UNIVERSITY OF INSUBRIA

Department of Theoretical and Applied Science

Ph.D. Thesis

to obtain the title of

Doctor of Philosophy

in

Computer Science
and
Computational Mathematics

Candidate
Alberto Arturo Vergani

Contributions in Computational Intelligence with Results in Functional Neuroimaging

Advisor: Elisabetta Binaghi

to be discuss on December 2018

Virtus non timet quod facit

Declaration

I, Alberto Arturo Vergani, declare that this Ph.D thesis entitled "*Contributions in Computational Intelligence with Results in Functional Neuroimaging*" was carried out by me for the degree of Doctor of Philosophy in Computer Science and Computational Mathematics under the guidance and supervision of Prof. Elisabetta Binaghi, Department of Theoretical and Applied Science, University of Insubria, Varese, Italy.

I declare that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person.

I also declare that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma.

Place: _____ (Alberto Arturo Vergani)

Date: _____

Acknowledgements

I would like to thank Prof.ssa Elisabetta Binaghi to have supervised me during the PhD studies at University of Insubria. She introduced me into the field of computational Intelligence with application in Neuroimaging and integrated me into the CRAIIM research group: in particular, I am so grateful for the expertise shared by Dr.ssa Valentina Pedoia, Dr.ssa Sabina Strocchi, Dr. Sergio Balbi and Dr. Renzo Minotto. I also thank for their scientific suggestions Dr.ssa Gloria Gonella, Dr.ssa Irene Biotti, Dr.ssa Giorgia Bigai, Dr. Samuele Martinelli, Dr. Andrea Costantino and Dr. Andrea Montalbetti, which I wish to all of them a rewarding professional carrier.

There are also many people from University that have earned my gratitude for their contribution to my time, directly and indirectly. They are professors, researchers, technicians and PhD co-workers from the Departments of Theoretical and Applied Science in Varese.

I am also thankful with Prof. Marco Donatelli, coordinator of the PhD program in Computer Science and Computational Mathematics, which promoted several stimulating experiences for PhD students.

The last but most important thanks to my *family* and *friends*.

The title of the thesis is Contributions in Computational Intelligences with Results in Functional Neuroimaging.

It is no a statement on a general outcomes, but an indication about new contributions that are presented in this thesis regarding the scientific framework of the Computational Intelligence (CI). There is also a specification that the methodological contributions presented have been applied in the context of the Functional Neuroimaging (FN).

Therefore, the title wants to orientate the readers on a major topic that is referred to contributions related to the field of Computer Science (CS), with also a declination of their potentialities in the experimental settings of Imaging Sciences (IS).

Abstract

This thesis applies computational intelligence methodologies to study functional brain images. It is a state-of-the-art application relative to unsupervised learning domain to functional neuroimaging. There are also contributions related to computational intelligence on topics relative to clustering validation and spatio-temporal clustering analysis. Specifically, there are the presentation of a new separation measure based on fuzzy sets theory to establish the validity of the fuzzy clustering outcomes and the presentation of a framework to approach the parcellation of functional neuroimages taking in account both spatial and temporal patterns. These contributions have been applied to neuroimages obtained with functional Magnetic Resonance Imaging, using both active and passive paradigm and using both in-house data and fMRI repository. The results obtained shown, globally, an improvement on the quality of the neuroimaging analysis using the methodological contributions proposed.

Contents

Chapter 1

Introduction

In the the Introduction chapter there are two sections:

- the part related to the Motivations 1.1, i.e., there is a sinthetical presentation of the principal motivations with also a more extended explanations about the arguments that sustain their statements;
- the part related to the Structure of the thesis 1.2, i.e., a conceptual frame that allow the reader to see macroscopically the core of the thesis topic, and that permits to select the specific parts if necessary.

1.1 Motivations

1.1.1 Principal motivations

Principal motivation to apply soft artificial intelligence principle to neuroimages The complexity[1] of the scientific brain representation[2] necessities approximated methodologies of explorations, i.e., algorithms, methods and research frameworks belonged to computational intelligence theories and techniques, specially the ones based on fuzzy theory and competitive learning.

Principal motivation to use model-free methods to recognize patterns in neuroimages The modularity conceptualization of the central nervous systems, principally adopted for the telencephalon (i.e., the cerebrum), favourites the adoption of families of theory-free methodologies, i.e., unsupervised machine learning methods that work without neither notions or assumptions on the physiology of the brain. The modularity property of a system is basically defined as subcomponents which can perform unique functions and can adapt or evolve to external demands. The modules, therefore, may be study with parcellation methods, e.g. clustering algorithms.

1.1.2 Explanation of the principal motivations.

The complexity of the brain One of the more sparkling but also mazy scientific object is the brain, and in particular the mammalian brain. Obviously, to speak the brain as the singular object is an abstracted and easy way to conceptualize the structures and the functionalities that underpin the qualities that differentiate human beings from not humans. The parts that shape the central nervous system structures in mammals are articulated in a multi-scale mode, ranging from sizes under the nanometres (e.g., the soma of neurons, synapse buttons, dendrites, etc.) or over some centimetres (e.g., nuclei, lobes, networks, etc.). Therefore, several approaches may be used to explore the properties of these parts and their physiological functions with the respect of multi-scaled spatio-temporal processes (e.g., physics, biology,

[1] The notion of complexity is referred to the modelling framework that defines a complex system in terms of its subcomponents and their interactions, which together form a network [82].

[2] The neuroimages are *de facto* a representations of the brain structures (and functionalities) made by acquisitions from a scanner and a sequential procedures able to process the al image processing steps (see the Functional Neuroimaging section 3 for more details

electro-physiology, neuroimaging, cognitive psychology, etc[3]). Many authors from several fields tried to represent the many-sided values of the neuroscientific research. A recent example is the one from Bassett and Gazzaniga [11] that stated the following considerations:

> The brain, we argue, can be understood as a complex system or network, in which mental states emerge from the interaction between multiple physical and functional levels. Achieving further conceptual progress will crucially depend on broad-scale discussions regarding the properties of cognition and the tools that are currently available or must be developed in order to study mind–brain mechanisms.

The authors claimed about the stratifications of disciplines that are necessary to study the complexity of the brain. More in details, they underlined the spatial and the temporal domains, i.e.,

- in the spatial domain, the brain has similar organization at multiple resolutions. The cells themselves are heterogeneously distributed within the brain [204], in addition to the spatial distributions of molecules inside neuronal and non-neuronal cells. There are vertical columns through the cortical layers of the brain that have around 100 neurons, that are roughly 30 microns in diameter and they form the anatomical structure of columns, other areas such subareas (e.g., V1), macro-areas (e.g., visual cortex), cortical lobes (e.g., the occipital lobe). According to recent evidences, the connectivity between those components are also heterogeneous [99, 11].

- in the temporal domain, the difference in scale is evident across the inherent rhythms of brain activity [10] that vary in frequency and relate to different cognitive functionalities. The highest frequency *gamma* band (more then 30 Hz) is thought to be necessary for the cognitive binding of information from several sensory and cognitive modalities [103], while the *beta* (12–30 Hz), *alpha* (8–12 Hz), *theta* (2–4 Hz) and *delta* (1–2 Hz) bands each modulate other distinct but complementary functions [245]. The human cognitive functions are characterize by a broad range of temporal scales: e.g., the pattern of neuronal connections in the brain changes with learning and memory through the process of

[3]Philosophy is also discipline that attempts to examine fundamental questions related the brain processes (e.g., consciousness, decisions and agency, mind-body problem, etc.) and the sub-field of the philosophy of science that studies this topics is called philosophy of neuroscience or more in in general philosophy of cognitive science (see [30, 237])

synaptic plasticity on both short (seconds to minutes) and long (hours to days to months) timescales [1]. Furthermore, the brain organization dynamically changes over multiple temporal and spatial scales: e.g., there are evidences in functional [148], structural [159] and connectivity [176] signatures of short-term development and longer-term ageing have been found with experimental measures.

Furthermore, Bassett and Gazzaniga [11] sustained arguments on the other important neuroscientific topics, as the i) modularity notions of the the brain organization, ii) the relation between structure and function and iii) the emergence[4] phenomenon. From the list above, the brain modularity is a property that has characterized the neural substrates to be investigate with the methods belonged to the discrete mathematics – in general – but with framework coming from graph theory and unsupervised learning theory[5] In particular, the modularity property of the brain is important for its similarity with the notion of brain parcellation, i.e.,

- The general concept of modularity is that the components of a system can be categorized according to their functions. Components that subserve a similar function are said to belong to a single module, whereas components that subserve a second function are said to belong to another module. Modularity can also be defined mathematically in terms of network organization [34, 159, 66]. Nodes that share many common links are said to belong to a module, whereas nodes that do not share many links are likely to be assigned to different module

- The concept of brain parcellation allows divide the spatial domain of the brain into a set of non-overlapping regions, i.e., the modules, that show some homogeneity with respect to information provided by one or many image modalities, e.g., regularities in the cyto-architecture, anatomical connectivity, functional connectivity, or task-related activation (cognitive functionality). Brain parcellations are therefore often the results from tailored clustering algorithms applied to brain images. As referred in [238], there are three strategies are commonly used to study function beyond the voxel parcellation: i) the use of anatomical

[4]Multiscale organization is one hallmark of complex systems and provides the structural basis for another defining phenomenon; this is the concept of emergence in which the behavior, function and/or other properties of the system (e.g. consciousness or the subjective features of consciousness (i.e., *qualia*) are more than the sum of the system parts at any particular level or across levels [83]. In fact, such system properties can emerge from complex patterns of underlying subsystems.

[5]This topic will be discussed in the Computational Intelligence section.

or functional regions of interest (ROIs), ii) the use of a brain atlas, and iii) the use of data-driven parcellations.

- the ROI-based analysis is a model-based approach used to focusing data analysis on some anatomical structures of interest. It consists in building a summary of the signal in a predefined region [184], classical performed with statistical signal processing. The decision of the regions to be analyses could be based on predefined experiments [221]. The obvious limitation of ROI-based analysis is that the signal present outside the region under consideration is ignored a priori; as a consequence, the results depend heavily on the choice of this ROI, which may not fit well the new data. In the hypothesis testing framework, the smaller number of tests performed may, however, increase the power of the analysis.
- the Brain atlases approach come into play to provide a set of ROIs that cover the brain volume (among many others see e.g.,[168, 244, 224]. An atlas generally accounts for a certain state of the knowledge of the brain structures (anatomically, functionally or based on connectivity), from which well-defined entities can be distinguished. In other words, an atlas represents a certain labeling of brain structures. Often this labeling is linked to an ontology representing the current knowledge [61, 42]. In spite of their obvious usefulness, existing atlases are limited in two regards: first, there exist currently many different atlases, but they are mutually inconsistent [27]; second, a given atlas unfits the data well[6].
- the Brain parcellations are based on data-driven methods. They do not have an *a priori* ontology (e.g., brain structure labels), but they may much better investigate the functional measurements, i.e., they allow to provide a better model of the BOLD signal [67, 226, 239, 134, 133]. Only after the brain parcellation it useful to adopt an appropriate labelling of these parcels (e.g., the modules), choosing the most adequate atlas. Data-driven parcellations

[6]Atlas misfits depend to image characteristics and processing procedures that have evolved from the atlas modelling. But also because a given study deals with a population that is not well represented by the subjects used to construct the atlas. Furthermore, the information of interest is not-so-fine mapped into the atlas. The atlas misfits are often evident when it used to map brain functionalities, e.g., many anatomical atlases have large frontal brain regions, whereas researchers would rather separate into smaller parts that have more precise functional roles, or, e.g. there are atlases that do not cover appropriately some brain portions and then it is necessary to manipulate them to build a more comprehensive atlas model (see the Image Processing outcome 4.1 in the Thesis Results section for a practical example of this hybridization procedure).

can be derived from various image modalities reflecting different neurobiological information, for instance T1 images with anatomical information ()e.g., gyro-sulcal anatomy [51, 122], post-mortem in vitro receptor autoradiography for cyto-architecture [62, 64], anatomical connectivity [211] with diffusion imaging, or functional features with BOLD data). In this thesis, we focus on the latter, that we call functional parcellations. These modular parcellations is used either from both resting-state functional Magnetic Resonance Images (rs-fMRIs) [26, 44, 116, 265] and from activation data [67, 134, 133, 177]. There are also usage from meta-analyses framework [61]. In summary, the model free approach is better then model-based since simplified the models and reduce the assumptions[7].

Finally, the brain is a complex objects that is defined as a multi-scaled spatio-temporal union of structural and functional modules investigated by neuroimaging methods. It needs the usage of computational techniques that permits to explore and to extract information with model free approach, basically from the computer science domain. In the next paragraph will be explained why the specific type of methodologies adapted to handle the complexity of the brain (images) is the artificial intelligence, in particular the soft paradigm.

The paradigms of Artificial Intelligence The definition of intelligence has poly-semantic meaning, i.e., there are many possible statements depending on the disciplines used to study the structures of the intelligence, e.g., biology, neurophysiology, psychology, anthropology, sociology, economy, linguistics and computer science (but the list could be go ahead). In a broader sense, it is possible to distinguish the natural intelligence from the artificial intelligence. The position of this PhD thesis regards the artificial intelligence (AI) practically defined as the non-natural intelligence. Therefore, the term intelligence has the general meaning attributed to computer science context.

In the Encyclopaedia of Artificial Intelligence [223], Shapiro defines the AI with this statement:

> Artificial Intelligence is a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behaviour, and with the creation of artefacts that exhibit such behaviour.

[7]This argument will be more discussed in the Functional Neuroimaging section, specially in the part related to Post-Processing of fMRI data 3.2.2

Therefore, according to Shapiro's definition, the intelligence of a machine is more related to its outside performance, rather than its internal (psychological or cognitive) processes.

Furthermore, according to McCarthy's opinions[8], the branches of AI are – roughly – the following:

- logical AI, i.e., intersection between formal thinking and computer programming;
- search, i.e., the explorations of alternatives taking account a large number of possibilities to address a particular problem;
- pattern recognition, i.e., the machine ability to identify a on objects (real or abstracted) with a just-known similar stereotype;
- representation, i.e., the formal reference that associates real things with mathematical significance;
- inference, i.e., the argumentative step that allow to link premises with conclusions;
- common sense knowledge and reasoning, i.e., the processes to manipulate knowledge with the human mind perspective;
- learning form experience, i.e., the AI approach or paradigm established on the concept of connectionism and artificial neural networks;
- planning, i.e., a sequence of actions that consent to achieve a goal (or multi-goals) with the notions regarding facts about the world;
- epistemology, i.e., the philosophical discipline oriented to study the knowledge to advance in the problem solving attitude;
- ontology, i.e., the philosophical discipline oriented to defined the types of existing things;
- heuristic, i.e., a system of trying to determine the content imbedded in a program;
- genetic programming, i.e., the coding paradigm centred on the evolutionary rules.[126].

[8]See the webpage on this topic for more details.

Artificial Intelligence is also distinguished in symbolic and non-simbolic paradigms. Briefly, the main idea behind the difference between the paradigms[9] is the quality of the knowledge representation. The subsimbolic AI try to map the imitate the process of human cognition, e.g., perception, learning, pattern recognition, language production and translation, game plays, moral decision, emotional understanding and so on, using an approach based on connectionism (i.e., distributed form of knowledge representation), whereas the symbolic AI try to do with system that explicitly manage semantic symbols and rules. For more detailed and historical considerations about their difference see the Nillson's book [185].

In this thesis, as mentioned in the motivation above, the main approach is the usage of subsimbolic AI, also known as soft computing or computational intelligence. The core concept is related to connectionism, that is a principle used to model the neural networks, that are an example of soft computing since they solve problems which cannot be solved with logical certainty, and where an approximate solution is often sufficient. ANN where presented in the Forties by McCulloch and Pitts with their pioneer paper titles *A logical calculus of the ideas immanent in nervous activity* [169], and with other research that started to idealize other neural networks concepts, e.g., Rosenblatt [213], Minsky and Papert [179] and the learning models proposed by Hebb [102], and Rumelhart that made many contributions on this field of application, in particular with topic in mathematical psychology and parallel distributed processing [217, 218].

