Insubria University
Department of Theoretical and Applied Sciences (DiSTA)
PhD Thesis in Computer Science
XXXIII Cycle of Study

Analysis of fMRI Exams Through Unsupervised Learning and Evaluation Index

Advisor: Prof. Elisabetta Binaghi
Scientific Advisor: Dr. Alberto Arturo Vergani
External Reviewers: Dr. Perrinet Laurent, and Dr. Marco Marcon

Candidate
Samuele Martinelli

December 2020

to my wife

Contents

1 Introduction **7**
1.1 Motivation 7
1.2 Neuroimaging analysis through fMRI 9
1.3 Clustering Validation 11

2 Background **15**
2.1 FMRI 15
2.1.1 Bold Signal 16
2.1.2 Hemodynamic Response 17
2.1.3 FMRI Exam 18
2.1.4 Preprocessing FMRI Data 20
2.2 Machine Learning 23
2.2.1 Supervised learning 23
2.2.2 Unsupervised learning 24
2.2.3 Reinforcement learning 24
2.3 Neural Networks 26
2.4 Clustering 27
2.4.1 Uses of Clustering 29
2.4.2 Clustering algorithms 29
2.5 CVIs 37
2.5.1 Fukuyama Sugeno 38
2.5.2 Xie Beni 38
2.5.3 Compose Within and Between scattering Index 38
2.5.4 WSJ 39
2.5.5 PBM 40
2.5.6 FPBM 41
2.5.7 Davies Bouldin 41
2.5.8 Soft Davies Bouldin 41
2.6 ANFIS 43
2.6.1 Fuzzy Logic 43
2.6.2 Fuzzy Inference 43

2.6.3 NEURO-FUZZY Networks . . . 44
2.6.4 ANFIS architecture . . . 44

3 Unsupervised Learning in fMRI Analysis 49
3.1 Resting State FMRI Functional Connectivity Analysis Using Soft Competitive Learning Algorithms . . . 50
3.1.1 Premise . . . 50
3.1.2 Introduction . . . 51
3.1.3 Data . . . 53
3.1.4 Methods . . . 53
3.1.5 Result . . . 54
3.1.6 Discussion and Conclusion . . . 62
3.2 Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms . . . 63
3.2.1 Premise . . . 63
3.2.2 Introduction . . . 63
3.2.3 Materials and Methods . . . 65
3.2.4 Results . . . 67
3.2.5 Discussion . . . 68
3.2.6 Conclusion and Future Works . . . 78

4 CVIs and Evaluation Methodologies Analysis for fMRI 81
4.1 Comparison of validity indexes for fuzzy clusters of fMRI data . . . 82
4.1.1 Premise . . . 82
4.1.2 Introduction . . . 82
4.1.3 Clustering Problem and Fuzzy C-Means Algorithm . . . 84
4.1.4 Cluster Validation Indexes . . . 86
4.1.5 Experiments and Results . . . 89
4.1.6 fMRI dataset . . . 89
4.1.7 Experiments . . . 90
4.1.8 Discussion and Conclusions . . . 91
4.2 Comparison of Validity Indexes for Fuzzy Clusters of fMRI data by Using Adaptive-Network Fuzzy Inference Systems . . . 93
4.2.1 Premise . . . 93
4.2.2 Introduction . . . 93
4.2.3 Clustering problem . . . 95
4.2.4 Cluster validation index . . . 95
4.2.5 ANFIS . . . 99
4.2.6 Experiments and Results . . . 101
4.2.7 Experiments . . . 109

4.2.8 Results 113
4.2.9 Discussion of results and conclusions. 119

5 Conclusion and Future Works 121
5.1 Conclusion 121
5.2 Future Works 123

Acknowledgment

First of all, I would like to thank my supervisor Prof. Elisabetta Binaghi, who allowed me to have this extraordinary experience. Thanks to her, I learned what it means to do research, and she has provided me with constant support over the years. Thank you very much!

I want to thank my colleague Dr. Alberto Arturo Vergani with whom I collaborated during these three years, and who also was scientific supervisor of this thesis, thanks to him and his help, I was able to complete this experience.

A sincere thanks go to all my colleagues, in particular to Gloria Gonella, Christian Rondanini and Federico Daidone who shared this route with me, and with whom I have shared many experiences. Thanks to all professors, researchers, technicians and PhD co-workers from the Departments of Theoretical and Applied Science in Varese.

I also want to thank my thesis reviewers, Dr. Perrinet Laurent, and Dr. Marco Marcon for their instructive input and comments on my work.

A special thanks go to my family, who believed in me and in the choice to pursue a doctorate, helping and supporting me over the years.

Finally, the most important thanks goes to Gloria, who became my wife during this PhD course! To her, who shared with me the whole university and doctoral path, who supported me, but, above all endured me, and without her this adventure would never have started.

1
Introduction

1.1 Motivation

In the last few years, the clustering of time series [78, 33] has seen significant growth and has proven effective in providing useful information in various domains of use. This growing interest in time series clustering is the result of the effort made by the scientific community in the context of time data mining.

In the past decade, there has also been a field in the biomedical area that has attracted considerable interest and is the branch of functional neuroimaging. Functional neuroimaging is based on the use of technologies capable of measuring brain metabolism, to analyse and study the relationship between the activity of certain brain areas and specific brain functions, moreover, the functional neuroimaging is a tool of primary importance in cognitive neuroscience and neuropsychology. The data analyses obtained from these technologies are based on the study of time series. Starting from these premises, one part of this thesis work was developed based on the analysis of data based on time series through clustering techniques and comparison of various clustering algorithms.

The time series analysed come from fMRI (functional magnetic resonance imaging) examinations; other techniques use time series, such as EEG (electroencephalography), MEG (magnetoencephalogram) and even PET (positron emission tomography), but fMRI is the technique that has taken the most important position in the field of neuroimaging. Functional magnetic resonance imaging (fMRI) in recent years has been widely used to study the functional activities, and cognitive behaviours of the brain based on a stimulus [52] or during the resting state [63, 23]. To search for neuroscientific patterns within fMRI data, various statistical methods have been used [16, 38, 80, 54, 10]. For task-based analyses, the most widely used model is the general linear model (GLM) [24, 81], while for resting-state analyses one of the most widely used algorithms is the independent component analysis (ICA) [59].

The idea is becoming increasingly widespread that the analysis of the two different types of fMRI exams can generate a better understanding of the organisation and origin of the cognitive functioning of the brain. However, there are some problems to be addressed; the first is the variability of fMRI signals between brain scans and between individuals, in fact, it is difficult to derive coherent fMRI activation patterns across different brains and populations due to the enormous variability between individuals [87]. Another problem is the high dimensional of the data turns out to be a problem, especially in the comparison between different subjects, and finally there is also the problem of noise management in fMRI signals which may be due to hardware problems (the instability of the scanner, experiment design deficits, high field susceptibility) but also problems due to possible patient head movement and other factors unrelated to the [87] exam. Technics as GLM, ICA, and other statistical techniques appear to be in difficulty in managing these issues. Given the complexity of the field of study, to overcome the limits of statistical techniques and also to have a different analysis, in the scientific community there has been an increase in fMRI analyses through unsupervised learning models (clustering). These models given them characteristic of being guided by data without the need for any truth,and for their ability to manage large volumes of data and resistance to errors, they are candidates as an excellent alternative to the more classic methods of statistical analysis.

For these reasons, the first phase of the thesis focused on the study of the data obtained from fMRI exams carried out in task-based and resting state mode, using and comparing different clustering algorithms: Self-Organizing map (SOM) [47], the Growing Neural Gas (GNG) [28] and Neural Gas (NG) [58] which are crisp-type algorithms, a fuzzy algorithm, the Fuzzy C algorithm, was also used (FCM) [6]. The evaluation of the results obtained by using clustering algorithms was carried out using the Davies Bouldin evaluation index (DBI or DB index) [14].

Clustering algorithms being unsupervised learning models, they suffer the problem of how to carry out their evaluation, the most used technique is that of using evaluation indexes. The evaluation indexes assess the clustering carried out through internal metrics, usually the compactness and separation of the clusters created are evaluated. There are different indexes of validation and, based on their characteristics, they may be more or less suitable for the evaluation of clustering performed on certain types of data. Since fMRI clustering analysis is relatively new, there is still a lack of literature on validation indexes studies for this analysis, so the second part of this thesis focuses on the study of assessment indexes within fMRI analyses, using eight different indexes. The indexes analysed are the Fukuyama-Sugeno index (FSI) [29], the Xie-Beni index (XBI) [82], the Pakhira-Bandyopadhyay-Maulik (PBMI) index [60] in the crisp version and its fuzzy version (FPBMI) [60], the Rezaee-Lelieveldt-Reiberuno (RLRI) index [66], Wang-Sun-Jiang (WSJI) [77], the aforementioned Davies Bouldin (DBI) [14] index, and the soft version of the Davies Bouldin (SDBI) [73] index. The analysis of these indexes also implies an evaluation of the performance of the various indexes; typically an index is evaluated based on whether it is able to recognise the correct number of partitions into which a reference data set is divided, therefore only the performance excellent are recognised. Given the complexity of the data and its characteristics of heterogeneity, the evaluation of the excellent results may not be sufficient alone, therefore for the evaluation of the indexes a metric was introduced that takes into account the sub-optimal results. Wanting to extend the evaluation of sub-optimal results with other metrics an adequate methodology was needed; this was done through the use of the adaptive neuro-fuzzy inference system (ANFIS) [11]. In the next chapters, the themes just exposed will be deepened.

The two main topics of this thesis will be introduced in the next chapters, the analysis of fMRI exams and the evaluation of clustering. The first topic will be detailed in the Chapter 2.1which will explain in detail the usefulness, execution and processing of an fMRI exam. Giving a very concise definition, an fMRI exam is an exam that allows us to highlight brain activities, both during actions and in rest situations.

Clustering evaluation is the second topic of this thesis, clustering, which will be explained in detail in the Chapter 2.4, is a technique that allows to group data, according to their characteristics. To evaluate the goodness of the grouping there are specific techniques, but none of these is already consolidated for the study of fMRI exams, furthermore, the assess of evaluation techniques is still an open field of research.

1.2 Neuroimaging analysis through fMRI

The analysis and interpretation of neuroimaging data often require the division of the brain into several regions, with similar characteristics, regardless of whether they are defined in the brain volume or instead are on the cortical surface.

The most commonly used techniques for data studies are the regions of anatomical or functional interest (ROI) [62], the brain atlases, the data-based plots, these techniques can also be used in combination with each other.

The ROI-based analysis is a technique to focus data analysis on some structures of interest, in this case, a unique element is created that represents the signal in a predefined region, the regions can be chosen based on previous knowledge, and having very different dimensions, a region could also be made up of a handful of voxels. The results depend heavily on the choice of ROI.

Brain atlases, depending on the atlas, provide a mapping that can cover the entire volume of the brain divided into ROI. An atlas, therefore, reports a knowledge of brain structures (anatomically, functionally or based on connectivity) often derived from an ontology derived from previous knowledge in which the regions are clearly distinguished, to which labels are assigned. However, atlases have limits due precisely to their characteristics, as mentioned above, atlases are created based on previous knowledge, and currently, there are many different atlases. Still, they are not all consistent with each other, moreover, an atlas is a standard representation of a brain, but this representation does not always perfectly adapt to the data to be analysed. There are various factors that can influence, a common reason is the atlas processing strategies, which are different from those used in the image to be analysed. Another example concerns the frontal brain regions which are often very large in atlases, while it would be preferable to have a greater division into subgroups with well-defined roles.

The parcellations based on the data, unlike the brain atlases, do not reflect a predefined ontology of the brain structures but allow to represent a better model of the signal given by the signal itself. The parcellations through statistical techniques allow the reduction of data, the creation of models of brain connectivity, of physiological parameters, and it is useful for group analysis, and the comparison of multiple or multivariate models. This is particularly useful for analysing groups of subjects, as this step can reduce the data size by several orders of magnitude while retaining most of the information of interest.

In recent years, fMRI neuroimaging analysis has seen a significant increase, with both statistical and data-oriented brain fragmentation analyses. In fMRI analyses typically the aim is to identify which areas of the brain are activated following a stimulus. To establish which areas have been activated, a particular signal called BOLD signal is analysed, the analysis of this signal is determined for the experimental activity. In the statistical analysis of the acquired data, after a pre-processing, the voxels with greater intensity are identified. The images obtained, representing the activated regions, are called statistical parametric map. The GLM (General Linear Model) [24] represents the most used and robust method for obtaining statistical activation maps from fMRI images.GLM models the time series as a linear combination of several different signal components, and checks whether activity in a region of the brain is systematically linked to one of the known input functions.

Generalised linear models are a generalisation of the more classic linear model in the context of linear regression. While in the traditional linear model it is assumed that the endogenous variable is normally distributed, in the context of generalised linear models the endogenous variable does not have a predetermined distribution, so it can be distributed like any random variable.

The generalised linear models were formulated in order to standardise several other statistical models within a single model, including the linear model, logistic regression and poissonian regression. The result obtained is a volumetric Statistical Parametric Map (SPM) [26]. Importantly, spatial information is not exploited with the GLM. The standard methods of statistical analysis of the fMRI signal require high accuracy of the fMRI signal determined by the execution of the tasks provided in the experimental design. These methods are based on the execution of statistical tests that classify a voxel as active if its signal has a temporal trend corresponding to an ideal waveform that reflects the experimental design. But this hypothesis is a strong assumption and is often not reflected in reality. Data-driven analyses represent a complementary and / or alternative approach to the statistical analysis of the signal's temporal trend. Data-driven analyses are considered exploratory because the data structure is explored in search of activations that can be linked to the task performed.

One of the most used techniques in the data-driven analysis is clustering that allows identifying sets of voxels whose activity varies over time in a similar and distinctly different way from that of other sets sets of voxels whose activity varies over time in a similar way and in a distinctly different way from that of the other sets. Voxels belonging to each cluster need not be spatially close together. Analysis techniques through clustering are becoming the predominant method, given the ability to produce new information. Since clustering is an unsupervised learning technique, selecting the correct algorithm is a non-trivial problem, the choice of the model may depend on the context of the study and previous knowledge.

Given the growing interest in the study of fMRI neuroimaging through clustering in this thesis, several fMRI examinations have been analysed both with clustering techniques and with statistical techniques. For the analysis performed through clustering, different algorithms have been analysed, with different characteristics. A fundamental topic concerning clustering will now be introduced, the problem of clustering evaluation

1.3 Clustering Validation

Clustering is an unsupervised classification method that divides the input space into clusters. The goal of a clustering algorithm is to perform a partition in which the objects within a cluster are as similar as possible to each other and at the same time as dissimilar as possible from the elements present in the different clusters. This data-driven approach is suitable for identifying natural structures in a data set, and for this reason, it is widely used in many fields such as biology, model recognition, computer security, and image processing.

As clustering belongs to the unsupervised learning family, once the partitions have been obtained it is not possible to know whether the proposed partition is the most correct. This aspect is certainly not of secondary importance, indeed first of all, there is no universal clustering algorithm, in fact, the choice of algorithm is often taken based on the type of data and previous experience. This is because each clustering algorithm processes the partitions following different functions, and therefore two algorithms can create different partitions; in addition to this, rarely the clustering algorithms are able to independently determine the number of natural clusters in the data, which therefore must be fixed a priori, and this determines the division of clusters. Given that the choice of the number of clusters is not a simple task, the most used approach is to execute the clustering algorithm several times, changing the number of clusters to be created each time, the number of clusters is usually chosen through some heuristics. After the generation of the various clustering, the problem remains of identifying which partitioning is best suited to the data, to overcome this problem, techniques for the evaluation of data partitioning have been introduced, these techniques are called cluster validation.

Cluster validation is a very complex branch, still under development, which pays for the lack of a strong theoretical background which does not happen in other fields. In clustering validation, we can distinguish three main strands, internal validation, external validation, relative validation.

External validation: validate a partition by comparing it with the correct partition; therefore it is applied in experiments for which the correct partition of the data is known, or in which it is possible to know it a posteriori through experts. Internal validation: the techniques related to internal validation, are not based on the knowledge of the truth, but they evaluate the characteristics of the partitions, they are the most used techniques because typically clustering is used in situations where there is no previous knowledge. The features that are usually analysed by internal validation techniques are the "Compactness", i.e. how close/similar the elements within a cluster are, and the "Separation" that is how distant/dissimilar are the elements present in different clusters. Relative validation: the relative validation is the most recent approach, in this case, the validation is based on the stability. Validation takes place by comparing the partition with the partitions obtained, always on the same dataset, by other clustering algorithms or by the same algorithm, but also using different values for the internal parameters.

It is very difficult to make a direct comparison between the techniques mentioned above. In this thesis, we have dedicated ourselves to the analysis of clustering through internal evaluations, as they reflect more the real cases, this analysis has been performed through clustering evaluation index (CVI). In the literature it has been shown that as for the clustering algorithms, there is no optimal CVI compared to others, some perform better than others for some types of data, and therefore there is a need to have guidelines on which CVI to use for some data fields. Few papers provide suggestions on which CVI is optimal for a specific search field. Most of the research work on CVIs concerns the proposal of a new CVI, which is compared with others. Often these comparisons are made on benchmark dataset or synthetic dataset created expressly for comparison.

This thesis will analyze different CVIs using fMRI image data, given the lack of specific studies for this field in the literature, and a method to evaluate the different CVIs will be proposed. In most of the works in the literature, the evaluation of a CVI is performed using a data set of which the correct partitioning is known, through a clustering algorithm several partitions are created in which the number of clusters is varied. Once executed, the partitions are evaluated by the CVI and if the latter evaluates as better partitioning the one with a number of partitions corresponding to the real one, the CVI is positively evaluated. With the aim of overcoming this methodology that rewards only excellent results, we have proposed a methodology that evaluates not only the excellent performance but also the sub-optimal ones obtained by CVIs.

2
Background

In this chapter we will first explain how an fMRI exam is carried out, and what are the properties that allow the measurement of brain activities and their transformation into images. In this way it is possible to understand the complexity of the data, and consequently the inherent difficulty in analyzing this type of data. The concepts of machine learning will then be exposed to better understand the choices made for the analysis, the potential of the algorithms used, and the use of techniques for evaluating the results.

2.1 FMRI

Functional magnetic resonance imaging (fMRI) [12] is a non-invasive and in vivo technique for studying brain activity. This technique, through the help of a magnetic field and thanks to the characteristics of the cerebral blood flow, manages to provide images of brain activities. FMRI is a technique that, given its relevance, has involved various interdisciplinary teams of researchers in neuroscience, medicine, physics and computer science, in an attempt to address several crucial aspects during the acquisition, analysis and interpretation of the data.

Before proceeding with a more in-depth exposure on fMRI, it is good to explain the two fundamental phenomena on which the fMRI is based on, the BOLD signal and the hemodynamic response [4].

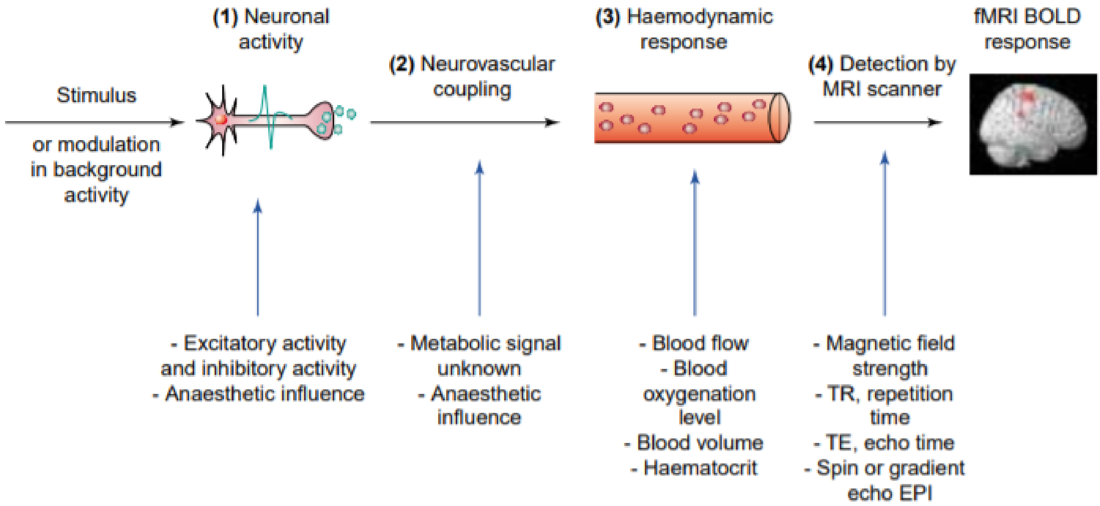

Figure 2.1: The BOLD signal consists of several parts: (1) the neuronal response to an underlying stimulus or modulation; (2) the relationship between neuronal activity and the starter of a hemodynamic response; (3) the hemodynamic response itself; and (4) how this response is detected by an MRI scanner.[4]

2.1.1 Bold Signal

Changes in the neural activity of the brain are associated with a change in energy needs: the greater the functional activity of a part of the brain, the higher its metabolism and, consequently, its energy demand. The increase in neuronal electrical activity leads to a higher demand for energy by neurons and consequently of oxygen. To satisfy this need for oxygen, the blood flow must increase, and it does so thanks to the dilation of the capillaries and/or the increase in the speed of the blood. However, the increase in local blood flow occurs to a greater extent than the increase in oxygen consumption; the result of this decoupling is an increase in the concentration of oxyhemoglobin (oxygenated hemoglobin) with reduction of the deoxyhemoglobin (free or reduced oxygen hemoglobin) on the venous side of the capillary bed. Hemoglobin has different magnetic properties depending on whether it is linked to oxygen or not. The oxygenated hemoglobin is diamagnetic, i.e. it has no unpaired electrons and has zero magnetic moments, while the deoxygenated hemoglobin is paramagnetic; therefore, it has unpaired electrons and non-zero magnetic moment. Fully deoxygenated blood has a magnetic susceptibility of 20% higher than fully oxygenated blood. Functional magnetic resonance imaging exploits this magnetic property of hemoglobin, which is used as an endogenous contrast medium. The introduction of material with magnetic susceptibility within a magnetic field causes the phase shift of the nuclear spins, creating decay of the transverse magnetization, linked to the temporal variable T2 * (transverse relaxation time and is a measure of the time taken by the spin protons to de-synchronize). The deoxygenation of the blood, by changing the magnetic susceptibility, causes the MR (magnetic resonance imaging) signal to vary, and the measured value of T2 *, which increases as the deoxygenation decreases. The amplitude of this effect increases with the square root of the intensity of the static magnetic field. In fact, the relative decrease in the concentration of deoxyhemoglobin in the regions where there is neuronal activity causes an increase in the T2 * parameter which in turn leads to an increase in the intensity of the image. The increase in blood flow following neural activity causes an incoming flow of blood into the venous system greater than the outgoing one, causing an increase in blood volume. In the small veins, the initial increase in volume is characterised by the presence of deoxygenated hemoglobin, which is eliminated first by the capillaries. This increase causes a loss of the MR signal, the initial dip. The increase in the acquired signal is due to a disproportionate increase in blood flow compared to actual needs, which leads to a local increase in the concentration of oxygenated hemoglobin and a relative decrease in the concentration of deoxygenated hemoglobin. The subsequent undershoot is caused by an imbalance between the rates at which flow and blood volume return to baseline.

2.1.2 Hemodynamic Response

The change in the measured MR signal, triggered by neuronal activity, is called hemodynamic response (HDR). The form of the hemodynamic response varies over time, from region to region and is closely connected to the stimulus that generated it, persistent stimuli will increase its amplitude, while more prolonged duration stimuli will prolong its duration. Since the changes in the BOLD signal are dominated by hemodynamics, and not directly by the activity of neurons, the increase in the signal following a neural activation is slow and protracted over time. It is essential to consider that the activity of each individual neuron is almost impulsive (temporal duration in the order of milliseconds), while the duration of the hemodynamic response, thanks to the cascade of events that occur during an activation, is about 10 second. Determining the exact relationship between neuronal events and the shape of the hemodynamic response is, therefore, very difficult since the dynamics that characterise them are different. After an initial latency, metabolic demands following neuronal activity increase above basal levels, resulting in an increase in the flow of oxygenated blood. If the supply of oxygenated blood in the area exceeds its extraction, there is a decrease in the oxygenated hemoglobin inside the voxel. About 2 seconds after the beginning of the nervous activity, the acquired signal increases above the basal value, this occurs in correspondence with the increase in blood flow beyond the real needs with consequent increase in the ratio between oxygenated and non-oxygenated hemoglobin. The signal reaches a peak value after about 5 seconds and, if neuronal activity persists over time, it stops at a constant value forming a plateau. At the end of the neuronal activity, there is a post-stimulus undershoot due to the fact that the blood flow decreases much faster than the blood volume causing a decrease in the amplitude of the fMRI signal up to a value lower than that of the baseline. As the blood volume returns to normal levels, the signal reaches the baseline. It is important to underline that the BOLD signal relative to a voxel reflects the total amount of deoxygenated hemoglobin present. Still, it is also influenced by the noise coming from different sources; this noise often reaches amplitudes of the same order of magnitude as the measured signal. It should also be remembered that the MR signal derives almost entirely from the hydrogen protons contained in the tissues, disturbed by the presence of the paramagnetic contrast medium, deoxyhemoglobin.

2.1.3 FMRI Exam

During an fMRI exam, the Blood-Oxygen-Level Dependent (BOLD) signal is measured, relating to the hemodynamic changes that occur after greater neural activity, this activity is measured and generates a large amount of data, unfortunately, this data contains noise. These data are structured as magnetic resonance imaging (MRI) sequences, each consisting of several evenly distributed voxels, each voxel is associated with a portion of the brain, the portions are of equal size to each other like voxels.

FMRI Data Acquisition

The data collected during an fMRI experiment consist of a sequence of individual magnetic resonance images; these are acquired to study oxygenation patterns in the brain. Therefore, to understand the nature of the fMRI data and how these images are used to infer neuronal activity, the acquisition of single MR images must first be studied. To build an image, the subject is placed inside by the fMRI machine which has a large electromagnet inside. The magnet has a powerful magnetic field, typically between 1.5-7.0 Tesla (1 Tesla = 10,000 Gauss, terrestrial magnetic field = 0.5 Gauss), which allows aligning hydrogen (H) atoms in the brain. A hydrogen atom is formed by a proton and an electron that rotates around it. Typically their position is random, but in the presence of a strong magnetic field, they align and absorb energy. In the moment which magnetic signal is interrupted, the protons return to their original position, dispersing all the stored energy. The latter is recorded by special reels placed inside the machine. With this recording, you get the basic MR signal. Through the particular coils placed inside the machine, the inhomogeneity of the magnetic field is checked, so that each measurement can be represented as the spin density in a single point of the frequency domain (k-space). To reconstruct a single fMRI image, it is necessary to sample a large amount of individual measurements of space k; the exact number depends on the resolution of the desired image. For example, to reconstruct a 64×64 images, 4096 measurements are needed, each sample in a coordinate different from the others in the k-space. From this, we understand that the time needed to acquire an image is directly related to its spatial resolution. Changes in brain hemodynamics in response to neuronal activity affect the local intensity of the MR signal. Therefore, a sequence of properly acquired brain images allows you to study changes in brain function over time. An fMRI study, therefore, consists of a series of brain volumes collected in rapid succession. The temporal resolution of the acquired data will depend on the time between the acquisitions of each individual volume. From this, we deduce that the intensity of each voxel over time is linked to changes in cerebral hemodynamics, in reaction to neuronal activity and can, therefore, be used to locate where the neural activity takes place. There are two main classes of fMRI design experiments:

- Task Based
- Resting State

Task Based A task-based fMRI exam allows you to detect the functional areas of the brain that are involved in motor or cognitive functions; this is usually done by asking patients to perform different tasks during an fMRI exam. All patients must perform several tests synchronised with the acquisition of images. There are two main classes of fMRI design experiments:

- Block Designs
- Event Related

Block Designs In a block design, the different experimental conditions are separated in time intervals called blocks. Usually, such an experiment is carried out in the following way, proceeding with the task of interest (for example, touching the finger) during an experimental block (A), and going to rest during a control block (B). The A-B comparison can be used to compare the signal differences between the two conditions. In general, increasing the length of each block will result in a larger response during the activity. This increases the separation of the signal between the blocks, which in turn leads to higher detection power. Examples of block design tasks are:

- Object Naming: The patient must think of the name of the object, whose image is presented
- Finger Tapping DX: The patient must repeatedly touch the thumb of his right hand with the index finger of the same hand
- Finger Tapping SX: The patient must repeatedly touch the thumb of his left hand with the index finger of the same hand

- Phonemic Fluency: The patient must think of all the words, which begin with a letter that is presented to him
- Generation of Verbs: The patient must think of the verb that expresses the action represented in a drawing

Event Related In event-related designs, the stimulus consists of short events (e.g. quick flashes of light) whose timing can be randomised, unlike the previous method, the rest time is short and can be of different length. Precisely because of this characteristic, it is little used because, given the nature of the measurement that takes place through the hemodynamic response, these random and close intervals lead to very nuanced and difficult to interpret results.

Resting State The brain is always active, even in the absence of a specific task, this is called resting state (RS). In recent magnetic resonance studies, it has been observed that brain activity at rest is not random noise, but reflects a functional organization of the brain, and the functional constituent networks are organised in a very coherent way. This has created a new avenue for neuroimaging research. To perform a resting fMRI exam, the patient only needs to sit still without performing any activity.

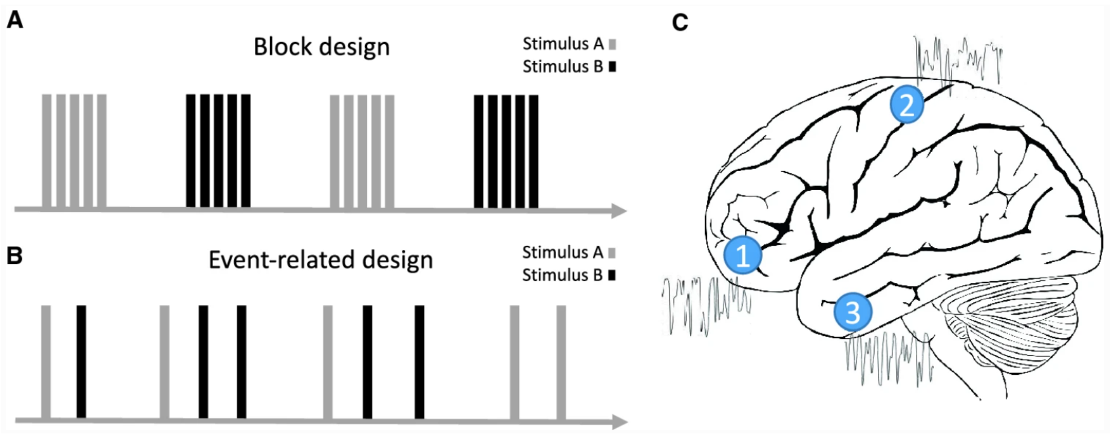

Figure 2.2: A represents a Block-design task-based protocol with two different tasks, the task and rest time are equal. B a represents Event-related task-based protocol with two tasks, the task has a limited duration and the rest time is random. C represents a Resting-state protocol, to analyze the cerebral activities at rest.[67]

2.1.4 Preprocessing FMRI Data

After the acquisition, the fMRI data is generally subjected to a series of preprocessing steps aimed at removing the artefacts and validating the model hypotheses. The main objectives are to minimise the influence of data acquisition and physiological artefacts, to standardise the positions of brain regions across subjects to obtain greater validity and sensitivity in the analysis. The main phases of fMRI preprocessing are:

- Slice Timing Correction
- Motion Correction
- Coregistration and Normalisation
- Spatial Smoothing

Slice Timing Correction When analysing three-dimensional fMRI data, it is generally assumed that the whole brain is measured simultaneously. In fact, since the volume of the brain is made up of several slices that are sampled in sequence and therefore at different times, similar temporal trends from different slices will be temporarily shifted relative to each other. Suppose that three voxels contained in three adjacent slices have the same temporal profile. Due to the fact that they are sampled at different time points, the corresponding measured time periods appear different. The slice timing correction involves moving the trend of each voxel so that it can be assumed that they are measured simultaneously. This can be done using Fourier theorem interpolation to correct differences in acquisition times.

Motion Correction An important issue involving the fMRI study is the correct management of any subject movement that could have taken place during data acquisition. Even small amounts of head movement during an experiment can be a major source of error if not treated properly. When movement occurs, the signal from a particular voxel will be contaminated by the signal from nearby voxels and the resulting data can be rendered useless. Therefore, it is of great importance to accurately evaluate the amount of movement and to use this information to correct the images. The first step in correcting movement is to find the best possible alignment between the input image and some target image (for example, the first image or the average image). A rigid body transformation involving six variables is used as parameters. This allows the translation of the input image (moved in the x, y and z directions) and its rotation (altered roll, pitch and yaw) to match the destination image. Usually, the matching process is performed by minimising a cost function (e.g. sums of square differences) that assesses the similarity between the two images. Once the parameters that achieve optimal realignment are determined, the image is resampled using interpolation to create the correct new voxel values. This procedure is repeated for every single volume of the brain.

Coregistration and Normalisation Functional MRI data in comparison with other techniques have a better spatial resolution, but in absolute values they are typically of low spatial resolution and provide relatively little anatomical detail. Therefore, it is common to map the results obtained from functional data onto a high-resolution structural MR image for presentation purposes. The process of aligning structural and functional images, called coregistration, is typically performed using a rigid body (using 6 parameters) or an affine transformation (using 12 parameters). For the analysis of multiple subjects in a group, it is important that each voxel is located within the same brain structure for each subject. Of course, individual brains have different shapes and characteristics, but there are regularities shared by each non-pathological brain. Normalisation attempts to record the anatomy of each subject in a standardised stereotaxic space defined by a template. In this context, a rigid transformation of the body is inappropriate due to the intrinsic differences in the brains of the subjects. On the contrary, it is common to use nonlinear transformations to match local characteristics. Subsequently, mapping is used to resample the input image so that it fits the target image. The main advantages of data normalisation are that spatial positions can be reported and interpreted consistently, the results can be generalised in a larger population, and the results can be compared with other studies.

