli could be decoded from brain patterns of fMRI time samples from both OCD patients and healthy controls. As a result, neurobiological markers were identified for a reliable diagnosis of OCD.

The second analysis performed used a searchlight approach to predict a subject's diagnostic status based on local brain patterns. To do so, the researchers used discriminating volumes to separate fear-inducing pictures from neutral ones and disgust-inducing from neutral pictures, for each subject. These discriminating volumes were determined in pattern recognition analysis from the first analysis mentioned above. The researchers accomplished a diagnostic accuracy of 100% for three specific regions (two coordinates of the right orbitofrontal cortex and on in the left caudate nucleus) for the separation of fear-inducing vs. neutral pictures. For the diagnostic classification based on discriminating volumes calculated for the separation of disgust-inducing vs. neutral pictures no significant results were obtained.

Focusing on the 100% accuracy result, which is the optimal result, it is possible to infer that discriminating volumes for the separation of fear-inducing vs. neutral pictures are an adequate measure to perform this classification task. Considering the present work's results, none of the results presented above are comparable to the ones from this study performed by Weygandt and colleagues, because the dataset used in both cases is of different nature. The present work includes FC data as well as demographic values, besides structural volumes. As for Weygandt study, fMRI data was used in the first analysis, which allowed the determination of brain regions of interest, whose volumes were then used for the second analysis. One aspect that coincides for both studies is the use of SVM to perform the classification task.

Another study regarding OCD patients' identification has been developed by (Soriano-Mas et al., 2007). Here, Soriano and colleagues assessed the feasibility of identifying OCD patients through whole-brain anatomical alterations.

After computing the whole-brain pattern of structural difference between OCD patients and healthy controls, an expression value representative of this difference was calculated for each subject to express their similarity to typical OCD anatomical alterations. Afterwards, a univariate statistical analysis was performed to characterize the whole-brain pattern of differences between the two groups. The resulting statistical map of the between-group differences was then multiplied by the GM image of each subject to obtain their individual expression values. The average value was calculated for each group, and classification was performed by the calculation of the distance between each subject's individual value and each one of the groups' mean expression value. In other words, for each subject, two distance measures were calculated, for both OCD and healthy controls mean value. These distances were then converted into complementary probabilities of being an OCD patient or a healthy control. OCD patients were considered wrongly classified when the complementary probability of being an OCD patient was less than 0.5.

In order to evaluate performance, two methods were implemented by these researchers. First, leave-one-out cross-validation was performed, by omitting one subject at a time from the original study sample and then computing a new statistical map from the remaining sample. For the subject outside of the dataset, the expression value was determined and the classification probabilities were calculated. This resulted in an accuracy value of 93.1%. Here, this value was determined as the average of sensitivity and specificity. Secondly, new subjects were used to evaluate the performance of the classification approach designed by these researchers. The expression values were calculated for the new subjects but the statistical map used was the same. This method resulted in and accuracy of 76.6%.

Although having a considerably good result in the cross-validation evaluation, this approach shows to be less adequate when tested against new data. This approach conducted by Soriano is also based on structural data and uses a different classification method than the ones used in the present work, so no fair comparison can be made.

In 2013, Fekete and colleagues conducted a study for the combination of fMRI complex network measures and machine learning (Fekete et al., 2013). This study also introduced a novel

learning approach called block diagonal optimization (Bdopt). Classification was applied to two resting-state fMRI datasets, schizophrenic vs. healthy controls and awake vs. sleep. First, the average timeseries was extracted from 116 automated anatomical labelling ROIs. Afterwards, the lagged correlations and partial correlations ranging from ±3TR were computed, the matrices were thresholded and then seven networks were constructed. Several network measures were calculated: assortativity, closeness centrality, characteristic path length, clustering coefficient, global efficiency, transitivity, local efficiency, modularity and small-world ration for the global measures. Regarding the local ones: node betweenness centrality, degree, node characteristic path length, node clustering coefficient and node global and local efficiency were calculated. As one can see, these metrics include the ones used for the present work.

Feature selection was implemented with a squared two-sample t-statistic approach. Next, several classifier were implemented for both datasets. Within the dataset comprising schizophrenic patients and healthy controls, the SVM classifier's accuracy is of 65%, with 80% specificity and 43% sensitivity. The remaining classifiers had better accuracy results, with 94% for both RFE and Recursive Composite Kernels (RCK) and with 100% for Bdopt. Regarding the second dataset, with awake vs. sleep subjects, SVM's accuracy was of 64%, specificity of 76% and sensitivity of 52%. In its turn, Bdopt's accuracy was of 91%, specificity of 87% and sensitivity of 95%. By comparing these results to the performance results achieved in the present work, one can see that they are similar, regarding the common classifier, SVM. However, in spite of being similar, both these SVM implementations have differences that can be crucial to the final outcome. For instance, the parameters weren't described in (Fekete et al., 2013) but they are most likely to not be the same as the ones used on the present work. The feature dataset was not the same, in Fekete's work more network measures were included and no brain regional volumes were part of the dataset. Moreover, the construction of the correlation matrices was different, because Fekete's work included lagged and partial correlations. All these variances define the final performance result, so this comparison cannot be considered absolute.