In the Computational Intelligence section many of these concepts will be develop in more details.

[9]Nowadays the symbolic and non-symbolic paradigms are two faces of the same medals, i.e., the Artificial Intelligence science, and they are used together in a non-exclusive way. So they are complementary and - depending the application - researchers select the appropriate theoretical framework. But, when will be the time that one approach dominates the other could be the case that is called, in philosophy of science, the scientific revolution, or in other terms, the so-called paradigm shift claimed by Kuhn [129, 128].

1.2 Structure of the thesis

The thesis is shaped with five chapters, including introduction and conclusions, and one appendix. For the two main chapters there are the adoption of this terminology to divide the contents:

- the cornerstones, i.e., the concepts that are fundamentals to frame the ones presented in the backgrounds section;
- the backgrounds, i.e., the concepts that have been practically adopted in the researches, or that in general are the state-of-the-art about the the topics in computational intelligence and in functional neuroimaging;
- the contributions and the concise results, i.e, the innovations proposed in this theses, that regards i) the are ideas related to soft computing and artificial intelligence based methodology, and ii) that are applications of the concepts proposed in the context of the functional neuroimaging data analysis.

More in details, the structures of the thesis is the following:

- the introduction 1, i.e., where there are the presentation of the motivations that explain the usage of computational intelligence methodologies in the context of functional neuroimages; there are the principal motivations presentation and the extended explanation about these motivations 1.1
- the chapter of Computational Intelligence (CI) 2, i.e., where there are conceptual divisions with cornerstones, backgrounds and contributions related to CI; in particular, there are the following substructures:
  - cornerstones 2.1, i.e., the computational intelligence definition, the learning theory and the sets theory (crisp and fuzzy ones);
  - backgrounds 2.2, i.e., the definition of clustering analysis, the fuzzy algorithms, the competitive learning algorithms, the validation measures and the spectral analysis;
  - contributions 2.3, i.e., the soft version of a crisp separation measures and the new framework to cluster spatio-temporal data;
- the chapter of Functional Neuroimaging (FN) 3, in particular there are the following substructures:
  - cornerstones 3.1, i.e., the differentiation between uni-modal functional neuroimaging and multi-modal functional neuroimaging;

  - backgrounds 3.2, i.e., an overview of the functional Magnetic Resonance Imaging (fMRI) and the procedures used to do pre-processing and post-processing the fMRI data
  - concise results 4.4, i.e., the sequence of outcomes presented in a light manner, with the possibility to extend their topic with the Thesis Results section;

- the chapter with the PhD Thesis Results4, i.e., where there are the outcomes obtained with the computational intelligence methodologies applied to functional neuroimages; in particular the results presented regard the following areas:
  - the Image Processing, i.e. the usage of computational procedures able to manipulate neuroimages 4.1;
  - the Medical application, i.e. the usage of computational procedures with the aim to study clinical subjects and healthy subjects that have had an fMRI clinical exam with active paradigm 4.2 or an fMRI cognitive experiments both with passive paradigm and active paradigms 4.3.

- the chapter with Conclusion5, i.e., where there the final propositions that collect the concepts and outcomes achieved, and the presentation of the future works;

- the Appendix part 6, i.e., where there are lists of images, algorithms exposed in the computational intelligence chapter, and the atlases used for the processing of the neuroimages.

Chapter 2

Computational Intelligence

The chapter on Computational Intelligence (CI) has three sections:

- the part on the cornerstones related to the CI topics used for the thesis 2.1, i.e., important definitions, notions on learning theory and sets theory;
- the part on the specific backgrounds used for the thesis 2.2, such as the general presentation of the clustering algorithms, the declination with competitive and fuzzy ones, the validation measures used to test the clustering outcomes and the spectral analysis techniques adopted in the thesis;
- the part on the contributions proposed in the computational intelligence domain 2.3, that in particular regard the a new soft separation measures and a new clustering framework for spatio-temporal analysis.

2.1 Cornerstones

The section cornerstones is divided in the three subsections:

- the part on the definition of the computational intelligence 2.1.1, with some outlines on its specificities and connections with other disciplines;
- the part that describes the learning theory in computer science 2.1.2, that regards the qualities of learning, e.g., supervised, semi-supervised, active, reinforcement, unsupervised one;
- the part that shows the sets theory 2.1.3, which regards the differences on the crisp sets theory and the fuzzy sets theory.

2.1.1 Computational Intelligence

A machine is intelligent when learns knowledge from data. This computerized ability is defined as Computational Intelligence (CI). Some authors refer this name to the synonymous of soft computing, but others are not in agreement.

The scientific ground where CI operates is the one where other discipline are useless. The nature-inspired methodologies have the abilities to solve complex problems hard to solve with classical computational models, e.g., problems that contain stochastic information, problems that have uncompleted data or that have noise or vagueness, and also problems that are difficult to be converted in digital form, such as the 0-1 discretization, often used in many computer science domains.

The main features that have machine able to define as intelligent is the capacity to work like a human reasoning system, that have the ability to handle knowledge with incomplete structures or inexact/distort truth representation about the objects considered [225], i.e.,

- the skill to process information like natural language is manage by human-like logic (multivalue logics or many-value logics) [186];
- the skill to learn from a samples of data and to generalize to know new ones with inference or with novelty recognition;
- models that mimes the natural selection able to compute knowledge as an evolutionary processes. [115].

A more operative but formal definition is the one contained in the Springer's Handbook of Computational Intelligence, edited by Kacprzyk and Pedrycz in 2015 [115], that it is referred with the citation coming from the Constitution of IEEE CIS[1], i.e.,

> The Field of Interest of the Society is the theory, design, application, and development of biologically and linguistically motivated computational paradigms emphasizing neural networks, connectionist systems, genetic algorithms, evolutionary programming, fuzzy systems, and hybrid intelligent systems in which these paradigms are contained (article I, Section 5)

Kacprzyk and Pedrycz claimed, also, that the above definition could be extended remarking that computational intelligence is a broad and diverse collection of nature inspired methodologies and apporocheas, and tools and

[1]Intitute of Electrical and Electronics Engineers - Computational Intelligence Society

techniques that are meant to be used to model and solve complex real problems in which the traditional approaches[2] based on strict and well-defined tools and techniques are either not feasible or not efficient. Therefore, the usage of intelligent machines allow to cope with the complexity of a problem using the smartness that has human-like models or - more generally - the nature-expired models.

A general framework that refers on the computational intelligence is the computational learning theory, i.e., a mathematical and computer science background used to study the machine learning algorithms, where the computation complexity is one of the most struggled topic (e.g., the $P \neq NP$ millennium problem).

In this section, some details about computational intelligence will be presented. In particular, the next subsections will be on general learning theory 2.1.2 and the sets theory 2.1.3. This two di theories are joined since they they have many relations: e.g., learning models classify new objects according to a specific definition of set. Therefore, the semantic of the learning process depends on the theory that defined the object learnt[3].

2.1.2 Learning Theory

Learning is the general ability that have systems to progressively improve the execution on a predefined task. Mitchell defines the machine learning ability as the capacity of a machine M to improve its capacity to solve the task T given a set of operations O able to evaluate its performance P [180].

These learning systems are not explicitly programmed, i.e., there are no explicated control structures, such as IF-THEN rules and similar. The learning processes are the results of statistical computation that allows to computer machine to learn [2], but also with mathematical connections related to optimization, operative research and inverse problem theory. The machine learning is a wide categories in which lie sub-disciplines that have different approach on learning, i.e., data mining is the one that is principally based on the unsupervised learning algorithms, whereas pattern recognition is the one mostly based on the supervised learning algorithms. Those differentiation are fundamental for the learning theory. In particular, there are the following types of learning approach:

- the supervised learning is the type of learning that uses the data truth,

[2]Traditional approaches are hard mathematical modelling, optimization, control theory, stochastic analysis, etc.

[3] To look about computational learning theory see the online book by Goodfellow, Benio and Curville [90].

i.e., the labels that allow the system to recognize or or to infer those labels given specific sets of inputs (i.e. the main tasks are the classification and the prediction).

- the semi-supervised learning is a particular type of supervised learning with the difference that it used only incomplete sets of labels, e.g., with datasets that have many missing values.
- the active learning is the type of learning in which the machine has to decide what kind of labels needs to learn and to optimize its decision with some budgets avalaible.
- the reinforcement learning is the type of learning with many similarity with the psychological learning, i.e., it is based on the difference between rewards and punishments that gives to the machine a dynamical feedback, e.g., systems that drives vehicles in a civil context or system that play games in sportive area are often based on this kind of learning.
- the unsupervised learning is the type of learning without data truth, i.e. without labels or patterns associated to specific input samples. It means that the unsupervised learning allows the machine to learn from data same possible regularities that shape the data itself. In other words, those kind of learning algorithms without data-model truth are used to explore possible data-driven truth. Applications of unsupervised learning is clustering, that is a family of computational method used to mine, to discover, patterns or labels, directly from the data with which the machine learns.

The above mentioned differentiation on the types of learning are principally based on the notions of data truth. More precisely, the data truth is known in the computer science domain as the ground truth, i.e., the empirical knowledge about a phenomena that allows to infer, given some data inputs as a premises, some outputs as a conclusions. When it is perfectly known is the case of the supervised learning, *viceversa*, when it is perfectly unknown is the case of unsupervised learning; the other kinds of learning are - globally - seeble as a mediation between these extremes. Furthermore, with the ground truth noted is possible to evaluate the learning algorithm during its learning processes: it is possible to design a procedure able to measure the errors with respect to the correct goals to achieve, i.e., the so-called confusion matrix, that is a special contingence table where is measured the confusion, the errors rates, about the systems investigated. In particular, there are selected indexes that are used for this operation, e.g., the sensitivity (or recall

or true positive rate), the specificity (or selectivity or true negative rate), the precision (or the positive predictive value) and the accuracy (see [180, 58, 90] for their mathematical details). Also, other important indexes are the ones noted in the statistical domains, e.g., the true positive, the true negative, the false positive and the false negative[4].

2.1.3 Sets Theory

The learning theory regards algorithmic processes, or steps, that have as results some considerations about the identification of regularities between mathematical objects. This means that once it has been identified the property that have in common many of them, the pattern identification defined geometrical boundaries shaping those objects. Frames, boundaries, geometrical definitions, classes, clusters, partitions, and so on, are names that referred on a general abstract concept, i.e., the sets. The following paragraphs will present the basic knowledge about crisp sets and fuzzy sets. Their have properties in common, but also differences that motivate the appropriated theoretical framework to use in computational intelligence applications.

Crisp Sets Cantor first proposed the set theory in the nineteenth century. The notions he defined have had a decisive role in the all branches of mathematics and applied sciences [32]. The next paragraph will expose same basic notions on crisp set theory. In particular, there will be the exposition of fundamental definitions about sets, the operations that allow to manipulate sets, the relations that permit to associate them and special kind of relations that are very important in many type of applications.

Concept of sets In the paragraph there are selected notions about sets regarding the definition of sets, the principle of extensionality, particular type of sets as the empty and the power ones, and the notion about the cardinality of a set.

- A set A is a collection of things a, e.g., ideas, numbers, objects, etc, that belong together for precise motivation. These objects are called elements of the set. Formally an element belong to a set is defined as $a \in A$, whereas if it does not belong to a set is defined as $a \notin A$. Sets can be described as an enumeration of elements in braces, e.g., $Animals = \{dolphins, dogs, cats, etc\}$ or $Integers = \{1, 2, 3, etc\}$ or $\mathbb{Q} = \{\frac{p}{q} | p, q \in \mathbb{Z} \wedge q \neq 0\}$.

[4]See the section on Validation Measure 2.2.4 for details about the evaluation of clustering results.

- The principle of extensionality for a sets states that two sets A and B are identical if and only if they have the same elements, i.e., $A = B \Leftrightarrow \forall x(x \in A \Leftrightarrow x \in B)$.

- The subset is defined as the following: if A and B are the sets and $\forall x(x \in A \Rightarrow x \in B)$ holds, then A is called a subset of B

- The empty set or void set is important and useful in many applications. It is defined as $\emptyset$ and it means the set that does not contain elements. Because of the principle of extensionality, there is only one empty set.

- The equality of sets is the characteristics for which two sets are equal if and only if both are subsets of each other, i.e., $A = B \Leftrightarrow A \subseteq B \wedge B \subseteq A$.

- The power set is the set of all subsets A of a set M. It is denoted as $P(M) = \{A|A \subseteq M\}$. Note that, if a set M has m elements, its power set $P(M)$ has 2^m elements. Note also that, for every set M there are $M, \emptyset \in P(M)$, i.e., M itself and the empty set are element of the power set of M.

- The cardinal number is the number of elements of a finite set M. It is denoted as $|M|$.

Operations and representations with sets Manipulate sets with appropriate operations is important. Represent them with symbolic tools helps to draw graphical hints about their structure. In the paragraph there are the definition of Venn diagram and the main sets operation, e.g., the union, the intersection, the disjoint, the complement, the Cartesian product, and the Algebra sets law (including De Morgan's rules [47, 48]).

- The Venn diagram is the graphical representation of sets and set operations [5]

- The union set or the union, i.e., $A \cup B = \{x \mid x \in A \vee X \in B\}$

- The intersection set or intersection (or cut set), i.e., $A \cap B = \{x \mid x \in A \wedge x \in B\}$.

- The disjoint sets are sets that have no elements shared, i.e., $A \cap B = \emptyset$

[5] See the following link to know more about Venn diagram.

- The complement set, or the complementary set, is formally $C_M(A) = \{x \mid x \in M \wedge X \notin A\}$. If one considers only the subsets of a given set M, the complementary set $C_m(A)$ of A with respect to M have all the elements of M not belonging to A.

- The difference of two sets is a set of the elements of A that does not belong to B, i.e., $A \setminus B = \{x \mid x \in A \wedge \notin B\}$.

- The Cartesian product of two sets is formally defined as $A \times B = \{(a, b) \mid a \in A \wedge B \in B\}$. The elements (a, b) of $A \times B$ are ordered pairs and the number of elements of a Cartesian product of two finite sets is equal to $card(A \times B) = (cardA)(cardB)$.

- The laws of set Algebra

  - associative laws, i.e., $(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$;
  - commutative laws, i.e., $A \cap B = B \cap A$ and $A \cup B = B \cup A$
  - distributive laws, i.e., $(A \cup B) \cap C = (A \cap B) \cup (B \cap C)$ and $(A \cap B) \cup C = (A \cup B) \cap (B \cup C)$;
  - adsorption laws, i.e., $A \cap (A \cup B) = A$ and $A \cup (A \cap B) = A$
  - idempotence laws, i.e. $A \cap A = A$ and $A \cup A = A$
  - De Morgan laws, i.e., $\overline{A \cap B} = \overline{A} \cup \overline{B}$ and $\overline{A \cup B} = \overline{A} \cap \overline{B}$
  - other laws, i.e., $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = M$, $A \cap M = A$, $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup M = M$, $\overline{M} = \emptyset$, $\overline{\emptyset} = M$ and, finally, $\overline{\overline{A}} = A$.

Relation and mappings Relations and mapping are important types of associations between sets. They allow to create more complex structures and to establish fundamental correspondences that linked sets.

- the n ary relations are correspondences between the elements of one or several sets. An n-ary relation or n-place relation R between the sets $A_1, \ldots, B_n$ is a subset of the Cartesian product of these sets, i.e., $R \subseteq A_1 \times \cdots \times A_n$. If the sets A_i, with $i = 1, \ldots, n$, are all the same set A, then $R \subseteq A^n$ holds and it is generally named an n ary relation in the set A. If $n = 2$, the correspondences are called binary relations, i.e., aRb.