Spatial Smoothing The spatial levelling process of an image is equivalent to applying a low pass filter First, smoothing can improve interindividual registration and overcome differences in spatial normalisation by blurring any residual anatomical differences. Second, the random field theory (RFT) assumptions, used regularly for multiple comparisons, are helpful. After these steps, the images can be processed for analysis.

2.2 Machine Learning

Machine learning is a sub-branch of artificial intelligence. The term Machine Learning (ML) was coined for the first time in 1959 by Arthur Samuel, and Tom Mitchell gave a formal definition: "A program is said to learn from a certain experience AND compared to a class of T tasks obtaining a P performance, if his performance in carrying out the T tasks, measured by the P performance, improves with the E experience". In short words, he says that if a machine improves the performance of a task thanks to past experience, it has learned. Unlike traditional algorithms where the programmer defines the parameters and data necessary for solving the activity, in the ML a specific model is not created to solve the problem, but it will be the model itself thanks to the training that adapts to the problem. The goal of ML is to extract useful knowledge from the data, so that it can support decision making. The algorithms used in the ML field, therefore, build a generalised model capable of making predictions and finding patterns in the dataset. ML techniques are applied in various sectors, in economics, biology, medicine, in personal life, virtual personal assistants, product recommendations, and autonomous driving are of great interest. ML learning processes can be classified into two macro-categories.

- Supervised learning
- Unsupervised learning
- Reinforcement learning

2.2.1 Supervised learning

The supervised definition comes from the fact that behind to the models that learn there is a "teacher" who has the task of rewarding or punishing the model on the basis of the results obtained in the learning phase. Supervised learning problems can be divided into problems of:

- Classification
- Regression

Classification This type of learning is based on the possibility of knowing the solution or truth of the problem to be solved. In supervised learning, data has a label that identifies its class to which it belongs. The dataset is generally divided into two sub-datasets: training and set. The training dataset is used for the model learning phase, once the training is finished the test one will be used to verify that the network is able to provide correct answers to unknown data, thus demonstrating generalisation skills. Sometimes the model can learn by heart, that is it has specialised too much on the data seen in training and therefore does not have the right degree of generalisation to evaluate new data correctly, some techniques are used in the training phase to overcome to this problem. The most used algorithms for classification analysis are logistic regression, the nearest neighbour, the supporting vector machines, the decision trees, the random forest and the neural networks.

Regression Regression models are used to identify a functional relationship between the variables that make up the dataset. The most used algorithms for regression analysis are linear and multilinear regression, support vector, random forest regressor, and neural networks [31, 13, 18].

2.2.2 Unsupervised learning

Learning without a supervisor is mainly used in contexts where the solution to the problem is not known. Contrary to supervised learning, it does not have the role of the “teacher” given the lack of truth; it would be impossible to reward or punish the machine. In this learning, the model has the mood to find relationships or patterns between the various data analyzed independently. Unsupervised learning can also be divided into two macro-categories

- Clustering
- Association rules

Clustering

Clustering algorithms are used to identify groups of homogeneous elements within a data set. The clusters contain highly homogeneous elements with each other and with a high degree of heterogeneity concerning the elements present in the other clusters. Clustering will be explained in detail in Chapter 2.4

Association rules

The goal of association rules [86] is to identify models and patterns that recur between combinations of objects. For example, they are used to determine suggested products based on those already purchased (the most famous is the beer-diaper scheme).

2.2.3 Reinforcement learning

Reinforcement learning [70] is a technique for the training of automatic learning models that aims to create models which, in order to achieve the objective, can perform actions based on the environment in which they are immersed. In reinforcement learning, artificial intelligence faces a situation similar to a game and learns through trial and error, in fact, the model gets rewards or penalties for the actions it performs, the ultimate goal of the model is to maximise the total reward. The intervention of the designer is limited to the choice of the reward policy; the model must independently understand how to perform the activity to maximise the reward. In the beginning, the model will start with totally random attempts, but as the training progresses, the choices made by the model will become increasingly more efficient.

2.3 Neural Networks

A neural network (NN) is a computational model that draws inspiration from the neural connections of the human brain[47]. Inspired by the human brain, this model is composed of neurons (also called nodes) that make up the processing units, and the connections that connect neurons to each other. Another characteristic that makes neural networks similar to a biological brain is that the NN is an adaptive system, it continuously changes its structure during the learning phase, based on information external or internal to the NN. A NN receives external signals through a layer of neurons called input neurons, each of the input neurons connected to numerous internal nodes, organised on several levels. Neurons can only perform simple tasks, typically they are activated when a certain threshold of the signal received at the input is exceeded. If a unit is activated, the signal will be carried to the other nodes connected to it through the connections. Based on the characteristics of the connections they can modify the signal, increasing or decreasing it, these characteristics are called weights. Based on the connection structure between neurons, neural networks can be divided into two categories which are:

- Feed-forward network.
- Recurrent (or feedback) network, in which there are feedback loops.

The most common type of feed-forward networks is the multilayer perceptron (MLP) [61]; the neurons are organised in layers that have unidirectional connections. Feed-forward networks [40] are “out of memory”, this means that the output returned by input not dependent on the previous states of the network. On the contrary, in recurring networks through a feedback loop, the output of one or more nodes will be presented to the network as new input, this will bring the network into a new state, this behaviour leads to a “memory” effect in which the output will depend on previous network states.

2.4 Clustering

Clustering is a process of grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of "similar" objects among them that are "dissimilar" to other cluster objects [37]. The clustering algorithms partition the elements based on their mutual distance, therefore whether or not to belong to a cluster depends on the distance of the element from the cluster. To carry out a clustering it is necessary to decide a priori some rules:

- What characteristics will be the basis of the grouping
- The choice of an adequate measure of the similarity between the objects
- The clustering algorithm to be used
- How to validate the results

Variable choice The choice of the variables to be clustered depends very much on what you want to achieve as a result. It is necessary to choose the most significant variables for identifying clusters, this phase is also called features selection. The features selection involves the reduction of the features, as only the selected features are used. The choice is made by searching for the features that best express the similarity between the objects belonging to the same cluster.

Similarity measure The measure of similarity is the technique with which the similarity between the objects of the set under consideration is assessed, this occurs through the "distance" that exists between the objects. The more two objects are similar, the more their distance tends to zero. There are various techniques for calculating the distance, some distance examples are Euclidean distance, Manhattan, Minkowski's distance, cosine similarity, and Pearson's correlation.

Clustering algorithms A clustering algorithm arises from the combination between the choice of the proximity/distance measure, and choose the criterion function with which to group the data, thus defining how to measure the level of similarity between the units. This implicitly decides the "types of groups" to research in the data [83]. A criterion function is intended to achieve the "best" grouping of data based on the data structure itself. A large number of clustering algorithms have been developed from various combinations of proximity measurement and criterion function. Although there are many types of clustering algorithms, there is no algorithm that can universally solve all problems. To choose the appropriate type of clustering it is important to carefully examine the characteristics of the problem, taking into account the hypotheses implicit in the method, the shape of the clusters, the partitioning structure, the hypotheses made based on the proximity/distance measurements and the criterion function. Clustering algorithms can be divided into hierarchical or partitioned, excluded and non-exclusive, total or partial [83].

- Hierarchical: Clusters are nested, and organised within a hierarchical tree
- Partitioning: objects are divided into clusters without any hierarchy. They are divided into two categories:
  - Exclusive: a datum always and only belongs to a cluster.
  - Non-exclusive: In non-exclusive clustering, data can belong to multiple clusters.
- Totals: Clustering is done on all data.
- Partial: Clustering takes place only on a specific portion of the data.

The most famous algorithm is also among the simplest clustering algorithms and is the K-means [42]. K-means is a partitioning algorithm that aims to minimise total intra-cluster variance. In K-means, each cluster is identified by a centroid or midpoint. The algorithm follows an iterative procedure, initially creating K partitions and assigning a centroid to each partition can be done randomly or using some heuristic information. After assigning the centroids for each cluster, the distances of each point from all the centroids are calculated, and the point is assigned to the nearest centroid. Then the centroids are recalculated, and are recalculated as the average of the values belonging to the cluster. This cycle is repeated until the algorithm reaches the convergence, the convergence occurs when, after calculating the new centroids, no point changes its cluster. Other clustering algorithms will be exposed to the development of this thesis.

Validation Validation is a fundamental aspect of the Cluster Analysis, because given a dataset, a clustering algorithm generates a subdivision of it, but generates a partition also when the data do not have any natural grouping, that is, it can create groups that do not represent anything. It follows that the success of the Cluster Analysis applications depends entirely on knowing whether the grouping model performed corresponds to a real structure or not. Furthermore, different approaches often lead to different solutions; and also for the same algorithm, the choice of input parameters or even the order of the units in the dataset can influence the final results. The use of valid evaluation criteria is important to provide the user with results with a certain degree of reliability. This validation should be objective and independent of the clustering algorithm chosen. Also, they must be useful for identifying the number of clusters present in the data, for assessing whether the clusters obtained are significant or are only an artefact of the algorithms, or for deciding which algorithms to use. There are three approaches to verify the validity of clustering [83] :

- external validation criteria, which measure how much the clusters identified correspond to class labels provided externally (previous knowledge).
- internal validation criteria, which measure how well a Clustering solution fits data. Data is the only information available.
- relative validation criteria, which measure the goodness of a Clustering solution by comparing it with the results obtained by other Clustering algorithms, or by the same algorithm, but using different parameter values.

2.4.1 Uses of Clustering

Clustering is used in many industries as an example in marketing for the division of customers into homogeneous groups, which is one of the fundamental strategies. Clustering analysis can also be used to classify the city based on the possibility that a service may be profitable. Clustering is also used in seismic studies where groups of epicentres of seismic events are formed, and checks are made to see if they are located along the faults of the continents. Other fields in which clustering is used are astronomy, biology, archaeology and many other sectors, besides, of course, the IT sectors where it is widely used [65]. In computer science, clustering is used for pattern recognition, spatial data analysis, image processing, on the web for document classification, and also for Weblog data analysis to discover groups characterised by common profiles of access and navigation and in many other areas. Clustering was adopted for the study of fMRI data, offering excellent performance, especially for the classification of brain areas, managing to manage the data, the number of voxels and the adequacy to the nature of the data. For these reasons, it is receiving a lot of consensuses in functional imaging field.

2.4.2 Clustering algorithms

In this thesis the following clustering algorithms have been used:

- Self Organizing Maps (SOM)
- Neural Gas (NG)
- Growing Neural Gas (GNG)
- Fuzzy C-Means (FCM)

The detailed description of the four algorithms is defined below.

Self Organizing Maps

Kohonen introduced Self-organized maps (SOM) [47] in the 1980s, and they have become very popular for their characteristic of respecting the spatial ordering while preserving the space of the input data which is reflected by order of the prototype vectors (cluster centroids). In SOMs, the data is mapped onto an n-dimensional grid of neurons, and this grid forms the output space. The neurons in the output layer are generally arranged in a one-dimensional or two-dimensional lattice and are all connected to each other; in most cases a rectangular or hexagonal distribution is chosen. The structure of the grid and the number of neurons present in it is chosen before the phase of education of the network. The neurons of the network can be selected randomly from the input data or be generated randomly with the same dimensionality as the input data. During the competitive learning process, at the network is presented with a random vector belonging to the input data; this will be compared with all the neurons of the network that will compete with each other. The winning neuron called Best Matching Unit (BMU), will be the one that, based on a discriminated function, will be the most similar to the input vector, so the BMU and the neighboring nodes will modify their weights through a fitting rule that will move them closer in space to the input vector. In the SOM the models that are close in the input space will be mapped into neurons that are close (or equal) in the output space. Thanks to this feature the SOMs are defined as "topology conservation", i.e. the surroundings are preserved through the mapping process, the characteristics of the input domain are also reported in the output domain, and since this happens without any prior knowledge about the input distribution, they have received the name self-organizing map. In SOM, three phases are distinguished:

- Competition
- Cooperation
- Synaptic Adaptation

Competition In the competition phase, an input vector is passed to the network, and the neurons in the network will begin to compete with each other by calculating their similarity values through a discriminant function. Based on this discriminating function that competition between neurons takes place. The neuron with the best value is declared the winner of the competition. Let $x = [x_1, x_2, ..., x_d]$ be a randomly selected input vector from the input space, where d represents the size of the input space. A prototype neuron j of the SOM output space can be defined as:

$$w_j = [w_{j,1}, w_{j,2}, ..., w_{j,d}]^T, \quad j = 1, 2, ..., C \tag{2.1}$$

Where C is the total number of neurons in the output layer, in other words, the number of the cluster. The similarity between the output neurons and the input vector can be calculated through a distance, and the neuron j that minimises this distance is selected as the winning neuron (BMU) and can be defined as:

$$BMU = min||x_t - w_j t|| \quad 1 \leq j \leq C \tag{2.2}$$

Where t represents discrete time coordinate.

Cooperation In the cooperative process, a topological neighbourhood is defined in which the winning neuron is the center and the cooperating neurons are the neighbours. The winning neuron will excite the neighbouring neurons more while the distant ones will have a lower excitation value. Let $h_{j,i}$ denote the topological neighbourhood centered on winning neuron i and $d_{j,i}$ denote the lateral distance between the winning neuron i and the excited neuron j. The topological neighbourhood $h_{j,i}$ can be a unimodal function of the lateral distance $d_{j,i}$ satisfying the following two distinct requirement:

- $h_{j,i}$ is symmetric about the maximum point defined by $d_{j,i} = 0$; in other words, it attains its maximum value at the winning neuron i for which the distance $d_{j,i}$ is zero.
- The amplitude of $h_{j,i}$ decreases monotonically with increasing lateral distance $d_{j,i}$, and decays to zero as $d_{j,i} \to \infty$.

Synaptic Adaptation In this part, the excited neurons (both BMU and its surroundings) change their internal weights through a discriminate function that arranges them through appropriate adjustments, in relation to the input pattern. The adjustments lead to better the response of the winning neuron to the subsequent application of a similar input model. The discriminating function is formalized as follows:

$$w_j(t+1) = w_j(t) + \varphi(t) h_{j,i}(t)(x(t) - w_j(t)) \tag{2.3}$$

Which is applied to all the neurons in the lattice that lie inside the topological neighbourhood of the winning neuron i,and where the parameter $\varphi(t)$ represents the learning-rate which is a value that gradually decreases with increasing time.

Algorithm Step Before training the network, you need to set some parameters such as:

- Number of neurons for the output layer.
- Select neurons from the input dataset or generate them randomly.

- Define the structure of the lattice and the size of the neighbourhood.
- Define the value of the learning-rate value, and of the maximum number of iterations.

Once the parameters have been set, we start training the network as follows:

1. Randomly select a vector x from input dataset.
2. Find the BMU by using the minimum-distance (see formula (2.2)).
3. Adjust the synaptic-weight vectors of all excited neurons by using the update formula (2.3).
4. Repeat from the beginning until no noticeable changes in synaptic weight output vectors, or maximum number of iterations is achieved.

Neural Gas

The Neural-Gas (NG) [58] learning algorithm is a variant of the Kohonen algorithm and has the advantage of converging more quickly. It was presented in 1991 by Thomas M. Martinetz. Unlike SOM, the neurons of the output layer are not organised on a two-dimensional grid, since the algorithm does not maintain any topological relationship, but is a standard network of two linear layers. As in the SOM, NG uses a soft-max adaptation for the weights, i.e.the weights of the non-winning neurons are also updated, but, unlike the SOMs, in NG a list is created with the neurons ordered according to the proximity of the winning neuron, and the weights are updated according to this list. The name Neural-Gas derives from the dynamics of the prototype vectors during the adaptation process, which distribute themselves like a gas within the data space. In NG as in SOM, it is necessary to choose a priori the number of neurons, therefore of clusters, to be used. NG uses parameters which are:

- ε which represents the size of the adaptation step.
- λ which represents the proximity range, i.e. the number of neural units that will be involved in updating the synaptic weights at each step of adaptation.
- T, which defines the maximum age of a connection between neurons.
- k_i is the classification index of the node in the ordered list of neurons (the node in the first position has the value 0).

Both ε and λ decrease as the iterations increase, λ, therefore, brings strong excitation to all neurons when the first input samples are presented. The resulting graph at the end of the training procedure represents the similarity, hence the neighbourhood relationship between the input data. Covering the data space with minimal representation error.

Algorithm Step Before training the network you need to set some parameters such as:

- Number of neurons for the output layer.
- Select neurons from the input dataset or generate them randomly, but consistent with the input dataset.
- Define ε, λ, T.
- Define maximum number of iterations.

Once the parameters have been set, start training the network as follows:

1. Randomly draw a vector x from the input dataset.
2. Calculate the similarity of each neuron with the input vector using the formula (2.2).
3. The ordered list of neurons is created based on their similarity from the input vector, each neuron is assigned its own k_i value;
4. Adjust the synaptic-weight vectors of all excited neurons by using the update formula:
$$w_j(t+1) = w_j(t) + \varepsilon e^{\frac{-k_i}{\lambda}}(x(t) - w_j(t)) \tag{2.4}$$
5. Initialise or update to 0 the connection between the first and the second winning neurons.
6. Increase the age of all the other connections in the neighbourhood of the winning neurons.
7. Remove the lateral connections between the winning neuron and the neurons in its neighbourhood with age greater than the value of T.
8. Repeat from the beginning until the number of iterations is achieved.

Growing Neural Gas

The Growing Neural Gas (GNG) [28] algorithm proposed by Fritzke manages to overcome the NG problem of having to choose in advance the number of neurons to use. GNG has the ability to add and remove nodes based on implementation choices. The GNG algorithm starts with two neurons in random positions in the input space; new neurons are periodically added by evaluating the local statistical measurements collected during the previous adaptation phases. Each γ input vectors, a new neuron is inserted between the neuron that has accumulated the highest local error and its neighbour with the highest error. Furthermore, GNG, in addition to adding nodes also eliminates the connections that appear to be below a predefined value, if a neuron remains without connections, it is removed. So the GNG continues to add and delete neurons leaving only those that are very active. The fundamental parameters of GNG are:

- n_{max} which represents the maximum number of neurons that the network can have.
- ε_b which represents the learning rate of the winning neuron.
- ε_n which represents the learning rate of the neighbourhood of winning neuron.
- T, the maximum age of the connection.
- γ, parameter that defines how often it is possible to include a new neuron in the network.
- Δ_e which represents the local error variable.
- α and d, constants that influence the decrease of the error variables.

The performance of the GNG algorithm is heavily influenced by these parameters, and their optimal setting turns out to be a very complex operation.

Algorithm Step Before training the network you need to set some parameters such as:

- Define n_{max}, ε_b, ε_n, γ, T, α, d.
- Define maximum number of iterations.
- Select two neurons from the input dataset or generate them randomly, but consistent with the input dataset.

Once the parameters have been set,start training the network as follows:

1. Randomly draw a vector x from input dataset.
2. Calculate the similarity of each neuron with the input vector using the formula (2.2).
3. The ordered list of neurons is created based on their similarity from the input vector, select the first and second winning neurons, called s_1, s_2.
4. Update the age of the connection between the two winning neurons to 0, and increase the age of all connections emanating from s_1.

5. Add the squared distance between the input signal and the nearest unit in input space to local error variable.

6. Move the weights of the winning neuron, and of its direct topological neighbours by using the update formulas:

$$w_{s_1}(t+1) = \varepsilon_b(x(t) - w_{s_1}(t)) \tag{2.5}$$

$$w_i(t+1) = \varepsilon_n(x(t) - w_i(t)) \quad \forall i \in s_1 \;\; neighbourhood \tag{2.6}$$

7. If the number of input signals is an integer multiple of parameter γ, then insert new node as follow:

   - Determine the neuron q as having the largest accumulated error and determine f, a direct neighbour farthest from q.
   - Interpolate a new unit r between q and f:

   $$w_r = \frac{(w_q + w_f)}{2} \tag{2.7}$$

   - Insert connections between r and q and between r and f. Also, remove the connections between q and f.
   - Decrease the error variable of q and f.
   - Update the connections among these three neurons and their error variables Δ_e(multiplying them with the α parameters).

8. Remove the lateral connections between the winning neuron and the neurons in its neighbourhood with age greater than the value of T.In case a neuron remains without connections, remove it.

9. Decrease all error variables Δ_e (multiplying them with the d parameter).

10. Repeat from the beginning until the number of iterations is achieved.

Fuzzy C-Means

The classic clustering techniques create partitions that have an N $\rightarrow$ 1 correspondence with data between patterns and partitions (more data goes into a single cluster). The fuzzy approach [6] modifies this concept in order to have an $N \rightarrow N$ correspondence by associating each data to each cluster through the use of a membership function [85]. So a fuzzy clustering algorithm will not provide crisp data segmentation's but will be a softer (fuzzier) segmentation. For this reason, data assignment in a cluster is described with a membership degree between 0 and 1. The unsupervised FCM algorithm is a technique based on the objective function. The algorithm, which is a generalisation of the least-squares method, works by minimising the following objective function.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ji}^m ||x_i - z_j||^2 \tag{2.8}$$

where:

- m is the fuzzy factor, and is any real number greater than 1, this parameter indicates the overlap of the generated clusters; if it tends to 1 the clusters will be slightly overlapped, the more it tends to infinity and the more the clusters overlap.
- K is the number of cluster decided by the user.
- N is the number of data contained in the input dataset.
- x_i is the i-th data belonging to the input dataset.
- z_j is the j-th centroid belonging to the clusters, a centroid is the prototype vector that represents all the data belonging to the same cluster.
- μ_{ji} is the degree of membership of x_i in the cluster j.

Algorithm Step Before running the algorithm, you need to choose the following parameters:

- Define K.
- Define m.
- Define ε, represents the minimum update value, below this value, the algorithm ends.
- Define maximum number of iterations.
- Initialise $U = [\mu_{ji}]$ matrix at 0, U matrix represents the fuzzy partition matrix, returned as a matrix with K rows and N columns. Element in position $U(j, i)$ indicates the degree of adhesion of the i-th element in cluster j. The sum of the column values is one.
- Random generation of K centroids.

Once the parameters have been set,the algorithm proceeds as follows:

1. Calculation of centroids which is performed with the following formula:

$$z_j = \frac{\sum_{i=1}^{N} \mu_{ji}^m x_i}{\sum_{i=1}^{N} \mu_{ji}^m} \tag{2.9}$$

   in the first step this point is skipped since the centroids have just been created randomly.

2. U matrix update:

$$\mu_{ji} = \frac{1}{\sum_{k=1}^{K} \left(\frac{||x_i - z_j||^2}{||x_i - z_k||^2} \right)}^{\frac{2}{m-1}} \tag{2.10}$$

3. The J_m object function is calculated through the formula 2.8.

4. If the maximum number of iterations is reached or if $J_m(t) - J_m(t-1) \leq \varepsilon$ the algorithm ends otherwise it repeats from the first step.

2.5 CVIs

Clustering algorithms with different cost functions provide different solutions. No universal best choice exist. For most clustering algorithms, the number of clusters to use for dividing the dataset is not known a priori, and thus the latter is set as an experience-based parameter or on heuristic base. Once you have selected the correct clustering algorithm and the number of clusters, you need to tackle the evaluation of the clustering results. In conditions where there is no preliminary information on the data, the best evaluation method is internal indexes [15]; these indexes can be used to choose the best clustering algorithm and the optimal number of clusters, without the need for a priori information on the dataset.
Indexes that use internal validation are called the Cluster Evaluation Index (CVIs). Typically, CVIs measure Compactness, i.e. how similar the data are within a cluster, and Separation, i.e. how dissimilar the elements present in the different clusters are, precisely because they evaluate these values they are called internal evaluation measures.
Eight of the most famous or innovative CVIs were chosen for the analyses performed, while crisp and fuzzy CVIs were chosen. Among the CVIs chosen, one part maximises the evaluation function negatively (-), i.e. the more the returned value tends to 0, the better the evaluation, others maximise the evaluation function positively (+), i.e. the more the returned value tends to infinity the better the rating. The CVIs will now be presented in detail.

2.5.1 Fukuyama Sugeno

The index proposed by Fukuyama and Sugeno(FSI)[29] exploits cohesion and separation. The first term of the formula is a measure of compactness, while the second term is a degree of separation between each cluster and the mean of the centroids of the cluster($\overline{z}$). FSI is a fuzzy and negative (-) index:

$$FSI = \sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{m}||x_j - z_i||^2 - \sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{m}||z_i - \overline{z}||^2 \tag{2.11}$$

Where K is the number of clusters, N the number of elements present in the dataset, μ_{ij} is the value of the fuzzy membership matrix corresponding to the j-th element of the dataset in the i-th cluster, m is the fuzzyfication value used in the algorithm. Furthermore, x_j and z_i represent the j-th element of the dataset and the centroid of the i-th cluster, finally, $\overline{z}$ is a vector representing the mean of all K centroids.

2.5.2 Xie Beni

The Xie and Beni index (XBI) [82] is a negative (-) and fuzzy index and is also a ratio type index, which measures the average fuzzy compactness within the cluster compared to the minimum separation between clusters.

$$XBI = \frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{2}||x_j - z_i||^2}{N\left(min_{i\neq k}\{\|z_i - z_k\|^2\}\right)} \tag{2.12}$$

Where K is the number of clusters, N is the number of elements present in the dataset, μ_{ij} is the value of the fuzzy membership matrix corresponding to the j-th element of the dataset in the i-th cluster. Furthermore, x_j and z_i represent the j-th element of the dataset and the centroid of the i-th cluster,and z_k is a centroid different from z_i but always belonging to K.
The numerator of the equation, which represents the least squares error, indicates the compactness of the fuzzy partition, while the denominator indicates the separation between clusters. An optimal partition should have a low value for compactness and a high value for separation.

2.5.3 Compose Within and Between scattering Index

The Compose Within and Between scattering Index (CWBI) [66], alternatively called the RLR Index (RLRI) deriving from the name of its authors, is a negative (-) and fuzzy index that evaluates the average compactness and separation of fuzzy clustering using the sum of two functions.

$$CWBI = \alpha Scat(K) + Dis(K) \tag{2.13}$$

Where $Scat(K)$ represents the average scattering for the K clusters and is defined as:

$$Scat(K) = \frac{\frac{1}{K}\sum_{i=1}^{K}||sigma(z_i)||}{||\sigma(X)||} \tag{2.14}$$

where $||x|| = (x^T \cdot x)^{\frac{1}{2}}$ and $\sigma(X)$ represent the variance of the input data and is represented as:

$$\sigma(X) = \frac{1}{N}\sum_{j=1}^{N}(x_j - \overline{x})^2 \tag{2.15}$$

Where N is the number of elements present in the dataset K number of cluster, x_j represent the j-th element of the dataset, $\overline{x}$ is a vector representing the mean of all data input. Also $\sigma(z_i)$ indicates the fuzzy variation of the i-th cluster,which is defined as:

$$\sigma(z_i) = \frac{1}{N}\sum_{j=1}^{N}\mu ij(x_j - z_i)^2 \tag{2.16}$$

Where μ_{ij} is the value of the fuzzy membership matrix corresponding to the j-th element of the dataset in the i-th cluster, and z_i the centroid of the i-th cluster. The smaller the value of $Scat\,(K)$, the better the compactness of the clustering.
The distance function Dis(K) measuring the separation between clusters,and is defined as:

$$Dis(K) = \frac{D_{max}}{D_{min}}\sum_{i=1}^{K}(\sum_{k=1}^{K}\|z_i - z_k\|)^{-1} \quad i \neq k \tag{2.17}$$

Where $D_{max} = max(\|z_i - z_k\|)$ and $D_{min} = min(\|z_i - z_k\|)$.The smaller the value of $Dis(K)$, the better the separation of clusters. The parameter called α is a weighting factor equal to $Dis(K_{max})$. Where K_{max} in case of single evaluation corresponds to K, but in case of comparing different clusterings on the same input dataset, K_{max} represents the clustering whit the highest K value among the various clustering.

2.5.4 WSJ

Inspired by the CWBI, the WSJ index (WSJI) [69] is a negative (-) and fuzzy index, uses a linear combination of averaged fuzzy compactness and separation to measure clustering, which is defined as:

$$WSJI = Scat(k) + \frac{Sep(K)}{Sep(K_{max})} \tag{2.18}$$

Where $Scat(K)$ is given by equation (2.14), $Sep(K)$ denotes the between-cluster separation, which is defined as:

$$Sep(K) = \frac{D_{max}^2}{D_{min}^2} \sum_{i=1}^{K} (\sum_{k=1}^{K} \|z_i - z_k\|^2)^{-1} \quad i \neq k \tag{2.19}$$

Where $D_{max} = max(\|z_i - z_k\|)$ and $D_{min} = min(\|z_i - z_k\|)$. K_{max} in case of single evaluation corresponds to K, but in case of comparing different clusterings on the same input dataset, K_{max} represents the clustering whit the highest K value among the various clustering.

2.5.5 PBM

The PBM index (PBMI) [60] is named after its authors, is a positive (+) and crisp index, the index estimates the compactness within the cluster and the wide separation between fuzzy clustering clusters, and represented as follows:

$$PBM = \left(\frac{1}{K} \cdot \frac{E_1}{E_w} \cdot D_k \right)^2 \tag{2.20}$$

Where K is the number of clusters, E_1 is define as follows:

$$E_1 = \sum_{j=1}^{N} ||x_j - \overline{x}|| \tag{2.21}$$

Where N is the number of elements present in the input dataset, x_j represent the j-th element of the dataset, $\overline{x}$ is a vector representing the mean of all data input. D_k represents the max separation of clusters.

$$D_k = ||z_i - z_k|| \quad i \neq k \tag{2.22}$$

where z_i the centroid of the i-th cluster, and z_k is a centroid different from z_i but always belonging to K. E_w represents the compactness of the cluster.

$$E_w = \sum_{i=1}^{K} \sum_{j=1}^{N} ||x_j - z_i|| \tag{2.23}$$

2.5.6 FPBM

It is the fuzzy PBM (FPBMI)[60] version, and it is defined as follows:

$$PBM = \left(\frac{1}{K} \cdot \frac{E_1}{E_{fw}} \cdot D_k\right)^2 \tag{2.24}$$

Where E_1 is represented in the equation(2.21), D_k in the equation(2.22),K is the number of clusters, N the number of elements present in the input dataset, while E_{fw} is represented as follows:

$$E_f w = \sum_{i=1}^{K} \sum_{j=1}^{N} \mu ij ||x_j - z_i|| \tag{2.25}$$

Where μ_{ij} is the value of the fuzzy membership matrix corresponding to the j-th element of the dataset in the i-th cluster.

2.5.7 Davies Bouldin

The Davies-Bouldin Index (DBI) [14] is a negative (-) and crisp index, the index estimates the ratio of within-cluster compactness to between-cluster separation, which is defined as:

$$DBI = \frac{1}{K} \sum_{i=1}^{K} max(\frac{S_i + S_k}{||z_i - z_k||^2}) \quad i \neq k \tag{2.26}$$

Where K is the number of clusters, z_i the centroid of the i-th cluster, and z_k is a centroid different from z_i but always belonging to K, and S_i is represented as follows:

$$S_i = \frac{1}{C} \sum_{Kij=1}^{C} ||x_{Kij} - z_i||^2 \quad x_{Kij} \in K_i \tag{2.27}$$

Where C denotes the number of data points in the i-th cluster K_i, and x_{Kij} represents the j-th element in the i-th cluster K_i.

2.5.8 Soft Davies Bouldin

Soft Davies Bouldin index (SDBI) is the fuzzy version of DBI and is defined as:

$$SDBI = \frac{1}{K} \sum_{i=1}^{K} max(\frac{S_{Ui} + S_{Uk}}{||z_i - z_k||^2)} \quad i \neq k \tag{2.28}$$

Where K is the number of clusters, z_i the centroid of the i-th cluster,and z_k is a centroid different from z_i but always belonging to K, and S_{Ui} is represented as follows:

$$S_i = \frac{1}{C} \sum_{Kij=1}^{C} \mu_{ij}^m \cdot ||x_j - z_i||^2 \quad x_{Kij} \in K_i \tag{2.29}$$

Where C denotes the number of data points in the i-th cluster K_i, and x_{Kij} represents the j-th element in the i-th cluster K_i, and μ_{ij} represents the degree of belonging of x_{Kij} to the i-th cluster, m is the fuzzification coefficient.

2.6 ANFIS

In this chapter, we will explain the fundamental concepts on which ANFIS is based, its structure and the advantages and criticisms that this model entails.

2.6.1 Fuzzy Logic

Fuzzy logic [85] is a non-binary logic, the theorization of which was defined by Lotfi Zadeh in 1965. Binary (or classical) logic, deals only with what is completely true or completely false, while fuzzy logic extends its interest to what is not completely true and to what is probable or uncertain. Fuzzy logic is based on a linguistic approach, in which words or phrases in natural language are used instead of numbers. Thanks to his approach the fuzzy logic simplifies complex situations and concepts within traditional logic. Fuzzy logic uses fuzzy sets, which obey rules, structures and axioms quite similar to those of classical set theory, with the difference that an object can belong to several subsets at the same time. Membership in a subset, in fuzzy logic, is associated with a degree of membership. The set of inference rules that must be applied to a given system to obtain results through the use of fuzzy logic represents the fuzzy inference process.