6 Conclusions and Future Work

The aim of the present work was to investigate a series of machine learning algorithms and how to implement them to fMRI derived complex network measures. In order to do so a work plan was defined. Initially, some literature research was performed and several databases were investigated in order to find information on the three main study fields relevant for this work. After the gathering and analysis of Neuroimaging, Machine Learning and Graph Theory information, it was time to perform more practical tasks. The preprocessing of the resting-state fMRI images from an OCD study took place. Once the dataset was ready, the feature set was constructed by joining complex network measures, brain regional volumes and demographic data all together. Before implementing the machine learning algorithms chosen, it was required another preprocessing stage. Here, feature selection methods were applied to the data. Finally, cross validation was used to evaluate the classifiers' performance and average values of accuracy, sensitivity and specificity were calculated.

Given the nature of this problem, complex concepts and three different study fields to comprehend, there isn't much information on how to implement, which methods to use and which parameters produce the best outcome. Yet, this combination of neuroimaging and graph theory applied to machine learning has potential and intends to be very promising, whether in the development of tools to aid researchers or to support diagnosis and clinical decision systems for psychiatric pathologies.

In the present work, 5 classification algorithms – 3NN, 5NN, SVM, LDA and decision tree – were implemented by themselves, i.e., without feature selection, and combined with 3 feature selection approaches – RFE, forward selection and PCA. The work performed led to some knowledge about this amalgamation of concepts from machine learning, graph theory and functional connectivity data. In a general way, feature selection methods improved the performance results, except for the decision tree algorithm, which performed best with the complete dataset as input. SVM results can be considered within the expected values, by comparison with (Fekete et al., 2013). This might mean that the feature dataset used in this work is adequate to the classification task performed. As for the other classifiers that didn't perform very well, other features might produce different and better results.

The present work contributes with a small amount of knowledge on how to build the dataset, what works and doesn't work and how the classification processes must be built. In spite of being small, this knowledge is relevant because it's a starting point to further investigations. Without doubt, the bigger limitation of this work was time. It requires time to fully comprehend all the concepts involved, because they come from three distinct disciplines. Moreover, it takes time to implement and run the classification processes. Having more time to develop this work, more classifiers could have been tested.

The first suggestion would be to apply the obtained models to completely new datasets. After the acquisition of fMRI images of new subjects and application of the same preprocessing steps, the models can be tested in order to better assess their performance. These results are considered a good method to obtain feedback on the models created. If the results are satisfactory, it means models are a close representation of reality.

Another recommendation is to implement other classifiers, as well as other feature selection algorithms. Concerning the classifiers, the application of transductive conformal predictors described in (Nouretdinov et al., 2011) seems to be an interesting approach to be considered. It proposes a general probabilistic classification method to produce measures of confidence for MRI data. Another approach worthy of further investigation is the use of Bayesian network classifiers seen in (Morales et al., 2013) as well as a Maximum Entropy Approach (Phillips, Dudík, & Schapire, 2004).

Regarding the feature selection algorithms, the proposed ones are an evolutionary approach to optimize selection and the removal of correlated attributes. Another tactic is to select features using *a prior* approach. This can be performed by choosing features that are known beforehand to be related to the specific classification problem. For instance, in the present work, the volumes from brain regions could be selected in order to retain the ones that are most activated in OCD patients. As for the feature set building, other attributes could be considered. For instance, the actual values of FC, straight from the correlation matrix, could also be used as features.

To remove the noise from the fMRI images, the mentioned ICA-denoise method could be further explored. Another tool, called CompCor (Behzadi, Restom, Liau, & Liu, 2007), could also be used to remove confounds from WM and CSF signals.

Other environments could be experimented, outside of RapidMiner. The Python toolbox for Multivariate Pattern Analysis called PyMVPA (Hanke et al., 2009) seems to be an alternative

worthy of further investigation. This toolbox allows the application of classifier-based analysis techniques to fMRI datasets.

One more approach that deserves more attention is the implementation of a meta-algorithm that together with other learning algorithms can improve their performance. This technique is based on Bayes' Theorem and is known as Bayesian Boosting (Nock & Sebban, 2001).

References

A. Maidment, Seibert, J., & Flynn, M. (n.d.). Technical Advances in DR and CR. *Imaging Economics*, *17*(3), 45–51.

Abraham, R., Simha, J. B., & Iyengar, S. S. (2007). Medical Datamining with a New Algorithm for Feature Selection and Naive Bayesian Classifier. *10th International Conference on Information Technology (ICIT 2007)*, 44–49. doi:10.1109/ICIT.2007.41

Akthat, F., & Hahne, C. (2012). *RapidMiner 5 Operator Reference* (p. 990). Rapid-I GmbH.