- There are important properties of binary relations. The relation R is called

- reflexive, if $\forall a \in A aRa$
- irreflexive, if $\forall a \in A \neg aRa$
- symmetric, if $\forall a, bin A(aRb \Rightarrow bRa)$
- antisymmetric, if $\forall a, b \in A(aRb \wedge bRa \Rightarrow a = b)$
- transitive, if $\forall a, b, c \in A(aRb \wedge bRc \Rightarrow aRc)$
- linear, if $\forall a, b \in A(aRb \vee bRa)$

- the mappings or functions from a set A and B with notation $f : A \longrightarrow B$ is a rule to assign to every element $a \in A$ one element $b \in B$, which is called $f(a)$. A mapping f is a subset of $A \times B$, i.e., $f = \{(a, f(a)) \mid a \in A\}$. Note that, i) f is called injective if to every $b \in B$ at most one $a \in A$ with $f(a) = b$ exists; ii) f is called surjective, if to every $b \in B$ at least one $a \in A$ with $f(a) = b$ exists; iii) f is called bijective if f both injective and surjective [6].

Equivalence and order relations The most important classes of binary associations with respect to a set A are the equivalence and order relations. A binary relation R is defined as equivalence relation if R is reflexive, symmetric and transitive. The notation to define it is aRb or $a \sim b$, i.e., a is similar to b respect to R. A binary relation R is defined as order ordering relations if it is reflexive, antisymmetric and transitive. If R is linear, the relation is named linear ordering or chain. Note that in a linearly ordered set any two elements are comparable. The notation to defined the relation is $a \leq_R b$.

Fuzzy Sets The basic notions of crisp sets have been exposed in thepre-vious paragraphs, with also attention on the properties that regards those sets and their operations and relations. Crisp sets have the peculiarity to frame objects in classes that have unshared elements. But, this kind of peculiarity has been relaxed with fuzzy sets, *viceversa*, allow to frame objects in classes with shared elements. In the following paragraph, there will be presented the general notation and terminology of fuzzy sets, the operations that regard the fuzzy sets, the extension principle and the fuzzy sets with fuzzy membership functions. The main reference are [278] and [279]. Note that notions as fuzzy relations, projections and cylindrical fuzzy sets will no presented there. To deep those concepts see [279].

[6] For a bijective mapping $f : A \longrightarrow B$ there is the inverse relation $f^{-1} : B \longrightarrow A$

Notation and Terminology A fuzzy subset A of a universe of discourse U has a membership function $\mu_A : U \longrightarrow [0,1]$ which associates a number $\mu_A(u)$ in the interval [0,1] to each element $u \in U$. The grade of membership of u in A is $\mu_A(u)$. The set of points in U is the support of A at which $\mu_A(u)$ is positive. The supremum of $\mu_A(u)$ is the support of A. A cross-over point of A is a point U whose grade of membership in A is 0.5.

Following the notation presented in [279], a nonfuzzy finite set $U = \{u_1, \ldots, u_n\}$ will be expressed ad $U = u_1 + u_2 + \cdots + u_n$ or $U = \sum_{i=1}^{n} u_i$. Note that the operation + means the union rather then the arithmetic sum. Therefore, U is the union of its constituent singletons. The extension of the previous definition of U is

$$A = \mu_1 u_1 + \cdots + \mu_n u_n \longrightarrow \sum_{i=1}^{n} \mu_i u_i, \tag{2.1}$$

where μ_i, with $i = 1, \ldots, n$, is the grade of membership of u_i in A. If the u_i are numbers, there might be ambiguity regarding the identity of the μ_i and u_i components of the string $\mu_i u_i$. For this motivation, Zadeh [279] employed a separator symbol such as / for disambiguation, i.e.,

$$A = \mu_i/u_i + \cdots + \mu_n/u_n \longrightarrow A = \sum_{i=1}^{n} \mu_i/u_i. \tag{2.2}$$

From the definition of the union expressed in 2.1, if the $u_i = u_j$, then the formalization of the union is expressed as

$$\mu_i u_i + \mu_j u_j = (\mu_i \vee \mu_j) u_i. \tag{2.3}$$

When the support of a fuzzy sets is a continuum rather then a finite set, the 2.1 becomes

$$A = \int_U \mu_A(u)/u, \tag{2.4}$$

with the meaning that $\mu_A(u)$ is the grade of membership of $u \in A$ and the integral denotes the union of a the fuzzy singletons $\mu_A(u)/u$ with $u \in U$.

A fuzzy set A is normal if its height is unity, i.e., sup $\mu_A(u) = 1$.

A fuzzy subset of U may be a subset of another fuzzy or nonfuzzy subset of U, i.e.,

$$A \subset B \Leftrightarrow \mu_A(u) \leqslant \mu_B(u) \quad u \in U. \tag{2.5}$$

Operation on fuzzy sets According to Zadeh, the basic operation which can be performed on fuzzy sets are the following:

- The complement of A is denoted by $A' = \int_U = [1 - \mu_A(u)]/u$. The complement is a negation. Therefore, if A is a fuzzy set, then not-A is interpreted as A'.

- The union of fuzzy sets A and B is denoted by $A + B = \int_U [\mu_A(u) \vee \mu_B(u)]/u$. The connective or is the union operation, therefore, A *or* B means $A + B$.

- The intersection between two sets A and B is denoted by $A \cap B = \int_U [\mu_A(u) \wedge \mu_B(u)]\ u$. The connective and is the intersection operation, therefore A *and* B means $A \cap B$.

- The product of A and B is denoted by $AB = \int_U [\mu_A(u)]^\alpha\ u$.

- The convex combination[7] of $A_1, \ldots, A_n$ is a fuzzy set A whose membership function is expressed by $\mu_A = w_1\mu_{A_1} + \cdots + w_n\mu_{A_n}$ where $+$ denoted the arithmetic sum.

- The Cartesian product of n subsets of U_n is denoted by $A_1 \times \cdots \times A_n$ and is defined as a fuzzy subset of $U_1 \times \cdots \times U_n$ whose membership is formulated by $\mu_{A_1 \times \cdots \times A_n}(u_1, \ldots, u_n) = \mu_{A_1}(u_1) \wedge \cdots \wedge \mu_{A_n}(u_n)$

- The fuzzification transforms a non-fuzzy set into a fuzzy set or increasing the fuzziness of a fuzzy set. Therefore, a fuzzifier F is applied to a fuzzy subset A of U yealds a fuzzy subset $F(A; K) = \int_U \mu_A(u)K(u)$, where the kernel $K(u) = F(1/u; K)$ with $1/u$ is the singleton.

The extension principle [8] The extension principle is a fundamental concept that allows the domain of the definition of a mapping or a relation to be extended from points in U to fuzzy subsets of U. More formally, if f is a mapping from U to V, and A is a fuzzy subset of U expressed as $A = \mu_1 u_1 + \cdots + \mu_n u_n$, then the extension principle states that $f(A)) = f(\mu_1 u_1 + \cdots + \mu_n u_n) \equiv \mu_1 f(u_1) + \cdots + \mu_n f(u_n)$. Then, the image of A under f can be deduced from the knowledge of the images of $u_1, \ldots, u_n$ under f.

Fuzzy sets with fuzzy membership functions Zadeh pointed up that fuzzy sets with fuzzy membership functions are motivated by the close relation which exists between the concept of a linguist truth with truth-values

[7] The convex combination is able to represent linguistic hedges as essentially or typically, that modify the weigths associated with components of a fuzzy set

[8] In probability theory the analogy with the extension principle is the expression of probability distribution induced by mapping (see ref 30 in [279])

(e.g., true, quite true, very truw, more or less true, etc) and fuzzy sets in which the grades of membership are specified by linguistic term (e.g., low, medium, high, very low, not low, not high, etc.).

Therefore, if A is a fuzzy subset of a universe of discourse U, and the value of the membership function μ_A of A are allowed to be fuzzy subsets of the interval [0,1], then to differentiate such fuzzy sets from those considered previously, Zadeh nominated them as type 2 fuzzy set.

A type 1 membership functions are mappings from $U \longrightarrow [0, 1]$.

More generally, a type n fuzzy set, with $n = 2, 3, \ldots$, if its membership function ranges over fuzzy sets of type $n - 1$. The membership function of a fuzzy set of type 1 ranges over the interval [0,1].

2.2 Backgrounds

The section backgrounds is divided in five subsections:

- the part on the Clustering Analysis 2.2.1, i.e., there are the presentations of the notions related to clustering and formulation of the clustering problem;
- the part on Fuzzy Algorithms 2.2.2, i.e., there are the illustrations of the clustering procedures based on fuzzy theory, in particular the Fuzzy Type 1 C-Means and the Fuzzy Type 2 C-Means clustering methods;
- the part on Competitive Algorithms2.2.3, i.e., the expositions of the clustering procedures based on competitive learning theory, specially the Neural Gas (NG), the Growing Neural Gas (GNG) and the Self Organizing Maps (SOM);
- the part on Validation Measures 2.2.4, i.e., the indexes useful to validate the outcomes of a clustering procedures, taking in account the more relevant methods proposed in the literature, and their association with the significance evaluation of the clusters found (p-value index);
- the part on Spectral Analysis 2.2.5, i.e., the notions related to eigenvectors and eigenvalues analysis of the covariance matrix, with declination on Principal Component Analysis and Non-linear Fuzzy Robust PCA; there are also discussion on the usage of Random Matrix Theory tools for the features selection;

2.2.1 The Clustering Analysis

In this subsection there will be presented the general considerations about the clustering analysis and the formal definition about what is the clustering problem, how it could be approached using different sets theory and an example of the one of the most known clustering algorithm.

Generalities In data mining and machine learning, clustering is an exploratory techniques that allows to classify *ex novo* a given sets of measures. As mentioned in the Learning Theory section 2.1.2, clustering is an unsupervised learning family of algorithms that allows to find patterns, categories, labels, within data that do not have a ground truth knowledge associated. This classification[9] method was introduced in the thirties by Driver and Kroeber [55], Zubin [283], Tryon [243] and in forties by Cattell [41]. Then, the discipline grew up developing several algorithms for different aims and technicalities [10] .

Clustering is divided in two main types: the partitive clustering[11] and the hierarchical clustering. The idea that differentiates the types is following: if the aim of a researcher is to find independent groups of data, without the salience to assume some multi-level classes, the best choice is to use partitive techniques, whereas if the aim of researcher is to find stratification in patterns, then the best choice is to use hierarchical clustering[12]

The clustering analysis Clustering, or partitioning, is a computational [13] methodology used in many fields to draw knowledge on the distribution of raw data and the informative pattern that are within [100, 138]. The main objective of clustering is to collect into groups, i.e., clusters, a given input data taking account a similarity measures predefined, e.g., algebraic distance [189]. Formally, let $X = \{x_1, x_2, \ldots, x_N\}$ be a given input dataset with N points, K is the number of clusters, also known as patterns, within the data.

[9]The term *classification* in machine learning is referred principally for the task to classify points with supervised learning techniques, i.e., when a pattern learnt has been recognized by a machine. But, also, some authors use the term classify for machines that do clustering with the meaning to perform the first-time classification. In this paragraph, it is adopted this latter sense.

[10]See [111] for an historical overview and [215] for a discussion on different clustering approaches

[11]The partitive clustering is the one adopted within this PhD thesis.

[12]For more details about hierarchical clustering see [180, 57].

[13]Clustering theory may be define as in ill-posed problem because a given data-set can be partitioned in many ways without a clear criterion to prefer one cluster over another (see works by Stramaglia et al

The aim of clustering is to compute a partition matrix $U(X)$ of the data in order to define a partition $C = \{C_1, C_2, \ldots, C_k\}$, with the criteria that similar points have high similarity and dissimilar points have high dissimilarity, i.e. the similar points are close as possible, while the dissimilar points are far as possible. The partition matrix is denoted ad $U = [\mu_{ij}]$, $1 \leq i \leq K$, $1 \leq j \leq N$, where μ_{ij} is the grade of membership of point x_j to cluster $C_i (i = 1, \ldots, K)$.

It is important to note that clustering can be performed in two radically different ways: crisp and fuzzy.

- The crisp clustering uses a logical framework based on classic sets theory[14], i.e. any one point of a given dataset belongs to only one class of clusters. It means that exist a pattern, and only one pattern, for each points. Formally, the membership value μ_{ij} has value equal to 1 if $x_j \in C_i$, otherwise $\mu_{ij} = 0$.

- The fuzzy clustering is based on the fuzzy sets theory[15], i.e., a point might may belong to one or more clusters with a certain grade o membership. The partition matrix $U(X)$ is represented as $U = [\mu_{ij}]$, where $\mu_{ij} \in [0, 1]$. Therefore, the crisp clustering is a special case of fuzzy clustering if the membership of a point to a specific cluster is either 0 or 1. Furthermore, there exist methods to hard the fuzzy clustering, e.g., selecting for each points the cluster with associated the largest grade of membership.

An example of crisp clustering algorthm One of the most important crisp algorithm is the k-means algorithm. It was introduced by MacQueen in the Sixties [158] in the context fo signal processing as method for vector quantization. It works with k predefined clusters used to partition n points, i.e., the observations, with the inclusion criteria based on the their similarity with a prototype. In features space, the position of the k centroids is used to evaluate the membership of each points, i.e., one point is within a cluster in which has nearest mean.

More formally, given a set of observation $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$, where each observation is point in a d -dimensional features space, the k-means clustering partitions the n observations into k classes $C = \{C_1, C_2, \ldots, C_k\}$ according to the minimization of the variance within the cluster using an iterative procedure that minimize the squared error of the following objective function:

[14] See the section on Sets Theory 2.1.3 § Crisp Sets for more details.
[15] See the section of Fuzzy Sets Theory 2.1.3 § Fuzzy Sets for more details.

$$J = \sum_{i=1}^{n} \sum_{c=1}^{C} ||X_{i(c)} - A_c||^2 \tag{2.6}$$

where $||...||^2$ is a specified distance function, n are the data points, k are the clusters, $X_{i(c)}$ are the points in cluster c, A_c are the centroids of clusters c.

The fuzzy algorithms will be exposed in the Fuzzy Algorithms section 2.2.2.

Clustering as ill-posed problem According to Hadamard's definition [98] a problem is said to be ill-posed if at least one of the following criteria is not satisfied:

- a solution exists,
- the solution is unique,
- its behaviour depends continuously with the data.

If the above condition are completely satisfied, the problem is defined as well-posed.

A subclass of ill-posed problem are the inverse problems, i.e., an approach that wants to discover causes of an observed effects, e.g., in helioseismology that one wants to determine the structure of the sun by measurements from heart or space, in image restoration that one wants to determine unavailable exact image from an available contaminated version, or in medical imaging where one wants to discover physiological causation related to experimental paradigm.

Clustering is an ill-posed problem since it is lacking at least of the property to have a unique solution, because there are many possible clusters that can make interesting classifications of the input data. Specially, clustering is an inverse problem as its aim is to find patterns, i.e., causes/categories that explain the experimental data measures.

According to Bezdek [17, 19], there are other issues concerning the clustering analysis: 1) the *a priori* assumption that datasets have clusters, 2) the computation method to find clusters, 3) the evaluation procedure to verify the clusters found.

Therefore, merging the requirements to solve the illness defined by Hadamard and the issues proposed by Bezdek, the inverse clustering problem could be address with the following strategy:

- the existence of a solution must be an assumption, i.e., if the goal is to cluster a dataset, the experimentalist need to assume that there are at least two clusters;

- the uniqueness of a solution must be a decision, according to some optimal evaluation methodology, e.g., using a measure able to define what is the optimal clustering result, e.g., using the Davies-Bouldin index before mentioned;

- the matrix condition number must be lowered. Usually the condition number of experimental matrix is very high. Therefore, it is necessary to improve the quality of the matrix to have a more stable solution, i.e., small changes in the matrix should be associated to small changes in the results. More in detail, a mathematical problem with low condition number is said well-conditioned (i.e., $K(A) = 1$), whether if it has high condition number is said bad-conditioned (i.e., $K(A) >> 1$; e.g., singular matrices have $K(A) = \infty$). Specifically, the definition of condition number used for this operation is the 2-norm condition number for inversion of the matrix A, i.e.,

$$K(A) = \frac{\lambda_{max}}{\lambda_{min}} \tag{2.7}$$

  where λ_{max} is the largest eigenvalue of the matrix A and λ_{min} is the smallest eigenvalue of the matrix A [201]

  The matrix condition number decreases using signal processing (e.g., Empirical Mode Decomposition) and data reduction techniques (e.g., Principal Component Analysis 2.2.5).