2.6.2 Fuzzy Inference

Fuzzy inference [68, 55] is the mapping process starting from an input space and arriving at an appropriate output space using fuzzy rules. These rules are not represented by complicated mathematical models but by simple linguistic expressions, which are converted into mathematical formalism with the "if-then-else" language of fuzzy logic. Typically, the rules consist of two parts, "if" and "then", the antecedent (if) which defines the condition and the consequent (then) which represents the action. It is also possible to define an additional "else" rule which is used when none of the rules has been satisfied. Inference can be defined as the procedure through which the fuzzy output is determined starting from the antecedents. There are several models of fuzzy inference; the most used is Sugeno [68]. The Sugeno method has several sub-methods, but the most used ones are only of order 0 and order 1. In the Sugeno method of order 0, each rule results in a crisp value of the fixed output variable, in the Sugeno inference method of order 1, however, the value of the consequent is not a fixed value, but is a linear combination of the inputs. For example, if the inputs of the fuzzy system are P and T, the output for each rule is given by the following formula:

$$Z = a + bP + cT \tag{2.30}$$

where a, b and c are coefficients that vary from rule to rule.

2.6.3 NEURO-FUZZY Networks

Both neural networks and fuzzy algorithms solve problems of their domain without a mathematical model. The neural network consists of a black box, it is difficult to find logical rules to describe the final states, and the learning process which can take a long time and even come to converge, but is tolerant to changes in input and structure. The fuzzy method undergoes the problem of complexity in the definition of fuzzy rules as the number of parameters increases, but it is optimal as a basis for decision support. A neuro-fuzzy model has the characteristics to be able to fill the shortcomings of the fuzzy systems on the one hand and of the neural ones on the other. Fuzzy inference, unlike neural networks, does not acquire knowledge from input-output relationships but needs heuristic rules. The advantage of a neuro-fuzzy network over a neural one consists in the fact that the structure can be represented through linguistic rules. The nodes that make up a neuro-fuzzy network do not have weights, as in a system based on a neural network, and the training of the network takes place with back-propagation algorithms. ANFIS adaptive neuro-fuzzy inference models acquire knowledge from data using the operating algorithms typical of neural networks, representing them using fuzzy rules. Basically, these are neural networks structured on different levels which, starting from the correlated inputs and outputs, autonomously generate fuzzy rule systems that guide the output construction process. As in fuzzy logic, the final result is linked both to the fuzzy rules and to the membership functions. Membership functions can be of various types as for fuzzy logic.

2.6.4 ANFIS architecture

The neuro-fuzzy models use training data to acquire knowledge of the data to be analysed; the neuro-fuzzy algorithms allow to calibrate the membership functions of the fuzzy inference by training an NN. To carry out the training, it is necessary to define a matrix consisting of input parameters, an output value and a number of eras. The higher the number of epochs, the more detailed the output result will be. The ANFIS [43] models acquire knowledge from data through the operating algorithms typical of neural networks, representing them through fuzzy rules. The ANFIS structure can be divided into five different levels, which allow for the autonomous generation of fuzzy rule systems, based on the input data and the expected outputs.

A neuro-fuzzy model is characterised by a five-level network, which will now be detailed.

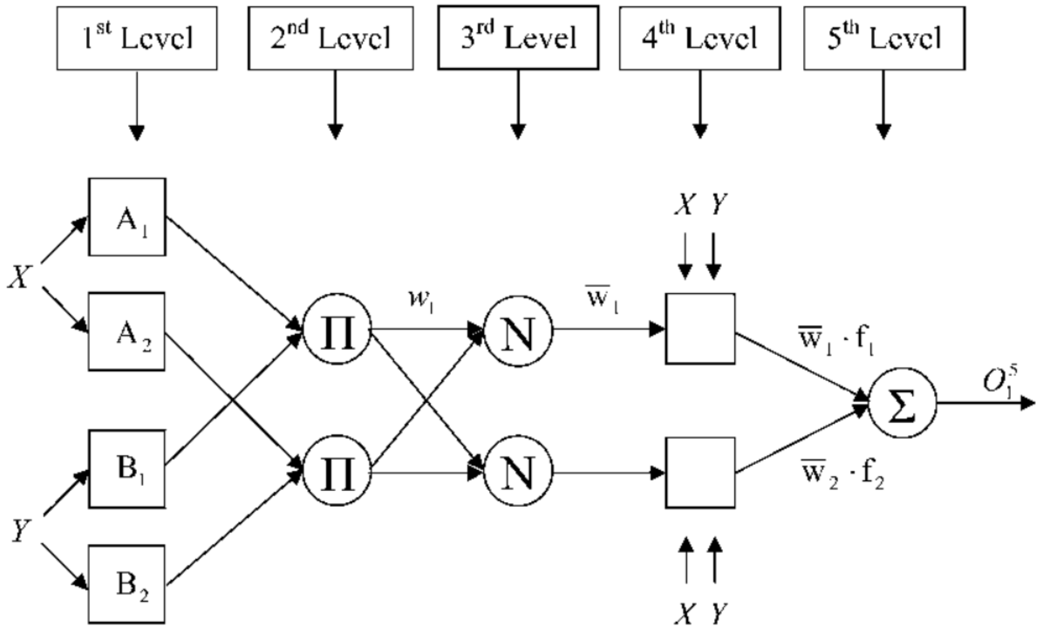

Figure 2.3: Organization of the ANFIS architecture

Description of the five layers

As an example we assume that the fuzzy inference system under consideration has two inputs x and y and one output z (Figure 2.3). Suppose the rule base contains two fuzzy if-then rules of the type of Takagi and Sugeno.

$$\begin{aligned} Rule1 &= \text{If } x \text{ } is \text{ } A_1 \text{ } and \text{ } y \text{ } B_1, \qquad \text{then } f_1 = p_1x + q_1y + r1 \\ Rule2 &= \text{If } x \text{ } is \text{ } A_2 \text{ } and \text{ } y \text{ } B_2, \qquad \text{then } f_2 = p_2x + q_2y + r2 \end{aligned} \tag{2.31}$$

Layer 1: This is the layer that deals with the Fuzzification of the input. Nodes are adaptive so that they can be modified based on the membership function. Every node i in this layer is a square node with a node function:

$$O_i^1 = \mu_{Ai}(x) \tag{2.32}$$

where x is the input to node i and Ai is the linguistic label associated with this node function. O_i^1 is the membership function of Ai that defines the degree to which the given x satisfies the quantifier Ai. The most common choice is the one in which $\mu_{Ai}(x)$ assumes a bell shape with maximum equal to 1 and minimum equal to 0, like the generalised bell function For example, with the Gaussian membership function, the node is defined by the following formula:

$$\mu_{Ai}\left(x\right)=\frac{1}{1+\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \tag{2.33}$$

where a_i, b_i, ci is the parameter set. As the values of these parameters change, the bell-shaped functions change accordingly, thus exhibiting various forms of membership functions on the Ai language label. Parameters in this level are referred to as premise parameters.

Layer 2: The nodes of the second level, on the other hand, incorporate the antecedents of the fuzzy rules. Within these nodes, only the logical AND operation is performed between the active inputs, for this reason they are called multiplier nodes. For example:

$$w_i=\mu_{Ai}(x)\times\mu_{Bi}(y) \quad i=1,2 \tag{2.34}$$

where w_i stands for weight. In ANFIS language, weights are usually called "Firing Strengths" of a rule.

Layer 3: In this layer the weights are normalised with respect to all the other weights for this reason they are called normalised fire force. They are described by the following formula:

$$\overline{w_i}=\frac{w_i}{w_1+w_2} \quad i=1,2 \tag{2.35}$$

Layer 4: The nodes of the fourth layer are adaptive and incorporate the consequent of the rules. Each node accepts as input all the input variables at the first level and the corresponding weight coming from the previous level.the nodes are represented by the following formula:

$$O_i^4=\overline{w_i}f_i=\overline{w_i}\left(p_ix+q_iy+r_i\right) \tag{2.36}$$

where $\overline{w_i}$ is the output of layer 3, and p_i, q_i, r_i is the parameter set. Parameters in this layer will be referred to as consequent parameters.

Layer 5: In this layer there is only one node, called or adder, because it calculates the overall output as the sum of all incoming signals, described by the following formula:

$$O_1^5=\sum_i\overline{w_i}f_i=\frac{\sum_i w_if_i}{w_i} \tag{2.37}$$

So, this is how any ANFIS produces an output from its fuzzy network by inserting inputs. However, there is no specific way to tell which is the optimal number of rules and membership functions to have a good approximator, without being too heavy in terms of calculation. The determination of the forms and quantity of the membership functions depends on the design choices made by the user. The initial fuzzy model can be derived systematically through some techniques so that you have a rough structure to start. The most popular methods are grid partitioning and subtractive clustering.

3
Unsupervised Learning in fMRI Analysis

In this section, we will present the results of the analysis of fMRI exams using unsupervised learning techniques. The results achieved are organised and set out below order:

- Analysis of resting-state fMRI exams using soft and competitive learning algorithms.i.e., results on the comparison of competitive algorithms, i.e. results concerning the application of a large family of unsupervised learning algorithms to an fMRI dataset in a resting state.

- Task-based fMRI exam analysis with fuzzy and competitive algorithms, i.e. results concerning the application of the clustering algorithm with the specific intent of evaluating the different results of cross clustering on task-oriented fMRI dataset.

3.1 Resting State FMRI Functional Connectivity Analysis Using Soft Competitive Learning Algorithms

3.1.1 Premise

This research work was presented at a Joint Congress of the 15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering and 3rd Conference on Imaging and Visualization held in Lisbon (Portugal) on March 2018. It that won the Taylor and Francis prize for the section "Imaging and Visualization" at the 15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering and 3rd Conference on Imaging and Visualization.This paper it is the result of collaboration with other authors, and presented in other works [1].The parts for which I have most contributed concern the selection, evaluation, implementation and optimization of the algorithms used, and the management of the experimental part.

The main topic regarding this work is resting state fMRI data analysis for functional connectivity explorations is a challenging topic in computational neuroimaging. Several approaches have been investigated to discover whole-brain data features. Among these, clustering techniques based on Soft Competitive Learning (SCL) have been shown effective in providing useful information in various contexts. However, although significant achievements have been reached, these techniques still present critical aspects that require further investigations. We selected three clustering algorithms, i.e. Self-Organizing Maps (SOM), Neural Gas (NG) and Growing Neural Gas (GNG), to study the intrinsic functional properties of images coming from a shared repository of resting state fMRI experiments (1000 Functional Connectome Project, i.e. Oxford dataset). To compare the functional connectivity based on soft clustering, we calculated the Seed Based Linear Correlation (SBLC) to study the Default Mode Network (DMN) functionality, i.e. we found that Precuneus L/R has the higher Correlations Coefficients with its controlateral part and with the posterior division of Cingulate Gyrus. The differences among the three soft clustering algorithms adopted were measured basing on Jaccard Similarity Coefficient (JSC), whereas the quality of clusters has been evaluated with Davies-Bouldin Index (DBI).

The optimal clustering computation was with 2 partitions for all the algorithms. We obtained the following results: a) clusters differentiated the amplitude of BOLD signals for both Males and Females, i.e. low level signal vs high level signal; b) clusters also differentiated the quality of seed-based correlations, i.e. strong (positive) associations vs weakly associations. These multivariate outcomes highlighted the complementary usage of clustering algorithms with statistical signal processing: the first made the partitions, the last explain the partitions.

3.1.2 Introduction

The main goal of our study is to integrate different methodologies useful to discover and to explore the inner properties of brain signals, with application to resting state BOLD time series in healthy subjects. The motivation of our study emerges in relation to a recent work published by Biswal et al. [8], that highlighted a universal architecture of functional connections in the brain resting state networks, with age and sex as significant determinants. The specific goal of our analysis was to evaluate the following points: 1) if there is a between gender functional variability, i.e. if there is a statistical BOLD signal difference between males and females, 2) if there is a within gender functional variability, i.e. if male and female exams have different spread, and 3) if there is a confirmation of some interesting functional connectivity networks. In particular, our intent was to extend this conclusion with clustering algorithms to find similarities in time series (signatures) or in activation patterns (exams).

Approaching the general fMRI signal processing with data-driven methods is a challenging application (cfr. works by Lachiche [49] and Liao [78] for an overview and Vergani et al [74] for our recent work with clustering techniques); data-driven methods are used as well in the resting state paradigm in functional neuroimaging (cfr. papers by Margulies [56], Van Den Heuvel [72, 39], Lee [50], and Wang [77]). For this purpose, we adopted as data driven methods the soft competitive learning algorithms to explore the natural partitions of the data (cfr. [19] and [21]). We choose Self Organizing Map (SOM), Neural Gas (NG) and Growing Neural Gas (GNN), that are a soft class of unsupervised artificial neural networks.

SOM are models initially proposed by Kohonen [47] and they are widely used because they allow the representation of data in a low-dimensional space, preserving the topological properties of the entrance space. SOMs are single-layer feedforward neural networks where output neurons are organised into low-dimensional grids (typically 2D or 3D spaces). The number of clusters that will be created is defined a priori.

NG is an alternative approach to SOM networks [17, 58] . The name derives from the fact that the neurons in the data space are moved as particles of a gaseous element, all negatively charged. Neurons repel each other, occupying the surrounding space, but they are attracted by areas of high data density as if the latter are positively charged particles. The NG algorithm is part of the soft competitive learning family, where not only the winning unit is adapted after the presentation of an input data, but also the remaining units. Unlike the SOM algorithm, no fixed topology is imposed on the network, the neurons are not arranged on the grid (they are free in space). Learning is performed according to a leaky learning strategy, by updating not only weight vectors of the winner neurons but also weight vectors of all losing neurons with a smaller rate that decreases in function of the increasing distance with the current input data.

GNG algorithm is an extension of NG and it was developed by Fritzke [27, 28]. Given a certain distribution of input data in the real domain, GNG incrementally creates a graph, or a network of nodes, where each node in the graph has a position in Rn. GNG is an adaptive algorithm because if the distribution of input data changes over time, GNG can adapt, that is to move the nodes in order to adapt to the new distribution. In this graph the number of nodes is increased incrementally starting from two initial nodes. The nodes are considered neighbours if they are connected by an edge, and the neighbourhood information is maintained during the execution of the algorithm basing on a variant of the standard Competitive Hebbian learning (CHL). The big difference compared to SOM and NG is that it is not necessary to establish previously the number of a priori nodes (clusters) since the nodes are added incrementally during execution. An edge is associated to each node that, through a vector, represents the position in the node space. The edge has an associated age variable and a local error variable that has the purpose to indicate the insertion point of a new node. GNG is an algorithm with many parameters and it is complex. Its strong point is the adaptation of nodes that can also be deleted. This allows to free users of the burden of choosing a priori the number of clusters. The weak point is the difficulty in finding the optimal value for the all the parameters involved.

Keeping in mind the peculiarities these clustering algorithms that have, we want to understand with more details the features of the elements partitioned. In other word, we decided to complement the clustering outcomes with the classical signal processing methodologies adopting methods able to analyse the temporal dynamic of the BOLD brain signals and the spatial features related to specific regions of interest. Also, we studied the functional associations across the brain regions about the spontaneous signal fluctuations depending by the resting state fMRI experimental paradigm.

Therefore, we first approach the resting state signals with standard tools for image processing making filtering, motion correction, standard registrations, labelling and data reduction procedures. Then, we integrated measures of strength/weakly signals association to investigate – in general – the cross correlations between all-ROIs with all-ROIs, and – precisely – the cross correlation with all-ROIs and two seeds: Left and Right Precuneus, that it is a bilateral region that has a role of central hub in the so-called Default Mode Network (DMN) (cfr. the historical work by Biswal [7, 9] and for the anatomo-functional details about the brain resting network cfr. Raichle et al [63, 64] or Utevsky et al [71] and also the work by Van Den Heuvel where he shown alternative to DMN [44] and work by Iraji [41] for technicalities about the resting state connectivity-domain analysis).

The general aim of this study is to address the functional connectivity problem in the resting state neuroimaging using both classical signal processing methods and soft clustering techniques. In the next sections, we propose and justify the type of data we have selected from a repository specialised in resting state functional neuroimages; then we present methods, results, discussion and conclusions.

3.1.3 Data

Within the NITRC repository (https://www.nitrc.org/) and the 1000 Functional Connectome Project (http://fcon_1000.projects.nitrc.org/), we selected the Oxford dataset with 22 healthy subjects (12M /10F; ages 20-35). The fMRI parameters were the following: TR = 2, slices = 34, time-points = 175, magnet = 3 [T]. The selection of this dataset is motivated by the nice age balance and the small age spread that have the subjects. Furthermore, we selected this dataset because was one used by Biswal et al [8] to discover resting state functional properties and their gender determinants. Therefore, our approach is also a confirmatory data analysis.

3.1.4 Methods

The methods we used followed this pipeline: image processing, data reduction, statistical analysis and algebraic measurements, functional connectivity investigations with linear correlation and soft competitive clustering techniques.

Image processing

Image processing was done with the functions for resting state image analysis contained in the software FSL [44]: we did spatial filtering with 3 [mm] of smoothing, frequency filtering with a high pass filter having 1/100 [Hz] as cutoff frequency, motion correction and standard registration with a reference atlas MNI152 (2mm). The ROIs labelling was done with the Harvard-Oxford atlas with 96 lateralized labels.

Data reduction for both females and males,

We did temporal signal reduction, spatial signal reduction, and whole brain signal reduction; i.e. the temporal data reduction was done with the extraction of mean and standard deviation of BOLD signals according to each time points, whereas the spatial reduction was the same but according to each atlas ROIs; the whole brain reduction is the global average of mean and standard deviation obtained by temporal reduction, with the aim to have two macro-signals, one for Females and one for Males.

Statistics analysis

To investigate if females and males are samples coming from different populations, we tested the mean and the standard deviation of the whole brain signals with parametric (one-way ANOVA) and non-parametric test (Kruskall-Wallis); we choose both kind of tests because we have globally 22 subjects and some ANOVA assumptions are difficult to sustain; therefore, we preferred to compare the parametric results with the non-parametric outcomes, that had mild assumptions.

Algebraic distances to compute metrics

we needed another step of data reduction: we averaged the spatial reduction results to have one value for each ROI, i.e. each exam became a vector with 96 components; then we measured how far are exams in vectorial forms from each other using Euclidean metric and Manhattan/Taxi-cub metric.

Correlations

To study the brain functional connectivity, we used the Pearson's linear Correlation Coefficient (CC) applied to all-ROIs versus all-ROIs and applied to seed versus all-ROIs; the seeds we used were Precuneus Left and Right, according to the anatomical architecture of Default Network Mode, as a model for the brain resting state paradigm; we selected only the higher or the lesser seed correlation results according to specific cut-off, i.e. $CC > 0.8$ or $CC < -0.8$ and $-0.2 < CC < 0.2$, respectively, in order to evaluate strong (positive/negative) associations and weakly/absent associations with the seed.

Clustering

To investigate brain resting functionality using unsupervised learning algorithms, we used Self Organizing Map (SOM), Growing Neural Gas (GNG) and Neural Gas (NG), i.e. to set GNG, we adopted 10 nodes, 1000 iterations, $\lambda = 2$, $\varepsilon_b = 0.0005$, $\varepsilon_n = 0.00001$, $\alpha = 0.05$,$\delta = 0.995$,age-node = 60. To set NG, we used 2 nodes, 500 iterations, $t_{max} = 8000$,$\varepsilon_{initial} = 0.90$, $\varepsilon_{final} = 0.50$, $\lambda_{initial} = 10$, $\lambda_{final} = 1$, $T_{initial} = 5$, $T_{final} = 10$. To evaluates the optimal partitions with the three clustering methods, we adopted Davies-Bouldin separation measure [14] and we compared the selected partitions with Jaccard similarity measures.

3.1.5 Result

The results we obtained regard the statistical descriptions of resting state fMRI data, the algebraic measures of fMRI in vectorial forms, the functional connectivity studied with the Pearson linear coefficient correlation in both the conditions (all-ROIs vs all-ROIs and Seeds vs all-ROIs) and the general outcomes of the three clustering techniques with their comparison with the classical statistical signal processing approach.

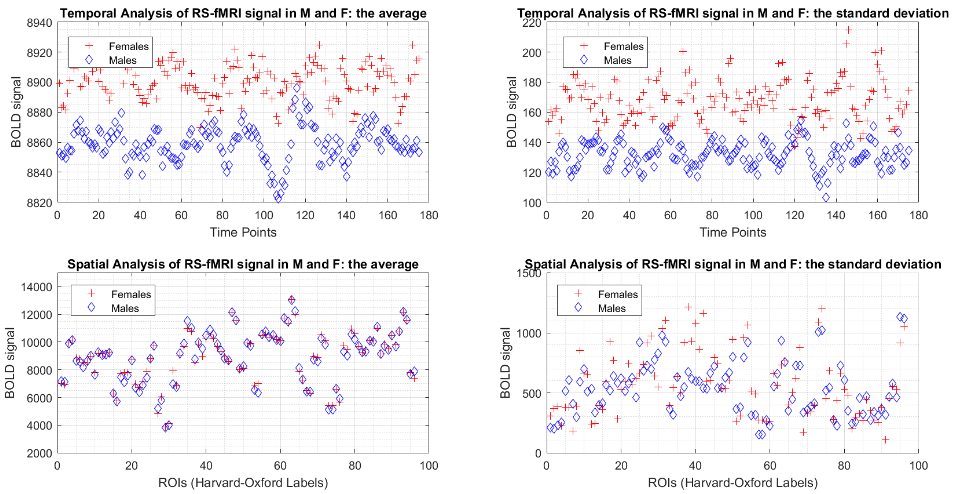

Figure 3.1: Plots of the average and the standard deviation of BOLD signals in Males and Females in the temporal (175 time-points versus BOLD signals) and spatial domains (96 ROIs versus BOLD signals). Both the average and the standard deviation of BOLD signals in the temporal domain are quite different between gender: Females have higher values then Males; whereas, in the spatial domain, both Males and Females have similar average BOLD signals, but Females have more standard deviation of BOLD signal.

Statistical descriptions of fMRI data

The temporal analysis of RS-fMRI exams indicated that Females had the higher values for both the average and the standard deviation of the BOLD signals (3.1 - top); the spatial analysis revealed that Females and Males were similar for the average signals, but Females had more standard deviation (3.1 - bottom); both one-way ANOVA and Kruskall-Wallis test proved that there are statistical differences between males and females ($p-value \leq 0.05$) for the mean and the standard deviation of the whole brain RS-fMRI signals, i.e. Females have greater mean and a greater variance than Males (3.2 - top).

Algebraic measures of fMRI data Euclidean and Manhattan/Taxi-cub distances estimated that Females are more far from each other than Males, for both the mean and the standard deviation of the measures (3.2 - bottom).

Clustering validations and outcomes comparison Davies-Bouldin separation index (DB) indicated that SOM, NG and GNG reached the best data partitions with two clusters (3.3 - top). The Jaccard index computed for all the pairs (NG vs GNG, SOM vs GNG, SOM vs NG) in both Females and Males showed that the more similar clustering was in Males and Females between NG and GNG (3.3 - bottom).

Functional Connectivity with Linear Correlation Pearson's linear Correlation Coef-

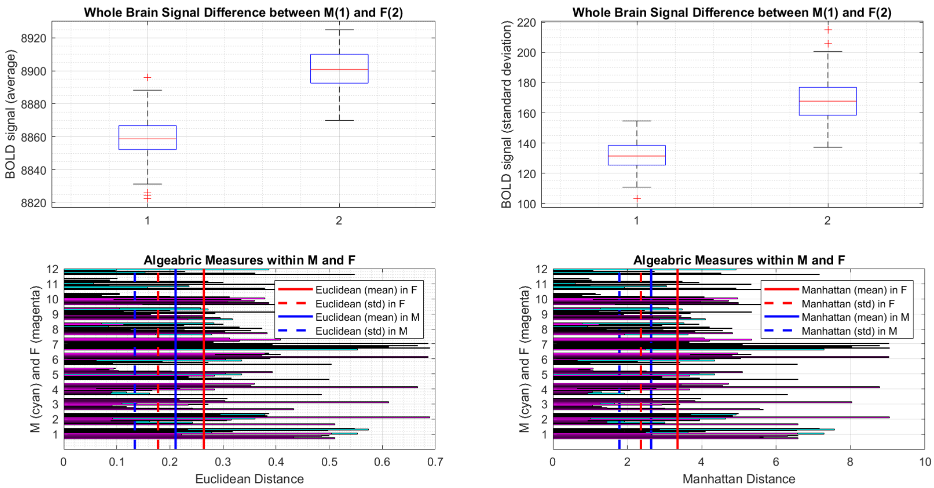

Figure 3.2: In the top part of the figure, there are the boxplots of whole brain signals in Males (1) and Females (2), showing the between gender statistical difference of the average (left) and the standard deviation (right) about the whole brain signals: in the both cases, Females have higher values then Males. In the bottom part of the figure, there are the within gender distance measures of the exams: with both Euclidean and Manhattan distances, Females have higher values then Males for the mean and the standard deviation of the distances computed.

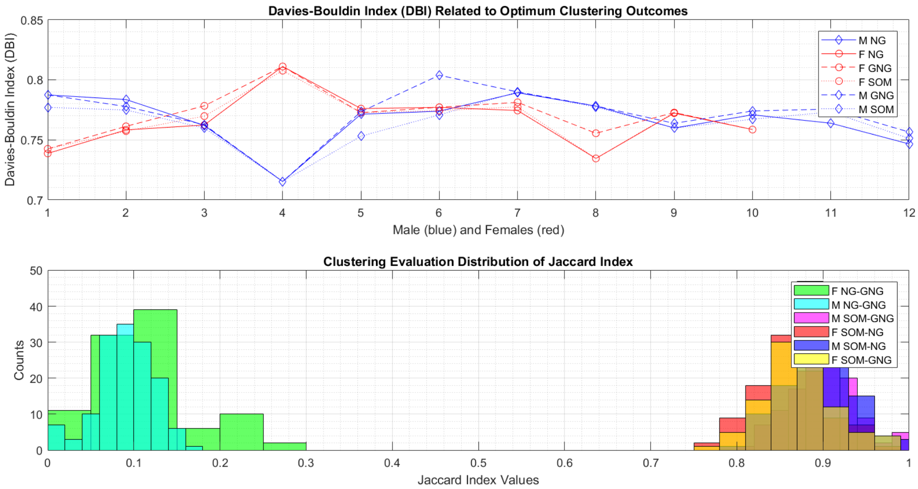

Figure 3.3: The top diagram shows the optimum Davies-Bouldin index associated to SOM, NG and GNG algorithms, for both Females and Males: all the DB are referred to 2 clusters as optimal clusters number for each algorithm; all the best indexes are under 1; the lesser is related to the fourth Males subject, and the higher to the fourth Female subject. The bottom histogram represents the discrete distribution of the Jaccard index computed for all the algorithm pairs: the more similar clustering outcomes are with the NG-GNG pairs in both Females and Males (Jaccard < 0.3), and the other clustering combinations are very different (Jaccard > 0.7)

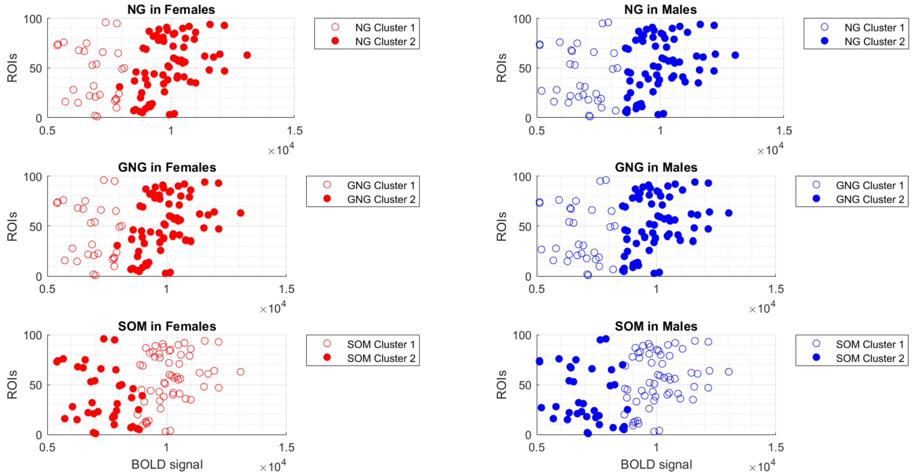

Figure 3.4: The scatter plots show the clustering outcomes for NG, GNG and SOM in Females (red points) and Males (blue points): the main result is that, using two clusters as optimum partitions number, the clusters detect the amplitude information of BOLD signals in Females and Males, i.e. clusters differentiate low levels and high levels in BOLD signals.

ficient (CC) applied to all-ROIs quantified that there is a difference between gender, i.e. Females had more negative correlations then Males, but they shared common positive correlations (3.4 – top – the left plot). Instead, the seed-based Correlation Coefficient (CC), with Left and Right Precuneus as seeds, shown that all the subjects had principally positive correlations, with a little presence of anti-correlation in Females (3.5 – top – the central and the right plot). Apart general comments about strength and weakly associations, it is remarkable the presence of the specific seeds weakly correlations, especially for the L/R Posterior Division of the Temporal Gyrus (ROI 29 and ROI 30) and for the L/R Anterior Division of Temporal Fusiform Cortexes (ROI 73 and ROI 74); also, it is remarkable the highest seeds positive correlation, especially for the L/R Cingulate Cortexes (ROI 59 and ROI 60) and with the controlateral part of the seed, the L/R Precuneus (ROI 61 and ROI 62) (3.5 - the correlation matrixes)

Functional Connectivity with Soft Competitive Clustering

The two optimal clusters distinguished the quality of correlations for both the seeds (Precuneus Left and Right) analysis in Females and Males, i.e. strength (positive) coefficients and weakly coefficients were always mismatched in separated clusters (3.6 and 3.7). We also noticed some overimposition for the central values of the coefficients. Furthermore, the two clusters discriminated the amplitude of BOLD signals, i.e. low

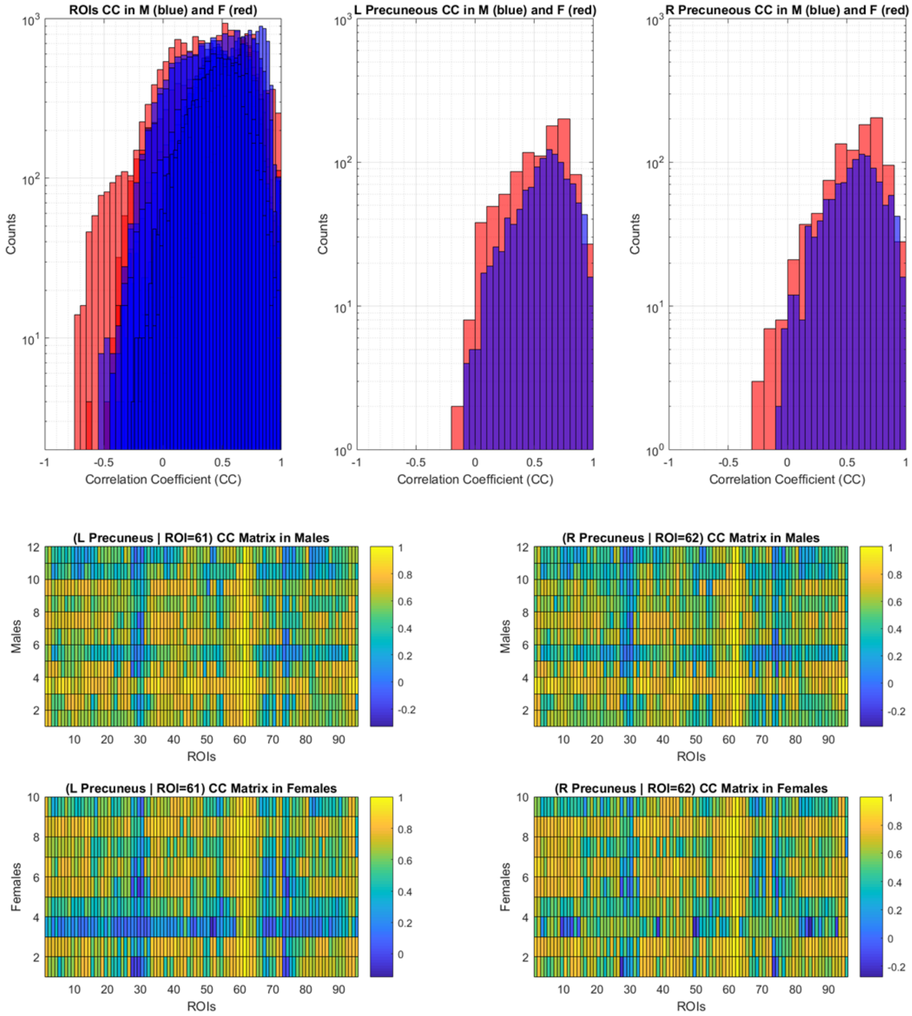

Figure 3.5: In the upper part of the figure, there are the Correlation Coefficients (CC) distributions between all-ROIs versus all-ROIs in Males and Females, and the correlation coefficients distributions between seeds (L/R Precuneus) versus all-ROIs. In the first discrete distribution, there are present positive and negative correlations, both in Males and Females; whereas, in the seeds based correlations distributions, there are less anti-correlations, with always, a bit gender difference. In the bottom part of the figure, there are the seed-based correlation matrix between Left and Right Precuneus in Males and Females: it is evident the strong correlation with many brain regions, as well as weak correlations with few brain regions. The highest positive correlations (CC ¿ 0.8) are with seed-controlateral region (ROI 61 and ROI 62) and L/R Posterior Division of Cingulated gyrus (ROI 59 and ROI 60). The lowest correlations (CC +/- 0.2) are with L/R Posterior Division of Temporal Gyrus (ROI 29 and ROI 30) and with the L/R Anterior Division of Temporal Fusiform Cortexes (ROI 73 and ROI 74).