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: basic principles. *Brain and Cognition*, *60*(3), 220–32. doi:10.1016/j.bandc.2005.11.009

APIM. (2012). Associação Portuguesa de Informática Médica. Retrieved August 12, 2014, from http://apim.med.up.pt/index.php?option=com_content&view=article&id=63&Itemid=77&lang=pt

Baumgartner, R., Ryner, L., Richter, W., Summers, R., Jarmasz, M., & Somorjai, R. (2000). Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magnetic Resonance Imaging*, *18*(1), 89–94. doi:10.1016/S0730-725X(99)00102-2

Beckmann, C. (2005). Investigations into resting-state connectivity using independent component analysis. *... of the Royal ...*, *1*, 1–16. Retrieved from http://rstb.royalsocietypublishing.org/content/360/1457/1001.short

Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1457), 1001–13. doi:10.1098/rstb.2005.1634

Behzadi, Y., Restom, K., Liau, J., & Liu, T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, *37*, 90–101. doi:10.1016/j.neuroimage.2007.04.042

Berge, C., & Doig, A. G. (1962). *The theory of graphs and its applications* (Vol. 6). Methuen London.

Biswal, B. B. (2012). Resting state fMRI: a personal history. *NeuroImage*, *62*(2), 938–44. doi:10.1016/j.neuroimage.2012.01.090

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541.

Boyd, A. (n.d.). Graph Theory. Retrieved October 03, 2014, from http://www.uh.edu/engines/epi2467.htm

Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping*, *6*(5-6), 373–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9788075

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, *10*(3), 186–98. doi:10.1038/nrn2575

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews. Neuroscience*, *13*(5), 336–49. doi:10.1038/nrn3214

Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, *14*(3), 140–51. doi:10.1002/hbm.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*, 161–168. doi:10.1145/1143844.1143865

Castro, E., Gómez-Verdejo, V., Martínez-Ramón, M., Kiehl, K. a, & Calhoun, V. D. (2013). A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia. *NeuroImage*. doi:10.1016/j.neuroimage.2013.10.065

Chang, C., & Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and ...*, 1–39. Retrieved from http://dl.acm.org/citation.cfm?id=1961199

Cohen, B. M., Gruber, S. A., & Renshaw, P. F. (1996). Functional magnetic resonance imaging of schizophrenic patients and comparison subjects during word production. *Am J Psychiatry*, *1*, 53.

Cole, D. M., Smith, S. M., & Beckmann, C. F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Frontiers in Systems Neuroscience*, *4*(April), 8. doi:10.3389/fnsys.2010.00008

Comon, P. (1992). Independent component analysis. In J-L.Lacoume & Elsevier (Eds.), *Higher Order Statistics* (Vol. 36, pp. 29–38). Retrieved from http://hal.archives-ouvertes.fr/hal-00346684/

Cordes, D., Haughton, V., & Carew, J. (2002). Hierarchical clustering to measure connectivity in fMRI resting-state data. *Magnetic Resonance ...*, *541*(1995), 2002. Retrieved from http://www.sciencedirect.com/science/article/pii/S0730725X02005039

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *297*(20), 273–297. Retrieved from http://link.springer.com/article/10.1007/BF00994018

Costa, E. P., Postal, C., Lorena, A. C., & Ad, R. S. (2007). A Review of Performance Evaluation Measures for Hierarchical Classifiers.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270. doi:10.1016/S1053-8119(03)00049-1

Damoiseaux, J. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the …*, (2). Retrieved from http://www.pnas.org/content/103/37/13848.short

Damoiseaux, J. S., & Greicius, M. D. (2009). Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Structure & Function*, *213*(6), 525–33. doi:10.1007/s00429-009-0208-6

Data Mining and Rapid Miner. (2010). Retrieved September 20, 2014, from http://www.slideshare.net/rapidminercontent/rapidminer-data-mining-and-rapid-miner-3667259?related=2

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, *53*(1), 1–15. doi:10.1016/j.neuroimage.2010.06.010.Automatic

Devlin, H. (2007). What is fMRI? Retrieved from http://psychcentral.com/lib/what-is-functional-magnetic-resonance-imaging-fmri/0001056

DICOM to NIfTI conversion. (n.d.). Retrieved November 10, 2014, from http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html

Dosenbach, N., Nardos, B., Cohen, A. L., Fair, D. A., & Al., E. (2010). Prediction of individual brain maturity using fMRI. *Science*, *1358*(2010). doi:10.1126/science.1194144

Estévez, P., & Tesmer, M. (2009). Normalized mutual information feature selection. *Neural Networks, IEEE …*, *20*(2), 189–201. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4749258

Fekete, T., Wilf, M., Rubin, D., Edelman, S., Malach, R., & Mujica-Parodi, L. R. (2013). Combining classification with fMRI-derived complex network measures for potential neurodiagnostics. *PloS One*, *8*(5), e62867. doi:10.1371/journal.pone.0062867

FMRIB Software Library. (2014). *Michael Chappell*. Retrieved October 29, 2014, from http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslOverview

Fonseca, R. B. (2011). Plano Nacional de Saúde 2011-2016 " Tecnologias de Informação e Comunicação ," *2010*.