It is interesting to note that there exists a relation between the Principal Component Analysis and the K-Means clustering algorithm. The theorem formulated by Ding and He [54] indicated that PCA is not only a data reduction methodology, but it works as an equivalent solution of the clustering problem with K-Means algorithm. In particular they related the cardinality of the optimal patterns found by the algorithm with the cardinality of the best informative components subspace. For more details see the Spectral Analysis section 2.2.5.

2.2.2 Fuzzy Algorithms

In this subsection will be presented clustering algorithms based on fuzzy sets theory, in particular the one with fuzzy sets and other one with fuzzy sets that have fuzzy membership functions: they are fuzzy c-means type 1 and fuzzy c-means type 2

Generalities The V is a set of objects. The distance between the elements v_i is a function $d : V \times V \longrightarrow \Re$ with these three properties: $\forall x, y, z \in V$

$$d(x, x) = 0 \tag{2.8}$$

$$d(x, y) = d(y, x) \quad \text{commutative property} \tag{2.9}$$

$$d(x, z) \leq d(x, y) + d(y, z) \quad \text{triangular inequality} \tag{2.10}$$

Generally, the typical distance adopted is the Euclidean metric. Let $V = \boldsymbol{v} = (v_1, \ldots, v_n)$, then the Euclidean distance between $\boldsymbol{u}$ and $\boldsymbol{v} \in V$ is

$$d(\boldsymbol{u}, \boldsymbol{v}) = \sqrt[2]{\sum_{i=1}^{n} (u_i - v_i)^2}. \tag{2.11}$$

The *hard* cluster analysis wants to find a suitable partition $\wp_c$ of a set V into c subsets such that $\wp_c = \{R_j | j = 1, \ldots, c, \quad R_j \subset V\}$ that fulfils the conditions $\cup_{j=1}^{c} R_j = V$, and that $R_i \cap R_j = 0$, $i \neq j$, $1 \leq i$ and $j \leq c$. Also, the minimization of some functional $J(\wp_c)$ allows to get the informal condition that elements in the same cluster $R_j, j \in \{1, \ldots, c\}$ should be mutually closer to each other. In *fuzzy* cluster analysis, the $R_j \subseteq V$ are fuzzy sets and they have to fulfil the following two conditions (i.e. *fuzzy pseudopartirtion properties*: let $V = \{v_1, \ldots, v_r\}$, then

$$\sum_{j=1}^{c} R_j(v_i) = 1, \quad i = 1, \ldots, r \tag{2.12}$$

$$0 < \sum_{j=1}^{c} R_j(v_i) < r, \quad j = 1, \ldots, c. \tag{2.13}$$

The first condition means that any element $v_i \in V$ should belong to at least one cluster and the sum of all its membership degrees to all clusters should be equal to 1. The second condition regards that no cluster can be empty and that there must not exist a cluster that contain all elements of the set V in the maximal degree. The data for $fuzzy$ cluster analysis are real

numbers represented as a matrix with $o_1, \ldots, o_r$ that are objects on which attributes $\phi_1, \ldots, \phi_n$ are measured (e.g. o_r are pixels, voxels, instants, people, machines, etc, whereas ϕ_n can be physical or abstract quantities). Therefore, the data for the *fuzzy* cluster analysis can be characterized as r points in a n-dimensional space, where the elements $u_{r1}, \ldots, u_{rn}$, are coordinates of these points. Given a row vector $\boldsymbol{u_i} = (u_{i1}, \ldots, u_{in}), \quad i = 1, \ldots, r$, and given a matrix H of type (m, n), then H^T denotes its transposition, that is a matrix of type (n, m). Then, $\boldsymbol{u}_i^T$ is the column vector

$$\boldsymbol{u}_i^T = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{in} \end{bmatrix} \tag{2.14}$$

Fuzzy C-Means Type 1 and Fuzzy C-Means Type 2 The most used *fuzzy* clustering method is the *fuzzy c-means*. It is a generalization of the classical *K-means* algorithm independently discovered in several scientific fields [149]. The first version of the fuzzy c-means was presented by J.C. Dunn [59] and improved by J. Bezdek [17]. The majority of the fuzzy clustering methods are derived from the fuzzy c-means [186].

The most important parameter in the fuzzy clustering is the number c of cluster to be found. If clusters $\wp_c = \{R_1, \ldots, R_c\}$ are already known, where $R_j \subset \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$, for all $\wp_c$ it is necessary to compute the *center* as follow:

$$\boldsymbol{s}_j = \frac{\sum_{i=1}^{r} (R_j(\boldsymbol{u}_i))^m \cdot \boldsymbol{u}_i}{\sum_{i=1}^{r} (R_j(\boldsymbol{u}_i))^m}, \quad j = 1, \ldots, c, \tag{2.15}$$

where $m > 1$ is the fuzzy parameter to be chosen beforehand. Note that in the equation 2.15, the $R_j(\boldsymbol{u}_i)$ is the membership degree of the vector $\boldsymbol{u}_i$ in the fuzzy cluster R_j. The minimization of the functional $J_m(\wp_c)$ is the optimal solution of the fuzzy clustering analysis.

In the case of Fuzzy C-Means Type 1, the functional is the following:

$$J_m(\wp_c) = \sum_{i=1}^{r} \sum_{j=1}^{c} (R_j(\boldsymbol{u}_i))^m \cdot d^2(\boldsymbol{u}_i, \boldsymbol{s}_j), \tag{2.16}$$

where d is the distance selected for the computation.

The theoretical extension of fuzzy c-means is the Fuzzy C-Means Type 2. Their main difference is in the computation of the membership functions $R_j(\boldsymbol{u}_i)$, i.e., in the type 1 algorithm the fuzzy sets have crisp membership functions, whereas in the type 2 algorithm the fuzzy sets have fuzzy membership functions[16]. It important to enlighten that the notation for the mem-

[16]See the section on Fuzzy Sets Theory for more details 2.1.3

bership functions computation in the algorithm 1 is $R_j(\boldsymbol{u}_i)$, whereas for the algorithm 2 is $A_j(\boldsymbol{u}_i)$. More in details:

$$R_j^{(l+1)}(\boldsymbol{u}_i) = \sum_{k=1}^{c} \left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})} \right)^{-(\frac{2}{m-1})} \tag{2.17}$$

$$A_j^{(l+1)}(\boldsymbol{u}_i) = \left[\sum_{k=1}^{c} \left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})} \right)^{-(\frac{2}{m-1})} \right] - \left[1 - \left[\sum_{k=1}^{c} \left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})} \right)^{\frac{2}{m-1}} \right]^2 \right] \tag{2.18}$$

In the Appendix section there are both the algorithms presented with all their technical characteristics: FCM Type 1 1 and FCM Type 2 2.

2.2.3 Competitive Algorithms

A particular kind of unsupervised learning is the competitive learning, i.e. it is a variant of the Hebbian learning approach[17] where the nodes compete themselves to give the appropriate response to input data. Competitive learning algorithms are principally used for partitioning, vector quantization and to self-organize Kohonen maps.

According to Fritzke, there are a number of different but often mutually exclusive outcome to get by competitive learning systems, i.e. error minimization, entropy maximization, feature mapping and other goals (see Fritzke's webpage). But, globally, there are also a set of principles that govern the competitive learning rule, i.e. i) the starting setting of the neural network, ii) the synapses weights range, iii) an formalism that allows to select the winner node given a subsamples of input, a.k.a. a procedure able to select the neurons that won the competition, also known as the winner-take-all neuron. Once the neuron won, that neuron will detect a specific pattern in relation to specific input features value [2] [3].

An neural network architecture, that implement a learning rule based on competition, contains, in its layers, a special layer known as the competitive layer [6]. A similarity measure is compute between the each competitive neurons and the input data, i.e. $||\boldsymbol{x_n} - \boldsymbol{w}_m||$, where $\boldsymbol{w}_m$ is the vector of weights for the m-th competitive neuron and $\boldsymbol{x_n}$ is the vector of input features. After the similarity computation, the winner neuron m-th responses with a positive signal, whereas the looser neurons with zero value signals.

The competitive learning rule may have also some difference according to the quality of the learning, i.e. the hard competitive learning and the soft competitive learning.

- The hard competitive learning is a type of learning that regards methods where each input signal determines the adaptation of the winner unit. The methodologies differ if they perform either batch[18] or on-line update[19].

[17] The Hebbian learning ruls claims that the synaptic efficacy arises from a postsynaptic cell that is repetitively stimulated by a presynaptic cell. This rule explains the synaptic plasticity, i.e., the adaptation of brain neurons during the learning process. It was introduced in 1949 by the Donald Hebb [102]. The main statement about the Hebbian's rule is *"Cells that fire together wire together"* [153]. The intuition that have had Donald Hebb on the brain learning processes has similarities with the cell mechanism knwon in neurobiology as spike-timing-dependent plasticity [40].

[18] The batch updating algorithms evaluate all the inputs before to update themselves, e.g. the generalized Lloyd algorithm [68, 142, 149]

[19] The on-line updating algorithms update themselves after each input signal, e.g., the k-means algorithm [158]

- The soft competitive learning is subdivided in in methods that have fixed network dimensionality and methods that have unfixed network dimensionality:

  - the soft competitive algorithms with fixed network dimensionality are models which have a k size of network choose in advance, i.e., given a network that maps from a n-dimension input space to the k-dimension structure. This peculiarity allows to make a low dimensionality representation of the data useful for visualization goals. An example of this competitive algorithms are the Self Organizing Maps (SOM) [123], the Growing Cell Structures (GCS) [80], the Growing Grid (GG) and other methods [12];
  - the soft competitive algorithms without fixed network dimensionality are models which have no predefined network topology. There are models that have no fixed topology at all, e.g., the Neural Gas (NG) [166], while in other cases the dimensionality of the network is related to the data local dimensionality and it changes within the input space, e.g., the Growing Neural Gas (GNG) [81].

In the next paragraphs will be presented three competitive learning algorithms used in this thesis[20]: the Self Organizing Maps (SOM), the Neural Gas (NG) and the Growing Neural Gas (GNG).

These algorithms shares some notation, i.e.,

- each network is a set of N units $A = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_N\}$ and each unit c has a an associated reference vector[21] $\boldsymbol{w}_c \in \Re^n$;
- between the units of the network there is a set $C \subset A \times A$ of neighbourhood connections[22] that are unweighed and symmetric;
- for a unit c, the N_c is the set of its direct topological neighbours;
- the n-dimensional input signals are assumed to be generated according to a continuous statistical distribution $p(\boldsymbol{\xi})$ with $\boldsymbol{\xi}$ in $\Re^n$ or from a finite training dataset;

[20] See the Results section 4.

[21] The reference vector indicates the position of unit c in the input space;

- the reference vector is also named receptive field center.

[22] These connections are not related to the weighted connections found, e.g., the ones in multi-layer perceptrons [217] The connections are used to extend the adaptation of the winner to some of its topological neighbours in competitive learning methods.

- for a given input signal $\boldsymbol{\xi}$ the winner $s(\boldsymbol{\xi})$ among the units in A is the unit with the nearest reference vector, according to specific similarity measure, i.e., $s(\boldsymbol{\xi}) = \arg\min_{c \in A} ||\boldsymbol{\xi} - \boldsymbol{w}_c||$, whereby the $||\cdot||$ denotes the Euclidean norm;

- there are fundamental concepts from computational geometry that underlie the competitive learning, i.e., the Voronoi tessellation and Delaunay Triangolation. In this part, the details about these two mathematical topics are skepped. To deepen them, see [264] and [263].

Self Organizing Maps Algorithm (SOM) The SOM model is based on the work by Willshaw and Von der Malsburg [266] in the Seventies and by Kohonen in the Eighties [123]. The interesting topic about this model is the constrained two-dimensional grid a_{ij} that does not change during the self-organization. The SOM model is similar to the Neural Gas (NG) since a decaying neighbourhood range and the adaptation strength are utilized.

The distance used on the 2D grid measures the adaptation strength of a unit $r = a_{km}$ when the unit $s = a_{ij}$ wins. The similarity measure adopted is the L_1-norm, i.e., the Manhattan distance $d_1(r, s) = |i - k| + |j - m|$ for $r = a_{km}$ and $s = a_{ij}$. In the Nineties, Ritter proposed to use 2.19 to define the relative strength of the adaptation for an arbitrary unit r in the network [210], i.e.,

$$h_{rs} = \exp(\frac{-d_1(r, s)^2}{2\sigma^2}), \tag{2.19}$$

where $\sigma(t)$ is the standard deviation of the Gaussan distribution varies according to $\sigma(t) = \sigma_i(\sigma_f(\sigma_i))^{t/t_{max}}$.

To see the details about the Self Organizing Maps competitive method, see the algorithm 3.

Neural Gas Algorithm (NG) The Neural Gas[23] algorithm was presented by Martinetz and Schulten in the Nineties [166] and it stems from the Self Organizing Maps (SOM). NG is an artificial neural network that finds the optimal representation of data based on features vectors. There are different version of this algorithms, e.g., the Growing Neural Gas (GNG) [81], the Incremental Growing Neural Gas (IGNG) [199] and the Growing When Required Network (GWRN) [164].

To see details about the Neural Gas competitive method, see the algorithm 4.

[23]The algorithm is named *gas* since it behaves as physical gas during the adaptation process, i.e., the dynamics of the feature vectors distribute themeselves like a gas within the the data space.

Growing Neural Gas Algorithm (GNG) The Growing Neural Gas is a models introduces by Fritzke in the Nineties [78, 81]. It is a combination of the Growing Cell Structures (GCS) [79] and the competitive Hebbian learning [166]. During the self-organization process, the number of units in changes. During the adaptation, the model starts with few units and then it inserts the new ones successively. A measure to evaluate the local error during the adaptation process is used to determine where to insert a new units in the feature space, i.e., the new unit is inserted near the one that has the most accumulated error.

To see details about the Growing Neural Gas competitive method, see the algorith 5.

2.2.4 Validation Measures

In this subsection are introduced a broad family of cluster validity indexes, taking account also similarity measure between couple of classes, i.e., the Jaccard similarity, and the relevance of the p-value as measure of the significance of the difference between classes found.

Cluster validity indices The performance of a clustering algorithm depends mostly on the definition of the number of clusters used, that it means to achieve, in a clustering process, the optimal classes cardinality [57]. If the ideal number of clusters is not obtained, there are two opposite situation, i.e.,

- the classes are less then the actual number, therefore one or more heterogeneous classes would be merged into other groups;
- the classes are more then the actual number, therefore on or more homogeneous classes would be separated into other groups.

In both the above situations, the clustering classification will be far from a real categorization, and then the information in the raw data is incorrectly used by the algorithm [260]. To solve this circumstance, the algorithm need a validity index design to detect the optimal cluster number for a given input dataset [100]. A Clustering Validity Index (CVI) has two indicators: the compactness and the separation [138], i.e.,

- the compactness, that indicates the concentration of points that share the same cluster, i.e., the similarity measure is often a vectorial distance between each point and the clustering centre of gravity; the smaller the distance, the better the cluster compactness [138];
- the separation, that evaluate the degree of isolation among clusters, i.e., the similarity measure is often a vectorial distance between cluster centroids; the larger the distance, the stronger the clusters isolation [272].

A dataset is well partitioned if there is both high compactness and high separation. But, often the two indicators conflict [234], e.g., if there compactness is high, the separation is low and *viceversa*. Therefore, a *rationale* between the two indicators is needed to design a clustering validation index.

Researchers from several disciplines have proposed a many indexes in order to put in relation the two indicators. In the next sub-paragraphs will be presented several validation models and also a similarity measures to compare classes.