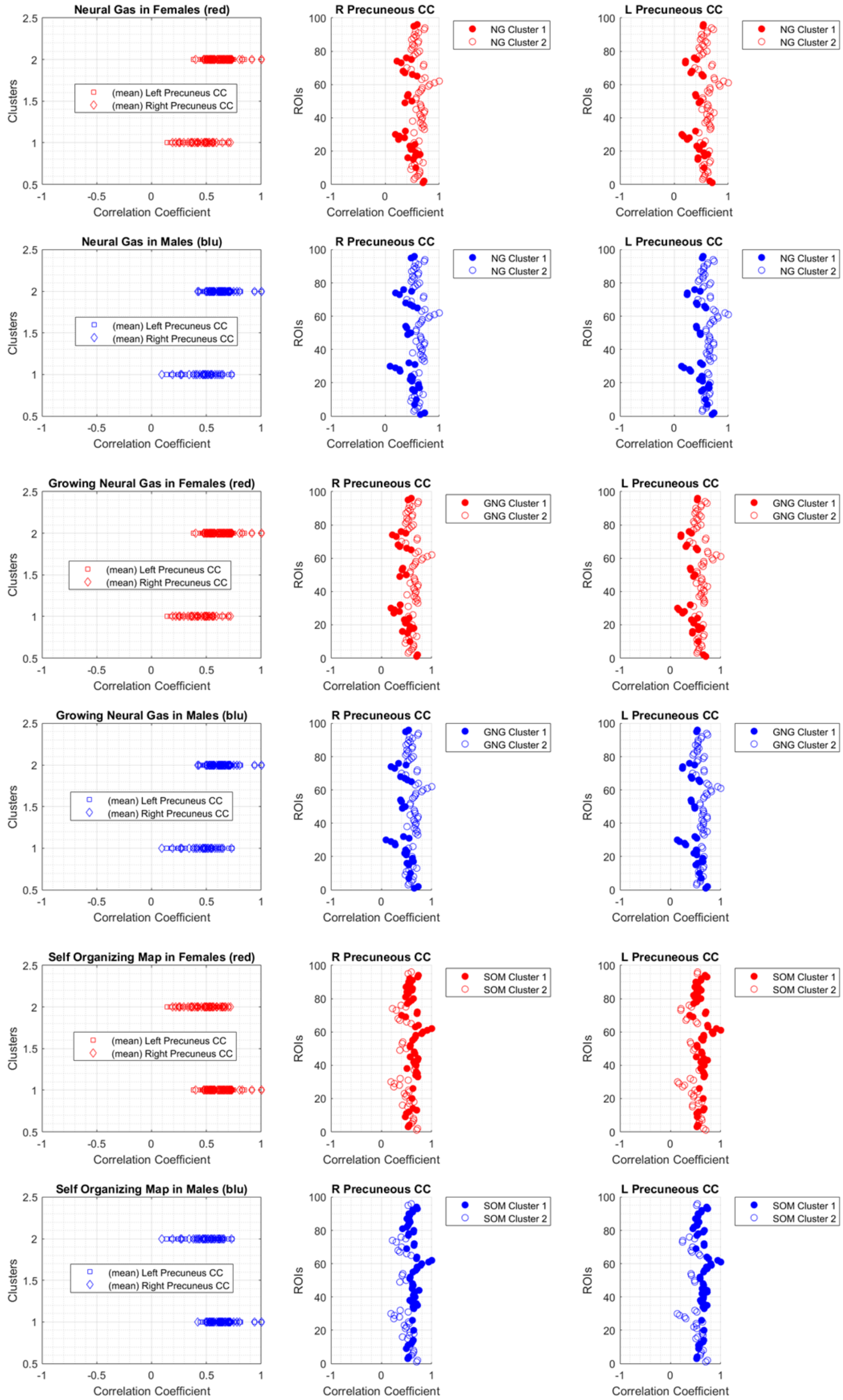

Figure 3.6: These figures represent the organization of the two optimal clusters in Males and Females in relation to the seeds based (L/R Precuneus) Correlation Coefficient. In the left column, it is reported the evidence that low and high correlations belong to different clusters, with some superimposition between them for the central values. The central and the right columns show the relation between ROIs, correlation coefficients and clusters, showing precisely that low and high correlations are within different clusters in Males and Females.

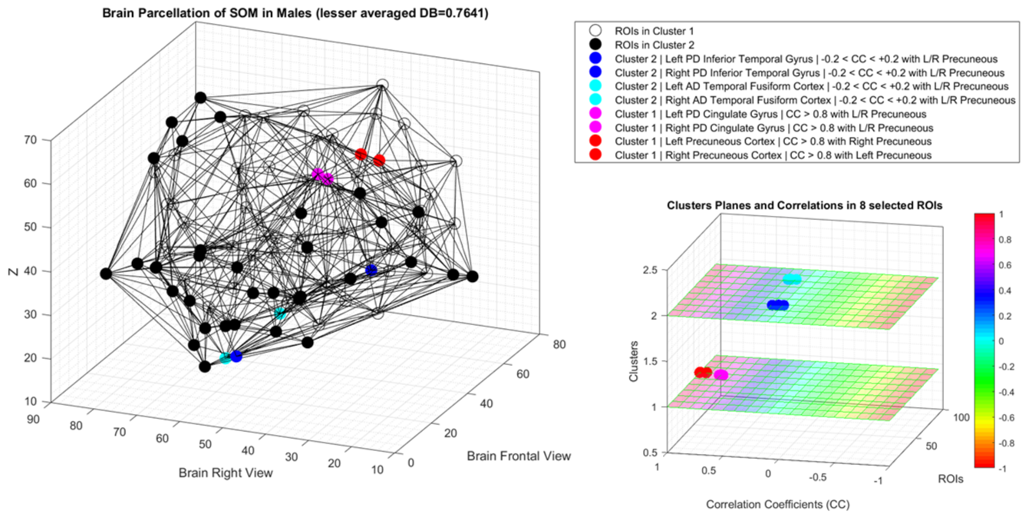

Figure 3.7: The image shows a detailed result of SOM algorithm in the Males case, with a specification about the higher and the lower correlations with both the seeds (Left and Right Precuneus), and their organisation within the two clusters. In the left side of the figure, the brain is represented with the centre-of-mass of each regions of interest (the 96 ROIs labelled with Harvard-Oxford atlas), that are filled or unfilled if they belong to cluster 1 or 2. The coloured circles are the height selected regions that are, respectively, the four with the higher correlations (hot colours), and the four with the lower correlation (cool colours). In the right side of the figure, there is a plot with the clusters planes in relation to their elements: the four higher correlations regions are on the plane of cluster 1, whereas the four lower correlation are on the plane of cluster 2. We choose to plot the brain parcellation relative to the SOM algorithm in Males because was the better optimized algorithm, having as a Davies-Bouldin (DB) separation measure a value equal to 0.7641 (the lesser value respect the other algorithms).

and high intensities were mismatched as well in both Females and Males (3.4).

3.1.6 Discussion and Conclusion

We obtained two types of results. The first-type result globally confirms the gender determinants in RS-fMRI functionality found by Biswal et al. Our approach is also related to the anatomo-functional correlation (L/F Precuneus - DMN) with other regions, and specially with their controlateral part and their associated bilateral cingulated regions; this is a confirmation because we found brain functionality noted in the scientific literatures [2]. We also added information about Male and Females peculiarities using algebraic distances to measure the within gender variability. Globally, we can claim that Females had more amplitude and more variability than Males. The second-type result regards the integration of clustering techniques, with classical statistical processing for signal analysis. We can affirm that, once the clustering algorithm differentiates data in different clusters, it became necessary to explicate the inner property that determine their inclusion: in our case, we found that clusters differentiated the intensity of the brain signal, i.e. low versus high level BOLD signal, and clusters also differentiated the quality of the brain functional connectivity, i.e. strength (positive) associations versus weak associations. There are limitations of this study: the number of samples we adopted are limited – 22 subjects – but, interestingly, the statistical tests demonstrated that the two subsamples (10 Females and 12 Males) are significantly different (they did not come from the same population, and then they are nice comparable but not useful to infer general population properties). Next study will attempt to address these limitations, using also other soft competitive learning algorithms to cluster fMRI data, e.g. using fuzzy algorithms to better handle the complexity of the of Resting-State fMRI data.

3.2 Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms

3.2.1 Premise

This research work was presented at 6th Computational Modelling of Objects Presented in Images Congress held in Cracow (Poland) on July 2018. This paper is the result of collaboration with other authors, and presented in other works [1]. The parts for which I have most contributed concern the selection, evaluation, implementation and optimization of the algorithms used, and the management of the experimental part. In this work, we used model free methods to explore the brain's functional properties adopting a partitioning procedure based on crossed-clustering. We selected Fuzzy C-Means (FCM) and Neural Gas (NG) algorithms to find spatial patterns with temporal features and temporal patterns with spatial features. We applied these algorithms to a shared fMRI repository of face recognition tasks. We matched the classes found and our results of functional connectivity analysis with partitioning of BOLD signal signatures. We compared the outcomes using the just known model-based knowledge as likely ground truth, confirming the role of Fusiform Brain Regions. In general, partitioning results show a better spatial clustering than temporal clustering for both algorithms. In the case of temporal clustering, FCM outperforms Neural Gas. The relevance of brain subregions related to face recognition were correctly distinguished by algorithms and the results are in agreement with the current neuroscientific literature.

3.2.2 Introduction

In functional Magnetic Resonance Imaging (fMRI) there are two kinds of approaches to data analysis: model-based methods and model-free methods. The main difference between the methodologies is that the first one needs *a priori* knowledge about the functional data structures, whereas the second one does not need any assumptions related the images to be investigated. The main model-based approach to fMRI data is the Statistical Parametric Maps (SPMs) approach introduced by Friston [25].

The main model-free models are the Analysis of Principal Components (PCA) or the Analysis of Indipendent Components (ICA) (for an overview see [53]). In addition, there are other model-free techniques to explore fMRI data properties that allow to classify functional patterns, such as the clustering algorithms, that are a class of computational models used to find the *natural groupings* of input features [20]. Several kinds of separation methodologies based on different theoretical framework are proposed in the literature [42]. Generally, clustering is divided in crisp or soft partitioning: crisp classes have unshared elements (e.g., k-Means Algorithm), whereas soft classes have elements that could be shared with more then one class (e.g., fuzzy sets based algorithms [6]). The soft properties in clustering have a wide meaning that encompasses not only the data multi-membership feature, but also (in computational learning theory) the competitive learning approach that is used by unsupervised algorithms to adapt themselves on the data to be clustered (e.g. self-organizing maps). Using this double meaning of the *soft clustering category*, we selected **Fuzzy C-Means (FCM)** and **Neural Gas (NG)** algorithms, where the first one is a soft algorithm in terms of multi-class properties and the second one is a soft algorithm in terms of competitive learning rule [79].

Clustering techniques applied to fMRI time series data are a interesting approaches to explore brain functional properties [78, 50, 74]. Partitioning works grouping image voxels together based on how much they are alike in relation to some measure (distances, correlation, etc.), that probes how their intensity profiles in time are similar. In more details, let n denote the number of scans in a fMRI experiment, and let K be the number of voxels in each volume: the dynamics of each voxel $\mu \in \{1, ..., K\}$, are the signal values $\{\mathbf{x}^{\mu}(1), ..., \mathbf{x}^{\mu}(n)\}$ that can be modeled as a vector $\mathbf{x}^{\mu}(i) \in \Re^{n}$ in the n-dimensional (Euclidean) feature space of possible time series. Each of the these points is partitioned into clusters based on the similarity of their intensity profile in time. Therefore, the principal approach to fMRI clustering is to cluster spatial features (i.e., brain regions) that have similar temporal patterns (i.e., the brain functional signatures). In other words, the procedure to cluster functional images has the goal to find common functional structures in different Region of Interest (ROIs).

In this work, we also aim to find functional structures in temporal features that have similar spatial patterns: we named these objects as "Times of Interest" (TOIs), that in computational terms mean to cluster the experimental blocks related to each brain volume. The assumption is that the exploration of TOIs allows to find properties related to the peculiarity of each block of the experimental design . E.g., imagine a block design structured as TASK-REST stimuli alternations; the TOIs clustering could allows finding spatial structures that are similar to TASK or REST blocks, whereas the classic ROIs clustering allows finding in different brain regions temporal structures that are similar. We named this global procedure **cross-clustering** that want as it finds spatial patterns in the temporal features (TOIs) and temporal patterns in the spatial features (ROIs). The postclustering procedure we applied performs a statistical evaluation of the obtained clusters. We used parametric and nonparametric tests to study whether the classes are statistically different with the aim to investigate the numerical properties that distinguished the clustering outcomes. Furthermore, we compared the classes computed by the algorithms using Jaccard similarity index. Also, we compared the functional connectivity of the fMRI scans within various subjects in order to find useful information to be associated with adopted clustering techniques. In the next section, we present the dataset we selected to be clustered, the computational methodologies we used, and the results we obtained. We conclude with a discussion and possible future works.

3.2.3 Materials and Methods

Data

We selected the dataset proposed by Wakeman and Henson [76] available in **openfmri** repository (https://openfmri.org/dataset/ds000117). This dataset contains face recognition task paradigms applied to 16 healthy and young subjects. The study of Wakeman and Henson showed functional peculiarity along the fusiform regions in the brain temporal and occipital parts. Keeping in mind these features, we clustered the subjects paying special attention on the clustering outcomes related to eight fusiform cortexes using Harvard-Oxford labels. According to these labels, they are ROIs 73:74 L/R Temporal Fusiform Cx Anterior Divisions, ROIs 75:76 L/R Temporal Fusiform Cx Posterior Divisions, ROIs 77:78 L/R Temporo-Occipital Fusiform Cx, and ROIs 79:80 L/R Occipital Fusiform Cx. Before clustering the images, we performed preprocessing with FSL standard tools [44]: such as spatial and temporal filtering, motion correction, standard registration (with MNI152 reference), and time-series extraction as per the meaning of Harvard-Oxford Atlas using 96 lateralized labels.

Clustering Algorithms

We adopted two soft clustering algorithms to process the fRMI data: **Fuzzy C-Means (FCM)** and **Neural Gas (NG)**. We used as an input features of both Regions of Interest (ROIs) and Times of Interest (TOIs). We validated the optimal clustering using Davies-Bouldin index [14] and we compared the various clustering outcomes using the Jaccard similarity measure. We investigated the statistical difference of the clusters computed with parametric (One-way Analysis of Variance – ANOVA-1) and non-parametric method (Kruskall-Wallis) using the p-value as a decision criterion. Furthermore, we computed the brain functional connectivity with the Pearson Linear Correlation Coefficient across the ROIs: we wanted to analyse the overall dynamics of the subjects in relation to the task-oriented study, i.e., the face recognition paradigm, in order to obtain similar results about the brain face processing as the ones in the literature specialised in experimental neuroscience (see [34]).

Fuzzy C-Means Fuzzy C-means is a clustering method that allows to each element of a dataset to be a part of more than one cluster with a respective degree of membership [6]. The algorithm is based on the optimisation of the following objective function:

$$J_m = \sum_{i=1}^{D} \sum_{j=1}^{C} \mu_{ij}^{m} ||X_i - A_j||^2 \tag{3.1}$$

For the $||...||^2$ we chose a distance, D is the number of the data points, C is the number of the clusters, m is the partition exponent controlling the fuzzy overlap between clusters that determines how fuzzy the boundaries between the clusters are ($m > 1$; in the computational set-up we adopted $m = 2$). X_i are the points, A_j are the centroids of the clusters and μ_{ij} is the degree of membership of X_i in the jth cluster (i.e., for a given cluster j, $\sum_{i=1}^{N} \mu_{ij} = 1$). FCM working randomly initialising the membership values μ_{ij} and then calculating the cluster centers. It repeats the updating of μ_{ij} until the objective function J_m reaches the minimum value. FCM is a useful clustering algorithm that allows to handle the *softness* of datasets. In many cases, if is necessary to take into account the possibility that data-points could be within more then one class, for example when they represent complex objects without a crisp natural shape (e.g., in the case of fMRI functional partitioning); so, the use of FCM could be an appropriate choice for the task-based paradigm, when periodically the brain functionally varies due to the stimulation of the rapidly changing experimental blocks.

Neural Gas The Neural Gas (NG) algorithm for clustering analysis is a vector-quantization approach inspired by the Self Organizing Map [58]. The method converges quickly to low error rate, and also has lower distortion value than the Kohonen's feature map. It uses the gradient descent method as optimisation solver. NG utilized a neighbourhood-ranking of the reference vectors $\mathbf{w}_i$ for the given data vector $\mathbf{x}$. The learning model is determined by the formula:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \epsilon(t) \exp\{-k_i(\mathbf{x}, \frac{\mathbf{w}_i}{\lambda}\}(\mathbf{x}(t) - \mathbf{w}_i(t)) \tag{3.2}$$

Here $k_i = 0, ..., N-1$ are the rank indexes describing the ranking of the neighbourhood of the reference vectors $\mathbf{w}_i$ to the data vector $\mathbf{x}$ in a decreasing order; N is the cardinality of the neural network units that update their synapses for each iteration; the step-size variable $\epsilon \in [0, 1]$ tunes the ranges of modification. Referred to Kohonen algorithm, NG has the advantage that it does not need a prespecified network and it produces topologically-preserving maps. Prototype classification is based on distance. It updates the prototypes and assigns the data to the prototype closest to it. When the number of maximum iterations is reached, the algorithm terminates; otherwise, it goes through the next iteration.

3.2.4 Results

In this section, we describe the obtained results, which are illustrated in Figures 3.8 trough 3.21. We discuss the results in the Section 3.2.5.

Figure 3.8 shows the empirical distributions of the overall Correlation Coefficients (CC) of every subject and the global correlation matrix represented as a mean of the correlation matrices of each subject.

Figure 3.9 shows, first, the best Davies-Bouldin Indexes (DBI) computed for clusters ranging from 2 to 20 for both FCM and NG algorithms applied on both ROIs and Time Points. Figure 3.10 illustrates spatial semantics of FCM and NG in the case of 2-classes computed as the statistical mode of the subjects (i.e., if a region frequently appears in a cluster i, than it belongs to the cluster i in the representation). Table 3.1 gives details about he the best DBI configurations for the dataset.

Figure 3.11 and Figure 3.12 display the results of the non-parametric (Kruskal-Wallis) and parametric (ANOVA 1-way) statistical tests applied to the optimal FCM and NG algorithms in both the spatial and the temporal configuration. The figures demonstrate that in the spatial configurations of both FCM and NG, the p-value was always less then 0.05. Hence, we can reject the hypothesis that all clusters have all the same means. In contrast, the p-value computed in the temporal configuration clustering for either FCM or NG, is not always less then 0.05, as the worst behaviour is in the case of NG algorithm.

In the Figure 3.13 and 3.14, the results are almost the ones in Figure 3.11 and 3.12.

In general, spatial clustering (i.e., ROIs partitioning) is always statistically significant (p-value less then 0.05) for both NG and FCM in the optimal case as well as in the case of 2-classes. Temporal clustering (i.e., TOIs partitioning) has lesser p-value, but it is not always significant; according to the non-parametric test, it is significant only in the case of FCM, where p-value is around 0.05.

Figure 3.15 shows the comparison of the Regions of Interests (ROIs) clustering of 2-classes using Jaccard matrix. The comparison is between subjects in case when applying NG and FCM algorithms. Under the matrices, the distributions of the Jaccard distance values are shown. In general, the distributions of the similarity of the subjects are uniform, expect when using the FCM algorithm. The explanation of this fact is not related to the quality of the clustering itself, but is rather rooted in the initialisation procedure of the algorithms we used.

Figure 3.16 shows the comparison of the Times of Interest (TOIs) clustering with 2-classes using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices, the distributions of the Jaccard distance values are shown. In general, the similarity distributions between subjects are uniform (for NG) and middle-centred (for FCM). The main result is that none of the outcomes seems to be in agreement between clustering algorithms. Instead, it is interesting to note that the empirical distributions of the indexes vary widely from almost unimodal to multi-modal distributions.

Figure 3.17 shows the Jaccard's distance of the the NG and FCM results using 2-classes. The ROIs clustering configuration exhibits almost uniform distribution of the values, while the TOIs clustering configuration exhibits unimodal (gaussian-like) distribution.

In sum, Figure 3.15, 3.16 and 3.17 demonstrate that when clustering algorithms perform the computations in a different way, it is hard to compare the results. This does not mean that the clustering methods miscalculated the data, but rather that the initialisation procedures used by the algorithms lead to assigning different classes to the same pattern found among the subjects.

Figures 3.18, 3.19, 3.20 and 3.21 shows the results related to one-subject analysis. We selected specifically the FCM 2-clusters for the Subject 12 because it has good performance in spatial and temporal configurations according to both the parametric and the non-parametric tests. Figure 3.18 shows the graph-based topological representation that gives an idea about the parcellation. In the top-left part of Figure 3.19 there are details about the relation between ROIs and clusters: in particular, following the horizontal lines, it is notable that the Fusiform regions (ROI from 73 to 80) are splitted half and half in the Cluster 1 and the Cluster 2; this peculiarity is due to the different BOLD signal amplitude that allows to discriminate the membership to clusters, as indicated in the centre plot of Figure 3.19. The difficulty to distinguish temporal pattern along the time points is evident in the right plot of Figure 3.19, where there is no clear data structure that explains the two partitions, though the statistical tests for temporal clusters have always a p-value less then (or around) 0.05 for the Subject 12. Specially, Figures 3.20 and 3.21 show the brain voxels parcellation of Subject 12 clustered with FCM using the optimal configuration. The algorithms computed 14 clusters that grouped differently all brain portions. Figure 3.21 shows the Fusiform regions superimposed to the brain parcellation in Figure 3.20: the eight Fusiform cortexes belong to different clusters according to their BOLD values: i.e., L/R Temporo-Occipital Fusiform Cx and L/R Occipital Fusiform Cx have greater BOLD values than the L/R Temporal Fusiform Cx Anterior and Posterior Divisions. Clusters 2, 6 and 8 contain the Fusiform ROIs with greater activation's, Cluster 3, 13, 14 contain the ones with the less activation's.

3.2.5 Discussion

The results in Figure 3.8 are related to the global functional connectivity analysis that reflects the presence of some variability in the single subject correlations (left figure), i.e., more than half subjects had positive correlations during the task, whereas few ones had negative correlations. In particular, the mean correlation matrix (right figure) shows great correlation in specific submatrixes, e.g. the submatrix that regards the ROIs

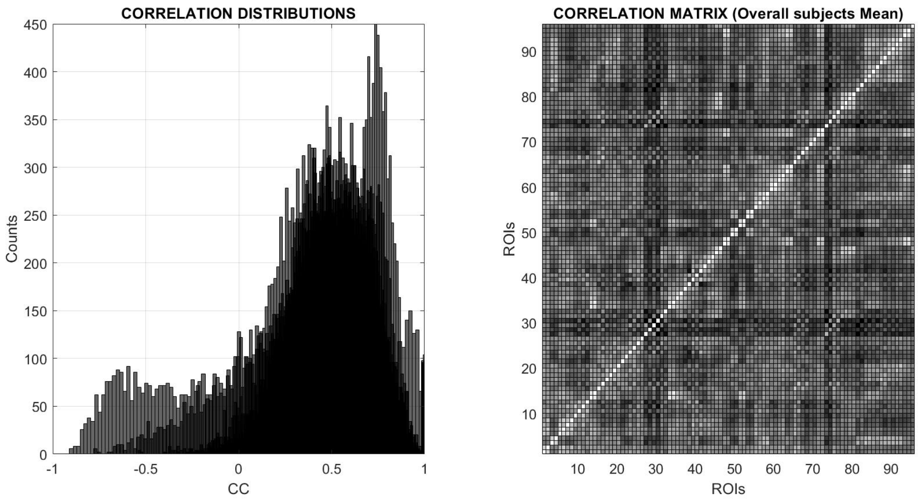

Figure 3.8: The Figure shows the information related to the correlation coefficients. The left plot is the empirical distribution of the all correlation coefficients (CC) computed for every subject. the right plot is the mean correlation matrix computed for all subjects' correlations)

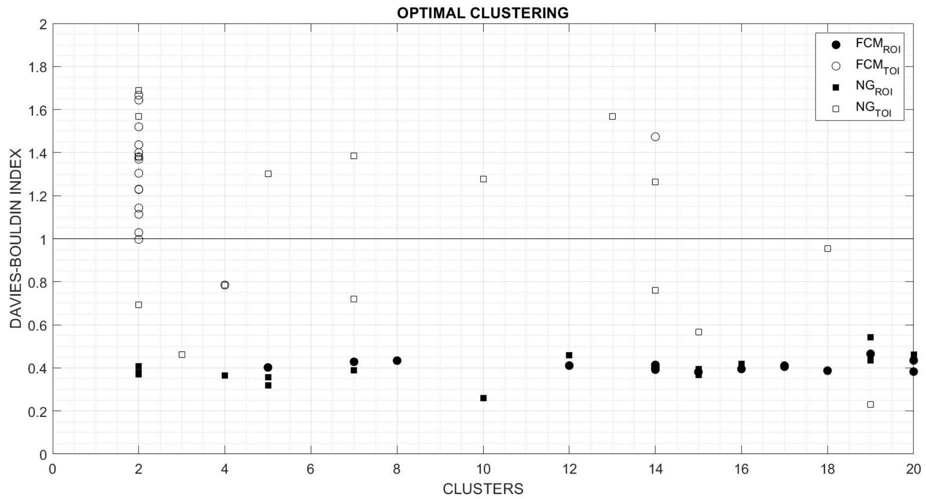

Figure 3.9: The figure shows the plot of the optimal clusters *versus* the lesser Davies-Bouldin index for clustering with FCM and NG algorithms, that differentiated by the Regions of Interest (ROIs) configuration and the Time Of Interest (TOIs) configuration. Globally, the partitioning of ROIs with both FCM and NG had lower Davies-Bouldin index values then the partitioning of TOIs.

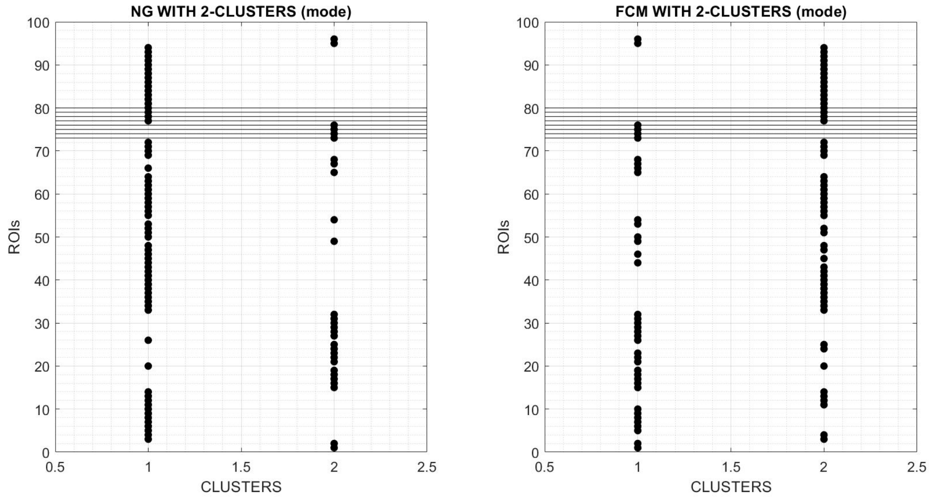

Figure 3.10: The figure shows the more frequent ROIs partitioning in the special case of clustering with 2 groups for NG and FCM for all the subjects; the horizontal line from ROIs 73 to 80 are the ones related to the eight Fusiform regions, i.e., they are ROIs 73:74 L/R Temporal Fusiform Cx Anterior Divisions, ROIs 75:76 L/R Temporal Fusiform Cx Posterior Divisions, ROIs 77:78 L/R Temporo-Occipital Fusiform Cx and ROIs 79:80 L/R Occipital Fusiform Cx.

Table 3.1: The table describes the Davies-Bouldin (DB) index computation for each subject differentiated for clustering (FCM or NG) and inputs (ROIs or TOIs).The values presented are the lesser DB associated with the corresponding number of clusters.

Sub	**DB**	**FCM**$_{ROI}$	**DB**	**NG**$_{ROI}$	**DB**	**FCM**$_{TOI}$	**DB**	**NG**$_{TOI}$
1	0.47	19	0.46	12	1.37	2	0.95	18
2	0.40	5	0.37	15	1.23	2	1.38	2
3	0.39	18	0.42	16	1.23	2	0.69	2
4	0.41	14	0.41	2	1.14	2	1.57	2
5	0.40	14	0.46	20	1.44	2	1.30	5
6	0.41	17	0.37	2	1.64	2	1.39	7
7	0.40	14	0.32	5	1.47	14	0.46	3
8	0.43	20	0.39	15	1.52	2	1.28	10
9	0.40	16	0.36	5	1.00	2	0.76	14
10	0.43	7	0.26	10	0.78	4	0.57	15
11	0.43	8	0.38	2	1.40	2	1.69	2
12	0.39	14	0.36	4	1.11	2	0.78	4
13	0.41	17	0.43	19	1.03	2	0.72	7
14	0.38	15	0.39	7	1.38	2	1.27	14
15	0.38	20	0.45	19	1.67	2	1.57	13
16	0.41	12	0.54	19	1.30	2	0.23	19

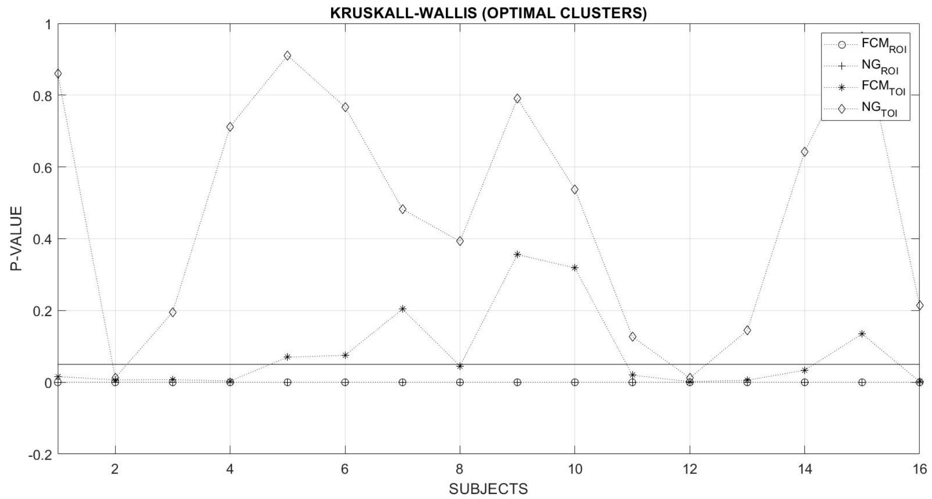

Figure 3.11: The figure shows the results of non-parametric tests (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) inputs in case of the optimal configurations. The black line is the significance level 0.05. The values under the black line allow to reject the null hypothesis of the test. ROIs clustering with both FCM and NG have statistically different clusters for all the subjects with non-parametric tests. TOIs clustering is globally near the significative criterion only with FCM algorithms.

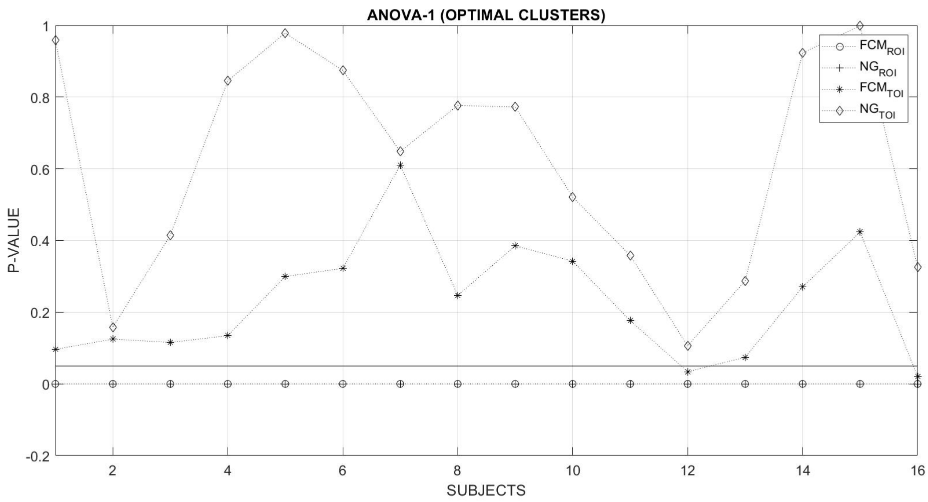

Figure 3.12: The figure shows the results of the parametric test (ANOVA-1) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of the optimal clusters configurations. The black line is the significance level 0.05. The results were similar to Figure 3.11, but with a higher p-value for the TOIs clustering with FCM algorithm.

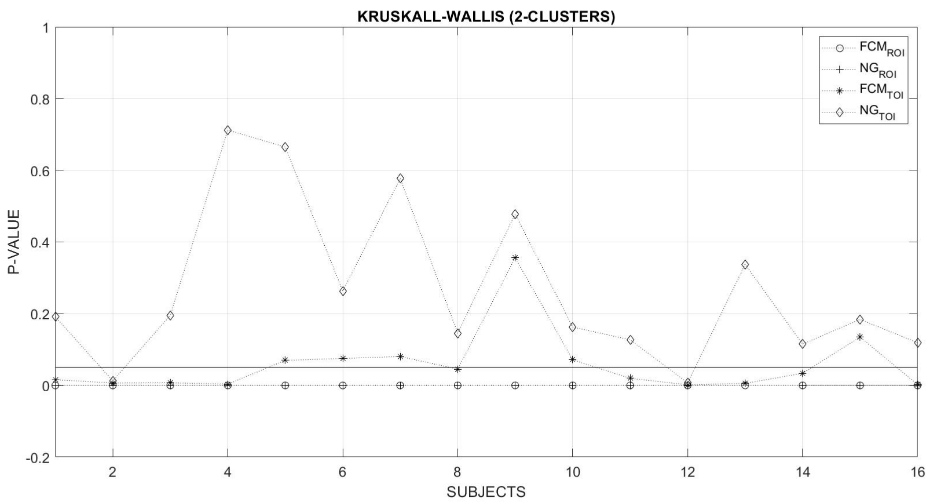

Figure 3.13: The figure shows the results of the non parametric test (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) configurations in the special case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 3.11 and Figure 3.12, but with lower p-value for the TOIs clustering with FCM algorithm.