Friston, K., & Frith, C. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, *13*(5). Retrieved from http://cream.fil.ion.ucl.ac.uk/~karl/Functional connectivity.pdf

Friston, K., & Holmes, A. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, (2), 189–210. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/hbm.460020402/full

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. doi:10.1016/S1053-8119(03)00202-7

Fumiko Hoeft. (2008). Voxel-Based Neuroimaging Analysis.

George Casella, & Berger, R. L. (2001). *Statistical Inference* (p. 660). Duxbury Press.

Gould, T. A., & Edmonds, M. (2010). How MRI Works. Retrieved July 30, 2014, from http://science.howstuffworks.com/mri.htm

Graner, J., Oakes, T. R., French, L. M., & Riedy, G. (2013). Functional MRI in the investigation of blast-related traumatic brain injury. *Frontiers in Neurology*, *4*(March), 16. doi:10.3389/fneur.2013.00016

Greicius, M. D., Kiviniemi, V., Tervonen, O., Vainionpää, V., Alahuhta, S., Reiss, A. L., & Menon, V. (2008). Persistent default-mode network connectivity during light sedation. *Human Brain Mapping*, *29*(7), 839–47. doi:10.1002/hbm.20537

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 253–8. doi:10.1073/pnas.0135058100

Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4637–42. doi:10.1073/pnas.0308627101

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(1), 72–8. doi:10.1093/cercor/bhn059

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., & Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), e159. doi:10.1371/journal.pbio.0060159

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V, & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*(1), 37–53. doi:10.1007/s12021-008-9041-y

Haslinger, B., Erhard, P., Ka, N., Boecker, H., Rummeny, E., Schwaiger, M., & Conrad, B. (2001). Event-related functional magnetic resonance imaging in Parkinson ' s disease before and after levodopa. *Brain*, *124*, 558–570.

Haux, R. (2010). Medical informatics: past, present, future. *International Journal of Medical Informatics*, *79*(9), 599–610. doi:10.1016/j.ijmedinf.2010.06.003

Haynes, J.-D., & Rees, G. (n.d.). Decoding Mental States from Brain Activity in Humans. *Nature Reviews. Neuroscience*, *7*, 523–534. doi:10.1038/nrn1931

Helmenstine, A. M. (2014). Paramagnetism Definition. Retrieved July 15, 2014, from http://chemistry.about.com/od/chemistryglossary/g/paramagnetism-definition.htm

Hesselmann, V., Zaro Weber, O., Wedekind, C., Krings, T., Schulte, O., Kugel, H., … Lackner, K. J. (2001). Age related signal decrease in functional magnetic resonance imaging during motor stimulation in humans. *Neuroscience Letters*, *308*(3), 141–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11479008

Honey, C., & Sporns, O. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the …*, *106*(6), 1–6. Retrieved from http://www.pnas.org/content/106/6/2035.short

Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., Fransson, P., Skio, B., & Åden, U. (2007). Resting-state networks in the infant brain, *104*(39), 15531–15536.

Hsu, C., Chang, C., & Lin, C. (2003). A practical guide to support vector classification, *1*(1), 1–16. Retrieved from https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf

Ibrahim, T., Nekolla, S., Schreiber, K., Odaka, K., Volz, S., Mehilli, J., … Schwaiger, M. (2002). Assessment of coronary flow reserve: comparison between contrast-enhanced magnetic resonance imaging and positron emission tomography. *J AM Coll Cardiol*, *23*, 864–870.

Janecek, A., Gansterer, W., Demel, M., & Ecker, G. (2008). On the Relationship Between Feature Selection and Classification Accuracy. *FSDM*, (4), 90–105. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:On+the+Relationship+Between+Feature+Selection+and+Classification+Accuracy#0

Jo, H. J., Saad, Z. S., Simmons, W. K., Milbury, L. a, & Cox, R. W. (2010). Mapping sources of correlation in resting state FMRI, with artifact detection and removal. *NeuroImage*, *52*(2), 571–82. doi:10.1016/j.neuroimage.2010.04.246

Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, *248*, 243–248. Retrieved from http://193.62.66.20/spm/doc/papers/oj_event.pdf

Kornak, J., Hall, D. a, & Haggard, M. P. (2011). Spatially extended FMRI signal response to stimulus in non-functionally relevant regions of the human brain: preliminary results. *The Open Neuroimaging Journal*, *5*, 24–32. doi:10.2174/1874440001105010024

Leistedt, S. J. J., Coumans, N., Dumont, M., Lanquart, J.-P., Stam, C. J., & Linkowski, P. (2009). Altered sleep brain functional connectivity in acutely depressed patients. *Human Brain Mapping*, *30*(7), 2207–19. doi:10.1002/hbm.20662

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–99. doi:10.1016/j.neuroimage.2010.11.004

Li, K., Guo, L., Nie, J., Li, G., & Liu, T. (2009). Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics : The Official Journal of the Computerized Medical Imaging Society*, *33*(2), 131–9. doi:10.1016/j.compmedimag.2008.10.011