Indexes Following the differentiation made by Li et al [138], the cluster validation indexes are divided in simple ones and advanced ones. Simple indexes are the following

- Partition Coefficient (PC) 25, i.e.,

$$PC(K) = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^2 \tag{2.20}$$

- Partition Entropy (PE) 22, i.e.,

$$PC(K) = -\frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^2 \log_2(u_{ij}) \tag{2.21}$$

- the Modified Partition Coefficient (MPC) 26, i.e.,

$$MPC(K) = 1 - \frac{K}{K-1}(1 - PC) \tag{2.22}$$

The advanced cluster validation indexes are the following:

- The Davies-Bouldin Index (DBI) [46] (see section 2.3.1)
- The Dunn Index (DI) [59, 60] computes the clusters separation with the minimum distance between clusters and computes the compactness with the maximum distance between each pair of within-cluster points, i.e.,

$$DI(K) = min_{1\leq p\leq K}\left(min_{s+1\leq 1\leq K-1}(\frac{(dis(C_p, C_q))}{max_{1\leq i\leq K}dia(C_i)})\right) \tag{2.23}$$

- The Calinski-Harabasz Index (CHI) [39] computes the separation with the distance B_K between centroid to the global centroid ($\overline{z}$) and it computes the compactness with the distance between each within-cluster point to its centroid, i.e.,

$$CHI(K) = \frac{B_K}{K-1}/\frac{W_K}{N-K} \tag{2.24}$$

- The Fukuyama and Sugeno Index (FSI) [233, 138]

$$FSI(K) = \sum_{i=1}^{K}\sum_{j=1}^{N}u_{ij}^m||x_j - z_i||^2 - \sum_{i=1}^{K}\sum_{j=1}^{N}u_{ij}^m||z_i - \overline{z}||^2 \tag{2.25}$$

- The Xie and Beni Index (XBI) [275]

$$XBI(K) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^2||x_j - z_i||^2}{N \cdot min_{i\neq k}\{||z_i - z_k||^2\}} \tag{2.26}$$

- The Kwon Index (KI) [131]

$$KI(K) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^2||x_j - z_i||^2 + \frac{1}{K}\sum_{i=1}^{K}||z_i - \overline{z}||^2}{min_{i\neq k}\{||z_i - z_k||^2\}} \tag{2.27}$$

- The Tang Index (TI) [236]

$$TI(K) = \frac{\sum_{min=1}^{K}\sum_{j=1}^{N}\mu_{ij}^2||x_j - z_i||^2 + \frac{1}{K(K-1)}\sum_{i=1}^{K}\sum_{k=1,k\neq i}^{K}||z_i - z_k||^2}{min_{i\neq k}||z_i - z_k||^2 + 1/K} \tag{2.28}$$

- The SC Index or ZLE Index[24] (the authors: Zahid, Limouri and Essaid [280])

$$SCI(K) = SC_1(K) - SC_2(K) \tag{2.29}$$

- The Compose Within and Between Scattering Index (CWBSI) [206], i.e,

$$CWBSI(K) = \alpha \cdot Scat(K) + Dis(K) \tag{2.30}$$

$$Scat(K) = \frac{1/K\sum_{i=1}^{K}||\sigma(z_i)||}{||\sigma(X)||} \tag{2.31}$$

$$Dis(K) = \frac{1}{\frac{D_{max}}{D_{min}}\sum_{i=1}^{K}\sum_{k=1}^{K}||z_i - z_k||} \tag{2.32}$$

- The WSJ Index (WSJI) [235], that is a derivation from CWBI, i.e.,

$$WSJI(K) = Scat(K) + \frac{Sep(K)}{Sep(K_{max})} \tag{2.33}$$

- The PBMF Index (PBMFI) [189]

$$PBMFI(K) = \frac{max_{i\neq k}\{||z_i - z_k||\} \times \sum_{j=1}^{N}\mu_{j1}||x_i - z_1||}{K\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^m||x_j - z_i||} \tag{2.34}$$

[24]For the formal definition of $SC_1(K)$ and $SC_2(K)$ see [138].

- The SVF Index (SVFI) [121]

$$SVFI(K) = \frac{\sum_{i=1}^{K} min||z_i - z_k||}{\sum_{i=1}^{K} max_{x_j \in C_i} \mu_{ij}^m ||x_j - z_i||} \tag{2.35}$$

- WL Index (WLI) [138]

$$WLI(K) = \frac{WL_n}{2 \cdot WL_d} \tag{2.36}$$

Clustering comparison Sometimes researchers need to compare the classes found within the same dataset. This practice is motivated to test if the repetition of a clustering gives the same results (i.e., a sort of stability control of results) and to test if different algorithms, given the same number of partitions, clustered in the same manner the different sample vectors.

To do this operation, there are many similarity group indexes, e.g., the Jaccard's index, that is an index introduced at the beginning of the XX century by Paul Jaccard [110]. He proposed a method to establish the similarity and the dissimilarity of sample sets, i.e., the size of the intersection on the size of the union. More formally:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{2.37}$$

where A and B are two sets and if they are empty the index is $J(A, B) = 1$. The complementary measure of the Jaccard's index is the Jaccard's distance[25], i.e.,

$$D_J(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}. \tag{2.38}$$

In this thesis, it is adopted the Jaccard distance to evaluate the similarity between clusters in different algorithms (see the Results section with healthy subjects 4.3).

P-Value, statistical tests and clustering validation The P-Value is a fundamental index in many scientific as well as technical or clinical disciplines. It has importance since with its size is possible to draw some considerations on the the results of a hypothesis testing. In the case of clustering analysis, the aim to measure a P-Value is to establish a precise statistical evaluation about the classes found, i.e., the clustering outcomes. In the next paragraphs, the definition of P-Value and a bit of hints about statistical tests that use P-Value will be exposed. Furthermore, a relation between clustering validity indexes and and P-Value will be argued.

[25] To deepen more mathematical details about sets and their distances, see [135]

P-Value The P-Value (or p-value) is a probability value associated to an experimental result to observe an equal or greater value of that result given the null hypotheses true [262]. For example, if an experiment is designed to measure the same variable in two times, with the expectation that the variable changes in time, it is reasonable to assume that the measures of the variable at time 1 and time 2 are different. Within a statistical hypotheses testing framework, the null hypotheses states that the measures are equal, whereas the alternative hypotheses states that the measure are not equal. In order to evaluate if the difference observed does not depend by random fluctuation, the probability to observe that difference given the null hypotheses true have to be very small. So, the p-value is this probability. Therefore, the smaller the p-value, the higher the significance of the experimental results.

More formally, if X is a random variable and if the null hypotheses is denoted with H_0, the p-value is given by $P(X \geq x|H_0)$ for a right tail test and $P(X \leq x|H_0)$ for a left tail test. The H_0 is rejected if the p-value is less the α, i.e., a level of significance, that in many cases is 0.05m 0.01, 0005 or 0.001.

The basic idea behind the p-value is the *reductio ad absurdum* reasoning, i.e., if it is false one argument, then it is true its complement. But, in few cases there are an exact duality, i.e., if it is false an argument A, the it is true its complement (all the ones that are not A) and in particular that the not-A is only B. In many cases, there are a lots of possible alternatives. So, it is important to set an experimental design able to handle multiple testing of hypotheses, in order to exclude in every p-value computation, the useless alternatives until is found the right one [187] .

Sometimes the usage of the p-value is not correct since there are misunderstanding on it. Often, it is confused with the probability of the hypothesis given the data, or the probability of that the hypothesis is true, and so on. This misunderstanding is known as the p-value fallacy [91, 43]

Statistical Tests The random variables used to compute the p-value with the reference of a null hypothesis is generally an outcome of a statistical test [273]. Each test has a statistic, i.e., a function that put in relation data, that follows a specific distribution. In many case is assumed to be a Normal distribution or similar. Tests are divided in parametric and non-parametric ones. The first have rigid assumptions, i.e., independence of observations, normality distribution of residuals, and the homogeneity of the variances. The second are free about this assumptions [181, 214].

In this theses are adopted both since the quality of data does not allow to take single position about the type of tests. They are parametric Analysis

of Variance test (ANOVA) [104] and the non-parametric Kruskall-Wallis test [127]. The details about both tests are skipped. To look formalism and implementation about the test, see these two links: ANOVA and Kruskall-Wallis.

The relation between P-Value and Clustering Validation The clustering validity indexes are used to evaluate an optimality of the clustering according to specific measure. Indicators that allow a measure to probe the clustering outcomes are the separation and the compactness. The p-value, instead, as exposed above, is a probability value to observe data, i.e., a statistical measure, given the null hypothesis true. If the p-value is very small, one can reject the null hypothesis and accept alternatives. In the case of clustering validity indexes the goal is to define optimal classes, whereas in the case of p-value (or in the usage of statistical tests) the goal is to define significative results.

The relation between p-value and clustering index is an ordinal relation. Once a dataset is clustered, in optimal sense according to specific index, these clusters have to need elements that are from different populations. Therefore, if a clustering algorithm found optimal classes, it is important to evaluate also if in these classes there are measures that are different without random fluctuations. For example, in the case of a dataset that is shape by two variables, i.e. two features, and the measures about these two variables regard at least three different quality of objects, e.g., different fruits, different stones, different stars, etc, then it is reasonable that the clustering of the samples in the features space results in - at least - three classes according to validity indexes. But also, if the three classes found are correctly individuated, they may contain samples that are statistical different according to specific test.

In the section of Results there is this comparison between validation indexes and p-value. In particular see 4.3.

2.2.5 Spectral Analysis

In this subsection are presented the terminologies related to spectral analysis, the principal component analysis (PCA), the non-linear fuzzy robust principal component analysis (NFRPCA) and the random matrix tools able to select the features obtain from the PCA and NFRPCA.

Terminologies

In the next paragraphes are defined the notions of random variable, the covariance measure and the correlation coefficients. They are statistical concepts that underpin the ones related to spectral analysis methodologies.

Random Variable The concept of random variables is peculiar in the meaning of the Principal Component Analysis, and in particular with the computations that regard the statistical measures such as covariance or correlation. Formally, a random variable X is a real-valued variable defined as a function between $X : \Omega \longrightarrow E = \Re$. On a measurable set $S \subseteq E$, the probability that X takes is

$$P(X \in S) = P(\omega \in \Omega | X(\mu) \in S). \tag{2.39}$$

Covariance Measure The covariance is a statistical measure of the joint variability of two random variables [214]. More precisely, for a population study, given two variables X and Y their covariance is defined as:

$$COV(X, Y) = E[(X - E[X])(Y - E[Y])], \tag{2.40}$$

where $E[X]$ is the expected value of X and $E[Y]$ the expected value of Y. For a vector $\boldsymbol{X} = [X_1, X_2, \ldots, X_n]^T$ with n jointly distributed random variables, its covariance matrix is defined as

$$\Sigma(\boldsymbol{X}) = COV(\boldsymbol{X}, \boldsymbol{X}). \tag{2.41}$$

IF $COV(X) = 0$ the random variables are uncorrelated. But, if two variables are uncorrelated, it does not imply that they are independent, because the covariance is a measure of the linear dependence between two variables, i.e., if two variables have non-linear dependence, their covariance is equal to zero. Instead, if $COV > 0$, then the random variables have a direct relation, whereas if $COV < 0$ the random variables have inverse relation. Therefore, the sign of the covariance refers the tendency of the linear relationship between random variables.

Correlation Coefficients The normalization of the covariance measure is the correlation coefficient, also know as the Pearson's correlation coefficient (see Pearson's works [192, 191]). The normalization is the *ratio* between the covariance of two random variables and the product between the standard deviation of each variable, i.e., the correlation coefficient indicated with $\rho(X, Y)$ is formally

$$\rho(X, Y) = \frac{COV(X, Y)}{\rho(X) \cdot \rho(Y)}, \tag{2.42}$$

where the $\rho(X)$ and $\rho(Y)$ are the standard deviation of the random variables X and Y, i.e.,

$$\rho(X) = \sqrt[2]{E[(X - \mu)^2]}, \tag{2.43}$$

with $\mu = E[X]$, that is the expected value of the random variables X. The same formulation regards the variable Y.

Principal Component Analysis (PCA)

The principal component analysis (PCA)[26] was ideate by Pearson in the first decade of XX century [193] and independently discovered by Hotelling in the Thirties [106]. It is an unsupervised linear transformation technique that is widely used in many fields, e.g., science, engineering, biomedical domains, with the specific aim to reduce the dimensionality of a system [205].There are also other applications[27] for PCA, e.g., exploratory data analysis and de-noising of signals.

Basically, PCA identifies patterns in data based on the correlation matrix of the features. The analysis of the principal component aims to find the direction of the maximum variance in high dimensional data and projects it onto a new subspace with equal or fewer dimensions than the original system. Therefore, the principal components are the orthogonal axis of the new subspace and they can be understood as the directions of the maximum variance given the constrain that the new feature axes ae orthogonal[28] to each other [205].

[26] To look more details about the mathematical foundation of PCA and its several applications see the book by Jollife [114].

[27] This secondary usage is the one adopted in this thesis to study the functional MRI images. To jump on the related results exposed in the Results section see 4.3).

[28] In this spectral section is not presented the Independent Component Analysis (ICA), that it an other methods to study the latent information of a system. But, it is important to note that the main difference between ICA and PCA is that the component of ICA are required to be independent rather then orthogonal as in the PCA. See Lindquist's paper fore more information [143].

Formally, given a d-dimensional features space, with a $\boldsymbol{X}$ dataset where $\boldsymbol{x} = [x_1, x_2, \ldots, x_d] \in \Re^d$ is a sample vector, $\boldsymbol{W}$ the $d \times k$-dimensional transformation matrix that allows to map a sample vector $\boldsymbol{x}$ onto a k-dimensional feature subspace, there are the following fundamental steps to run the PCA algorithm:

1. standardize the d-dimensional dataset, i.e., convert the raw-score in z-score according to mean μ and the standard deviation σ of the random variables considered;

2. compute the covariance matrix, i.e., the measure of a joint variability of random variables (see the above paragraph on the covariance and correlation computation);

3. decompose the covariance matrix into its eigenvectors and eigenvalues, i.e., applied a mathematical method to factorize a (real or complex) matrix (see the next paragraph);

4. select the k eigenvectors that correspond to the k largest eigenvalues, where k is the dimensionality of the feature subspace ($k \leq d$), i.e., according to some formal criteria, include (or exlcude) specific eigenvectors and their related eigenvalues (see the section on features selection with random matrix theory 2.2.5)

5. construct a projection matrix $\boldsymbol{W}$ from the top eigenvectors;

6. transform the d-dimensional input dataset $\boldsymbol{X}$ using the projection matrix $\boldsymbol{W}$ to have the new k-dimensional feature subspace

Single Value Decomposition The single value decomposition is a method to diagonalize matrices[29]. The concept of diagonalization is critical to understand the *core* structures of a physical (or biomedical) system. In the process of diagonalization, the correct coordinates, i.e., the basis functions, are revealed that reduce the given system to its low dimensional kernel. More formally, it is important to note that SVD is based on linear algebra concepts [32] and that there is an primary theorem that affirms that:

> Every matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ has a singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$, with $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ is unitary, $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ is unitary and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal. Furthermore, the singular values σ_j are

[29]The SVD is not the only method to diagonalize matrices. See the book from Kuts to look other methods [130].

uniquely determined, and, if $\boldsymbol{A}$ is square and the σ_j distinct, the singular vectors $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are uniquely determined up to complex signs (complex scalar factors of absolute value 1) [130].

More intuitively, the key idea behind the diagonalization procedure is that exist an *ideal* basis in which a square and symmetric matrix can be diagonalized so that, in this basis, all redundancies are removed and the largest variances of a measurements are ordered, i.e., the system, once it is diagonalized, is written in term of its principal components or in a proper orthogonal decomposition.

Contextualizing a bit this mathematical topic in functional neuroimaging, the single value decomposition for the principal component analysis permits to determine the spatial patterns that regard the greatest amount of variability in the time series of a neuroimages. According to Lindquist[30] [143], this can be done applying the single value decomposition of the data matrix, i.e.,

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T, \tag{2.44}$$

where $\boldsymbol{U}$ is a $T \times T$ unitary matrix, $\boldsymbol{S}$ is a $T \times N$ diagonal matrix with positive elements, and $\boldsymbol{V}$ is a $N \times N$ unitary matrix. The columns of $\boldsymbol{U}$ are the weighted sum of time series from different voxels, instead the columns of $\boldsymbol{V}$ are the voxel weights needed to build the components in $\boldsymbol{U}$. Therefore, $\boldsymbol{U}$ are the temporal components and $\boldsymbol{V}$ the spatial component of the data matrix. The variability explained by components are the elements of $\boldsymbol{S}$ and they are ordered from the higher to the lower.