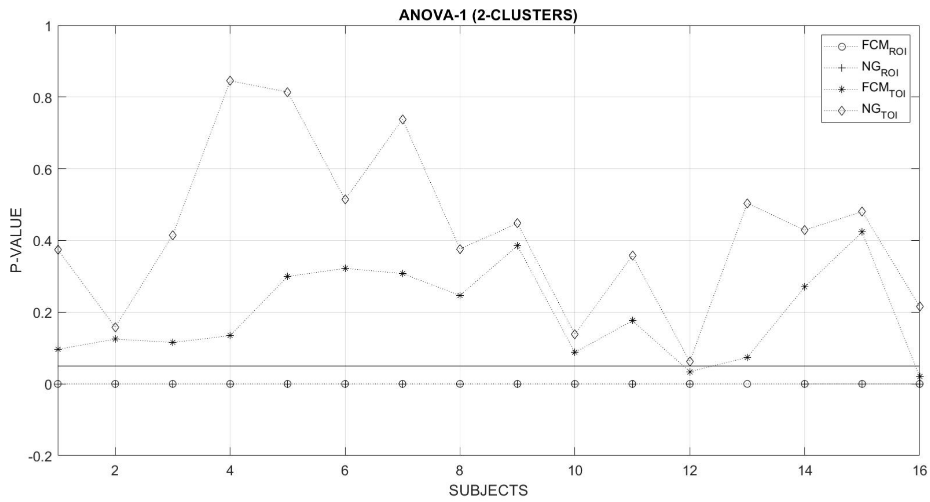

Figure 3.14: The figure shows the results of the parametric (one-way ANOVA) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 3.13, but with a higher p-value for the TOIs clustering with FCM algorithm.

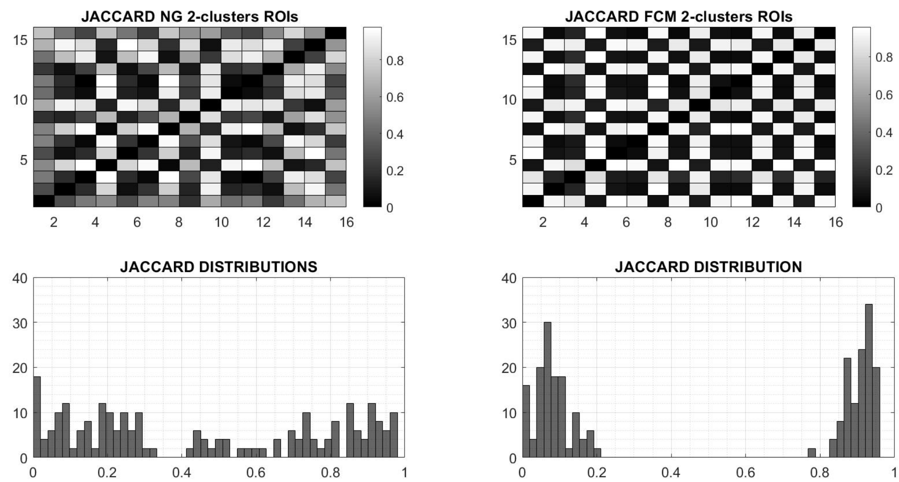

Figure 3.15: The figure shows the comparison of Regions of Interests (ROIs) clustering of two groups using Jaccard matrix. The comparison is between subjects in case of applying NG and FCM algorithms. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between the subjects is uniform and without big differences, expect when using the FCM algorithm.

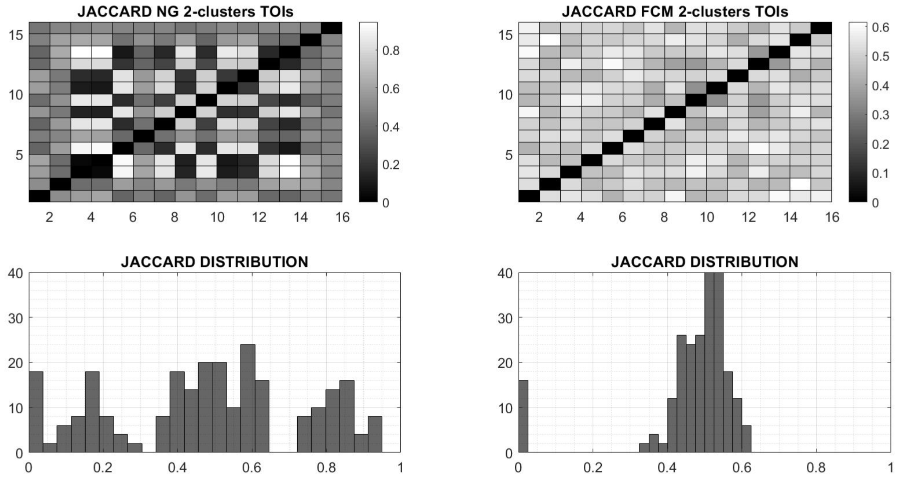

Figure 3.16: The figure shows the comparison of Times of Interest (TOIs) clustering with two groups using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between the subjects is uniform (for NG) and middle-centered (for FCM).

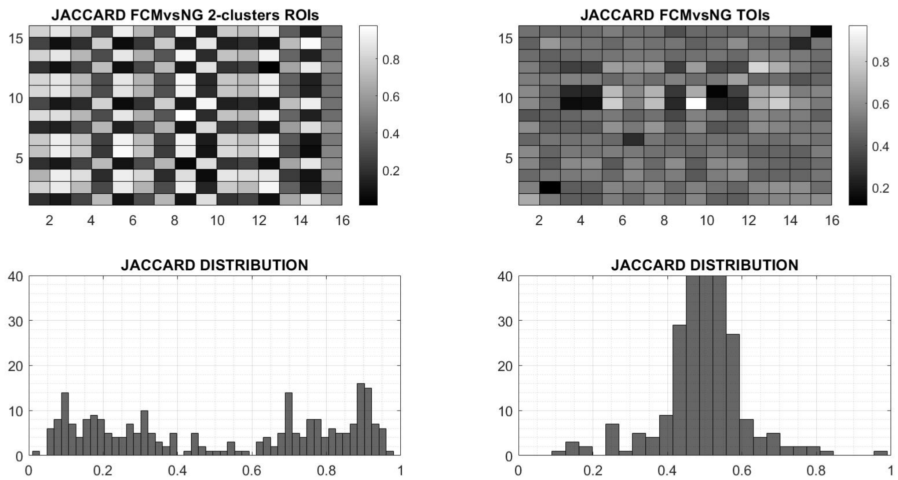

Figure 3.17: The figure shows the Jaccard Matrices in both the cases of two classes clustering with Regions of Interest (ROIs) and Times of Interest (TOIs). The comparison is between subjects in cases of NG and FCM. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between subjects is weakly bimodal for the ROIs clustering (bottom left) and middle-centred for the TOIs clustering (bottom right).

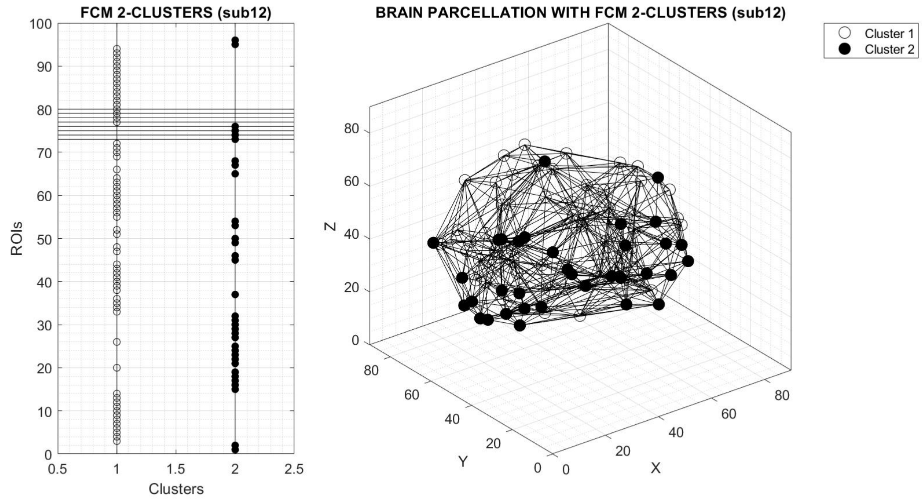

Figure 3.18: The figure shows the brain parcellation based on a graph model and the Regions of Interest (ROIs) organisation partitioned with two clusters. In the left plot the horizontal line indicates the Fusiform Cortexes (ROIs 73:80) and their clusters. In the right image the nodes are the 96 centroids according to Harvard-Oxford atlas. In general, the inferior regions were clustered in the Class 2 and the superior regions were clustered in the Class 1. The figure is referred specifically to the Regions of Interest (ROIs) of Subject 12 clustered with FCM.

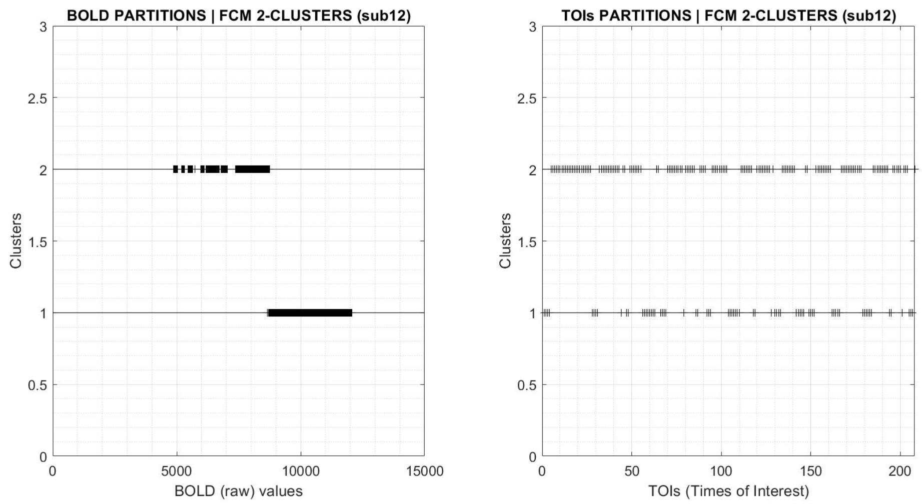

Figure 3.19: The figure shows the BOLD values partitioning among the two clusters and the Times of Interests (TOIs) partition among the two clusters. In the left plot, clusters exhibit the BOLD amplitude and in the right plot the clusters seem to exhibit spatio-temporal patterns. The figure refers specifically to Subject 12 clustered with FCM.

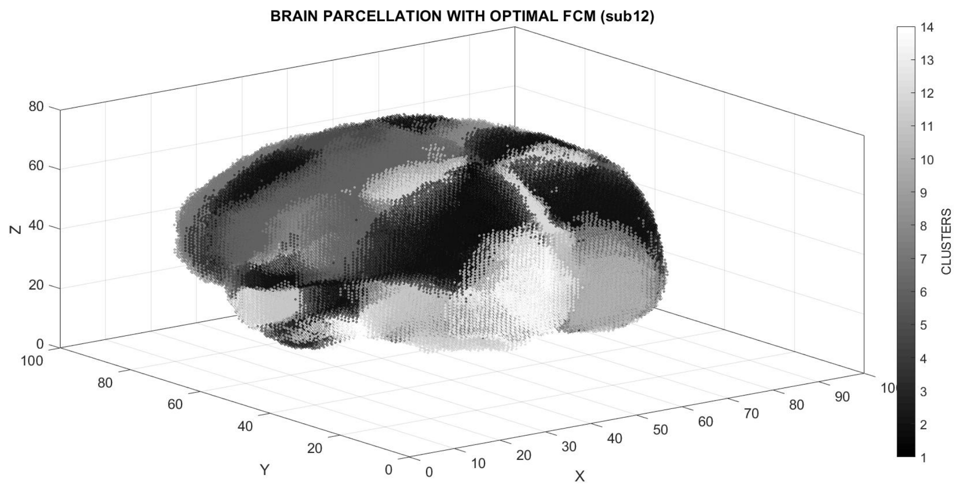

Figure 3.20: The figure shows the brain voxels parcellation of Subject 12 clustered with FCM using the optimal configuration (with the lesser Davies-Bouldin index). There are 14 clusters that covered all brain voxels (cfr. Figure 3.21 for more details).

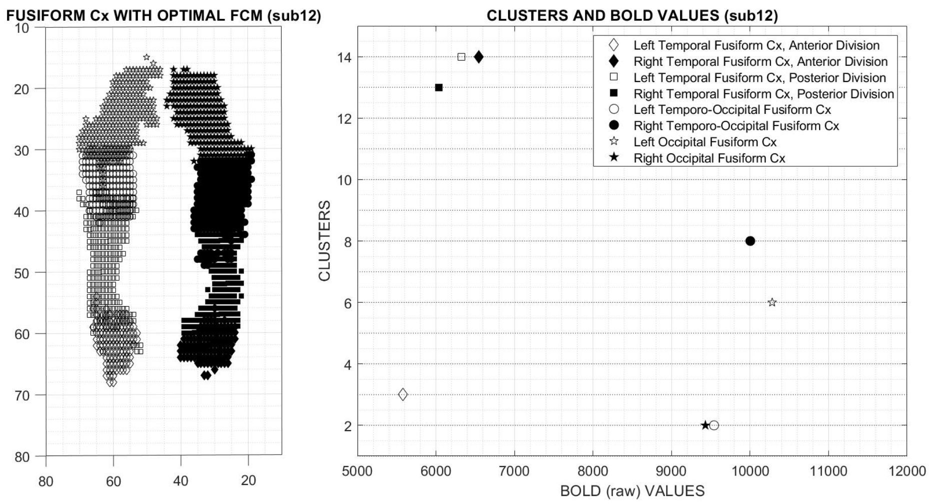

Figure 3.21: The left plot shows the substructures of the Fusiform regions (the legend is the same of the right plot). In the right plot there are the eight Fusiform regions distributed to different clusters according to their BOLD values: L/R Temporo-Occipital Fusiform Cx and L/R Occipital Fusiform Cx have greater BOLD values than the L/R Temporal Fusiform Cx Anterior and Posterior Divisions. Precisely, clusters 2, 6 and 8 contain the Fusiform ROIs with greater activations, whereas clusters 3, 13, 14 the ones with lesser activations.

from 73 to 80 includes the Fusiform brain regions (using the Harvard-Oxford labels, they are ROIs 73:74 L/R Temporal Fusiform Cx Anterior Divisions, ROIs 75:76 L/R Temporal Fusiform Cx Posterior Divisions, ROIs 77:78 L/R Temporo-Occipital Fusiform Cx and ROIs 79:80 L/R Occipital Fusiform Cx). This submatrix positive correlation is in agreement with the selective importance for the face recognition task of the Fusiform regions shown in the results of Wakeman and Henson [76].

The results presented in Figure 3.11, Figure 3.12, Figure 3.13 and Figure 3.14 reveal the important evidence that optimal clustering and 2-group clustering are both statistically different for the spatial configuration in the case of both FCM and NG for both the parametric and non-parametric test. Knowing that the two spatial clusters are sufficient to be statistically different is helpful for the comparison of the optimal clustering outcomes between subjects, that in our case leads a different number of clusters for subjects (see. Table 3.1). The use of 2-clusters (or a fixed-clusters) classifications allows to easly compare subjects for post-clustering analysis. Furthermore, spatial clustering has in general good properties, but temporal clustering does not have the same quality, due to the difficulty to find statistically different clusters, albeit FCM outperformed NG, but not always with p-value less than 0.05, as better results were achieved for 2-clusters setting.

The results presented in Figure 3.15 , Figure 3.16 and Figure 3.17 are regarding to the variability observed in the clustering results. Jaccard similarity matrices highlight huge differences within and between clustering algorithms. This fact could be explained with the random initialisation labels; then, although the clustering reached the optimal configuration, it could be the case that the i label is not assigned to the i pattern in all the subjects. In other words, the same pattern in different subjects could be labelled sometimes with label i or with label j.

Figures 3.18 and 3.19 present the results in the special case of the clustering computed for the Subject 12. It had the best behaviour for both the parametric and the non-parametric tests. The figures show the topological graph-based parcellation of FCM clustering in 2-group, where many inferior regions were clustered in the Class 1, whereas many superior regions were clustered in Class 2. The two spatial clusters differed two BOLD signal macro-levels, but the two temporal clusters do not have easily distinguishable differences, although they have in fact statistical differences probed by the non-parametric test.

Subject 12 was also studied with a voxel-based parcellation that has found 14 clusters with FCM in the case of ROIs clustering. Precisely, considerations about results in the Figure 3.20 and Figure 3.21 refer to the Fusiform regions classification. The eight regions involved in the Fusiform bilateral portions were correctly distinguished in zones with more activation that the others. This clustering evidences detailed the role of brain substructures particularly related to the Face Recognition task, confirming the specialized nature for the Fusiform cortexes, according, for example, to the results of Wakeman and Henson [76] that have shared the data we processed, and also the main related works (cfr. the seminal paper by Kanwisher [45], or the recent findings by Ghuman [30] and Grill-Spector [34])

3.2.6 Conclusion and Future Works

In this chapter we adopted a cross-clustering approach to fMRI data with the aim to cluster both spatial and temporal patterns, given that the main information related to brain activity, in the case of task-based paradigm, stems from both the anatomical regions with their BOLD temporal signatures and the ON/OFF blocks during an experiment in which they exhibit brain spatial response. More specifically, we processed fMRI images from a repository of images for face recognition. The selected images depict with 16 healthy subjects that did a Face Recognition Task. We investigated spatial (ROIs - Regions Of Interest) and temporal (TOIs - Times Of Interest) features, using Fuzzy C-Means (FCM) and Neural Gas (NG) algorithms to find similar and structured patterns. We validated the optimal clustering using the Davies-Bouldin index and we compared the different subjects outcomes using with Jaccard measure. We used parametrical and non-parametrical statistical tests to evaluate whether the differences between the classes are significant throughout clusters, using the p-value as a decision criterion. Also, we employed the functional connectivity analysis to explore the brain BOLD co-relations activities. This procedure is useful to understand the ROIs clustering meaning as it associates with the functional properties referred as task-based paradigm. The results showed that, in general, ROIs clustering was performed easier than TOIs clustering by either of the algorithms. However, in the case of TOIs clustering, which is more complex then the spatial one, Fuzzy C-Means method outperformed the Neural Gas Method, based on the statistical significance test.

This study has some limitations. The first one is the random assignment of starting clusters for each of the algorithms. The second limitation is the absence to known block paradigm that matches the TOIs clustering. This is a theoretical limitation, because it depends on the amount of information available in a repository. The first limitation is more important because it sheds light why when two clusters are used, the algorithm marks the same pattern with different labels (see Jaccard matrix). The results indicate that even if there is a correct classification of statistically different clusters, there isn't an easy way to compare the same class with objects from different datasets. In other words, there is no clustering consistency in terms of labels names for the subjects. This peculiarity could be overcome using a linguist procedure based on a formal description able to combine the labels of similar patterns. As future task, we plan to develop a translation procedure able to merge under one name different labels that are associated to similar patterns. Furthermore, we will investigate the cross-clustering more in detail. In this study, we clustered ROIs and TOIs and we tested the classes properties *uncoupled*. Using other statistical tests, e.g., the two-way ANOVA or Friedman test, we can evaluate whether the $coupled - classes$ are significantly different, i.e., we plan to test whether the clusters of ROIs combined with the clusters of TOIs have elements that are significantly different. This procedure would allow to find spatial patterns that are associated statistically with temporal patterns and *viceversa*. Combining the clustering results is a more precise exploration of the brain during task-paradigm, where the main features are both spatial dynamics (the regional signatures) and temporal dynamics (the ON-OFF blocks paradigm).

4

CVIs and Evaluation Methodologies Analysis for fMRI

In this section we will present the results of the analysis carried out on different evaluation indexes used on clustering performed on fMRI exams. A methodology for the evaluation of evaluation indexes will also be presented. The results achieved are organized and reported below:

- Analysis of the evaluation indexes with fMRI data, and introduction of a metric for the evaluation of the indexes in sub-optimal cases.
- Analysis of evaluation indexes with fMRI data and benchmark data, introduction of a new methodology through the use of ANFIS for the evaluation of indexes in sub-optimal cases.

4.1 Comparison of validity indexes for fuzzy clusters of fMRI data

4.1.1 Premise

This research work was presented at the ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing held in Porto. The main topic related to this work is a comparison analysis of eight representative fuzzy and crisp clustering validation indexes. This paper is the result of collaboration with other authors. For this work I have dedicated myself to the research design, to the choice, implementation and optimisation of the algorithms and indexes used, to the management of the experimental part and to the conception of the metric.

In computational neuroimaging, the analysis of functional Magnetic Resonance Images (fMRIs) using fuzzy clustering methods is a promising data driven approach to explore brain functional connectivity. In this complex domain, accurate evaluation procedures based on suitable indexes, able to identify optimal clustering results, are of great values strongly affecting the validity and interpretation of the overall fMRI data analysis. Salient aspects of the proposed strategy are the use of the widely adopted fuzzy c-means algorithm as underlying fuzzy clustering algorithm, the use of resting state fMRI data from the NITRC repository and the comparison and evaluation of eight of the most famous or innovative clustering evaluation indexes using clustering performed on fMRI exams; furthermore, the assessment of the indexes took into account the suboptimal performance provided by the indexes.

4.1.2 Introduction

Data Clustering is one of the widely used methods to explore data in several domains. It utilizes only the statistical information inherent in the data without human supervision [21]. Fuzzy clustering computes degrees of membership of a single data to multiple clusters. In computational neuroimaging, the analysis of functional Magnetic Resonance Images (fMRIs) using fuzzy clustering methods is a promising data driven approach to explore brain functional connectivity. fMRI data have a complex content that regards both spatial and temporal information: the spatial ones are related to the mapping of brain regions that have common topological properties, whereas the temporal ones are referred to the detection of brain signal changes in correspondence to specific experimental times (see 4.1). In this context, clustering techniques find homogeneous spatio-temporal patterns without relying on any model of functional response are considered in principle more accurate than model-based methods when dealing with fMRI data analysis under complicated experimental conditions [39, 79]. Clustering algorithms perform a partition of the complex fMRI content in homogeneous groups. Finding an optimised partition is a sophisticated task: not all the fMRI patterns are separable in distinguished crisp parcels since some of them could share common properties, as in the case of extended brain networks that vary the coactivation of different brain modules during an experimental task. Thus, the natural dynamic of the neuronal structures must be managed properly by clustering algorithms that should be able to handle both simple regularities of well-known patterns related to low-level active tasks and complex irregularities of partially-known patterns related to high-level active tasks or self-referred passive paradigms. Clustering has an important role in fMRI passive studies allowing to investigate the neurophysiological resting state that has debated biomarkers [39] and also evidence-based differences related both to gender and age [8].

In this context, accurate evaluation procedures based on suitable indexes able to identify optimal (and suboptimal) clustering results are of great values strongly affecting the validity and interpretation of the overall fMRI data analysis which is still a controversial task in neuroimaging. Among the varied methods used for fMRI data clustering, fuzzy c-means [6] is certainly the most popular method [39, 50, 73, 32]. An important issue in cluster analysis is the cluster validation aimed to measure how well the clustering results reflect the structure of the data set. For this purpose a large number of clustering validation indexes (CVIs) have been proposed in literature [60, 29, 66, 69, 82, 14, 73] to detect the optimal cluster number for a given dataset on the base of a balancing between the two opposite criteria of compactness within each cluster and separation between them. Several studies have been developed to investigate and compare the effectiveness of fuzzy and crisp CVIs in appropriately determining the number of clusters and measuring the goodness of clusters themselves produced by diverse algorithms [3, 36]. Despite several achievements obtained, guidelines resulting from these general studies have not yet been adopted with large consensus and validation indexes are often selected basing on individual experience and/or arbitrary criteria. Critical aspects arise also in fMRI data analysis where clustering techniques are usually validated using external criteria based on prior knowledge about the data, whenever possible, or using internal different indexes depending on individual studies.

The problem can be addressed by proposing comprehensive comparison studies oriented to specific clustering algorithm and specific application domains in such a way that resulting guidelines are applicable in future studies. Proceding from these considerations, in this work we focus the attention on validation of fuzzy clustering of fMRI data and develop a comparison analysis of a set of representative fuzzy and crisp CVIs. Salient aspects of the proposed strategy are the use of the widely adopted fuzzy c-means (FCM) algorithm as underlying clustering algorithm and the use of resting state fMRI data from the NITRC repository [46]. The remaining part of the paper is organised as follows: Section 2 describes the clustering problem and the soft algorithm chosen to approach its solution, Section 3 lists the indexes used to validate the clustering results, Section 4 describes the general experimental procedure, the datasets used and the results obtained. Section 5 reports both the discussion of the results and the conclusions.

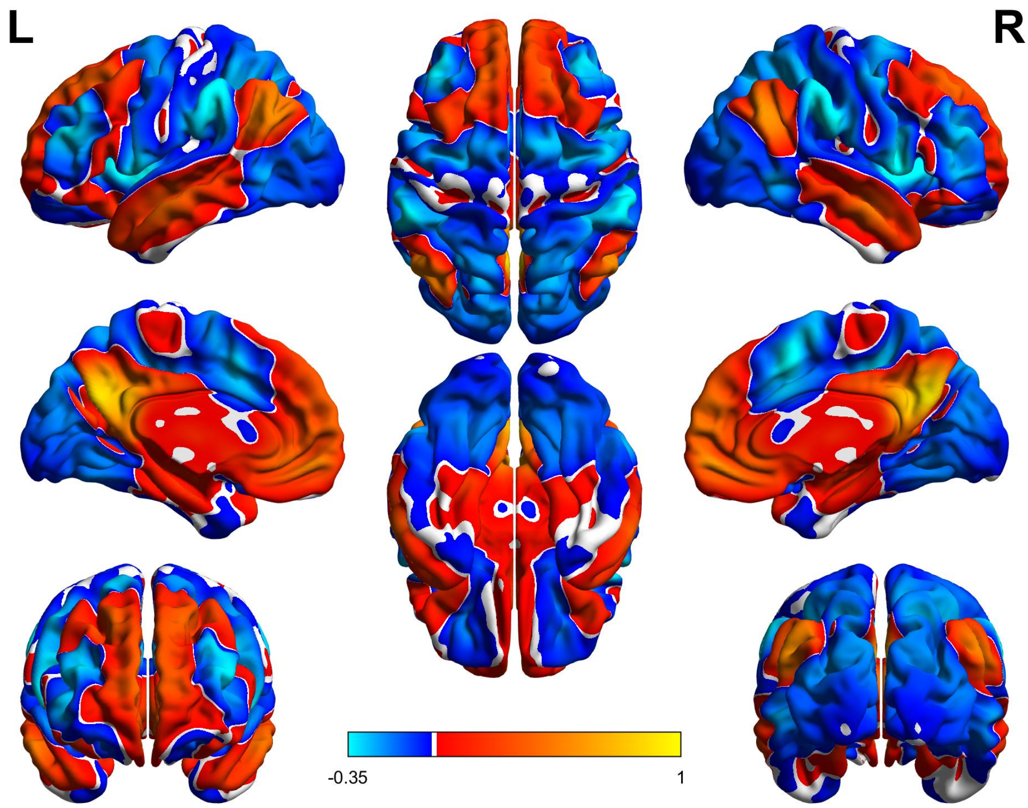

Figure 4.1: This image displays resting-state functional connectivity as linear correlation for the seed region in a sample of 1,000 subjects. The seed chosen is the Precuneus (X/Y/Z MNI152 coordinates: 2 -60 30), that is the main core of the Default Mode Network (DMN), a candidate biomarker for the fMRI resting state studies. In the images, the Precuneus is in the zone with the highest functionality (yellow color).

4.1.3 Clustering Problem and Fuzzy C-Means Algorithm

The purpose of clustering is to partition a given set of data into groups (clusters) following a predefined criterion. These groups contain data that have both high similarity within clusters and high dissimilarity between the other clusters [21].

Let $X = \{x_1, x_2, \dots, x_n\}$ a given dataset (with n elements), and let $C = \{c_1, c_2, ..., c_K\}$ the set of cluster, where K is the desired number of clusters. Regardless of the criterion chosen for the partition, the purpose of clustering is to develop a partition matrix of size $K \times n$ denoted as $U = [\mu_{ij}]$, with $i = 1, 2, \dots, K$ and $j = 1, 2, \dots, n$, where μ_{ij} is the grade of membership of point x_j to cluster c_i.

In crisp clustering, each point in the specified dataset belongs to a single cluster class. Then $\mu_{ij} = 1$ if $x_j \in c_i$, otherwise $\mu_{ij} = 0$. Instead, in fuzzy clustering, a point can be associated with more than one cluster, potentially also to all clusters, with a certain degree of membership, and the partition matrix in this case is represented as $U = [\mu_{ij}]$, where $\mu_{ij} \in [0, 1]$ indicates the degree of membership of the j-th element to the i-th cluster.

The FCM algorithm proposed by Bezdek [6] is used for the data analysis in a non-supervised way in several fields. The purpose of the FCM algorithm is to create vectors called centroids that minimise the value of the function J_m that is given by the sum of the intra-cluster quadratic error. J_m it is defined as:

$$J_m = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \tag{4.1}$$

where

- $m > 1$ is the exponent of the element of the fuzzy partition matrix to adjust the degree of fuzzy overlap.
- z_i is the centre of the i-th cluster.
- μ_{ij} is the degree of membership of x_j to the i-th cluster.
- $||...||$ is the Euclidean norm between a point and the corresponding cluster center.

The FCM algorithm performs the following steps:

1. Randomly initialise the U matrix.
2. Calculate the cluster centroids with the following formula :

$$z_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m (x_j)}{\sum_{j=1}^{n} (\mu_{ij})^m} \tag{4.2}$$

3. Update μ_{ij} according to the following formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{K} (\frac{||x_j - z_i||^2}{||x_j - z_k||^2})^{\frac{2}{(m-1)}}} \tag{4.3}$$

4. Calculate the objective function J_m
5. Repeat steps 2-4 until J_m improves less than the prefixed threshold or until the specified maximum number of iterations is reached.

4.1.4 Cluster Validation Indexes

The use of a clustering algorithm must be complemented with the use of a validation index to detect the optimal cluster number for a given input dataset. A clustering validity index has two indicators: the compactness and the separation [51]. The compactness indicates the concentration of points that share the same cluster. The separation, evaluates the degree of isolation among clusters. A dataset is well partitioned if there is both high compactness and high separation. But often the two indicators conflict, e.g., if the compactness is high, the separation is low and *viceversa*. Therefore, a *rationale* between the two indicators is needed to design a clustering validation index.

The aim of the present work is to identify suitable CVIs for fMRI Clustering studies among a set of representative and widely used crisp and fuzzy indexes. A total of eight indexes is considered and their formal definition given below.

- The Pakhira Bandyopadhyay Maulik Index (PBMI) [60]. It evaluates the product between compactness and separation and its optimal value is towards the maximum. It is formalised as

$$PBMI(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K\right)^2 \tag{4.4}$$

where K is the number of clusters used, i.e., $K = \{k^{'}, k^{''}, \ldots, k^{K}\}$, the $E_K = \sum_{k=1}^{K} E_k$ holds such that the compactness is defined as crisp functional

$$J(U, Z) = E_k = \sum_{n=1}^{N} u_{nk}||x_n - z_k|| \tag{4.5}$$

where $U(N) = [u_{nk}]_{N\times K}$ is the binary partition matrix of the clustered data and the crisp separation is formalized as

$$D_k = \max_{k',k''}^{K} \left\{||z_{k'} - z_{k''}||\right\} \tag{4.6}$$

with $z_{k'} \neq z_{k''}$ (that are different centroids). Note that x_n is the n-th point in the dataset and z_k is the center of the k-th cluster. $E_1 = \sum_{n=1}^{N} ||x_n - z_1||$ z_1 is the centroid calculated on all points of the dataset

- The FPBMI is the fuzzy version of the index proposed by Pakhira et al. [60]. It evaluates the product between compactness and separation and its optimal value is towards the maximum. It is formalised similar as in the equation (4.4), except for the compactness of all clusters that it is defined as fuzzy functional, i.e.,

$$J_m(U,Z) = E_k = \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||x_n - z_k|| \tag{4.7}$$

and E_1 that defined the fuzzy compactness of the cluster 1, i.e. $E_1 = \sum_{n=1}^{N} u_{n1}^m ||x_n - z_1||$. Both contain the membership value u_{nk}, where $U(N) = [u_{nk}]_{N\times K}$ is the fuzzy partition matrix of the clustered data.

- The Fukuyama Sugeno Index (FSI) [29]. It computes the difference between fuzzy compactness and fuzzy separation and its optimal value is towards the minimum, i.e.,

$$FSI(K) = \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||x_n - z_k||^2 - \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||z_k - \overline{z}||^2 \tag{4.8}$$

in which the $\overline{z}$ is the mean of all Z centroids and the u_{nk} is the membership value of the n-th point in the k-th cluster, and m is the fuzzy exponent.