Liang, X., Wang, J., Yan, C., Shu, N., Xu, K., Gong, G., & He, Y. (2012). Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional MRI study. *PloS One*, *7*(3), e32766. doi:10.1371/journal.pone.0032766

Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., … Jiang, T. (2008). Disrupted small-world networks in schizophrenia. *Brain : A Journal of Neurology*, *131*(Pt 4), 945–61. doi:10.1093/brain/awn018

Long, D., Wang, J., Xuan, M., Gu, Q., Xu, X., Kong, D., & Zhang, M. (2012). Automatic classification of early Parkinson's disease with multi-modal MR imaging. *PloS One*, *7*(11), e47714. doi:10.1371/journal.pone.0047714

Marques, P., Soares, J. M., Alves, V., & Sousa, N. (2013). BrainCAT - a tool for automated and combined functional magnetic resonance imaging and diffusion tensor imaging brain connectivity analysis. *Frontiers in Human Neuroscience*, *7*(November), 794. doi:10.3389/fnhum.2013.00794

Martino, F. De, Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58. doi:10.1016/j.neuroimage.2008.06.037

MathWorks. (n.d.). MatLab. Retrieved from http://www.mathworks.com/products/matlab/

Morales, D. a, Vives-Gilabert, Y., Gómez-Ansón, B., Bengoetxea, E., Larrañaga, P., Bielza, C., … Delfino, M. (2013). Predicting dementia development in Parkinson's disease using Bayesian network classifiers. *Psychiatry Research*, *213*(2), 92–8. doi:10.1016/j.pscychresns.2012.06.001

Moreau, A. (2014). FreeSurfer Recommended Reconstructor. Retrieved October 02, 2014, from https://surfer.nmr.mgh.harvard.edu/fswiki/RecommendedReconstruction

Murphy, K., Birn, R. M., & Bandettini, P. a. (2013). Resting-state fMRI confounds and cleanup. *NeuroImage*, *80*, 349–59. doi:10.1016/j.neuroimage.2013.04.001

Murphy, K., Birn, R. M., Handwerker, D. a, Jones, T. B., & Bandettini, P. a. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage*, *44*(3), 893–905. doi:10.1016/j.neuroimage.2008.09.036

National High Magnetic Field Laboratory. (2014). Retrieved October 01, 2014, from http://www.magnet.fsu.edu/usershub/scientificdivisions/nmr/mri.html

NetworkX - High productivity software for complex networks. (2014). Retrieved November 05, 2014, from https://networkx.github.io/

Ng, A., & Jordan, M. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in Neural Information Processing ....* Retrieved from https://papers.nips.cc/paper/2020-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf

Nitime: Time-series Analysis for Neuroscience. (n.d.). Retrieved June 14, 2014, from http://nipy.org/nitime/

Nock, R., & Sebban, M. (2001). A Bayesian boosting theorem. *Pattern Recognition Letters*, *22*(3-4), 413–419. doi:10.1016/S0167-8655(00)00137-9

Norman, K. a, Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30. doi:10.1016/j.tics.2006.07.005

Nouretdinov, I., Costafreda, S. G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., & Fu, C. H. Y. (2011). Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. *NeuroImage*, *56*(2), 809–13. doi:10.1016/j.neuroimage.2010.05.023

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872. doi:10.1073/pnas.87.24.9868

Ogawa, S., Tank, D., MEnon, R., Ellermann, J. M., Kim, S.-G., Merkle, H., & KAmil Ugu. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the ...*, *89*(July), 5951–5955. Retrieved from http://www.pnas.org/content/89/13/5951.short

Onias, H., Viol, A., Palhano-Fontes, F., Andrade, K. C., Sturzbecher, M., Viswanathan, G., & de Araujo, D. B. (2013). Brain complex network analysis by means of resting state fMRI and graph analysis: Will it be helpful in clinical epilepsy? *Epilepsy & Behavior : E&B*. doi:10.1016/j.yebeh.2013.11.019

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1 Suppl), S199–209. doi:10.1016/j.neuroimage.2008.11.007

Phillips, S., Dudík, M., & Schapire, R. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First ....* Retrieved from http://dl.acm.org/citation.cfm?id=1015412

Pujol, J., Soriano-Mas, C., Alonso, P., Cardoner, N., Menchón, J. M., Deus, J., & Vallejo, J. (2004). Mapping structural brain alterations in obsessive-compulsive disorder. *Archives of General Psychiatry*, *61*(7), 720–30. doi:10.1001/archpsyc.61.7.720

Radua, J., & Mataix-Cols, D. (2009). Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *The British Journal of Psychiatry : The Journal of Mental Science*, *195*(5), 393–402. doi:10.1192/bjp.bp.108.055046

Raichle, M. E., MacLeod, a M., Snyder, a Z., Powers, W. J., Gusnard, D. a, & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 676–82. doi:10.1073/pnas.98.2.676

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*(3), 1059–69. doi:10.1016/j.neuroimage.2009.10.003