SVD applied to principal component analysis is useful because i) it allows to potentially reveal the nature of the observed signal by finding its linearly independent sources, and ii) it decompose the signal ordering the components according to their weight, that is useful to simplify data or to filter out unwanted components. Nonetheless, some systems are not linear, or take linearity as assumption, and then the PCA could not process all the complexity of the data, as the functional neuroimages are (see the section 3 for extended information). It is noted that the brain is a dynamical system with several non-linearities [173]. Therefore, a method able to handle this non-linearities could process in a more realistic way the complexity of its functional structures. For this motivation, it has been selected a PCA method with the capacity to manage better the complexity of the brain signals. In the next part, it will be presented the Nonlinear Fuzzy Robust PCA Algorithm (NFRPCA).

[30] In this part is adopted the same Lindquist's notation.

Nonlinear Fuzzy Robust PCA Algorithm

Lukka [157] derived the Nonlinear Fuzzy Robust PCA (NFRPCA) algorithm from the linear one introduced by Yang and Wang[31] [277]. They transformed the crisp PCA objective function (eq. 2.45) in a fuzzy PCA objective function (eq. 2.46, i.e.,

$$E = \sum_{i=1}^{n} u_i e(x_i) + \eta \sum_{i=1}^{n} (i - u_i), \tag{2.45}$$

$$E = \sum_{i=1}^{n} u_i^{m_1} e(x_i) + \eta \sum_{i=1}^{n} (i - u_i)^{m_i}, \tag{2.46}$$

where in both 2.45 and 2.46 $X = x_1, x_2, \ldots, x_n$ is the dataset, $Uu_i | i = 1, \ldots, n$ is the membership set, η is the threshold and $u_i \in [0, 1]$. Whereas, only for 2.46) $m_1 \in [1, \infty]$. The u_i is the membership value of x_i belonging to the data cluster and $(i - u_i)$ is the membership value of x_i belonging to the noise cluster and m_1 is the fuzziness variable. Note that $e(x_i)$ measures the error between x_i and the class centre, and this idea is very similar as in the fuzzy c-means algorithm[32] [17]. The minimization of the crisp and fuzzy PCA optimization function is done with the gradient descent approach[33].

Features selection with Random Matrix Theory

The eigenvalue density of the empirical correlation matrix for uncorrelated independent identically distributed Gaussian variables follows the Marchenko-Pastur distribution [162] (or Wishart distribution [267]):

$$\rho(\lambda) = \frac{1}{2\pi r \lambda} \sqrt[2]{(\lambda_+ - \lambda) \cdot (\lambda - \lambda_-)}, \tag{2.47}$$

where the rectangularity parameter $r = N/T$ is the *ratio* between the number N of random variables and the sample size T in the data matrix, i.e., in the case of time-related random variables, N is the number of time-dependent variables and T are the time-points of each variables. According to 2.47, the eigenvalues of the normalized covariance matrix are located in a finite support $\lambda \in [\lambda_-, \lambda_+]$, in which the endpoints λ_- and λ_+ are $\lambda_\pm = (1 \pm \sqrt{r})^2$. Note that if $r \longrightarrow 0$, then T is much larger then N and the $\rho(\lambda)$

[31] See also the work by Xu and Yuilles [276] that inspired the work of Yang and Wang, i.e., Xu and Yuilles proposed an objective function with the consideration of outliers and with a PCA learning rules related to energy function.

[32] See the Fuzzy Clustering section 2.2.2.

[33] See the original work by Lukka for mathematical details [157]

became a *delta* function. When $r = N/T$ is finite, the eigenvalue density is smeared. But, when $N = T$, $r = 1$ is critical because the $\lambda_- = 0$.

The usage of the Marchenko-Pastur distribution of the eigenvalue of uncorrelated random variables is important since with this model is possible to select eigenvalues that above the range of $\lambda_\pm$, specially with the upper limit λ_+. In the section 4.3 there is an application of this RMT tools.

The relation between K-Means and Principal Component Analysis

Given $Y = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)$ as the standardize data matrix, the covariance of Y, ignoring the factor $1/n$, is $\Sigma_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T = YY^T$. The principal directions $\boldsymbol{u}_k$ and the principal components $\boldsymbol{v}_k$ are eigenvectors satisfying

$$YY^T\boldsymbol{u}_k = \lambda_k\boldsymbol{u}_k \tag{2.48}$$

$$Y^TY\boldsymbol{v}_k = \lambda_k\boldsymbol{v_k} \tag{2.49}$$

$$\boldsymbol{v}_k = Y^T\boldsymbol{u}_k/\lambda_k^{1/2} \tag{2.50}$$

Elements of $\boldsymbol{v}_k$ are the projected values of data points on the principal directions $\boldsymbol{u}_k$.

According to the work by Ding and He there is relation between K-Means clustering algorithm and the Principal Component Analysis (PCA). In their seminal paper [54], they proved that principal components are actually a continuous solution of the cluster membership indicators for the K-Means clustering method. It means that the PCA dimension reduction automatically performs data clustering according to the K-Means objective function $J_(K)$ with K clusters. Also, they linked the number of clusters K with number of principal directions found by PCA, i.e., the clustering structure is embedded in the $K - 1$ components. More formally, they proposed the following theorem [54])[34]

> When optimizing the K-Means objective function, the continuous solution for the transformed discrete cluster membership indicator vectors Q_{K-1} are the $K - 1$ principal components $Q_{K-1} = (\boldsymbol{v}_1, ..., \boldsymbol{v}_{K-1})$. The functional J_K satisfied the upper and lower bounds

[34]The proof of this theorem is the direct application of the well-known theorem by Ky Fan to optimization problem [63]

$$n\overline{\boldsymbol{y}^2} - \sum_{k=1}^{K-1} \lambda_K < J_K < n\overline{\boldsymbol{y}^2} \tag{2.51}$$

where $n\overline{\boldsymbol{y}^2}$ is the total variance and λ_k are the principal eigenvalues of the covariance matrix YY^T

Practically, if the correct number of class is three in a selected dataset (e.g., the Fisher's Iris dataset [65]), the most discriminative subspace is shaped by the first two eigenvector, because this PCA subspace is particularly effective for the K-Means clustering.

2.3 Contributions

The section contributions is divided in the two subsections:

- the part on the Soft Davies-Bouldin Index 2.3.1, i.e., a new separation measures based on fuzzy theory that is an extension of the so-called crisp and classic Davies-Bouldin index used for clustering validation;
- the part on the Crossed-Clustering framework 2.3.2, i.e., a new approach to study the spatio-temporal with clustering analysis taking in account both the spatial and the temporal components.

2.3.1 The Soft Davies-Bouldin Index

Conceptual foundations

Fuzzy principles A soft separation measure is useful to evaluate soft clustering analysis, that is based on the notions of fuzzy sets. More precisely, the definition of fuzzy sets is the following [278]:

> let X is a space of points and x a generic element of X, a fuzzy sets (or class) A in X is characterized by a membership (or characteristic) function $f_A(x)$ which associates each point in X a real number in the interval $[0, 1]$, where the value of $f_A(x)$ at x is the grade of membership of x in A.

Then, soft clustering allow to assign to each element of a dataset a degree of membership for each cluster found; as extension, there are these two propositions:

a. elements belong to all clusters,

b. elements have a membership value to each clusters.

After soft clustering analysis, the standard procedure is to harden the results to obtain a crisp partition of data, and then validate and conduct post clustering analysis. In literature, there are a various methods available to defuzzify results from soft clustering (see Leekwijck and Kerre for an extended overview [137]); generally, there are five methods that are frequently used: they allow to crisp the fuzzy results selecting, along the range of all membership values of an element, the smallest, the middle, the centroids, the bisector, the largest, and, finally, (the more used) the maximum.

This kind of hardening procedures could lose the *softness* that is peculiar to soft clustering outcomes, and in many cases, it is exactly the motivation that explain the use of fuzzy clustering instead the classic crisp clustering. Having in mind this noted soft characteristic, our intent is to evaluate fuzzy clustering results keeping all the grades of memberships that have each input feature. Therefore, in the part B of this section, we use the propositions *a* and *b* to define a soft separation measure as a reformulation of the classic Davies-Bouldin index (DB).

Separation measures As proposed by Davies and Bouldin, a general cluster separation measure should require little interaction by the user to set parameters, should be useful for hierarchical datasets, should be easy computable for big data, and should yield grounded results for high dimensional data.

With these premises, it is possible to introduce the concepts of metric measure as a distance function $d(X_i, X_j)$ and a dispersion measure as a deviation function $S(X_1, X_2, ..., X_m)$, where $X_1, X_2, ..., X_m \in E_p$, i.e. the p-dimensional Euclidean space. To define a cluster similarity measure, i.e. R, Davies and Bouldin proposed a function with the following properties:

- $R(S_i, S_j, M_{ij}) \geq 0$
- $R(S_i, S_j, M_{ij}) = R(S_j, S_i, M_{ij})$
- $R(S_i, S_j, M_{ij}) = 0$ iff $S_i = S_j$,
- if $S_i = S_j$ and $M_{ij} < M_{ik}$, then $R(S_i, S_j, M_{ij}) > R(S_i, S_k, M_{ik})$
- if $M_{ij} = M_{ik}$ and $S_j > S_i$, then $R(S_i, S_j, M_{ij}) > R(S_i, S_k, M_{ik})$, in which M_{ij} is the distance between centroids of clusters i and cluster j and S_i and S_j are their dispersion measures.

The heuristic meaning of the above properties is the following:

- the similarity function R is nonnegative;
- it is simmetrical;
- $R = 0$ only if their dispersion disappear;
- if the M_{ij} increases while Si and Sj remain costant, R decreases, and *viceversa*, if M_{ij} is constant while Si and Sj increase, R increases as well.

In the next paragraph, we first present in details the classic Davies-Bouldin index (DB) for crisp clustering and then its soft version for fuzzy clustering named Soft Davies-Bouldin index (SDB).

The classic Davies-Bouldin index (DB) the separation measure proposed by Davies and Bouldin for crisp clustering is defined as the system-wide average similarities of each clusters. Using the same notation of the authors, the index is formally:

$$\overline{R} = \frac{1}{C} \sum_{i=1}^{C} R_i \tag{2.52}$$

where C is the number of clusters and R_i is the $max_{j \neq i}\{R_{i,j}\}$. $\overline{R}$ is the system-wide average of the similarity measures of each cluster with its most

similar cluster; therefore, in relation to the algebraical properties that shape the DB index, the better the clusters, the lesser the average similarity. With $R_{i,j}$ is indicated the *ratio* between the average similarities of each clusters with its most similar cluster and the distance between the centroids of the clusters. More in details: given the average similarities S_i and Sj of the cluster i and the cluster j, and given $M_{i,j}$ the distance between the centroids of the cluster i and j, $R_{i,j}$ is defined as:

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \tag{2.53}$$

where, formally, S_i is :

$$S_i = \left\{ \frac{1}{T_i} \sum_{j=1}^{T_i} ||X_j - A_i||^q \right\}^{\frac{1}{q}} \tag{2.54}$$

in which T_i is the number of vectors in the clusters i, X_j are the vectors in i and A_i is the centroid of the cluster i. When $q = 1$, S_i is the average Euclidean distance of vectors to the centroids of i, whereas if $q = 2$, S_i is the standard deviation of metric about the samples in cluster in relation with its centroid. S_j is defined equivalently. Instead, the distance between centroids is formalized as:

$$M_{i,j} = \left\{ \sum_{k=1}^{N} ||a_{ki} - a_{kj}||^p \right\}^{\frac{1}{p}} \tag{2.55}$$

where a_{ki} is the kth component of the n-dimensional vector a_i, that is the centroid of cluster i. $M_{i,j}$ is the Minkowski metric of the centroids i and j; then, if $p = 1$ is the city block/Manhattan distance, whereas if $p = 2$ is the Euclidean distance between centroids; note that if $p = q = 2$, $R_{i,j}$ is the Fisher similarity measure computed between cluster i and cluster j.

The Soft Davies-Bouldin Index (SDB) to define the Soft Davies-Bouldin index (SDB), we have to take in account the two proposition a and b exposed in the part A of this section, that are:

a. elements belong to all clusters;

b. elements have a membership value to each clusters;

therefore, the average distance measures have to handle the average membership degrees. Technically speaking, the crisp *ratio* R_{ij} in the equation 2.53 became the fuzzy *ratio* R_{ij}^f, i.e.

$$R_{ij}^f = \frac{S_i^f U_i + S_j^f U_j}{M_{ij}^f} \tag{2.56}$$

where the crisp Si became the fuzzy S_i^f:

$$S_i^f = \left\{ \frac{1}{N} \sum_{j=1}^{N} ||X_j - A_i^f||^q \right\}^{\frac{1}{q}} \tag{2.57}$$

As specified for the equation 2.54, if $q = 1$, S_i^f is the average Euclidean distance of vectors to the centroids A_i^f of cluster i, and if $q = 2$, it is the standard deviation of cluster i. The difference in equation 2.57 respect the equation 2.54 is that the vectors X_j are always all the vectors in the dataset (see proposition a), whereas in the equation 2.54 they are only the ones in the cluster i. Therefore, the average is always computed on N, i.e. the all points in the dataset. The formulation for S_j^f is equivalent. In the equation 2.56, M_{ij}^f is the Minkoswki distance computed between the centroid A_i and A_j. Instead, the average membership values computed for the cluster i is U_i and for the cluster j is U_j. Formally,

$$U_i = \frac{1}{N} \sum_{j=1}^{N} \mu_{ij} \tag{2.58}$$

where μ_{ij} is the membership value of vector j to cluster i. The formulation for U_j is equivalent. Adapting the equations 2.53 and 2.54 with the equations 2.56 and 2.57, we could merge the properties about the clustering separation measure, mentioned by Davies and Bouldin, with the principle a and b about fuzzy clustering. Therefore, the equation 2.52 became the following:

$$\overline{R}^f = \frac{1}{C} \sum_{i=1}^{C} R_i^f \tag{2.59}$$

where C are the clusters and R_i^f is the $max_{j \neq i}\{R_{i,j}^f\}$. The $\overline{R}^f$ is the system-wide average of the soft similarity measures of each cluster with its most similar cluster; the better the fuzzy clustering, the lesser is the value of $\overline{R}^f$.

Theoretical framework Recently, Huapeng Li *et al* proposed a taxonomy of Cluster Validity Indexes (CVIs) [138]. They globally classified the measures using two indicators: *compactness*, i.e. the within clusters distances [260], and *separation*, i.e. the between clusters distances [274]. The

ideal case is when a dataset is partitioned with high compactness and high separation, but often they are mutually struggled [234]. Furthermore, they specially distinguished the subgroup of CVIs useful for fuzzy clustering in *simple* and *advanced* forms. The simple form considers only the membership degrees to a class of the data, whereas the advanced ones considered both the fuzzy degrees and the geometrical properties of the data to be clustered. In Table 2.1 are listed the measures classified by Huapeng Li *et al*, with the distinction between simple and advanced forms and the type of optimality value that have indexes (i.e the heigher (+) the better or the lower (-) the better). In relation the criteria used by the authors to make the taxonomy, it is reasonable to insert the Soft Davies-Bouldin index (SDB) in the class of CVIs that have advanced form.