- The Rezaee Lelieveldt Reider Index (RLRI) [66], also known as Compose Within and Between scattering Index (CWBI). It is the sum of compactness and separation and its optimal value is towards the minimum. RLRI assesses the average compactness and separation of fuzzy clustering by using the sum of two functions, i.e.,

$$RLR(K) = \alpha Scat(K) + Dis(K), \tag{4.9}$$

where α is a weighting factor equals to $Dis(K_{max})$ (the $Dis(K)$ with the maximum cluster number), and $Scat(K)$ that is the clustering compactness measure defined as

$$Scat(K) = \frac{\frac{1}{K}\sum_{k=1}^{K} ||\sigma^2(z_k)||}{||\sigma^2(X)||} \tag{4.10}$$

with $||x|| = (x^T \cdot x)^{1/2}$. Note that $\sigma^2(X)$ denotes the variance of all the dataset X and $\sigma^2(z_k)$ is the fuzzy variance of cluster k. The $Dis(K)$ is the clustering separation measure defined as

$$Dis(K) = \frac{D_{max}}{D_{min}} \sum_{k=1}^{K} \left[\sum_{k=1}^{K} ||z_{k'} - z_{k''}|| \right]^{-1} \tag{4.11}$$

with $z_{k'} \neq z_{k''}$ (different k centroids) and with Dis_{max} and Dis_{min} are the clustering separation with the maximum and minimum cluster number respectively.

- The Wang Sun Jiang Index (WSJI) [69]. It is the sum of compactness and separation and its optimal value is towards the minimum. It derived from the RLRI, adopting a linear combination of average fuzzy compactness and separation to evaluate clustering outcomes, i.e.,

$$WSJI(K) = Scat(K) + \frac{Sep(K)}{Sep(K_{max})} \tag{4.12}$$

where the separation $Sep(K)$ is differently defined as in equation 4.11, i.e.,

$$Dis(K) = \frac{D^2_{max}}{D^2_{min}} \sum_{k=1}^{K} \Big[\sum_{k=1}^{K} ||z_{k'} - z_{k''}||^2 \Big]^{-1}. \tag{4.13}$$

Instead, the $Scat(K)$ is the defined as in equation (4.10).

- The Xie Beni Index (XBI) [82]. It is the *ratio* between compactness and separation and its optimal value is toward the minimum. It measures the average within cluster fuzzy compactness *versus* the minimal value of the between-clusters separation, i.e.,

$$XBI(K) = \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} u^2_{nk} ||x_n - z_k||^2}{N \cdot min_{k\prime \neq k\prime\prime} \{||z_{k\prime} - z_{k\prime\prime}||^2\}} \tag{4.14}$$

  with $K = \{k\prime, k\prime\prime, \ldots, k_K\}$ is the number of clusters used, N the number of data points, u_{nk} the membership values associated to the points n and a cluster k, the z_k is the centroid of a generic cluster k.

- The Davies Bouldin Index (DBI) [14]. It is the *ratio* between crisp compactness and separation and its optimal value is towards the minimum., i.e.,

$$DBI(K) = \frac{1}{K} \sum_{k=1}^{K} \max \Big\{ \frac{S_{k'} + S_{k''}}{||z_{k'} - z_{k''}||} \Big\} \tag{4.15}$$

  with $k\prime \neq k\prime\prime$ (different k centroids) and $S_{k'}$ the crisp clustering compactness of the $k' = k-$th cluster defined as

$$S_{k'} = \Big(\frac{1}{N_{k'}} \sum_{x_n \in k_i} ||x_n - z_{k'}||^2 \Big)^{1/2} \tag{4.16}$$

  where $N_{k'}$ is the cardinality of the cluster $k^{'}$.

- The SDBI is the soft (fuzzy) version of DBI [73]. It is the *ratio* between the fuzzy compactness and the separation and its optimal value is towards the minimum. It is defined as

$$SDBI(K) = \frac{1}{K} \sum_{k=1}^{K} \max \Big\{ \frac{S_{k'}\overline{U_{k'}} + S_{k''}\overline{U_{k''}}}{||z_{k'} - z_{k''}||} \Big\} \tag{4.17}$$

where the fuzzy compactness $S_{k'}$ is the defined as follow

$$S_{k'} = \Big(\frac{1}{N}\sum_{x_n \in N} ||x_n - z_{k'}||^2\Big)^{1/2} \tag{4.18}$$

in which N is the cardinality of the used datasets, whereas the $\overline{U_{k'}}$ is the average of the membership values for the cluster $k^{'}$ (note that $k^{'}$ and $k^{''}$ are different clusters).

4.1.5 Experiments and Results

Performances of the eight indexes introduced in Section 3 are evaluated using clustering results obtained by processing fMRI datasets with different configuration of FCM algorithm and comparing the optimal number of clusters indicated by the indexes with those indicated by the available ground truth.
Taking as an example a single subject and a single CVI, the experiment is performed as follows:

- Step 1: the dataset is passed in the FCM algorithm input initially set with the value $K = 2$ and the value $m = 1.1$.
- Step 2: the algorithm creates K clusters and calculates the J_m value. This step is repeated 200 times, and the clustering with the lower J_m value is saved.
- Step 3: The clustering obtained in Step 2 is passed in input to the CVI that carries out the evaluation, and the evaluation is saved.
- Step 4: You return to Step 1, setting the K value as $K = K + 1$, and this is repeated until $K <= 10$, if $K > 10$ go to the next step.
- Step 5: A normalisation is applied to the evaluations through z-scores to make them comparable with those obtained from the other indices.
- Step 6: You go back to Step 1, setting the value $K = 2$ again, and setting the value m as $m = m + 0.1$, and this is repeated until$m <= 2.5$, if $m > 2.5$ go to the next step.
- Step 7: For each of the 15 tables containing the evaluations from $K = 2$ to $K = 10$, parameter E is calculated first with a two-class truth, then with a 4-class truth.

4.1.6 fMRI dataset

From the NITRC repository [46] and 1000 Functional Connectome Project, we selected the Beijing dataset with 187 healthy subjects (73M /114F; ages 18-25; all righthanded). The subjects did a resting state experimental paradigm with eyes closed. The fMRI parameters were the following: TR = 2, slices = 33 acquired with interleaved ascending procedure, time-points = 225, magnet = 3 [T]. The selection of this dataset is motivated by the specific age range and because it was used by Biswal et al. [8] to discover resting state functional properties and their gender determinants. The brain resting state measured with fMRI has a bunch of possible biomarkers that allow researchers to build a likely ground truth (or experimental-based ground truth). The common knowledge about those biomarkers are presented in [39, 8, 50]. Since we want to get an empirical ground truth to validate the indexes, we defined it taking in account the acquired common knowledge about resting state fMRI biomarkers, obtaining a two classes ground truth and a four classes ground truth. The first has two labels associated to the presence/absence of regions related to the so-called Default Mode Network (DMN) [22] and the second has four labels associated to regions part of DMN and other three candidate resting networks, i.e., the Visual Network (VN), the Sensory/Motor Network (SMN) and the Other Resting Networks (ORN) (the last one encompasses all the regions that are not classified as DMN, VN or SMN).

4.1.7 Experiments

Two experiments have been developed by using fMRI data. In the first experiment two classes of truth are considered: what is DMN network and what is not. In the second experiment, 4 classes are considered: DMN network, VN network, SMN network and other resting networks. The FCM algorithm was configured with number of clusters $K = 2, 3, .., \sqrt{n}$ and weighting exponent m =1,1.2,. . . ,2.5. To improve robustness in the evaluation, each FCM implementation was executed 200 times for each configuration and clustering result having the lowest Jm value was considered for the CVIs evaluation. In both the experiments the eight CVIs were applied to evaluate clustering results obtained by the allowed FCM implementations distinguished by the different values of K and m parameters. To enable the quantitative comparison analysis, CVIs values were normalised taking into consideration the fact that some indexes designate the optimal number of clusters by using the maximum value, while the others the minimum value. In particular, the z-score normalisation has been implemented in a positive way for the indexes that minimise their optimal value, and in a negative way for the indexes that maximise their optimal value. After normalisation, the indexes indicated the number of optimal clusters with the lowest value, making them to be well comparable. Table 1 illustrates the CVIs values resulting from the evaluation of clustering fMRI dataset by FCM with $m = 2$ and i ranging from 2 to 10.

Table 4.1: Values of CVIs resulting from the evaluation of clustering fMRI dataset by FCM with m=2 and K ranging from 2 to 10.

Index	K=**2**	K=**3**	K=**4**	K=**5**	K=**6**	K=**7**	K=**8**	K=**9**	K=**10**
FPBMI	1.66	1.59	-1.66	-0.12	1.95	-0.44	0.62	1.62	-0.12
PBMI	1.02	0.89	-1.19	-0.47	0.69	-0.67	0.93	1.12	-0.44
FSI	0.75	0.81	-0.08	-0.53	0.07	-0.64	0.93	-0.42	-0.54
WSJI	-0.27	0.02	0.02	-0.44	0.05	-0.35	1.09	-0.95	-0.61
XBI	-0.73	-0.90	0.11	-0.46	-0.81	-0.40	-0.61	-0.63	-0.41
RLRI	-0.82	-0.98	0.65	-0.31	-0.96	-0.11	-0.57	-1.05	-0.63
DBI	-0.92	-0.97	0.93	-0.07	-1.03	0.28	-0.94	-0.29	0.46
SDBI	-0.67	-0.46	1.21	2.43	0.02	2.35	-1.44	0.61	2.30

To summarise the set of results generated and develop systematically a comparative evaluation of CVIs, we introduced a measure E defined as:

$$E = |ni - nr| \tag{4.19}$$

where ni is the optimal number of clusters designated by the index, nr the number of cluster by reference. Table 2 illustrates performance of E values of the 8 CVIs, computed as average of E values obtained varying parameter m in the two experiments mentioned above.

Table 4.2: Mean and variance of E values for the 8 index evaluating clustering of fMRI data with 2 (Experiment 1) and 4 (Experiment 2) reference classes, the CVIs are in ascending order based on the E mean.

Index	Experiment 1		Index	Experiment2	
	E mean	Var		E mean	Var
FSI	0.58	0.13	WSJI	0.78	0.08
RLRI	0.65	0.18	RLRI	1.35	0.17
WSJI	1.28	0.10	FSI	1.47	0.08
SDBI	1.54	1.50	SDBI	1.55	0.49
DBI	4.76	5.02	DBI	3.31	2.35
XBI	5.50	0.40	XBI	3.69	0.39
PBMI	6.22	0.42	PBMI	4.22	0.42
FPBMI	6.36	0.25	FPBMI	4.36	0.25

4.1.8 Discussion and Conclusions

In this work the performance of eight well-known CVIs was quantitatively evaluated by using the FCM algorithm to process fMRI data. The use of the selected dataset allows to investigate the behaviour of CVIs under two different levels of organising data in two and four reference classes. The results obtained are preliminary but useful to suggest guidelines for a reliable use of cluster evaluation indexes and to contribute to a proper use of data driven, clustering techniques in the complex and more and more investigated brain function evaluation domain. Looking into the details of the results listed in Table 4.2, we noticed that RLRI, WSJI and FSI gained the top three positions in both the experiments even if with a different internal order. This fact leads to the conclusion that each one of them is able to both mediate between different characteristics of cluster structures and efficiently create a balance between compactness and separation. It was found also that widely used indexes such XBI, DBI and PBMI showed values considerably lower than the three indexes mentioned above. The major differences between the two sets of CVIs lie in the formalisation of separation component that plays an important role when dealing with clusters allocated closely as probably in case of fMRI data, and in the management of the two measures (compactness and separation) in the case of RLRI and WSJI is the sum of the two components, FSI subtraction while XBI, DBI, SDBI apply the ratio and FPBMI, PBMI the product. The novel SDBI index showed better values than crisp standard version and gained a position just below the top three positions. The XBI, PBM, FPBM, and DBI indexes seem to be more suitable for contexts in which data distribution with little overlap is hypothesised, or in which cluster compactness is preferred.

Main conclusions obtained by our experimental work are consistent with results obtained in previous works [51] while considering the different experimental strategies and different domains. However caution must be exercised when applying results to other fMRI contexts taking into account the variability and complexity of these data and the different processing strategies. Future work contemplates a refinement of the metric adopted in the comparison to include other evaluation criteria and the use of a broader set of fMRI data with different levels of complexity and inter-cluster overlap, to obtain which would be results more robust and extensible to other similar contexts.

4.2 Comparison of Validity Indexes for Fuzzy Clusters of fMRI data by Using Adaptive-Network Fuzzy Inference Systems

4.2.1 Premise

This research work is the extension of the paper presented at the ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing held in Porto [57]. This paper is the result of collaboration with other authors. For this work I have dedicated myself to the research design, to the choice, implementation and optimisation of the algorithms and indexes used, to the management of the experimental part, to the conception of metrics and to the training of the ANFIS model. This work focused on the introduction of new parameters for the evaluation of CVIs, and on the introduction of ANFIS to carry out the evaluation of clustering evaluation indexes. Furthermore, synthetic and benchmark datasets were also used for the evaluation of CVIS in addition to fMRI datasets.

In computational neuroimaging, functional magnetic resonance imaging (fMRI) analysis using fuzzy clustering methods is arousing a lot of interest, providing very interesting results. In a very complex context such as functional neuroimaging, index-based assessments play a fundamental role in obtaining a correct analysis. In this article, a comparative analysis of eight validation indexes chosen from the most famous / innovative will be exposed. The fuzzy c means algorithm, already widely used in the fuzzy analysis of fMRI exams, will be used as the clustering algorithm for the analysis of fMRI exams. For the evaluation of the indexes, metrics will be introduced, which will make it possible to provide an evaluation in which even sub-optimal results obtained by the indexes will have a weight. Furthermore, ANFIS will be used with the aim of creating rules for the evaluation of the indexes, and obtaining an evaluation through a linguistic label. Furthermore, we wanted to expand this study on indices to different domains, and we did so through the use of synthetic and benchmark dataset, as well as fMRI dataset, which remain the main interest.

4.2.2 Introduction

Data clustering is one of the most used data analysis methods to explore data in different domains, is a methodology that uses only the statistical information inherent in the data without the need for human supervision [21].
The grouping carried out by clustering can be of the fuzzy or crisp type, in the crisp case the algorithm assigns a datum to a single cluster. In contrast, in the fuzzy case, a datum can belong with different degrees of belonging to several clusters.

Given its flexibility, clustering is used in several domains, including marketing, insurance, city-planning, earthquake studies, and computational neuroimaging. In functional magnetic resonance imaging (fMRI) analysis, the use of fuzzy clustering techniques is showing great promise. Thanks to its characteristic of being a data-based approach, it appears to be a suitable approach to explore the functional connectivity of the brain. An fMRI exam provides both spatial information and temporal information, spatial information is related to the mapping of brain regions, while temporal information is related to detecting changes in the brain signal at specific experimental times.

The results of clustering techniques in fMRI analyses are more accurate than the results obtained from model-based methods, especially when it comes to analyses performed under complicated experimental conditions [39, 79].

Given the complexity of fMRI data, finding the optimal partition for clustering is not an easy task. Not all fMRI models are easily separable, as some of them may share common properties, an example being the extended brain networks that vary the coactivation of different brain modules during an experimental activity. Clustering must therefore overcome the challenge of managing the natural dynamics of neuronal structures, this implies the management of the regularities of known low-level patterns, but also the irregularities related to high-level active tasks or passive self-referential paradigms.

Clustering plays an important role in passive fMRI studies, it has allowed to investigate the neurophysiological resting state with the discussion of biomarkers [39], and to highlight evidence-based differences relating to both sex and age [8].

Among the various algorithms for analysing fMRI data, fuzzy c-mean (FCM) [6] is certainly the most popular method [39, 50, 73, 32] for its ability to handle complex data. One of the main problems in clustering analysis is represented by cluster validation. Clusters are created by the algorithm by aggregating similar elements, but to evaluate whether this aggregation is optimal, an external agent is needed. There are various techniques for the evaluation of clustering, but the most used is that of the cluster validation indexes (CVIs) [60, 29, 66, 69, 82, 14, 73], these algorithms evaluate the clustering carried out through the computation of Compactness (the degree of similarity between the elements belonging to the same cluster) and of the Separation (the degree of dissimilarity between the elements of the various clusters).

This is a very important topic, numerous studies have been conducted to analyse and compare the performance of the various CVIs [3, 36]. Despite the numerous studies there are still no rules that provide based on the type of data, which index is the most appropriate, so the validation indexes are often selected on the basis of individual experience, based on arbitrary criteria or using both. These criticalities also arise in the analysis of fMRI data, in which clustering techniques often find validation through previous knowledge of the data, or on the basis of studies using internal indexes.

To overcome this problem, a comparative study was carried out on the performance of CVIs both in the fMRI field and in more general areas through synthetic and benchmark data. For the evaluation of CVIs the FCM algorithm will be used and metrics will be introduced that will allow the evaluation not only of the optimal performance of the CVIs, but also of the sub-optimal ones. An adaptive neuro-fuzzy inference system (ANFIS) will be used to combine the evaluations obtained from the metrics, and have evaluations in linguistic labeling form.

4.2.3 Clustering problem

Clustering is part of unsupervised learning techniques. The clustering technique consists of dividing a specific set of data into subgroups called clusters; clusters are created following a predefined policy. Clustering being an unsupervised technique, does not need previous knowledge on the data but is driven by data.

The clusters contain data that appear to have a high similarity to each other, and at the same time, present a high dissimilarity with the data present in the other clusters [21]. Clustering can be formally defined as follows: Let $X = \{x_1, x_2, \ldots, x_n\}$ a given dataset (with n elements) and $C = \{C_1, C_2, ..., C_K\}$ the collection of clusters, where K is the desired number of clusters. Regardless of the partitioning policy used, clustering requires the creation of a partition matrix of size $K \times n$ defined as $U = [\mu_{ij}]$, with $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots, n$, where μ_{ij} is the grade of membership of x_j point to C_i cluster.

Clusters can be crisp type or fuzzy type, in a crisp clustering, each element of the specified data set will belong to a single cluster. So $\mu_{ij} = 1$ if $x_j \in C_i$, otherwise $\mu_{ij} = 0$. In a fuzzy clustering, each element of the dataset can also be associated with all clusters, and for each cluster to which it is associated it, will have a degree of membership. The partition matrix in this case is represented as $U = [\mu_{ij}]$, where $\mu_{ij} \in [0, 1]$ indicates the degree of belonging to j -th to cluster i -th.

There are several clustering algorithms for both crisp clustering and fuzzy clustering; in this paper, we chose to carry out the analyses using a fuzzy clustering algorithm, the Fuzzy C-Mean (FCM) [6] already used in other research on fMRI data.

4.2.4 Cluster validation index

The use of a clustering algorithm also involves the use of a criterion for the validation of the clustering performed. There are three types of criteria to evaluate the partitioning performed by the clustering algorithm:

- External criteria: in which the performance of a clustering is measured by comparing a priori information.

- Internal criteria: they measure the performance of a clustering using only the data

- Relative criteria: the results of two or more clustering algorithms applied to the same dataset are compared.

In this paper, we will use the internal criterion through clustering validation indexes (CVIs). The CVIs adopt two indicators: i) Compactness and ii) Separation [51] to carry out their evaluation. The Compactness scale probes the density of the elements within the calculated clusters, in other words, it is the measure used to estimate the similarity between the elements sharing the same cluster. The Separation probes the isolation between the computed clusters, in other words, it measures the diversity of the elements of a cluster compared to the elements present in the other clusters. A grouping result should be defined as optimal if there is a high Compactness and a high Separation.

Therefore, a dataset is said to be optimally split if both Compactness and Separation are satisfied. The disadvantage is that very often, the two indicators conflict, for example if the Compactness is high and the separation is low and *vice versa*. Therefore, a *rationale* between the two indicators is needed to design a clustering validation index.

The eight CVIs used in this work will now be exposed, and these CVIs use different *rationale* to make their evaluations.

- The Pakhira Bandyopadhyay Maulik Index (PBMI) [60]. It evaluates the product between compactness and separation and its optimal value is towards the maximum. It is formalized as

$$PBMI(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K\right)^2 \tag{4.20}$$

  where K is the number of clusters used, i.e., $K = \{k^{'}, k^{''}, \ldots, k^{K}\}$, the $E_K = \sum_{k=1}^{K} E_k$ holds such that the compactness is defined as crisp functional

$$J(U, Z) = E_k = \sum_{n=1}^{N} u_{nk} ||x_n - z_k|| \tag{4.21}$$

  where $U(N) = [u_{nk}]_{N \times K}$ is the binary partition matrix of the clustered data and the crisp separation is formalised as

$$D_k = \max_{k', k''}^{K} \left\{ ||z_{k'} - z_{k''}|| \right\} \tag{4.22}$$

  with $z_{k'} \neq z_{k''}$ (that are different centroids). Note that x_n is the n-th point in the dataset and z_k is the center of the k-th cluster. $E_1 = \sum_{n=1}^{N} ||x_n - z_1||$ z_1 is the centroid calculated on all points of the dataset

- The FPBMI is the fuzzy version of the index proposed by Pakhira et al. [60]. It evaluates the product between compactness and separation and its optimal value is towards the maximum. It is formalised similar as in the equation (4.20), except for the compactness of all clusters that it is defined as fuzzy functional, i.e.,

$$J_m(U,Z) = E_k = \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||x_n - z_k|| \tag{4.23}$$

and E_1 that defined the fuzzy compactness of the cluster 1, i.e., $E_1 = \sum_{n=1}^{N} u_{n1}^m ||x_n - z_1||$. Both contain the membership value u_{nk}, where $U(N) = [u_{nk}]_{N\times K}$ is the fuzzy partition matrix of the clustered data.

- The Fukuyama Sugeno Index (FSI) [29]. It computes the difference between fuzzy compactness and fuzzy separation and its optimal value is towards the minimum, i.e.

$$FSI(K) = \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||x_n - z_k||^2 - \sum_{k=1}^{K}\sum_{n=1}^{N} u_{nk}^m ||z_k - \overline{z}||^2 \tag{4.24}$$

  in which the $\overline{z}$ is the mean of all Z centroids and the u_{nk} is the membership value of the n-th point in the k-th cluster, and m is the fuzzy exponent.

- The Rezaee Lelieveldt Reider Index (RLRI) [66], also known as Compose Within and Between scattering Index (CWBI). It is the sum of compactness and separation and its optimal value is towards the minimum. RLRI assesses the average compactness and separation of fuzzy clustering by using the sum of two functions, i.e.,

$$RLR(K) = \alpha Scat(K) + Dis(K), \tag{4.25}$$

  where α is a weighting factor equals to $Dis(K_{max})$ (the $Dis(K)$ with the maximum cluster number), and $Scat(K)$ that is the clustering compactness measure defined as

$$Scat(K) = \frac{\frac{1}{K}\sum_{k=1}^{K} ||\sigma^2(z_k)||}{||\sigma^2(X)||} \tag{4.26}$$

  with $||x|| = (x^T \cdot x)^{1/2}$. Note that $\sigma^2(X)$ denotes the variance of all the dataset X and $\sigma^2(z_k)$ is the fuzzy variance of cluster k. The $Dis(K)$ is the clustering separation measure defined as

$$Dis(K) = \frac{D_{max}}{D_{min}} \sum_{k=1}^{K} \left[\sum_{k=1}^{K} ||z_{k'} - z_{k''}|| \right]^{-1} \tag{4.27}$$

  with $z_{k'} \neq z_{k''}$ (different k centroids) and with Dis_{max} and Dis_{min} are the clustering separation with the maximum and minimum cluster number respectively.

- The Wang Sun Jiang Index (WSJI) [69]. It is the sum of compactness and separation and its optimal value is towards the minimum. It derived from the RLRI, adopting a linear combination of average fuzzy compactness and separation to evaluate clustering outcomes, i.e.,

$$WSJI(K) = Scat(K) + \frac{Sep(K)}{Sep(K_{max})} \tag{4.28}$$

where the separation $Sep(K)$ is differently defined as in equation 4.27, i.e.,

$$Dis(K) = \frac{D^2_{max}}{D^2_{min}} \sum_{k=1}^{K} \Big[\sum_{k=1}^{K} ||z_{k'} - z_{k''}||^2 \Big]^{-1}. \tag{4.29}$$

Instead, the $Scat(K)$ is the defined as in equation (4.26).

- The Xie Beni Index (XBI) [82]. It is the *ratio* between compactness and separation and its optimal value is toward the minimum. It measures the average within cluster fuzzy compactness *versus* the minimal value of the between-clusters separation, i.e.,

$$XBI(K) = \frac{\sum_{k=1}^{K} \sum_{n=1}^{N} u^2_{nk} ||x_n - z_k||^2}{N \cdot min_{k\prime \neq k\prime\prime} \{||z_{k\prime} - z_{k\prime\prime}||^2\}} \tag{4.30}$$

with $K = \{k\prime, k\prime\prime, \ldots, k_K\}$ is the number of clusters used, N the number of data points, u_{nk} the membership values associated to the points n and a cluster k, the z_k is the centroid of a generic cluster k.

- The Davies Bouldin Index (DBI) [14]. It is the *ratio* between crisp compactness and separation and its optimal value is towards the minimum., i.e.,

$$DBI(K) = \frac{1}{K} \sum_{k=1}^{K} \max \Big\{ \frac{S_{k'} + S_{k''}}{||z_{k'} - z_{k''}||} \Big\} \tag{4.31}$$

with $k\prime \neq k\prime\prime$ (different k centroids) and $S_{k'}$ the crisp clustering compactness of the $k' = k-$th cluster defined as

$$S_{k'} = \Big(\frac{1}{N_{k'}} \sum_{x_n \in k_i} ||x_n - z_{k'}||^2 \Big)^{1/2} \tag{4.32}$$

where $N_{k'}$ is the cardinality of the cluster $k^{'}$.

- The SDBI is the soft (fuzzy) version of DBI [73]. It is the *ratio* between the fuzzy compactness and the separation and its optimal value is towards the minimum. It is defined as

$$SDBI(K) = \frac{1}{K}\sum_{k=1}^{K} \max\{\frac{S_{k'}\overline{U_{k'}} + S_{k''}\overline{U_{k''}}}{||z_{k'} - z_{k''}||}\} \tag{4.33}$$

where the fuzzy compactness $S_{k'}$ is the defined as follow

$$S_{k'} = \Big(\frac{1}{N}\sum_{x_n \in N} ||x_n - z_{k'}||^2\Big)^{1/2} \tag{4.34}$$

in which N is the cardinality of the used datasets, whereas the $\overline{U_{k'}}$ is the average of the membership values for the cluster $k^{'}$ (note that $k^{'}$ and $k^{''}$ are different clusters).

4.2.5 ANFIS

ANFIS is a class of adaptive networks that incorporate both neural networks and fuzzy logic principles. In fuzzy logic, the control signal is generated by the activation of the rule base. In ANFIS, the rule base is selected using neural network techniques. Through the approximation of a non-linear system by setting IF-THEN rules inherited from the fuzzy logic, ANFIS is able to obtain excellent performance in various application areas, in fact ANFIS is defined as a universal estimator[48]. ANFIS draws its origins from the fuzzy logic theorem proposed by Zadeh [85] to describe complicated systems, this theorem, over the years has been used successfully in various fields, eg. manufacturing technique, decision making, pattern recognition, diagnostics, data analysis [11, 35].

ANFIS is essentially a rule-based fuzzy logic model whose rules are developed during the model training process trougth a data-driven process . ANFIS builds a fuzzy inference system (FIS) whose membership function parameters are derived from the training examples. An example of an ANFIS system will now be exposed, assuming that there are two inputs: x and y. Two fuzzy if–then rules for a first-order Sugeno fuzzy model can be expressed as follows

$$Rule1 : If(x\,is\,A_1)\,and\,(y\,is\,B_1)\,then\,(f_1 = p_1x + q_1y + r_1)$$

$$Rule2 : If(x\,is\,A_2)\,and\,(y\,is\,B_2)\,then\,(f_2 = p_2x + q_2y + r_2)$$

Where x and y represent the input data, A_1 and B_1 are the fuzzy sets, while f_i represents the output, which is within the fuzzy area specified by the fuzzy rule. The variables p_i, q_i and r_i correspond to the design parameters calculated during the training process.

In figure 4.2 the ANFIS architecture used to implement the two rules of the example, we note that it is divided into five layers that will now be analysed in detail.

Level 1: Input nodes; each node of this level assigns the degree to which the input

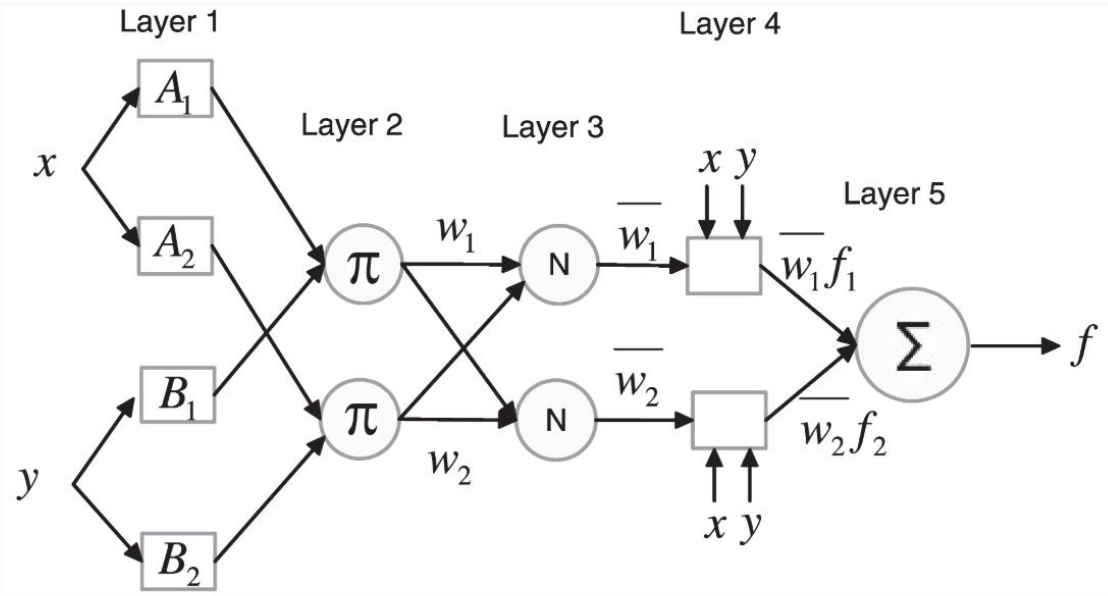

Figure 4.2: In the figure the ANFIS architecture with the five layers, the fixed type nodes are represented with squares, instead of the adaptive nodes with circles.

belongs to the membership function. These nodes generate the membership degrees to which the appropriate fuzzy sets belong to use the membership functions. They are determined as follows:

$$O_i^1 = \mu_{A_i}(x) \quad i = 1, 2 \tag{4.35}$$

$$O_i^1 = \mu_{B_{i-2}}(y) \quad i = 3, 4 \tag{4.36}$$

In the first level, all nodes are adaptive nodes,$\mu_{A_i}(x)$ and $\mu_{B_{i-2}}(y)$ adopt the chosen membership function,if for example a bell-shaped membership function was used, $\mu_{A_i}(x)$ is defined as follows:

$$\mu_{A_i}(x) = \frac{1}{1 + \left\{ \left(\frac{x - c_i}{a_i} \right)^2 \right\}^{b_i}} \tag{4.37}$$

where a_i, b_i and c_i are the parameters for bell-shaped functions. Analysing the second layer, we find nodes of a fixed type, which have the function of a multiplier, and are represented as follows:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(x) \quad i = 1, 2 \tag{4.38}$$

In the third level, we still have fixed nodes, which perform the function of normalising the nodes of the previous level, and are represented as:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{4.39}$$

In the fourth layer, the output of each node in this layer is calculated as the contribution of each of the rules, the nodes are adaptive and are defined as follows:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \quad i = 1, 2 \tag{4.40}$$

In the fifth and last level there is an adder node (S) which, as can be guessed, performs the sum of the values obtained from the previous layer. It is a static type node and the output it generates is calculated as follows:

$$O_i^5 = \sum_{i=1}^{2} \overline{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2} \tag{4.41}$$

Note that there are two adaptive levels in this ANFIS architecture, in the first layer where there are three modifiable parameters a_i, b_i, c_i, and in the fourth layer where we find the three modifiable parameters p_i, q_i, r_i. The parameters of the first layer a_i, b_i, c_i are called premise parameters, while the parameters of the fourth layer p_i, q_i, r_i are called consequent parameters. The learning algorithm aims to find the values of a_i, b_i, c_i and p_i, q_i, r_i that match ANFIS output to that seen in training.

ANFIS uses two different learning algorithms, back-propagation and hybrid methods, which aim to minimize the error between observed and predicted data [84].

4.2.6 Experiments and Results

For the experiments on the evaluation of CVIs, three different types of data sets were used, and the fuzzy c-mean was selected as the algorithm for the creation of the clusterings that will be evaluated by the CVIs. Three metrics will be introduced to allow evaluation of sub-optimal performances by CVIs, and an ANFIS model will be introduced to classify the evaluations obtained. Taking a single dataset and a single CVI as an example, the experiment is performed as follows:

- Step 1: the dataset is passed into the FCM algorithm input set with the value$K = 2$ and the value $m = 2.0$.

- Step 2: the algorithm creates K clusters and calculates the J_m value. This step is repeated 200 times, and the grouping with the lowest J_m value is saved.

- Step 3: The grouping obtained in Step 2 is passed as input to the CVI that performs the evaluation, and the evaluation is saved.

- Step 4: you go back to Step 1, setting the K value as $K = K + 1$, and this is repeated until $K <= 10$, if $K > 10$ you go to the next step.
- Step 5: A normalisation is applied to the evaluations through z-scores to make them comparable with those obtained from the other indices.
- Step 6: From the table containing the evaluations from $K = 2$ to $K = 10$, the parameters E_1, E_2, E_3 are calculated through the truth relating to the dataset, in the case of the fMRI data the truth is calculated as described in the paragraph.
- Step 7: The parameters E_1, E_2, E_3 are passed in input to ANFIS, which carries out the evaluation.