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics (Oxford, England)*, *23*(19), 2507–17. doi:10.1093/bioinformatics/btm344

Saini, S. (2004). Altered cerebellar functional connectivity mediates potential adaptive plasticity in patients with multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(6), 840–846. doi:10.1136/jnnp.2003.016782

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. doi:10.1147/rd.33.0210

Sato, J. R., Basilio, R., Paiva, F. F., Garrido, G. J., Bramati, I. E., Bado, P., ... Moll, J. (2013). Real-Time fMRI Pattern Decoding and Neurofeedback Using FRIEND: An FSL-Integrated BCI Toolbox. *PloS One*, *8*(12), e81658. doi:10.1371/journal.pone.0081658

Schowe, B. (2011). Feature Selection for high-dimensional data with RapidMiner. *... of the 2nd RapidMiner Community Meeting ....* Retrieved from http://www-ai.cs.uni-dortmund.de/PublicPublicationFiles/schowe_2011a.pdf

Schröter, M. S., Spoormaker, V. I., Schorer, A., Wohlschläger, A., Czisch, M., Kochs, E. F., ... Ilg, R. (2012). Spatiotemporal reconfiguration of large-scale brain functional networks during propofol-induced loss of consciousness. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(37), 12832–40. doi:10.1523/JNEUROSCI.6046-11.2012

Song, M., Zhou, Y., Li, J., Liu, Y., Tian, L., Yu, C., & Jiang, T. (2008). Brain spontaneous functional connectivity and intelligence. *NeuroImage*, *41*(3), 1168–76. doi:10.1016/j.neuroimage.2008.02.036

Song, S., Zhan, Z., Long, Z., Zhang, J., & Yao, L. (2011). Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PloS One*, *6*(2), e17191. doi:10.1371/journal.pone.0017191

Soriano-Mas, C., Pujol, J., Alonso, P., Cardoner, N., Menchón, J. M., Harrison, B. J., ... Gaser, C. (2007). Identifying patients with obsessive-compulsive disorder using whole-brain anatomy. *NeuroImage*, *35*(3), 1028–37. doi:10.1016/j.neuroimage.2007.01.011

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, *2*(2), 145–62. doi:10.1385/NI:2:2:145

Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., & Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(1), 92–9. doi:10.1093/cercor/bhj127

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics*, *1*(1), 3. doi:10.1186/1753-4631-1-3

Stephanie Watson. (2008). How fMRI Works. Retrieved July 30, 2014, from http://science.howstuffworks.com/fmri1.htm

Sui, J., Adali, T., Pearlson, G. D., & Calhoun, V. D. (2009). An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques. *NeuroImage*, *46*(1), 73–86. doi:10.1016/j.neuroimage.2009.01.026

Tohka, J., Foerde, K., Aron, A., & Tom, S. (2008). Automatic independent component labeling for artifact removal in fMRI. *Neuroimage*, *39*(3), 1227–1245. doi:10.1016/j.neuroimage.2007.10.013.Automatic

University of Virginia. (n.d.). University of Virginia - School of Medicine: Health Informatics Definition. Retrieved June 06, 2014, from http://www.medicine.virginia.edu/clinical/departments/phs/administrative-divisions/informatics/HealthInfDef-page

Van den Heuvel, M., Mandl, R., & Hulshoff Pol, H. (2008). Normalized cut group clustering of resting-state FMRI data. *PloS One*, *3*(4), e2001. doi:10.1371/journal.pone.0002001

Van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology : The Journal of the European College of Neuropsychopharmacology*, *20*(8), 519–34. doi:10.1016/j.euroneuro.2010.03.008

Van Wijk, B. C. M., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PloS One*, *5*(10), e13701. doi:10.1371/journal.pone.0013701

Vapnik, V. N. (1998). *Statistical learning theory*. Retrieved from http://books-make-you-smart.info/wp-content/uploads/pdfs/Statistical Learning Theory by Vladimir N Vapnik - An Excellent Overview.pdf

Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional MRI. *Frontiers in Systems Neuroscience*, *4*(June), 16. doi:10.3389/fnsys.2010.00016

Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., … Wang, Y. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Human Brain Mapping*, *30*(2), 638–49. doi:10.1002/hbm.20530

Warren Powel. (2008). *Approximate Dynamic Programming: Solving the Curses of Di- mensionality*.

Wee, C.-Y., Yap, P.-T., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. a, … Shen, D. (2012). Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. *PloS One*, *7*(5), e37828. doi:10.1371/journal.pone.0037828

Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., & Windischberger, C. (2009). Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. *NeuroImage*, *47*(4), 1408–16. doi:10.1016/j.neuroimage.2009.05.005

Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine Learning in Medical Imaging. *Signal Processing Magazine, IEEE*, *27*(4), 25–38. doi:10.1109/MSP.2010.936730

Weygandt, M., Blecker, C. R., Schäfer, A., Hackmack, K., Haynes, J.-D., Vaitl, D., … Schienle, A. (2012). fMRI pattern recognition in obsessive-compulsive disorder. *NeuroImage*, *60*(2), 1186–93. doi:10.1016/j.neuroimage.2012.01.064

Zwass, V. (n.d.). Encyclopaedia Britannica - Information System. Retrieved October 03, 2014, from http://global.britannica.com/EBchecked/topic/287895/information-system

Appendix

Correspondence between the brain areas used as nodes and their shortened version.