Table 2.1: Cluster Validity Indexes (CVIs) taxonomy

CVIs Forms	*Measures*	*Optimality value*
simple	Partition Coefficient (PC)	the heigher (+)
simple	Partition Entropy (PE)	the lower (-)
simple	Modified PC (MPC)	the heigher (+)
advanced	Davies Bouldin index (DB)	the lower (-)
advanced	**Soft DB (SDB)**	**the lower (-)**
advanced	Dunn Index (DI)	the heigher (+)
advanced	Caliski-Harabasz Index (CHI)	the heigher (+)
advanced	Fukuyama-Sugeno Index (FSI)	the lower (-)
advanced	Xie-Beni Index (XBI)	the lower (-)
advanced	Kwon Index (KI)	the lower (-)
advanced	Tank Index (TI)	the lower (-)
advanced	SC Index (SCI)	the heigher (+)
advanced	Compose WBS Index (WBS)	the lower (-)
advanced	WSJ Index (WSJI)	the lower (-)
advanced	PBMF Index (PBMFI)	the heigher (+)
advanced	SVF Index (SVFI)	the heigher (+)
advanced	WL Index (WLI)	the lower (-)

2.3.2 The Crossed Clustering Framework

Clustering algorithms are a family of computational methods that allow to do the brain functional parcellation. The main goals of these methodologies is to explore the fMRI data in order to find similar structures. Once it is individuated a structure, the element that are within have geometrical properties in common, such as the metrical proximity, i.e., the sample vectors that are close to a prototyping vector are include in the same cluster, that has as centroid the coordinates of the prototype one. The measure of the vectors similarity is probed by a distance, that in many case is based on the Euclidean norm. Given a specified optimization criteria for the algorithms and an internal validation index for the clusters found, the final results is a sets of labels that are found in the features space able to shape the sample points of the dataset [35].

Nevertheless, the main objective of clustering of fMRI data is to find the functional patterns, that are principally thought as spatial patterns, i.e., ROIs that have had a temporal similarities (a very common signature); therefore, the parcellation of functional images returns a collection of brain modules that shared a similar BOLD trend. The drawback of this classic approach is the it does not find temporal patterns, but instead spatial patterns.

The crossed-clustering framework is a solution of this problem. It regards a double clustering approach, one able to find the spatial patterns and the other one able to find the temporal patterns. This is a procedure designed for spatio-temporal data, as the fMRI dataset. The empirical definition of the crossed-clustering approach is the following:

> The crossed-clustering allows to find the spatial patterns in the temporal features spaces and the temporal patterns in spatial features space.

A typical fMRI images is a data matrix $X \in \Re^{nxm}$ where $n = (1, ..., 96)$, i.e., 96 ROIs, and $m = (1, ..., 208)$, i.e., 208 time points (aka: brain volumes or experimental blocks). In a matrix form, $X = (a_{nm})$:

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ a_{21} & \dots & a_{2m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{n,m=96,208} \end{pmatrix}$$

[35] This is an easy way to define what is the meaning of clustering of fMRI data, but see the Section on Computational Intelligence on Clustering Analysis 2.2.1 to deepen the topics).

In this case, the classic clustering algorithms approach is to operate the classification of ROIs that have had similar trends, i.e., the aim is to find the spatial patterns in the temporal features spaces. Therefore, the more the BOLD values are similar for all the time steps, the more likelihood the ROIs are within the same label.

Practically, if there are ROIs that are functionally less active during the task, and if they shared stability of the signals (the signature trend), then they may be clustered in the same class, i.e., the class of the less active ROIs. And, if there are ROIs that are functionally more active during the task, and if they shared stability of the signals (the signature trend), then they may be clustered in the same class, i.e., the class of the more active ROIs.

The operation described above is the classic perspective to do clustering of fMRI images,i.e., the spatial clustering. In the case of crossed-clustering framework there is also the aim to do the temporal clustering. To do this operation is necessary to make the transposition of the above mentioned fMRI data matrix a_{nm}, i.e.

$$a_{nm}^{T} = a_{mn} \tag{2.60}$$

where $a_{mn} = X^T \in \Re^{mxn}$, and $n = (1, ..., 96)$, i.e., 96 ROIs, and $m = (1, ..., 208)$, i.e., 208 time points (aka: brain volumes or experimental blocks). In a matrix form, $X^T = (a_{mn})$:

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{m,n=208,96} \end{pmatrix}$$

After transposition, there is a change in the shape of the data to be clustered: the features space became the ROIs dimensions, instead of the time steps. Therefore, temporal patterns are the sample vectors that are within the spatial (ROIs) space that have had the same fMRI BOLD values. In other words, in the case of classic spatial clustering, a sample vectors are part of the same class if the contribution of all the temporal features were similar, whereas in the case of temporal clustering, a sample vectors are part of the same class if the the contribution of the spatial features were similar.

This knowledge about the functional neuroimages is uncoupled i.e., the two patterns found are from two independent clustering operations that give two indications on the content of the spatio-temporal images. Therefore, it is useful a computational method able to join these uncoupled results, e.g., a possible solution to solve this task is a functional that allows to link the two patterns information.

But, it is tricky to elaborate this argument with more details since there are some hypothesis that could take a role, e.g., the usage of statistical methods able to control factors. The factors are the the two patterns, where the levels of the factors are their labels, e.g., the spatial factor is shaped with the spatial patterns levels and the temporal factor is shaped with the temporal patterns levels. Statistical methods may test if their common response variable (e.g., the BOLD signal value) could individuate significative coupling, i.e., if there are spatial patterns that have specific temporal patterns and, *viceversa*, if there are temporal patterns that have specific spatial patterns. Potential methods that could address this topic is the parametric balanced ANOVA two-way or the Friedman test that it is its non-parametric version.

In the Thesis Results chapter 4 there are the outcomes related the usage of the crossed-clustering framework, specially in section with healthy subjects 4.3. This new clustering approach has been adopted to both resting-state fMRI data and to task-oriented fMRI data.

In the Conclusions chapter 5 there are the presentation of the future works 5.2. In particular, one of them is to solve the issue relative to the un-coupled results of crossed-clustering .

Chapter 3

Functional Neuroimaging

The functional Neuroimaging (FN) chapter regards two sections:

- the chapter related to the cornerstones of Functional Neuroimaging, i.e., two subsections that describe the general investigation of functional neuroimaging, keeping in account the uni-modality and the multi-modalities perspective 3.1
- the chapter related to the backgrounds of Functional Neuroimaging adopted for this PhD thesis, i.e., there are two subsections where are presented first an overview of the functional MRI and, second, the post-processing methods and techniques used to manage the complexity of the fMRI data 3.2.

3.1 Cornerstones

Neuroimaging regards investigations of the brain with adequate spatial resolution to build maps of anatomy (structural neuroimaging) and activity (functional neuroimaging). The snapshots captured by functional neuroimaging modality can be used to predict spatio-temporal brain dynamics. To generate a signal that exceeds the noise base, the functional neuroimages requires the coherent activity of many neurons.

The Neural Population Models (NPMs) are designed to find the mass activity of neural groupings at mesoscopic scales. Nevertheless, to compare NPM predictions to data from neuroimaging it is necessary to include signal expression, i.e., a measurement function, that maps the NPM-generated activity onto the sensor recordings. Both unimodal and multimodal neuroimaging framework used such forward modelling pipelines to predict and analyze data.

3.1.1 Unimodal Investigation

Functional neuroimaging encompasses a wide variety of methods to detect and quantify brain functional activity [9, 74, 105]. The most popular methods nowadays are electroenelography (EEG) and magnetoencephalography (MEG) and also functional magnetic resonance imaging (fMRI) . There are other methods, e.g.m positron-emission tomography (PET), voltage-sensitive dye imaging (VSDi) , and functional near-infrared spectroscopy (fNIRS).

It is possible to differentiate neuroimaging and methods that directly record the activity of individual neurons (e.g., patch clamping) from the ones that indirectly measure the brain functions (e.g., the fMRI). Nevertheless, this boundary is fluid, e.g., the question regarding measurements obtained with a Multi-Electrode Array (MEA), that it is not clear if they belong to the category of neuroimaging or electrophysiology. In principle, the imaging investigation - or the imaging aspect - is the ability, for a machine, to map out spatially distributed activities. To have an image, the pre-processing stage is required to constructed an map for visualization. Accoring to Friston, the inference from image data are from statistical maps of signal variance, rather then the signals themselves [72]. Even if the neuroimaging is principally used for spatial activity mapping, there are also imaging applications that do not take in account the space, but the temporal pattern, e.g., the EEG patterns from single EEG electrode.

Another difference between neuroimaging modalities regards if they are invasive and noninvasive methods. To study the human brain activity *in vivo* in a simple manner, the non-invasive methods are necessary. But, of

course, there are also the invasive ones, primarily use for surgical treatment of epilepsy, e.g., the electrocorticograms (ECoG). Furthermore, imaging methods could be divided in another subcategories related to invasive and not-invasive methodologies, i.e., the ones that investigate the electrophysiological activity of the brain and in the ones that use the metabolic and hemodynamic response to investigate the neural activity . The important things is that, even if methods are oriented to measure different features of the brain, they could work together in the same time, e.g., the EEG and fMRI can be recorded concurrently. This is a justification on why the field of multimodal neuroimaging is currently very active [70], since simultaneous recordings are more precise to find selective brain states associated to, for example, the same experimental task. Therefore, the statistical modelling and the data analysis methods need to be develop according to the integration of EEG and fMRI, that will became a basis for the multimodal images in functional neuroscience [182].

3.1.2 Multimodal Investigation

Statistics dominates the field of multimodal neuroimaging [70, 232].

The methods used are complex and sophisticated, but they share a common objective, i.e., the data multi-modal co-registration, that combine source of activities measures in the same brain location. Therefore, if the different modalities reveal high functionality in the same spatio-temporal domain, the statistics will result significative there. And, *viceversa*, if both are low, also the their combination will be low. Nevertheless, if only one of them is high, the conclusions still depend on the methodological details used and the assumption made by the researcher about the phenomenon.

The above statements are intuitive, since they relies of the assumption that different neural states generate signals from different modalities, having, as co-active, a covariation properties. In any case, it is not clear if this *a priori* is always true, given the multi-layered structures that has the signal generated by neuroimaging devices (e.g., the EEG is driven by the balance between inhibitory and excitatory synaptic activity, but the fMRI BOLD is driven - principally - by the excitatory synaptic activity alone). Therefore, the combination of EEG-fMRI gives both high excitation and high inhibition, that has a large fMRI BOLD response buy a small EEG signal; and, in the other sense, the large inhibitory input due by the absence of excitation influences the EEG, but not the fMRI BOLD. The physiological states that may obtain high level signals in both the modalities is the one with large excitation without corresponding inhibition.

A proper generative model could be provided by the Neural Population

Models (NPMs), that will be useful for simultaneous recordings of multi-modalities data because the common inferred neural state can be used in several forms, and it provides the starting stage for many signal expressions related to different imaging devices. Nowadays, the research are focused on merging EEG/MEG and fMRI predictions[1]. Furthermore, see the series of papers that have been established the ability to fit auditory task data [7, 8], or specially the fitting of auditory data with visual task data [208], or the Alpha rhythm blocking and correlation with BOLD [229].

Deneux and Faugeras investigated how much information EEG and fMRI BOLD can contribute to combine signals [50]. Bojak et al. [28] studied the dependence of the predictions on the assumptions about brain connectivity. Ritter proposed a first complete and freely available software package for carrying out such investigations [220].

[1]See for example the review by Deco [49] regarding the investigation on resting-state oscillations and the functional networks associated (e.g., the Default Mode Network [202, 219])

3.2 Backgrounds

The background section is related to functional neuroimaging explaining the basic knowledge to understand a brain image. The section is divided in two subsections: the general overview part 3.2.1 and the image processing part 3.2.2.

3.2.1 Functional MRI: an overview

This subsection concerns an outline about some functional MRI marked topics,i.e., what is the fMRI data and what is an fMRI experiment. A paragraph on generalities about fMRI introduces the one related to fMRI data and the one related to experimental design.

Generalities Functional neuroimaging is a fundamental methods to explore the brain *in vivo*. There are a lot of imaging modalities that allow clinicians and researchers to study the physiological changes during a brain activation. These techniques have specific peculiarities and they provide a special perspective of the functionality of the brain. Neuroimaging modalities are focused on the activity of group of neurons rather then single cells. Some functional methods are based on studying the electro-magnetic properties of the brain, e.g. ElectroEncephalography (EEG) and MagnetoEncephaloGraphy (MEG); they have high temporal resolution (ms) and low spatial resolution. Other methods are based on studying the brain hemodynamics, e.g. functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET); they have high spatial resolution and low temporal resolution.

Functional MRI is a noninvasive technique for investigate the activation of the brain. A series of brain images are acquired during an fMRI experiment session. The variations in the measured signals between images are used to infer the brain functionality related to specific task.

The analysis of the fMRI data regard multiple objectives: the localization of the regions involved in the task, the determination of the distributed networks linked to brain functions, the predictions about physiological and pathological brain states. In order to achieve these goals, data analysts have to choose the adapt statistical methods.

The computational analysis of a fMRI sequence is challenging. The fMRI data consists of a number of uniformly spaced volume elements, i.e. the voxels, that is a partition of the brain in equally size boxes and the intensity of each voxel represents for that area the spatial distribution of the nuclear spin density. The changes in the brain hemodynamics, as a reaction

to neural activity, impact on the MR signal. Therefore the changes in voxel intensity across time are an indirect measure of the cells activations and this measurement can be used to infer the spatial and the temporal information regard some brain functionality. Furthermore, the data have complex spatio-temporal noise with a relative weak signal. A complete fMRI model of data is considered infeasible and statistician have to simplify this problem maintaining validity and power of statistical analysis.

The fMRI Data In this subparagraph relating to fMRI it presented some information about the acquisition of the fMRI data, the BOLD fMRI signal and the noises associated and the models able to formalize them.

The acquisition of fMRI Data

- Basic physics. Physics and statistics regard the acquisition of fMRI data [97] [143]. Basically, to acquire an image, the subject is places into the field of an electromagnet, that typically has a magnetic field between 1.5-7.0 Tesla[2]. The magnetic field aligns the hydrogen (H) atoms in the brain. A radio frequency pulse tips over the aligned nuclei in a slice of the brain. After the pulsation, the nuclei return to their original positions and this variation induce a current in a receiver coil, i.e. the current provides the MR signal. Each measurement of the signal can be expressed as the Furier transformation of the spin density at a single point in the frequency domain, i.e. the k-space. More formally, the measurement of the MR signal at jth time point of a period is

$$S(t_j) \approx \int_x \int_y M(x,y) \cdot e^{(-2\pi\imath(k_x(t_j)x + k_y(t_j)y))} dxdy, \tag{3.1}$$

  where $M(x,y)$ is the spin density at the point (x,y) and the ((),()) is the point in the frequency domain (k-space) at which the Furier transformation is measured at time $t_j = j\triangle_t$.[3] This formulation regards the sampling of a single two-dimensional (2D) slice of the brain, but most studies need the acquisition of a three-dimensional (3D) brain volume.

- Statistical properties in the MR signal. The time-dependent information expressed by 3.1 is measured over two channels and the raw

[2] 1 Tesla = 10,000 Gauss; see that Earth magnetic field = 0.5 Gauss.

[3] $j\triangle_t$ is the time of the jth measurement and $\triangle_t$ depends on the sampling bandwidth of the scanner (e.g. 250-1000 μs).

k-space data are complex numbers. The assumption is that both real and imaginary part is measured with independent normally distributed error. In the final stage of the Furier transformation, these complex values are divided into magnitude and phase components. Usually, only the magnitude portion of the signal adopted in the data analysis, while the phase portion is discarded (even if there are models that use both [216]). It is notable that the magnitude value follow the Rice distribution [96]. There are two particular case: when the siganl-to-noise *ratio* (SNR) is low (e.g. voxels outside the brain), its shape is like the Rayleigh distribution, but when the SNR is high (e.g. voxels inside the brain), its shape is close to Gauss distribution. To know the properties related these distributions, statisticians have to develop methods to estimate noise models and techniques to remove artefacts [143]. The changes in brain hemodynamics in consequence to neuronal activity impact the local intensity of the MR signal. Basically, an fMRI experiment is a series of ordered brain volumes. The temporal resolution of the scansions depends on the acquisition time between volumes. Generally, volumes with dimension $64 \times 64 \times 30$ (i.e. 131,072 voxels of 1 mm^3) are united with $T \approx 100 - 2000$ [s] separate time points during an experiment, and if we consider therefore the resulting data is around 100,000 time series of length T, often repeated for $M \approx 10-40$ subjects [143].