For ANFIS training, we proceeded as follows:

- Step 1: random creation of a dataset containing 1000 elements and three features corresponding to E_1, E_2, E_3.
- Step 2: evaluation of each element by an expert.
- Step 3: the creation of 30 different models by varying the internal parameters of ANFIS.
- Step 4: model training using750 data as training and 250 as test.
- Step 5: Comparison of the results of the various models, and choice of the model with the best performance

Fuzzy C-means

The FCM algorithm proposed by Bezdek [6] is used for the analysis of unsupervised data in various fields, and is one of the most used fuzzy clustering algorithms. The purpose of the FCM algorithm is to create vectors called centroids that minimise the value of the function J_m which is obtained by the sum of the quadratic error calculated between the points present within the cluster. J_m is therefore defined as defined as:

$$J_m = \sum_{j=1}^{N} \sum_{i=1}^{K} \mu_{ij}^{m} ||x_j - z_i||^2 \tag{4.42}$$

Where:

- $m > 1$ is the exponent of the element of the fuzzy partition matrix to adjust the degree of fuzzy overlap. The fuzzy overlap defines how blurred the boundaries between clusters are, the greater the value of me the greater the summary of boundaries

- x_j is the data point j-th.
- z_i is the center of the i-th cluster.
- K is the number of clusters required.
- N is the number of elements contained in the dataset.
- μ_{ij} is the degree to which x_j belongs to the i -th cluster.Considering a given point x_j, the sum of its cluster membership values must give 1.
- $\|...\|$ is the Euclidean norm between a point and the center of the corresponding cluster.

The FCM algorithm does the following step:

1. Randomly initialises the U matrix.
2. Calculate the centroids of the cluster with the following formula:

$$z_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m(x_j)}{\sum_{j=1}^{n}(\mu_{ij})^m} \tag{4.43}$$

3. Update μ_{ij} according to the following formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{K}(\frac{||x_j - z_i||^2}{||x_j - z_k||^2})^{\frac{2}{(m-1)}}} \tag{4.44}$$

4. Calculate the objective function J_m
5. If J_m improves less than the set value or if the specified maximum number of iterations has been reached, the algorithm ends, otherwise steps 2 to 4 are repeated.

Datasets

Three different types of data sets were used for this analysis of the index behavior. The first type is composed of fMRI datasets, this dataset is very complex and with a large number of elements. The second type is represented by the benchmark datasets, these datasets are very heterogeneous, they have different degrees of complexity and number of elements. Finally, synthetic datasets were used whose we can control distribution and truth classes. The three types of datasets used will be described below and a representation of the synthetic and fMRI data is shown in Figure 4.3. The methodology used to obtain truth classes from fMRI datasets will also be introduced.

fMRI Dataset

From the NITRC repository [46], two subjects, one female and one male, part of the Beijing dataset were randomly selected. Beijing dataset contains in total of 187 healthy subjects (73M /114F; ages 18-25; all righthanded). All the subjects did a resting state experimental paradigm with eyes closed. The fMRI parameters were the following: TR = 2, slices = 33 acquired with interleaved ascending procedure, time-points = 225, magnet = 3 [T]. We indeed selected this dataset because the specific age range and for its salience as benchmark dataset in functional neuroimaging (see for example the work by Biswal et al. [8] that explored the resting-state functional properties and their gender determinants). Moreover, the brain resting state measured with fMRI has a bunch of possible biomarkers (or experimental-based ground-truth). We then computed the likely ground truth for both the subjects by using the shared knowledge about those biomarkers [39, 8, 50].

In the figure 4.3-b are presented the resting state fMRI images of both the subjects, displaying one volume in three different sections.

Benchmark Datasets

There are five benchmark datasets (Cancer, Glass, Iris, SimpleClass, Thyroid); these are very different from each other as a data type, this choice was made to have the most heterogeneous datasets possible, so that the evaluation of the indexes made on them provides a non-specialised evaluation but more general, in order to see which index is best suited to different types of data distribution.

- The Cancer dataset consists of biopsies on benign and malignant tumours. The dataset is a 699x9 matrix where on the row we have biopsies and on the columns the features.

  The number of classes is two because they are divided into benign tumours and malignant moods.

- The Glass dataset can be used to classifies glass either as window or non-window depending on the glass chemistry. This dataset is a 214x9 matrix where on the rows we have the samples and on the columns the features.

  This dataset has two classes, Window glass or Non-window glass.

- The Iris dataset can be used that classifies iris flowers into three species. The dataset is represented by a 150x4 matrix where on the rows we have iris and on the columns the features.

  This dataset has three-class, iris setosa, iris virginica and iris versicolor.

- The SimpleClass dataset consists of two-dimensional vectors that represent geometric shapes. This dataset is represented by a matrix 1000x2 where on the rows we have the examples and on the columns the characteristics that represent the coordinates in a 2D space.

  This dataset has four class, four geometrical figures.

- The thyroid dataset has patients with a clinical report on thyroid function. The dataset is a 7200x21 matrix where on the rows there are the patients, and on the columns, the characteristics are composed of 15 binary and 6 continuous features.

  The classes, in this case, are three: Normal (Not hyperthyroid), Hyperfunction, Subnormal functioning.

Synthetic Datasets

Two distinct groups of synthetic data sets were created, the first to analyse the data overlap and the second to explore the dynamics present due to a large number of data and clusters.

- Overlap Dataset: to test the performance of CVIs in cases where the data is strongly overlapped, we initially created a dataset consisting of 1000 elements and 4 features belonging to two different classes, where the data does not overlap.

  In total, five datasets of 1000 elements and four features have been created starting from the first in which the data is perfectly divisible; we arrive at the fifth where the data overlap heavily. Therefore, starting from a dataset without overlapping, an iterative function was used, which allowed to reach a high degree of data overlap. In detail, starting from the first dataset, we have created a second of identical dimensions and always from two classes of truth. However, in this case, the elements had a slight overlap, successively we have created a third dataset of the same size where the data shows more overlap, and so for a fourth and fifth data set. These datasets were called:

  - Without overlap
  - Minimal overlap
  - Mean overlap
  - Medium-high overlap
  - High overlap

- High Cluster Number Dataset: to analyse the behaviour in the case of a large number of clusters, a synthetic dataset of 60000 elements has been created. This dataset is composed of 60000 points in two-dimensional space, therefore with two features. The points are distributed in 60 classes of 1000 points each equally distributed.

The synthetic datasets are presented in the figure 4.3-a.

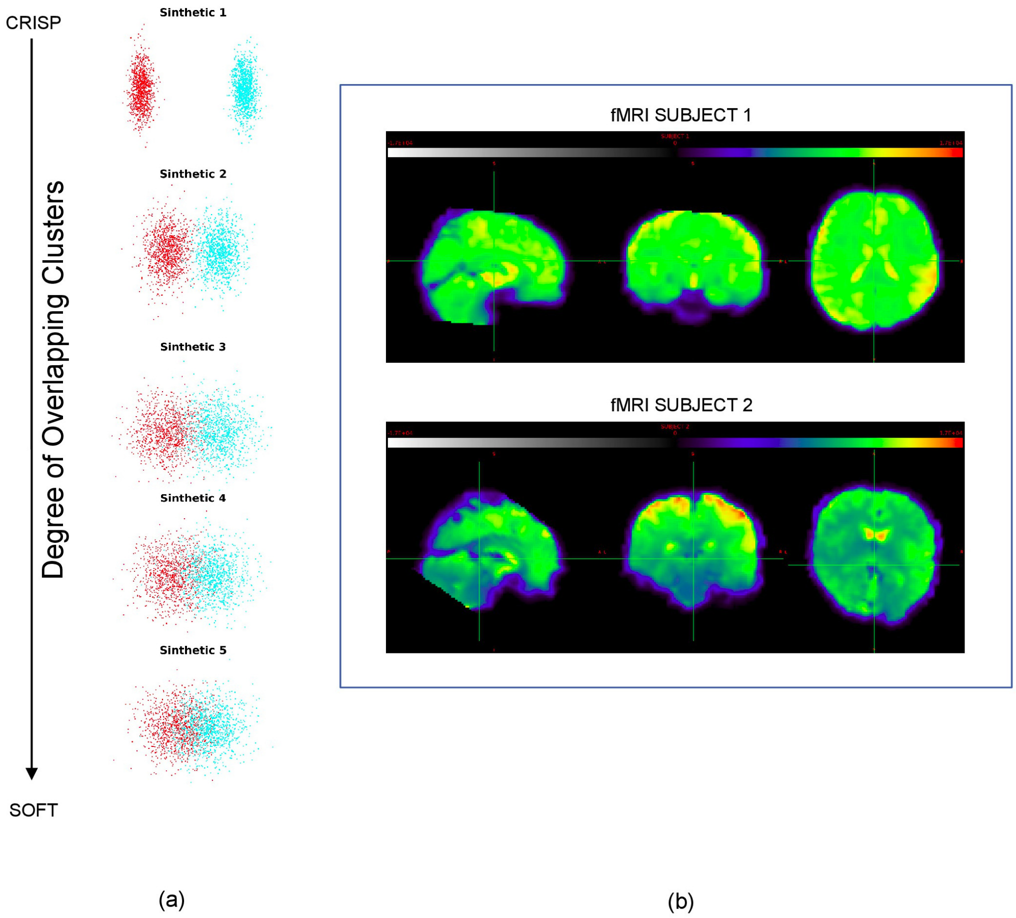

Figure 4.3: The image shows the changes in the synthetic datasets in their fuzziness (a) and the resting state fMRI brain volumes observed in three different sections (coronal, saggittal and medial) about the two subjects (b). In the (a) pictures, the five datasets were reduced via Principal Component Analysis and represented by the first two components.

Computation of experimental-based ground-truth

In order to estimate the experimental ground-truth for the clustering of the fMRI datasets, we defined the clustering centroids with the literature-based seed centroids. From the works by Lee et al. [50], Raichle et al. [63], Fox et al. [23] and Vincent et al. [75], we selected seven seeds that - given the actual knowledge acquired in Neuroimaging - are the candidate biomarkers for the resting state physiological functional activity of the brain, measured with fMRI scanners:

1. the Default Mode Network (DMN) (MNI152: x="-2", y="-60", z="44", i.e, the Precuneous Cortex in the Harvard-Oxford Atlas);
2. the Fronto-Parietal Control Network (FPCN) (MNI152: x="-40", y="30", z="38", i.e., the Left Middle Frontal Gyrus in Harvard-Oxford Atlas);
3. the Language Network (LN) (MNI152: x="-66", y="-34", z="-10", i.e,, the Middle Temporal Gyrus, poster division, in the Harvard-Oxford Atlas);
4. the Ventral-Attention Network (VAN) (MNI152: x="48", y="46", z="-6", i.e., the Right Frontal Pole in the Harvard-Oxford Atlas);
5. the Somato-Sensory Network (SSN) (MNI152: x="-44", y="-32", z="18", i.e., the Left Parietal Operculum Cortex in the Harvard-Oxford Atlas);
6. the Visual Network (VN) (MNI152: x="16", y="-102", z="4", i.e., the Right Occipital Pole);
7. the Dorsal Attention Network (DAN) (MNI152: x="62", y="4", z="28", i.e, the Right Precentral Gyrus in the Harvard-Oxford Atlas).

Taken the above seven seeds as static clustering centroids for the fuzzy c-means algorithm, we clustered the two fMRI datasets obtaining an experimental ground-truth, one for each subject. In this way, the likely ground-truth used to validate the clustering outcomes has both the generality of the neuroscientific literature and the speciality of the subjects. In Figure 4.4 the seeds are shown.

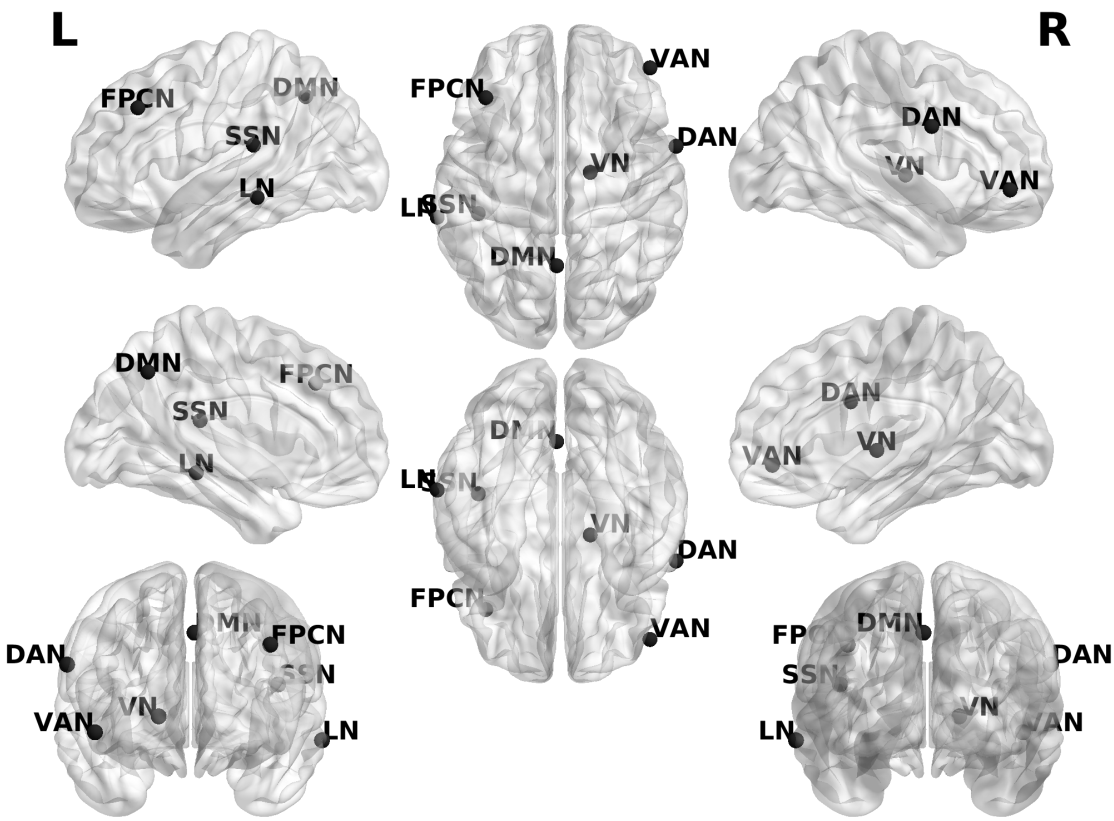

Figure 4.4: The image shows the localization of the seeds in a template brain. The seeds are the Precuneus for the Default Mode Network (DMN), the Left Middle Frontal Gyrus for the Fronto-Parietal Control Network (FPCN), the Left Middle Temporal Gyrus for the Language Network (LN), the Left Parietal Operculum for the Somato-Sensory Network (SSN), the Right Frontal Lobe for the Ventral Attention Network (VAN), the Right Precentral Gyrus for the Dorsal Attention Network (DAN) and Right Occipital Pole for the Visual Network (VN).

4.2.7 Experiments

For this experiment, the fuzzy c-means clustering algorithm was applied to each of the datasets described in Chapter 4.2.6. The FCM algorithm was used by setting the parameter $m = 2$ and varying the parameter of the desired clusters K from two to ten ($K = 2, 3, ...10$), to obtain nine clusters different for each data set. For the dataset "Data set with large number of clusters" the parameter K was varied from two to one hundred ($K = 2, 3, ...100$) to obtain ninety-nine clusters. The algorithm was repeated 200 times to improve clustering robustness, and this iteration was performed for each dataset and K value, where the clustering with the lowest J_m value was saved. For each dataset, the clustering with K equal to the number of truth classes of the dataset, as compared with the respective truth. In the results of the comparisons made, the minimum similarity was 86% and an average of about 91%. For reference datasets and synthetic datasets there was a known truth, for fMRI datasets the truth was derived experimentally as described in Chapater 4.2.6.

The clusterings obtained from each dataset were evaluated by the CVIs. The evaluation logic of the various CVIs are different, for some the evaluation is as much better as it gets closer to 0, for others the evaluation is as much better as it gets closer ad ∞.

In order to make the CVIs results comparable, a normalization of the data was therefore carried out, specifically it was the use of the z-score normalisation. The z-score technique has been implemented positively for indexes that minimize their optimal value, and negatively for indexes that maximize their optimal value. After normalization, for all indexes, the valuation is as much better as it approaches 0. Table 4.3 shows the evaluations of the indexes on a dataset after applying the zscore.

Table 4.3: In this table the evaluations of the indexes after applying the zscore of the synthetic dataset with medium overlap.

Index	K=**2**	K=**3**	K=**4**	K=**5**	K=**6**	K=**7**	K=**8**	K=**9**	K=**10**
FSI	-2,188	-0.918	-0,256	0,113	0,356	0.548	0,662	0,797	0,885
WSJI	-2,028	-1,231	0,393	0,378	1,107	0,119	0,136	0,262	0,862
XBI	-0,462	-0,452	-0,444	-0,252	-0,400	0,058	-0,305	-0,373	2,631
RLRI	-2,273	-0,161	1,018	0,769	1,027	-0,011	-0,242	-0225	0,099
FPBMI	2,101	0,585	0,454	0,338	-0,879	-0,161	-1,018	-0,515	-0,905
PBMI	-1,964	-1,271	-0,247	0,228	0,233	0,637	0,600	0.872	0,910
DBI	-1,399	0,739	1,824	0,729	0,430	-0,332	-0,720	-0,701	-0,570
SDBI	1,830	0,911	0,859	0,121	-0,347	-0,608	-0,816	-0,934	-1,017

The methodology for evaluating the performance of CVIs most widespread in the literature consists of a binary measurement, in which the index indicates correct clustering or fails. This methodology allows to evaluate only the optimal performances of CVIs. In this chapter, we want to overcome this binary methodology, and evaluate both excellent and sub-optimal performances. For this reason, three different parameters have been introduced. These parameters named E_1, E_2, E_3 measure in different ways the result proposed by the CVIs, and are described as follows:

The first parameter, E_1, represents the error made by the index expressed in number of clusters, and is described as:

$$E_1 = |n_i - n_r| \tag{4.45}$$

where n_i is the number of clusters contained in the clustering designated by the index, n_r is the number of classes from which the reference truth is composed. The second parameter, E_2,represents the difference between the evaluation obtained by clustering suggested by the index, and the evaluation obtained by clustering corresponds to the truth, and is described as:

$$E_2 = |v_i - v_r| \tag{4.46}$$

where v_i is the value calculated by the index corresponding to the optimal number of clusters designated by the index, v_r is the value calculated by the index corresponding to the number of clusters by reference.

The third parameter, E_3 it is calculated after sorting in ascending order all CVIs evaluation values for a given dataset. So E_3 corresponds to the difference between the first position and the position of the suggested clustering value, and is represented as:

$$E_3 = |1 - p| \tag{4.47}$$

where p represents the position of v_r in an ascending ordered vector, containing the values computed from the index of each grouping of a given dataset. In other words, how many other values does the index suggest before the one corresponding to the truth.

Once the parameters E_1, E_2, E_3 were calculated for each index and each dataset, they were given as input to ANFIS for the evaluation of the index performance. The choice to use ANFIS for the evaluation is due to the need to overcome the criticality related to methodology for combining the three indexes, to obtain a final evaluation. The choice of which criterion to use to combine the three parameters, and the weight to be given to each, turned out to be a non-trivial task. Furthermore, we believe that an evaluation using fuzzy rules is the most suitable for a complex field such as that of index evaluation. ANFIS, through a training performed with a training dataset, independently creates the fuzzy rules to carry out the evaluation. One of the fundamental parameters for the construction of an ANFIS model are the membership functions. There are a wide variety of forms of the membership functions: Gaussian, triangular, bell, moreover, these membership functions can also be modified, or even hybridized, to customize the shape to generate maximum precision better. Another essential parameter concerns the technique used for optimization the most used are backpropagation and hybrid.

To train the ANFIS model, a synthetic dataset was generated consisting of two thousand elements, with three characteristics for each element and with four truth classes. An expert evaluated the dataset, and each data was assigned a label with its evaluation. Fifteen hundred elements were used for the training phase, and the remaining five hundred elements were used for the testing phase. Several ANFIS models have been created by varying the forms of the membership functions and the optimization metric. For the experiments, the model with the highest degree of accuracy was selected, which was 93%, where the membership function was the Gaussian bell with back propagation. The labels used to evaluate the indexes are:

- *Excellent*: represents an excellent evaluation made by the index, therefore without errors.
- *Good*: represents an index that did not give an optimal rating but is still a suboptimal solution.
- *Insufficient*: represents an index that did not provide a rating or a suboptimal rating, an incorrect rating and far from the truth.
- *Poor*: represents an index that has given an incorrect evaluation far from the truth.

In addition to CVIS performance assessments on individual datasets, CVI performance assessments by dataset type were performed. To carry out the evaluation by dataset type, the average parameters E_{1m}, E_{2m}, E_{3m} were calculated for each index. The parameters E_{1m}, E_{2m}, E_{3m} were calculated as the average of the calculated E_1, E_2, E_3 parameters for each index based on their type fMRI, benchmark and synthetic dataset. For example, for the DBI index there is E_{1m} for fMRI datasets, one for benchmark datasets and one for synthetic datasets, the same for E_{2m} and E_{3m}. In the case of synthetic data sets, E_{1m}, E_{2m}, E_{3m}, were calculated by excluding the high-number dataset, this is because, given the diversity from the other sets of data. Furthermore, a global performance of the indexes was evaluated by calculating, similarly to what was done for the assessments on the types of datasets, the parameters E_{1mg}, E_{2mg}, E_{3mg}. They represent the values of the global parameters obtained by averaging the values E_1, E_2, E_3 obtained from the indexes on all datasets.The average ratings in the case of E_{1m}, E_{3m}, E_{1mg}, and E_{3mg} provide non-integer values. To bring them to integer values it was decided that any non-integer value should be carried to the next integer. The experiments were conducted both with integer values and with non-integer values, having obtained the same results, only the experiments with integer values will be reported. In the next chapter, the results of the evaluations provided by ANFIS will be presented, divided both by single dataset, type of dataset and the average of all the evaluations obtained by the CVIs. The experimental work process is schematized in figure 4.5.

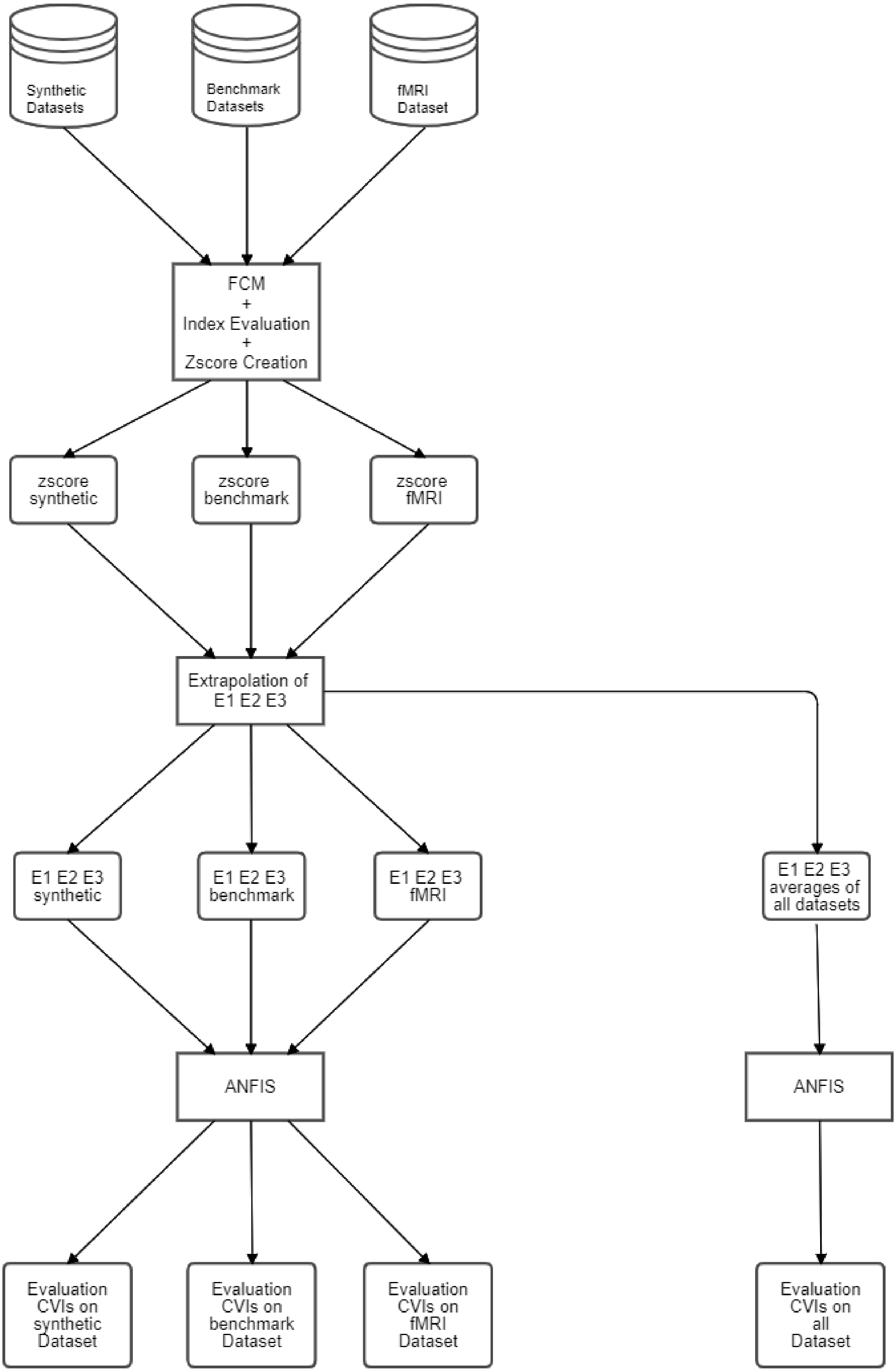

Figure 4.5: Workflow of the experiments

4.2.8 Results

In this paragraph all the evaluations on the CVIs performance will be shown divided by single data set. The evaluations will be shown in tabular form, where for each index, the values of E_1, E_2, E_3 and the evaluation assigned by ANFIS are reported. The same methodology was used to show the average evaluations carried out by type of dataset.

Evaluation of fMRI datasets

The evaluations obtained by the CVIs on single fMRI dataset are shown below

Table 4.4: This table represents the evaluation obtained of CVIs on the male subject of fMRI dataset

Index	E_1	E_2	E_3	Evaluation
FSI	2	0,387	2	*Good*
WSJI	3	0,962	2	*Good*
XBI	3	0,650	4	*Insufficient*
RLRI	3	0,691	1	*Good*
FPBMI	3	0.728	4	*Insufficient*
PBMI	3	0,851	4	*Insufficient*
DBI	5	2,864	5	*Poor*
SDBI	4	2,282	5	*Poor*

Table 4.5: This table represents the evaluation obtained of CVIs on the female subject of fMRI dataset

Index	E_1	E_2	E_3	Evaluation
FSI	2	0,568	1	*Good*
WSJI	3	0,789	1	*Good*
XBI	3	0,529	4	Insufficent
RLRI	1	0,703	1	*Good*
FPBMI	3	0.662	4	*Insufficient*
PBMI	3	0,591	4	*Insufficient*
DBI	5	2,970	6	*Poor*
SDBI	4	2,312	6	*Poor*

Evaluation of Benchmark datasets

The evaluations obtained by the CVIs on the single Benchmark datasets are shown below

Table 4.6: This table represents the evaluation obtained of CVIs on the Cancer dataset.

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	2	0.001	1	*Good*
XBI	0	0	0	*Excellent*
RLRI	1	0,014	2	*Good*
FPBMI	6	0.850	3	*Insufficient*
PBMI	3	1,037	2	*Good*
DBI	0	0	0	*Excellent*
SDBI	1	0,0006	1	*Good*

Table 4.7: This table represents the evaluation obtained of CVIs on the Glass dataset

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	2	0,667	2	*Good*
XBI	7	0,289	5	*Poor*
RLRI	1	0,780	2	*Good*
FPBMI	8	2,505	9	*Poor*
PBMI	4	0,045	3	*Insufficient*
DBI	1	0,728	2	*Good*
SDBI	8	2,548	7	*Poor*

Table 4.8: This table represents the evaluation obtained of CVIs on the Iris dataset

Index	E_1	E_2	E_3	Evaluation
FSI	1	0,860	1	*Good*
WSJI	0	0	0	*Excellent*
XBI	1	0,007	1	*Good*
RLRI	1	0,292	1	*Good*
FPBMI	3	1,064	7	*Poor*
PBMI	0	0	0	*Excellent*
DBI	1	1,169	1	*Good*
SDBI	2	2,206	7	*Poor*

Table 4.9: This table represents the evaluation obtained of CVIs on the Simple dataset

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	2	2,713	8	*Poor*
RLRI	0	0	0	*Excellent*
FPBMI	4	0,424	3	*Insufficient*
PBMI	0	0	0	*Excellent*
DBI	0	0	0	*Excellent*
SDBI	0	0	0	*Excellent*

Table 4.10: This table represents the evaluation obtained of CVIs on the Thyroid dataset

Index	E_1	E_2	E_3	Evaluation
FSI	3	0,783	2	*Good*
WSJI	3	0,609	2	*Good*
XBI	4	1,826	3	*Insufficient*
RLRI	2	0,587	2	*Good*
FPBMI	6	2,410	7	*Poor*
PBMI	3	1,424	4	*Insufficient*
DBI	0	0	0	*Excellent*
SDBI	6	1,010	5	*Poor*

Evaluation of Synthetic datasets

The evaluations obtained by the CVIs on the single Synthetic datasets are shown below

Table 4.11: This table represents the evaluation obtained of CVIs on the synthetic data set without overlap

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	0	0	0	*Excellent*
FPBMI	2	0.185	1	*Good*
PBMI	0	0	0	*Excellent*
DBI	0	0	0	*Excellent*
SDBI	4	0,052	1	*Good*

Table 4.12: This table represents the evaluation obtained of CVIs on the synthetic data set with minimal overlap

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	0	0	0	*Excellent*
FPBMI	8	3,333	8	*Poor*
PBMI	0	0	0	*Excellent*
DBI	0	0	0	*Excellent*
SDBI	8	2,256	7	*Poor*

Table 4.13: This table represents the evaluation obtained of CVIs on the synthetic data set with mean overlap

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	0	0	0	*Excellent*
FPBMI	6	3,119	8	*Poor*
PBMI	0	0	0	*Excellent*
DBI	0	0	0	*Excellent*
SDBI	8	2,847	8	*Poor*

Table 4.14: This table represents the evaluation obtained of CVIs on the synthetic data set with medium-high overlap

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	0	0	0	*Excellent*
FPBMI	8	3,380	8	*Poor*
PBMI	0	0	0	*Excellent*
DBI	5	0,280	2	*Good*
SDBI	8	2,924	8	*Poor*

Table 4.15: This table represents the evaluation obtained of CVIs on the synthetic data set with high overlap

Index	E_1	E_2	E_3	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	3	0,541	1	*Good*
FPBMI	8	3,024	8	*Poor*
PBMI	0	0	0	*Excellent*
DBI	6	2,019	7	*Poor*
SDBI	8	3,011	8	*Poor*

Table 4.16: This table represents the evaluation obtained of CVIs on the synthetic dataset with 60 clusters

Index	E_1	E_2	E_3	Evaluation
FSI	8	0.860	8	*Poor*
WSJI	13	0.997	9	*Poor*
XBI	15	1.105	6	*Poor*
RLRI	6	1.103	7	*Poor*
FPBMI	18	0.289	9	*Poor*
PBMI	18	2.000	12	*Poor*
DBI	10	0,500	38	*Poor*
SDBI	28	1.242	51	*Poor*

CVIs evaluation based on dataset type

The CVIs evaluation based on dataset type are shown below.

Table 4.17: This table represents the evaluations obtained by the CVIs considering all fMRI dataset

Index	E_{1m}	E_{2m}	E_{3m}	Evaluation
FSI	2	0,477	2	*Good*
WSJI	3	0,875	2	*Good*
XBI	3	0.589	4	*Insufficient*
RLRI	2	0,697	1	*Good*
FPBMI	3	1,390	4	*Insufficient*
PBMI	3	0,721	4	*Insufficient*
DBI	5	2.917	6	*Poor*
SDBI	4	2.297	6	*Poor*

Table 4.18: This table represents the evaluations obtained by the CVIs considering all Benchmark dataset

Index	E_{1m}	E_{2m}	E_{3m}	Evaluation
FSI	2	0,328	1	*Good*
WSJI	2	0,255	1	*Good*
XBI	3	0.967	4	*Insufficient*
RLRI	1	0,334	2	*Good*
FPBMI	6	1,450	6	*Poor*
PBMI	2	0,501	2	*Good*
DBI	1	0,379	1	*Good*
SDBI	3	1.152	4	*Insufficient*

Table 4.19: This table represents the evaluations obtained by the CVIs considering all synthetic overlap dataset, excluding the 60 cluster synthetic dataset

Index	E_{1m}	E_{2m}	E_{3m}	Evaluation
FSI	0	0	0	*Excellent*
WSJI	0	0	0	*Excellent*
XBI	0	0	0	*Excellent*
RLRI	1	0,108	1	*Good*
FPBMI	7	2.608	7	*Poor*
PBMI	0	0	0	*Excellent*
DBI	3	0,459	2	*Good*
SDBI	8	2,218	7	*Poor*

CVIs evaluation based on all dataset

CVIs evaluation based on all datasets, excluding the high-volume synthetic dataset, is shown below.

Table 4.20: This table represents the mean evaluation of CVIs on all datasets excluding the synthetic 60 cluster dataset

Index	E_{1mg}	E_{2mg}	E_{3mg}	Evaluation
FSI	2	0,268	1	*Good*
WSJI	2	0,376	1	*Good*
XBI	3	0,518	3	*Good*
RLRI	2	0,379	2	*Good*
FPBMI	6	1,816	6	*Poor*
PBMI	2	0,407	2	*Good*
DBI	3	1.251	3	*Good*
SDBI	5	1.889	6	*Poor*

4.2.9 Discussion of results and conclusions.

In this document, we wanted to evaluate the performance of eight evaluation indexes chosen from among the most famous or innovative. The CVIs were evaluated using three different types of datasets: fMRI, benchmark and synthetic, these datasets were chosen because we wanted an evaluation that took into account the sub-optimal performance of CVIs in the fMRI field, but also to extend the evaluation to more heterogeneous areas.