Left-Thalamus-Proper	LTP
Left-Caudate	LCAU
Left-Putamen	LPUT
Left-Pallidum	LPAL
Left-Hippocampus	LHIP
Left-Amygdala	LAMY
Left-Accumbens-area	LACC
Right-Thalamus-Proper	RTP
Right-Caudate	RCAU
Right-Putamen	RPUT
Right-Pallidum	RPAL
Right-Hippocampus	RHIP
Right-Amygdala	RAMY
Right-Accumbens-area	RACC
ctx_lh_G_and_S_frontomargin	LFM
ctx_lh_G_and_S_occipital_inf	LOCCI
ctx_lh_G_and_S_paracentral	LPC
ctx_lh_G_and_S_subcentral	LSC
ctx_lh_G_and_S_transv_frontopol	LTF
ctx_lh_G_and_S_cingul-Ant	LCANT
ctx_lh_G_and_S_cingul-Mid-Ant	LCMANT
ctx_lh_G_and_S_cingul-Mid-Post	LCMPOS
ctx_lh_G_cingul-Post-dorsal	LCPOSD
ctx_lh_G_cingul-Post-ventral	LCPOSV
ctx_lh_G_cuneus	LCUN
ctx_lh_G_front_inf-Opercular	LFIOPER
ctx_lh_G_front_inf-Orbital	LFIORB
ctx_lh_G_front_inf-Triangul	LFITRI

ctx_lh_G_front_middle	LGFMID
ctx_lh_G_front_sup	LGFSUP
ctx_lh_G_Ins_lg_and_S_cent_ins	LIGCENT
ctx_lh_G_insular_short	LIS
ctx_lh_G_occipital_middle	LOCCMID
ctx_lh_G_occipital_sup	LOCCSUP
ctx_lh_G_oc-temp_lat-fusifor	LLAT-FUSI
ctx_lh_G_oc-temp_med-Lingual	LMED-LING
ctx_lh_G_oc-temp_med-Parahip	LMED-PARA
ctx_lh_G_orbital	LORB
ctx_lh_G_pariet_inf-Angular	LPARIANG
ctx_lh_G_pariet_inf-Supramar	LPARISUP
ctx_lh_G_parietal_sup	LPARSUP
ctx_lh_G_postcentral	LPOS
ctx_lh_G_precentral	LPRE
ctx_lh_G_precuneus	LPREC
ctx_lh_G_rectus	LREC
ctx_lh_G_subcallosal	LSUBCALL
ctx_lh_G_temp_sup-G_T_transv	LTRANSV
ctx_lh_G_temp_sup-Lateral	LSUPLAT
ctx_lh_G_temp_sup-Plan_polar	LSPPOL
ctx_lh_G_temp_sup-Plan_tempo	LSPTEMP
ctx_lh_G_temporal_inf	LTINF
ctx_lh_G_temporal_middle	LTMID
ctx_lh_Lat_Fis-ant-Horizont	LFH
ctx_lh_Lat_Fis-ant-Vertical	LFV
ctx_lh_Lat_Fis-post	LFPOS
ctx_lh_Pole_occipital0	LPOCC
ctx_lh_Pole_temporal	LPTEMP
ctx_lh_S_calcarine	LCAL
ctx_lh_S_central	LCEN
ctx_lh_S_cingul-Marginalis	LCINMAR

ctx_lh_S_circular_insula_ant	LCIRINSANT
ctx_lh_S_circular_insula_inf	LCIRINSINF
ctx_lh_S_circular_insula_sup	LCIRINSSUP
ctx_lh_S_collat_transv_ant	LCOLLANT
ctx_lh_S_collat_transv_post	LCOLLPOS
ctx_lh_S_front_inf	LFINF
ctx_lh_S_front_middle	LSFMID
ctx_lh_S_front_sup	LSFSUP
ctx_lh_S_interm_prim-Jensen	LINJENS
ctx_lh_S_intrapariet_and_P_trans	LINTPAR
ctx_lh_S_oc_middle_and_Lunatus	LOCMIDLUN
ctx_lh_S_oc_sup_and_transversal	LOCTRANS
ctx_lh_S_occipital_ant	LOCCANT
ctx_lh_S_oc-temp_lat	LOCTLAT
ctx_lh_S_oc-temp_med_and_Lingual	LOCTMEDLING
ctx_lh_S_orbital_lateral	LORBLAT
ctx_lh_S_orbital_med-olfact	LORBMED-OLF
ctx_lh_S_orbital-H_Shaped	LORB-SH
ctx_lh_S_parieto_occipital	LPAROCC
ctx_lh_S_postcentral	LPOSCEN
ctx_lh_S_precentral-inf-part	LPRECINF
ctx_lh_S_precentral-sup-part	LPRECSUP
ctx_lh_S_suborbital	LSUBORB
ctx_lh_S_subparietal	LSUBPAR
ctx_lh_S_temporal_inf	LTEMPINF
ctx_lh_S_temporal_sup	LTEMPSUP
ctx_lh_S_temporal_transverse	LTTRANS
ctx_rh_G_and_S_frontomargin	RFM
ctx_rh_G_and_S_occipital_inf	ROCCI
ctx_rh_G_and_S_paracentral	RPC
ctx_rh_G_and_S_subcentral	RSC
ctx_rh_G_and_S_transv_frontopol	RTF