The results obtained in the study of CVI on the single fMRI datasets confirmed the results obtained in previous work by [57], always conducted on an evaluation of suboptimal results. In this last work, it was pointed out that the indexes using summations or differences to relate Compactness and Separation, offer better performance than indexes that use multiplications or ratios as a type of relationship between Compactness and Separation. Furthermore, the formalization of the Separation component also plays an important role; the group of indexes that obtained a good evaluation uses a very similar procedure for the calculation of the Separation.

In the study conducted on the benchmark datasets, a good overall performance was found for almost all CVIs, with the exception of the indexes of the FPBM, Xie-Beni and SDBi indexes,the good performance of most CVIs is not surprising, as in the literature some of the CVIs used had already been evaluated with benchmark datasets, even if using the classic methodologies. We believe that the low performances obtained by the FPBMI, Xie-Beni and SDB indices are linked to the choice of the value of the exponent belonging to a cluster. In FPBM and SDBI, the exponent parameter of the membership value of a datum to a cluster is represented by m. In the Xie-Beni index, the value 2 is used as the exponent of the membership value of a cluster, therefore in a similar way also to the FPBMI and SDBI indexes since the value of m is equal to 2.

Examining the results obtained in the synthetic datasets, we note, also in this case, the excellent performances by the CVIs except SDBI and FPBMI, we believe due, as in the case of the benchmark datasets, to the parameter m. The high-number synthetic dataset deserves a separate discussion, in this case, all the indexes obtained low evaluations; even though RLRI and FSI have values E_1, E_2, E_3 lower than the other indexes. The difficulties encountered by the indexes are due to the number of cluster to be evaluated, they were ninety-nine, it is easy with such a high number of clustering that more partitions provide good value of Compactness and Separation, and this increases the complexity of the evaluation by the CVIs.

The results of the evaluations on CVIs based on the type of dataset (Tables 4.17, 4.18, 4.19), provide a good summary of the evaluations given to the indexes. The results obtained in this study show limits to the evaluation of fMRI data due to the use of a limited number of exams, but the exams were nevertheless analyzed in-depth and with a data-driven approach. In general, another limit is given by the selection of the indices, from which other known indices and with different valuation methods have been excluded. However, we believe that the work presented may provide insights into the reliable use of cluster evaluation indexes, particularly in the complex and increasingly studied domain of FMRI data evaluation.

Finally, we would like to underline the importance of using ANFIS, which made it possible to solve the criticality of the combination of parameters, and which gives flexibility to the proposed methodology. It allows the extension to n-number of evaluation parameters by performing only retraining of the model. Thanks to ANFIS, innumerable metrics for evaluation can be introduced, including categorical ones. Hence, it is also possible to choose, based on the type of dataset, which metrics to use and then use the relative ANFIS model.

In the next works, the analysis on FMRI exams will be extended using a large number of datasets and introducing new indexes and new evaluation metrics.

5 Conclusion and Future Works

5.1 Conclusion

In this thesis a multitude of fMRI exams were analyzed, the exams were performed both in task-based mode and in resting-state mode, these exams were analyzed using unsupervised learning techniques (clustering), instead of the more widely used statistical techniques. Different clustering algorithms were used for the analyzes, both crisp and fuzzy, always referring to the field of fMRI data analysis, given the use of clustering techniques, clustering evaluation indexes were also studied. Various indexes have been analyzed, both crisp and fuzzy, and a new evaluation method has also been proposed, which allows to assess the clustering evaluation indices based on different parameters.

The analysis of fMRI exams is a very complex research field, until a few years ago they used almost exclusively statistical techniques, which have provided good results; but they have also shown some limitations, for this reason in recent years other analysis techniques have been developed in parallel with the statistical techniques. Among these new techniques, the ones that are gaining more interest thanks to the results provided are the techniques of unsupervised learning (clustering). Clustering techniques provide an analysis of fMRI exams different from the statistical one, thanks to their characteristic of being guided by the data without the need for any previous knowledge.

In the study of active fMRI exams, a cross-clustering approach was adopted to the data with the aim of grouping both spatial and temporal patterns. The spatial (ROI - Regions of interest) and temporal (TOI - Times of interest) characteristics were analyzed, parametric and non-parametric statistical tests were used to assess whether the differences between classes are significant in all clusters, using the value p as a decision criterion. This procedure is useful for understanding the meaning of ROI clustering as it is associated with functional properties called activity-based paradigm. Noteworthy is the behavior of the algorithms in the TOI analysis, the Fuzzy C-Means method has surpassed the neural gas method, based on the statistical significance test.

In the analysis of examinations carried out in resting state, there are confirmation of the gender determinants in RS-fMRI functionality found by Biswal et al., in addition, information on male and female peculiarities was also added through algebraic distances to measure gender variability. Globally it was noted that female subjects had greater amplitude and greater variability of basal signal than male subjects. Another result concerns the integration of clustering techniques with classical statistical processing for signal analysis. The algorithms created clusters based on the intensity of the brain signal, and also differentiated the quality of the brain's functional connectivity, that is, strong (positive) associations versus weak associations.

The use of clustering techniques also involves the evaluation of the clustering performed. The evaluation of clustering is a very complex research field, there are different methodologies to carry out the evaluation of clustering, the most used are the evaluation indexes. There are various evaluation indexes, in this thesis we wanted to analyze,in the context of fMRI analysis, the behavior of some of the most famous/innovative indexes .

To overcome the classic methodology of index evaluation, and to give a different analysis to a complex field such as clustering of fMRI exams, a new metric was introduced. This new metric allows, albeit to a limited extent, to give weight also to the sub-optimal performance of the valuation indices, while the classic methodologies, on the other hand, only evaluate the optimal performance of the indices. Valuation indexes use internal metrics to evaluate clustering, each index uses different metrics. The analyzes showed that the indexes that used sums or subtractions to combine the different metrics performed better than those that used multiplication or ratios to combine the metrics.

With the aim of extending the metrics for the evaluation of the indexes, an evaluation methodology has been introduced which uses an ANFIS system. ANFIS is used to autonomously create rules that allow, based on the input parameters, to evaluate the performance of the indexes through linguistic labels. This methodology was applied both to evaluate the performance of the CVIs inherent to the clustering evaluation of fMRI data, and to evaluate the performance of the CVIs inherent to the clustering evaluation on benchmark data.

The analysis of the indices in the fMRI field confirmed the best performance of the indices using adders and subtractions, for the combination of internal metrics. In the analysis of the benchmark data, all CVIs obtained good results, this result was desirable because often in the literature CVIs are tested through the use of benchmark datasets.

The new methodology proposed through the use of metrics for the evaluation of sub-optimal performances and the use of ANFIS has shown great potential. By introducing further metrics, it is possible to deepen the evaluations or focus on particular aspects of the evaluation of CVIs not taken into consideration in this thesis.

For further details, I remain at the conclusions of the works presented in this thesis 3.1.6,3.2.6,4.1.8,4.2.9.

5.2 Future Works

In the analyses carried out in this thesis, both the clustering algorithms and the CVIs adopted the Euclidean distance as a metric for the calculation of the similarities between the elements. In the literature, some works are emerging which support the use of correlation as a metric to calculate the similarity between elements for clustering, and it would seem it could be a valid alternative to the Euclidean distance [5].

A study is being developed in which we want to analyse this possibility. For this study, synthetic data are used to simulate fMRI exams, this choice was made to have maximum control over the data, and allow to evaluate the results obtained from both CVIs and clustering more precisely. The work will be divided into two interconnected parts in which both the clustering algorithms and the CVIs will be analysed. In the first part of the work, the performance of the clustering algorithms on synthetic fMRI data will be evaluated. The results obtained using both the Euclidean distance and the correlation will be compared with respect to a known truth.

In the second part, the performances of the CVIs will be evaluated, which will be declined both in the version with the use of both the Euclidean distance and the correlation. For the evaluation, the CVIs will receive as input the clustering generated by the clustering algorithms using both the Euclidean distance and the correlation. CVI evaluation will take place with the methodology expressed in Chapter 4.2, therefore with the use of ANFIS and new metrics could be introduced for a more complete analysis.

List of Figures

2.1 The BOLD signal consists of several parts: (1) the neuronal response to an underlying stimulus or modulation; (2) the relationship between neuronal activity and the starter of a hemodynamic response; (3) the hemodynamic response itself; and (4) how this response is detected by an MRI scanner.[4] 16

2.2 A represents a Block-design task-based protocol with two different tasks, the task and rest time are equal. B a represents Event-related task-based protocol with two tasks, the task has a limited duration and the rest time is random. C represents a Resting-state protocol, to analyze the cerebral activities at rest.[67] 20

2.3 Organization of the ANFIS architecture 45

3.1 Plots of the average and the standard deviation of BOLD signals in Males and Females in the temporal (175 time-points versus BOLD signals) and spatial domains (96 ROIs versus BOLD signals). Both the average and the standard deviation of BOLD signals in the temporal domain are quite different between gender: Females have higher values then Males; whereas, in the spatial domain, both Males and Females have similar average BOLD signals, but Females have more standard deviation of BOLD signal. 55

3.2 In the top part of the figure, there are the boxplots of whole brain signals in Males (1) and Females (2), showing the between gender statistical difference of the average (left) and the standard deviation (right) about the whole brain signals: in the both cases, Females have higher values then Males. In the bottom part of the figure, there are the within gender distance measures of the exams: with both Euclidean and Manhattan distances, Females have higher values then Males for the mean and the standard deviation of the distances computed. 56

3.3 The top diagram shows the optimum Davies-Bouldin index associated to SOM, NG and GNG algorithms, for both Females and Males: all the DB are referred to 2 clusters as optimal clusters number for each algorithm; all the best indexes are under 1; the lesser is related to the fourth Males subject, and the higher to the fourth Female subject. The bottom histogram represents the discrete distribution of the Jaccard index computed for all the algorithm pairs: the more similar clustering outcomes are with the NG-GNG pairs in both Females and Males (Jaccard < 0.3), and the other clustering combinations are very different (Jaccard > 0.7) 57

3.4 The scatter plots show the clustering outcomes for NG, GNG and SOM in Females (red points) and Males (blue points): the main result is that, using two clusters as optimum partitions number, the clusters detect the amplitude information of BOLD signals in Females and Males, i.e. clusters differentiate low levels and high levels in BOLD signals. 58

3.5 In the upper part of the figure, there are the Correlation Coefficients (CC) distributions between all-ROIs versus all-ROIs in Males and Females, and the correlation coefficients distributions between seeds (L/R Precuneus) versus all-ROIs. In the first discrete distribution, there are present positive and negative correlations, both in Males and Females; whereas, in the seeds based correlations distributions, there are less anti-correlations, with always, a bit gender difference. In the bottom part of the figure, there are the seed-based correlation matrix between Left and Right Precuneus in Males and Females: it is evident the strong correlation with many brain regions, as well as weak correlations with few brain regions. The highest positive correlations (CC ¿ 0.8) are with seed-controlateral region (ROI 61 and ROI 62) and L/R Posterior Division of Cingulated gyrus (ROI 59 and ROI 60). The lowest correlations (CC +/- 0.2) are with L/R Posterior Division of Temporal Gyrus (ROI 29 and ROI 30) and with the L/R Anterior Division of Temporal Fusiform Cortexes (ROI 73 and ROI 74). 59

3.6 These figures represent the organization of the two optimal clusters in Males and Females in relation to the seeds based (L/R Precuneus) Correlation Coefficient. In the left column, it is reported the evidence that low and high correlations belong to different clusters, with some superimposition between them for the central values. The central and the right columns show the relation between ROIs, correlation coefficients and clusters, showing precisely that low and high correlations are within different clusters in Males and Females. 60

3.7 The image shows a detailed result of SOM algorithm in the Males case, with a specification about the higher and the lower correlations with both the seeds (Left and Right Precuneus), and their organisation within the two clusters. In the left side of the figure, the brain is represented with the centre-of-mass of each regions of interest (the 96 ROIs labelled with Harvard-Oxford atlas), that are filled or unfilled if they belong to cluster 1 or 2. The coloured circles are the height selected regions that are, respectively, the four with the higher correlations (hot colours), and the four with the lower correlation (cool colours). In the right side of the figure, there is a plot with the clusters planes in relation to their elements: the four higher correlations regions are on the plane of cluster 1, whereas the four lower correlation are on the plane of cluster 2. We choose to plot the brain parcellation relative to the SOM algorithm in Males because was the better optimized algorithm, having as a Davies-Bouldin (DB) separation measure a value equal to 0.7641 (the lesser value respect the other algorithms). 61

3.8 The Figure shows the information related to the correlation coefficients. The left plot is the empirical distribution of the all correlation coefficients (CC) computed for every subject. the right plot is the mean correlation matrix computed for all subjects' correlations) 69

3.9 The figure shows the plot of the optimal clusters *versus* the lesser Davies-Bouldin index for clustering with FCM and NG algorithms, that differentiated by the Regions of Interest (ROIs) configuration and the Time Of Interest (TOIs) configuration. Globally, the partitioning of ROIs with both FCM and NG had lower Davies-Bouldin index values then the partitioning of TOIs. 69

3.10 The figure shows the more frequent ROIs partitioning in the special case of clustering with 2 groups for NG and FCM for all the subjects; the horizontal line from ROIs 73 to 80 are the ones related to the eight Fusiform regions, i.e., they are ROIs 73:74 L/R Temporal Fusiform Cx Anterior Divisions, ROIs 75:76 L/R Temporal Fusiform Cx Posterior Divisions, ROIs 77:78 L/R Temporo-Occipital Fusiform Cx and ROIs 79:80 L/R Occipital Fusiform Cx. 70

3.11 The figure shows the results of non-parametric tests (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) inputs in case of the optimal configurations. The black line is the significance level 0.05. The values under the black line allow to reject the null hypothesis of the test. ROIs clustering with both FCM and NG have statistically different clusters for all the subjects with non-parametric tests. TOIs clustering is globally near the significative criterion only with FCM algorithms. 71

3.12 The figure shows the results of the parametric test (ANOVA-1) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of the optimal clusters configurations. The black line is the significance level 0.05. The results were similar to Figure 3.11, but with a higher p-value for the TOIs clustering with FCM algorithm. 72

3.13 The figure shows the results of the non parametric test (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) configurations in the special case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 3.11 and Figure 3.12, but with lower p-value for the TOIs clustering with FCM algorithm. 72

3.14 The figure shows the results of the parametric (one-way ANOVA) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 3.13, but with a higher p-value for the TOIs clustering with FCM algorithm. 73

3.15 The figure shows the comparison of Regions of Interests (ROIs) clustering of two groups using Jaccard matrix. The comparison is between subjects in case of applying NG and FCM algorithms. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between the subjects is uniform and without big differences, expect when using the FCM algorithm. 73

3.16 The figure shows the comparison of Times of Interest (TOIs) clustering with two groups using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between the subjects is uniform (for NG) and middle-centered (for FCM). . 74

3.17 The figure shows the Jaccard Matrices in both the cases of two classes clustering with Regions of Interest (ROIs) and Times of Interest (TOIs). The comparison is between subjects in cases of NG and FCM. Under the matrices there is the distribution of the Jaccard distance values. In general, the similarity distribution between subjects is weakly bimodal for the ROIs clustering (bottom left) and middle-centred for the TOIs clustering (bottom right). 74

3.18 The figure shows the brain parcellation based on a graph model and the Regions of Interest (ROIs) organisation partitioned with two clusters. In the left plot the horizontal line indicates the Fusiform Cortexes (ROIs 73:80) and their clusters. In the right image the nodes are the 96 centroids according to Harvard-Oxford atlas. In general, the inferior regions were clustered in the Class 2 and the superior regions were clustered in the Class 1. The figure is referred specifically to the Regions of Interest (ROIs) of Subject 12 clustered with FCM. 75

3.19 The figure shows the BOLD values partitioning among the two clusters and the Times of Interests (TOIs) partition among the two clusters. In the left plot, clusters exhibit the BOLD amplitude and in the right plot the clusters seem to exhibit spatio-temporal patterns. The figure refers specifically to Subject 12 clustered with FCM. 75

3.20 The figure shows the brain voxels parcellation of Subject 12 clustered with FCM using the optimal configuration (with the lesser Davies-Bouldin index). There are 14 clusters that covered all brain voxels (cfr. Figure 3.21 for more details). 76

3.21 The left plot shows the substructures of the Fusiform regions (the legend is the same of the right plot). In the right plot there are the eight Fusiform regions distributed to different clusters according to their BOLD values: L/R Temporo-Occipital Fusiform Cx and L/R Occipital Fusiform Cx have greater BOLD values than the L/R Temporal Fusiform Cx Anterior and Posterior Divisions. Precisely, clusters 2, 6 and 8 contain the Fusiform ROIs with greater activations, whereas clusters 3, 13, 14 the ones with lesser activations. 76

4.1 This image displays resting-state functional connectivity as linear correlation for the seed region in a sample of 1,000 subjects. The seed chosen is the Precuneus (X/Y/Z MNI152 coordinates: 2 -60 30), that is the main core of the Default Mode Network (DMN), a candidate biomarker for the fMRI resting state studies. In the images, the Precuneus is in the zone with the highest functionality (yellow color). 84

4.2 In the figure the ANFIS architecture with the five layers, the fixed type nodes are represented with squares, instead of the adaptive nodes with circles. 100
4.3 The image shows the changes in the synthetic datasets in their fuzziness (a) and the resting state fMRI brain volumes observed in three different sections (coronal, saggittal and medial) about the two subjects (b). In the (a) pictures, the five datasets were reduced via Principal Component Analysis and represented by the first two components. 106
4.4 The image shows the localization of the seeds in a template brain. The seeds are the Precuneus for the Default Mode Network (DMN), the Left Middle Frontal Gyrus for the Fronto-Parietal Control Network (FPCN), the Left Middle Temporal Gyrus for the Language Network (LN), the Left Parietal Operculum for the Somato-Sensory Network (SSN), the Right Frontal Lobe for the Ventral Attention Network (VAN), the Right Precentral Gyrus for the Dorsal Attention Network (DAN) and Right Occipital Pole for the Visual Network (VN). 108
4.5 Workflow of the experiments 112

List of Tables

3.1 The table describes the Davies-Bouldin (DB) index computation for each subject differentiated for clustering (FCM or NG) and inputs (ROIs or TOIs).The values presented are the lesser DB associated with the corresponding number of clusters. 70

4.1 Values of CVIs resulting from the evaluation of clustering fMRI dataset by FCM with m=2 and K ranging from 2 to 10. 91
4.2 Mean and variance of E values for the 8 index evaluating clustering of fMRI data with 2 (Experiment 1) and 4 (Experiment 2) reference classes, the CVIs are in ascending order based on the E mean. 91
4.3 In this table the evaluations of the indexes after applying the zscore of the synthetic dataset with medium overlap. 109
4.4 This table represents the evaluation obtained of CVIs on the male subject of fMRI dataset 113
4.5 This table represents the evaluation obtained of CVIs on the female subject of fMRI dataset 113
4.6 This table represents the evaluation obtained of CVIs on the Cancer dataset. 114
4.7 This table represents the evaluation obtained of CVIs on the Glass dataset 114
4.8 This table represents the evaluation obtained of CVIs on the Iris dataset . 114
4.9 This table represents the evaluation obtained of CVIs on the Simple dataset 115
4.10 This table represents the evaluation obtained of CVIs on the Thyroid dataset 115
4.11 This table represents the evaluation obtained of CVIs on the synthetic data set without overlap 115
4.12 This table represents the evaluation obtained of CVIs on the synthetic data set with minimal overlap 116
4.13 This table represents the evaluation obtained of CVIs on the synthetic data set with mean overlap 116
4.14 This table represents the evaluation obtained of CVIs on the synthetic data set with medium-high overlap 116

4.15 This table represents the evaluation obtained of CVIs on the synthetic data set with high overlap . . . 117
4.16 This table represents the evaluation obtained of CVIs on the synthetic dataset with 60 clusters . . . 117
4.17 This table represents the evaluations obtained by the CVIs considering all fMRI dataset . . . 118
4.18 This table represents the evaluations obtained by the CVIs considering all Benchmark dataset . . . 118
4.19 This table represents the evaluations obtained by the CVIs considering all synthetic overlap dataset, excluding the 60 cluster synthetic dataset . . 118
4.20 This table represents the mean evaluation of CVIs on all datasets excluding the synthetic 60 cluster dataset . . . 119

Bibliography

[1] A.A.Vergani. Contributions in computational intelligence with results in functional neuroimaging. *PhD Thesis*, 2018.

[2] Jessica R. Andrews-Hanna, Jonathan Smallwood, and R. Nathan Spreng. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1):29–52, 2014.

[3] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona. An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46(1):243 – 256, 2013.

[4] Owen J Arthurs and Simon Boniface. How well do we understand the neural origins of the fmri bold signal? *TRENDS in Neurosciences*, 25(1):27–31, 2002.

[5] Michael R Berthold and Frank Höppner. On clustering time series using euclidean distance and pearson correlation. *arXiv preprint arXiv:1601.02213*, 2016.

[6] J.C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3):191–203, 1984.

[7] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, October 1995.

[8] B.B Biswal, M. Mennes, X. Zuo, S. Gohel, C. Kelly, S.M. Smith, C.F. Beckmann, J.S. Adelstein, R.L. Buckner, S. Colcombe, et al. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences*, 107(10):4734–4739, 2010.

[9] Bharat B. Biswal, Joel Van Kylen, and James S. Hyde. Simultaneous assessment of flow and bold signals in resting-state functional connectivity maps. *NMR in Biomedicine*, 10(4-5):165–170, 1997.

[10] Ed Bullmore, Jalal Fadili, Michael Breakspear, Raymond Salvador, John Suckling, and Michael Brammer. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research*, 12(5):375–399, 2003. PMID: 14599002.

[11] Fi-John Chang and Ya-Ting Chang. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources*, 29(1):1 – 10, 2006.

[12] Stuart Clare. Functional mri: methods and applications. *University of Nottingham*, page 155, 1997.

[13] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[14] D.L. Davies and D.W. Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, PAMI-1(2):224–227, 1979.

[15] L Jegatha Deborah, R Baskaran, and A Kannan. A survey on internal validity measure for cluster validation. *International Journal of Computer Science & Engineering Survey*, 1(2):85–102, 2010.

[16] Xavier Descombes, Frithjof Kruggel, and D Yves von Cramon. fmri signal restoration using a spatio-temporal markov random field preserving transitions. *NeuroImage*, 8(4):340–349, 1998.

[17] Evgenia Dimitriadou, Markus Barth, Christian Windischberger, Kurt Hornik, and Ewald Moser. A quantitative comparison of functional mri cluster analysis. *Artificial intelligence in medicine*, 31:57–71, 06 2004.

[18] Wenliang Du, Yunghsiang S Han, and Shigang Chen. Privacy-preserving multivariate statistical analysis: Linear regression and classification. In *Proceedings of the 2004 SIAM international conference on data mining*, pages 222–233. SIAM, 2004.

[19] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Willey & Sons, New Yotk, 1973.

[20] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification.* Wiley, New York, 2 edition, 2001.

[21] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification.* John Wiley & Sons, 2012.

[22] M.D. Fox, A.Z. Snyder, J.L. Vincent, M. Corbetta, D.C. Van Essen, and M.E. Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–9678, 2005.

[23] Michael D Fox, Maurizio Corbetta, Abraham Z Snyder, Justin L Vincent, and Marcus E Raichle. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, 103(26):10046–10051, 2006.

[24] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994.

[25] Karl J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

[26] Karl J Friston. Statistical parametric mapping. In *Neuroscience databases*, pages 237–250. Springer, 2003.

[27] Bernd Fritzke. Let it grow-self-organizing feature maps with problem dependent cell structure. In *Artificial neural networks*. Citeseer, 1991.

[28] Bernd Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 625–632. MIT Press, 1995.

[29] Yoshiki Fukuyama. A new method of choosing the number of clusters for the fuzzy c-mean method. In *Proc. 5th Fuzzy Syst. Symp., 1989*, pages 247–250, 1989.

[30] A. Ghuman, Nicolas M. Brunet, Yuanning Li, R. Konecky, J. Pyles, Shawn A. Walls, V. Destefino, Weiqi Wang, and R. M. Richardson. Dynamic encoding of face information in the human fusiform gyrus. *Nature communications*, 5:5672 – 5672, 2014.

[31] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson. Random forest classification of multisource remote sensing and geographic data. In *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1049–1052 vol.2, 2004.

[32] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger. A new correlation-based fuzzy logic clustering algorithm for fmri. *Magnetic Resonance in Medicine*, 40(2):249–260, 1998.

[33] Cyril Goutte, Peter Toft, Egill Rostrup, Finn A Nielsen, and Lars Kai Hansen. On clustering fmri time series. *NeuroImage*, 9(3):298–310, 1999.

[34] Kalanit Grill-Spector and Kevin Weiner. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature reviews. Neuroscience*, 15, 06 2014.

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

[36] I. Gurrutxaga, J. Muguerza, O. Arbelaitz, J. M. Pérez, and J.I. Martín. Towards a standard methodology to evaluate internal cluster validity indices. *Pattern Recognition Letters*, 32(3):505 – 515, 2011.

[37] John A Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., 1975.

[38] Niels Væver Hartvig and Jens Ledet Jensen. Spatial mixture modeling of fmri data. *Human brain mapping*, 11(4):233–248, 2000.

[39] M.P. Van Den Heuvel and H.E.H. Pol. Exploring the brain network: A review on resting-state fmri functional connectivity. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 20:519–34, 08 2010.

[40] John J Hopfield. Learning algorithms and probability distributions in feed-forward and feed-back networks. *Proceedings of the national academy of sciences*, 84(23):8429–8433, 1987.

[41] Armin Iraji, Vince Calhoun, Natalie Wiseman, Esmaeil Davoodi-Bojd, Mohammad Avanaki, Mark Haacke, and Zhifeng Kou. The connectivity domain: Analyzing resting state fmri data using feature-based data-driven and model-based methods. *NeuroImage*, 134, 04 2016.

[42] Anil Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31:651–666, 06 2010.

[43] Jyh-Shing Jang. Self-learning fuzzy controllers based on temporal back propagation. *Neural Networks, IEEE Transactions on*, 3:714 – 723, 10 1992.

[44] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. Fsl. *NeuroImage*, 62(2):782 – 790, 2012. 20 YEARS OF fMRI.

[45] Nancy Kanwisher Josh, Josh Mcdermott, and Marvin M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302–4311, 1997.

[46] D.N. Kennedy, C. Haselgrove, J. Riehl, N. Preuss, and R. Buccigrossi. The nitrc image repository. *Neuroimage*, 124:1069–1073, 2016.

[47] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, January 1982.

[48] Ashok Kusagur, SF Kodad, and BV Sankar Ram. Modeling, design & simulation of an adaptive neuro-fuzzy inference system (anfis) for speed control of induction motor. *International Journal of Computer Applications*, 6(12):29–44, 2010.

[49] Nicolas Lachiche, Jean Hommet, Jerzy Korczak, and Agnes Braud. Neuronal clustering of brain fmri images. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 300–305. Springer, 2005.

[50] Megan H. Lee, Carl D. Hacker, Abraham Z. Snyder, Maurizio Corbetta, Dongyang Zhang, Eric C. Leuthardt, and Joshua S. Shimony. Clustering of resting state networks. *PLOS ONE*, 7(7):1–12, 07 2012.

[51] H. Li, S. Zhang, X. Ding, C. Zhang, and P. Dale. Performance evaluation of cluster validity indices (cvis) on multi/hyperspectral remote sensing datasets. *Remote Sensing*, 8(4):295, 2016.

[52] David EJ Linden, David Prvulovic, Elia Formisano, Martin Völlinger, Friedhelm E Zanella, Rainer Goebel, and Thomas Dierks. The functional neuroanatomy of target detection: an fmri study of visual and auditory oddball tasks. *Cerebral cortex*, 9(8):815–823, 1999.

[53] Martin A. Lindquist. The statistical analysis of fmri data. *Statistical Science*, 23(4):439–464, Nov 2008.

[54] H. Luo and S. Puthusserypady. fmri data analysis with nonstationary noise models: A bayesian approach. *IEEE Transactions on Biomedical Engineering*, 54(9):1621–1630, 2007.

[55] Ebrahim H Mamdani and Sedrak Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*, 7(1):1–13, 1975.

[56] D. Margulies, J. Böttger, X. Long, Y. Lv, C. Kelly, A. Schaefer, Dirk Goldhahn, A. Abbushi, M. Milham, G. Lohmann, and A. Villringer. Resting developments: a

review of fmri post-processing methodologies for spontaneous brain activity. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23:289–307, 2010.

[57] Samuele Martinelli, Alberto Arturo Vergani, and Elisabetta Binaghi. Comparison of validity indexes for fuzzy clusters of fmri data. In João Manuel R. S. Tavares and Renato Manuel Natal Jorge, editors, *VipIMAGE 2019*, pages 169–178, Cham, 2019. Springer International Publishing.

[58] Thomas Martinetz and K. Schulten. A "neural-gas" network learns topologies. *Artificial neural networks*, 1:397–402, 01 1991.

[59] Martin J. McKeown, Tzyy-Ping Jung, Scott Makeig, Greg Brown, Sandra S. Kindermann, Te-Won Lee, and Terrence J. Sejnowski. Spatially independent activity patterns in functional mri data during the stroop color-naming task. *Proceedings of the National Academy of Sciences*, 95(3):803–810, 1998.

[60] M.K. Pakhira, S. Bandyopadhyay, and U. Maulik. Validity index for crisp and fuzzy clusters. *Pattern recognition*, 37(3):487–501, 2004.

[61] Sankar K Pal and Sushmita Mitra. Multilayer perceptron, fuzzy sets, classifiaction. *IEEE transactions on neural networks*, 3, 1992.

[62] Russell A Poldrack. Region of interest analysis for fmri. *Social cognitive and affective neuroscience*, 2(1):67–70, 2007.

[63] Marcus E. Raichle, Ann Mary MacLeod, Abraham Z. Snyder, William J. Powers, Debra A. Gusnard, and Gordon L. Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682, 2001.

[64] Marcus E. Raichle and Abraham Z. Snyder. A default mode of brain function: A brief history of an evolving idea. *NeuroImage*, 37(4):1083 – 1090, 2007.

[65] Tariq Rashid. Clustering. *Osmar R. Zaiane:" Principles of Knowledge Discovery in Databases*, 2011.

[66] R.M Rezaee, B.P.F. Lelieveldt, and J.H.C. Reiber. A new cluster validity index for the fuzzy c-mean. *Pattern recognition letters*, 19(3-4):237–246, 1998.

[67] Miriam Sklerov, Eran Dayan, and Nina Browner. Functional neuroimaging of the central autonomic network: recent developments and clinical implications. *Clinical Autonomic Research*, 29(6):555–566, 2019.

[68] Michio Sugeno. *Industrial applications of fuzzy control.* Elsevier Science Inc., 1985.

[69] H. Sun, S. Wang, and Q. Jiang. Fcm-based model selection algorithms for determining the number of clusters. *Pattern recognition*, 37(10):2027–2037, 2004.

[70] Ming Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning*, pages 330–337, 1993.

[71] Amanda V. Utevsky, David V. Smith, and Scott A. Huettel. Precuneus is a functional core of the default-mode network. *Journal of Neuroscience*, 34(3):932–940, 2014.

[72] M. P. van den Heuvel, R. Mandl, and H. H. Hulshoff Pol. Normalized cut group clustering of resting-state fmri data. *PLoS ONE*, 3, 2008.

[73] A.A. Vergani and E. Binaghi. A soft davies-bouldin separation measure. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2018.

[74] Binaghi E. Vergani A.A., Martinelli S. Cluster analysis of functional neuroimages using data reduction and competitive learning algorithms. *Lecture Notes in Computational Vision and Biomechanics*, 27(4):62–71, 2017.

[75] Justin L Vincent, Itamar Kahn, Abraham Z Snyder, Marcus E Raichle, and Randy L Buckner. Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of neurophysiology*, 100(6):3328–3342, 2008.

[76] Henson R.N. Wakeman D.G. A multi-subject, multi-modal human neuroimaging dataset. *Scientific data*, 2:150001, 01 2015.

[77] Yanlu Wang and T. Li. Analysis of whole-brain resting-state fmri data using hierarchical clustering approach. *PLoS ONE*, 8, 2013.

[78] T. Warren Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857 – 1874, 2005.

[79] A. Wismuller. Model-free functional mri analysis based on unsupervised clustering. *Journal of Biomedical Informatics*, 37:10–18, 2004.

[80] M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith. Fully bayesian spatio-temporal modeling of fmri data. *IEEE Transactions on Medical Imaging*, 23(2):213–231, 2004.

[81] Keith J. Worsley. An overview and some new developments in the statistical analysis of pet and fmri data. *Human Brain Mapping*, 5(4):254–258, 1997.

[82] X.L. Xie and G. Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(8):841–847, 1991.

[83] Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

[84] Kaan Yetilmezsoy, Merv Fingas, and Ben Fieldhouse. An adaptive neuro-fuzzy approach for modeling of water-in-oil emulsion formation. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 389(1-3):50–62, 2011.

[85] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, jun 1965.

[86] Foxiao Zhan, Xiaolan Zhu, Lei Zhang, Xuexi Wang, Lu Wang, and Chaoyi Liu. Summary of association rules. *IOP Conference Series: Earth and Environmental Science*, 252:032219, 07 2019.

[87] Shu Zhang, Xiang Li, Jinglei Lv, Xi Jiang, Lei Guo, and Tianming Liu. Characterizing and differentiating task-based and resting state fmri signals via two-stage sparse representations. *Brain imaging and behavior*, 10(1):21–32, 2016.