ctx_rh_G_and_S_cingul-Ant	RCANT
ctx_rh_G_and_S_cingul-Mid-Ant	RCMANT
ctx_rh_G_and_S_cingul-Mid-Post	RCMPOS
ctx_rh_G_cingul-Post-dorsal	RCPOSD
ctx_rh_G_cingul-Post-ventral	RCPOSV
ctx_rh_G_cuneus	RCUN
ctx_rh_G_front_inf-Opercular	RFIOPER
ctx_rh_G_front_inf-Orbital	RFIORB
ctx_rh_G_front_inf-Triangul	RFITRI
ctx_rh_G_front_middle	RGFMID
ctx_rh_G_front_sup	RGFSUP
ctx_rh_G_Ins_lg_and_S_cent_ins	RIGCENT
ctx_rh_G_insular_short	RIS
ctx_rh_G_occipital_middle	ROCCMID
ctx_rh_G_occipital_sup	ROCCSUP
ctx_rh_G_oc-temp_lat-fusifor	RLAT-FUSI
ctx_rh_G_oc-temp_med-LinguaL	RMED-LING
ctx_rh_G_oc-temp_med-Parahip	RMED-PARA
ctx_rh_G_orbital	RORB
ctx_rh_G_pariet_inf-Angular	RPARIANG
ctx_rh_G_pariet_inf-Supramar	RPARISUP
ctx_rh_G_parietal_sup	RPARSUP
ctx_rh_G_postcentral	RPOS
ctx_rh_G_precentral	RPRE
ctx_rh_G_precuneus	RPREC
ctx_rh_G_rectus	RREC
ctx_rh_G_subcallosal	RSUBCALL
ctx_rh_G_temp_sup-G_T_transv	RTRANSV
ctx_rh_G_temp_sup-Lateral	RSUPLAT
ctx_rh_G_temp_sup-Plan_polar	RSPPOL
ctx_rh_G_temp_sup-Plan_tempo	RSPTEMP
ctx_rh_G_temporal_inf	RTINF

ctx_rh_G_temporal_middle	RTMID
ctx_rh_Lat_Fis-ant-Horizont	RFH
ctx_rh_Lat_Fis-ant-Vertical	RFV
ctx_rh_Lat_Fis-post	RFPOS
ctx_rh_Pole_occipital	RPOCC
ctx_rh_Pole_temporal	RPTEMP
ctx_rh_S_calcarine	RCAL
ctx_rh_S_central	RCEN
ctx_rh_S_cingul-Marginalis	RCINMAR
ctx_rh_S_circular_insula_ant	RCIRINSANT
ctx_rh_S_circular_insula_inf	RCIRINSINF
ctx_rh_S_circular_insula_sup	RCIRINSSUP
ctx_rh_S_collat_transv_ant	RCOLLANT
ctx_rh_S_collat_transv_post	RCOLLPOS
ctx_rh_S_front_inf	RFINF
ctx_rh_S_front_middle	RSFMID
ctx_rh_S_front_sup	RSFSUP
ctx_rh_S_interm_prim-Jensen	RINJENS
ctx_rh_S_intrapariet_and_P_trans	RINTPAR
ctx_rh_S_oc_middle_and_Lunatus	ROCMIDLUN
ctx_rh_S_oc_sup_and_transversal	ROCTRANS
ctx_rh_S_occipital_ant	ROCCANT
ctx_rh_S_oc-temp_lat	ROCTLAT
ctx_rh_S_oc-temp_med_and_Lingual	ROCTMEDLING
ctx_rh_S_orbital_lateral	RORBLAT
ctx_rh_S_orbital_med-olfact	RORBMED-OLF
ctx_rh_S_orbital-H_Shaped	RORB-SH
ctx_rh_S_parieto_occipital	RPAROCC
ctx_rh_S_postcentral	RPOSCEN
ctx_rh_S_precentral-inf-part	RPRECINF
ctx_rh_S_precentral-sup-part	RPRECSUP
ctx_rh_S_suborbital	RSUBORB

ctx_rh_S_subparietal	RSUBPAR
ctx_rh_S_temporal_inf	RTEMPINF
ctx_rh_S_temporal_sup	RTEMPSUP
ctx_rh_S_temporal_transverse	RTTRANS