BOLD fMRI signal In this subparagraph will be presented the i) biophysical relation that exists with blood and oxygen, ii) the negative neural dip, as important physiological status during fMRI brain activation, iii) the limitation of the functional MRI scanner on spatial resolution and temporal resolution.

- The relation between blood and oxygen. To study local changes in deoxyhemoglobin concentration in the brain, fMRI is most performed using blood oxygenation level dependent (BOLD) contrast [188]. There are advantages for BOLD imaging with the difference between oxygenated and deoxygenated hemoglobin. Different local magnetic fields are produced with hemoglobin states thanks the diamagnetic and paramagnetic properties. In the case of deoxy-hemoglobin, the MR signal is suppressed, while in the case of oxy-hemoglobin is not. Areas of the brain, that are active during the execution of mental task, are refreshed by the cerebral blood flow that changes the local magnetic susceptibility and, consequently, the MR signal in the active brain regions. The changes in the blood oxygenation are studied with a series of properly

acquired MR images with which it is possible to infer brain activity. The hemodynamic response function (HRF) is referred to the evoked brain hemodynamic response to neural event. The increasing of the inflow of oxygenated blood to active brain regions is led to the increased metabolic demand due to neural activity. The decreasing in the concentration of deoxy-hemoglobin is led by increasing of the oxygen consumed; this variation increase the MR signal. This positive rise in signal has an onset approximately 2 seconds after the onset of neural activity and it continues 5–8 seconds after that neural activity has peaked [2]. The BOLD signal decreases to a below baseline level after reaching its peak level, which is sustained for 10-12 seconds. This effect is known as the post-stimulus undershoot, that is due to the rapidly decreasing of the blood flow than the blood volume, facilitating a greater concentration of deoxy-hemoglobin in the active brain regions.

- Negative neural dip. Many researches have shown evidence about the negative neural dip, that is a decrease in oxygenation levels immediately following neural activity, giving rise to a decrease in the BOLD signal in the first 1–2 seconds after the activation. This effect is called initial negative BOLD response [175, 161]. Interestingly, there are evidences that the negative neural dip is more localized within areas of neural activity than the subsequent rise which appears less spatially specific. For these motivation, the negative response has not been reliably ascertained and its existence remains debatable [150].

- Limitations in spatial resolution. MRI is an imaging technique that provides detailed anatomical scans of gray and white matter with high spatial resolution (i.e. below 1 mm^3), but the time to have detailed scans is high without real application for functional investigations. Rather, the spatial resolution is usually about of $3X3X5mm^3$, that is an image dimensions about $64x64x30$ sampled close to 2 seconds. fMRI has relatively high spatial resolution compared with other functional imaging modalities. Nevertheless, it is remarkable to notice that the high spatial resolution is limited by two main factors [143]:

  - the resolution of the data decreases after the spatially smoothing of the fMRI data before to do the analysis (see paragraph;
  - the inference on population needs the processing of groups of subjects that have different brain sizes and shapes. To compare data across subjects it is necessary a normalization procedure to warp the brains onto a standard template brain, that introduces spatial imprecision and blurring in the group data.

Standard and advances tools in data acquisition and preprocessing can help to improve the inferences in space (see the section 3.2.2). For more details and possible solutions to address the space limitations in fMRI see [150].

- Limitation in temporal resolution The repetition time (TR) is the temporal resolution of an fMRI study. It depends on the time between acquisition of each individual image. Usually in fMRI studies the TR ranges from 0.5–4.0 [s]. The neural activity takes place on the order of 10 [ms], therefore there is a disconnection between the cell activity and the fMRI temporal resolution. In any case, the computational analysis of fMRI data is principally focused on using the positive BOLD response to study the inherent cell activity. Therefore, the constrain factor in deciding the suitable temporal resolution is broadly not considered the speed of image acquisition, but rather then the speed of the evoked hemodynamic response to a neural event. The inference is based on oxygenation patterns taking place 5–8 seconds subsequently activation. So, TR values in the range of 2 [s] are mostly believe enough. Since the time-to-peak positive BOLD response appears in a larger time scale than the speed of brain processes, there is a possible risk of unknown confounding factors able to influence the order of time-to-peak relative to the order of brain activation in different ROIs. Therefore, it is not easy to evaluate the absolute timing of brain activation using functional MRI. Nevertheless, there are researches that have shown that the relative timing within a voxel in response to several stimuli could be accurately identified [174, 178]. There are as well hints that centering inference on features related to the initial dip can solve problems related to possible confounders [146]. Still, according to Lindquist [143], these kinds of researches need to increase the temporal resolution and the capacity to rapidly acquire data becomes more and more fundamental. Lastly, another solution of improving inferences in time is with an appropriate experimental design, e.g. estimating the hemodynamic response function at a higher temporal resolution than the TR meanwhile the start of repeated stimuli are jittered in time [45].

BOLD, noise and nuisance components The information within the functional brain images have a layered structures of different signals with added noise and nuisance (i.e. physiological but useless signals). They have to be manage as sources with specific computational approaches, taking account their inherited features or taking some assumptions about them. In this subparagraph there will be expose the formalisms able to model BOLD

signal as well as noise and nuisance.

- BOLD signal. The evoked BOLD response in fMRI is a nonlinear function of the results of cellular and vascular variations [258]. Its complexity complicates the possibility to fitly model its activity. The dynamic of the response is related to both the applied stimulus and the hemodynamic response to neuronal events. There are a lot of methods for modelling the BOLD signal and the inherent hemodynamic response function in the literature. The main difference between them is on the type of relationship between the stimulus and BOLD response. Lindquist differentiates between nonlinear physiological-based models, e.g. the Balloon model [36, 76, 209], which describes how the dynamics of cerebral blood volume and deoxygenation influences the BOLD signal, and models that have assumptions, e.g. linear time invariant models, that are based on the hemodynamic response function (HRF) as an impulsive response function to a task manipulation.

  - The nonlinear models are detailed mathematical and physical equations, but linear models have more robustness and capacity to represent noisy system. The assumption about the linear relationship between neuronal activity and BOLD response is common, where linearity implies that the magnitude and shape of the evoked hemodynamic response function is independent on any stimuli. There are studies that have shown that under particular conditions the BOLD response can be modelled as linear with respect to the stimulus [29], in particular if experimental events are at least 5 [s] spaced [178]. But there are also researches that have found that nonlinear effects can be quite large with stimuli spaced less than 2 [s]) [22, 258]. Therefore, the capacity to assume linearity is crucial, because it allows to model the relationship between stimuli and the BOLD response using a linear time invariant system, in which the stimulus is the input and the hemodynamic response function is the impulse response function.

    An example of nonlinear model is the Balloon model: it consists a set of ordinary differential equations (ODE) that model variations in blood volume, blood inflow, deoxyhemoglobin and flow inducing signal and how these variations impact the observed BOLD response. There are a drawbacks of these models, even if they are more biophysically plausible than their linear alternatives, e.g., a) the estimation of a lot of parameters, b) they only sometimes provide reliable estimates with noisy data, c) they provide an indirect

framework for performing inference, and d) globally, they are considered infeasible to perform whole-brain multi-subject analysis of fMRI data in cognitive neuroscience.

– In a linear system framework the signal at time t , $x(t)$, is modelled as the convolution of a stimulus function $v(t)$ and the hemodynamic response $h(t)$, that is,

$$x(t) = (v * h)(t), \tag{3.2}$$

where $h(t)$ is either assumed to have a canonical form, or alternatively modelled with a set of linear basis functions. Also, the timing and shape of the HRF are known to vary across the brain, within an individual and across individuals [2, 222]. Some of this variability depends on the inherent conformation of the vascular bed, which may explain why there are differences in the hemodynamic response function between brain regions with the same task for physiological reasons [249]. Other variability depends on the differences in the pattern of evoked neural activity in regions related to different functions but linked to the same task. Globally, the major deficit during the fMRI data analysis is that researchers usually assume a canonical hemodinamic response function [95], that could mismodel the signal in large amounts of the brain [151]. For this motivations, there has been a orientation toward a models that are both detailed but also able to manage noise and nuisance.

- Noise and nuisance signal. The measured fMRI signal is vitiated by random noise and various nuisance components that depend both to hardware/physical reasons and the subject/physiological motivations. They are thermal noise, scanner drift and physiological noise.

  – Thermal motion of electrons within the subject and the scanner causes the fluctuations in the MR signal intensity, that is a random noise independent of the experimental task. The total thermal noise is a linear function of the scanner field, i.e. the higher field strengths the more noise. Nonetheless, the noise have not spatial structure and its minimization is possible averaging the signal over multiple data points.

  – The scanner drift is another source of variability in the signal due to scanner instabilities. It is a slow changes in voxel intensity over time (low-frequency noise). The drift varies across space and the models of the fMRI data have to include it as source of variation.

– The physiological noise is also a set of BOLD alterations, e.g. patient motion, respiration and heartbeat, that cause fluctuations in signal across both space and time. The noise related to human physiology can often be modeled and its effects removed. But, heart-rate and respiration have periodic fluctuations that are difficult to manage. Note that according to the Nyquist criteria, it is necessary to sample at least twice as high as the frequency of the periodic function one is going to model. Therefore, if the TR is too low, as in many fMRI studies, there is the possibility to have problems with aliasing. In the case of alias noise, periodic fluctuations are distributed throughout the time course and they give rise to temporal autocorrelation. Noise in fMRI is usually modelled using AR(p) or ARMA(1, 1) process [200], where the autocorrelation is assumed to be motivated by unmodelled nuisance signal. The resulting error term corresponds to *white noise* [154], if these terms are right removed. Heart-rate and respiration can be estimated and included in the model for high temporal resolution studies, or isolated with a band-pass filter.

Modelling the fMRI Data There are a lot of shared objectives in the analysis of fMRI data, e.g. a) localizing regions of the brain related to a specific task, b) determining distributed networks that correspond to brain function, and c) predicting about psychological or pathological status. The basic concept for fMRI data analysis is that certain stimuli leads to changes in neuronal activity. The statistical analysis of fMRI data regards working with massive data sets (or big data sets[4]) that have a complicated spatial and temporal noise structure. Therefore, simplifications are required to balance computational feasibility with model efficiency.

- Single voxel model If the fMRI data consists of a brain volume with N voxels that is repeatedly measured at T different time points, and the experiment is repeated for M subjects, including in the model the BOLD response, various nuisance signal and noise, Lindquist proposed a model for fMRI activation in a single voxel for a single subject [143], i.e.,

[4]The meaning of Big Data here is refereed to the formal definition. The term is used to indicate the study and applications of data sets that are so complex; big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. There are five concepts associated to big data: i) volume, ii) variety, iii) velocity, iv) veracity (i.e., how much noise is the data) and v) value [228]

$$y_{ij}(t) = \sum_{g=1}^{G} z_{ijg}(t)\gamma_{ijg} + \sum_{k=1}^{K} x_{ijk}(t)\beta_{ijk} + \epsilon ij(t), \tag{3.3}$$

for $i = 1, \ldots, N$, $j = 1, \ldots, M$ and $t = 1, \ldots, T$. Here $z_{ijg}(t)$ represents the contribution of nuisance covariates at time t , including terms modelling the scanner drift, periodic fluctuations due to heart rate and respiration, and head motion. Similarly, $x_{ij}k(t)$ represents the task-related BOLD response (the signal of interest) corresponding to the k-th condition at time t. The terms β_{ij} k and γ_{ijg} represent the unknown amplitude of x_{ijk} and z_{ijg}, respectively, and $\epsilon_{ij}(t)$ the noise process. For more standard fMRI experiment, Linquist summarizes 3.3 as

$$y_{ij}(t) = \sum_{g=1}^{p} \gamma_{ijg} t^{g-1} + \sum_{k=1}^{K} \beta_{ijk} \int h_{ij}(u) v_k(t-u) du + \epsilon_{ij}(t), \tag{3.4}$$

where ϵ_{ij} is assumed to follow an AR(2) process. In matrix form the model is

$$\boldsymbol{y_{ij}} = \boldsymbol{Z_{ij}\gamma_{ij}} + \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{ij}, \tag{3.5}$$

where $\gamma_{ij} = (\gamma_{ij1}, \ldots, \gamma_{ijp})^T$, $\beta_{ij} = (\beta_{ij}, \ldots, \beta_{ijK})^T$, Z_{ij} is a $T \times p$ matrix with columns corresponding to the polynomial functions, and X_{ij} is a $T \times K$ matrix with columns corresponding to the predicted BOLD response for each condition. Furthermore, the model in 3.5 can be combined across voxels as follows

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\boldsymbol{B}_j + \boldsymbol{Z}_j\boldsymbol{G}_j + \boldsymbol{E}_j \tag{3.6}$$

The $\boldsymbol{Y}_j$ is a $T \times N$ matrix[5], where each column is a time series corresponding to a single brain voxel and each row is the set of voxels. The matrices $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ are the common design matrices used for each voxel. At last, $\boldsymbol{B}_j = (\beta_{1j}, \ldots, \beta_{Nj})$, $\boldsymbol{G}_j = (\gamma_{1j}, \ldots, \gamma_{N1j})$ and $\boldsymbol{Ej} = (\epsilon_{1j}, \ldots, \epsilon_{Nj})$. The vectorized variance of E is typically assumed to be separable in time and space and the spatial covariance is ignored

[5]Alternatively, the matrix $\boldsymbol{Y_j}$ is sometimes analyzed using multivariate methods. See the section on Functional MRI processing 3.2.2 for more details.

because is often assumed to be negligible compared to the temporal covariance. Lindquist notes that while 3.6 gives a framework for a full spatio-temporal model of brain activity, it is currently an infeasible alternative due to the hard computational cost to fit the model. Whereas, the model in 3.5 is applied to each voxel separately and spatial issues are incorporated only a later stage.

Experimental Design The design of an fMRI experiment regard the standard issues of psychological investigation and also topics about the data acquisition and the presentation of stimulus. The spacing and ordering of events is critical. The optimal experiment depends on the cognitive nature of the task, the capacity of the fMRI signal to capture the changes task related over time and the comparisons that research is interested in doing. Furthermore, the efficiency of the statistical analysis is directly related to the experimental design, therefore, it is important that the computational methods be carefully considered during the design procedure. The maximization of both statistical power and psychological validity is what a good experiment have to achieve. the ability to estimate the HRF and the ability to detect significant activation are performances that characterized the experimental solidity. To control issue related to anticipation, habituation and boredom, researcher usually adopt the randomness of the stimulus presentation, in order also to increase the psychological validity. An optimal experiment has a balance between the efficiency, detection power and randomness that at last regards the goals and the conditions that one is interested in investigate.

Globally, the experimental paradigm are divided in two main kind of procedures: the active paradigm and the passive paradigm. The first is used when the researcher want to establish the anatomo-physiological relation that occur during a specific task, whereas the second is adopted when the researcher want to study the baseline signal of the brain, without doing a specific task. The passive paradigm, also known as resting-state paradigm, is useful for the clinical investigation using fMRI in patient that have no the ability to relate with the experimentalist, e.g., pathologies that don't allow the subjects to understand the task to do as, for example, degenerative diseases or consciousness diseases.

In the next subparagraph will be exposed more in details the difference between active paradigm and passive paradigm. See the Result section 4 to look results with active paradigm and passive paradigm. In particular, for the active paradigm with clinical subjects see 4.2 and the active paradigm with healthy subjects see 4.3, whereas for the passive paradigm, only with healthy subjects, see 4.3 and 4.3.

Active Paradigm The active paradigm is a task oriented experimental study. The goal of this paradigm is to investigate the spatio-temporal relation between the execution of a specific task, e.g., perception, motion, reasoning, language, memory, etc, and the associated brain functional activations, i.e. the neural correlated of the experimental task. There are two kinds of active paradigm: the block design and event-related design. More in details:

- in the block design the experimental conditions are divided into extended time intervals (i.e., the blocks.), e.g., the classical finger tapping design has a finger block and a rest (control) block. The task–rest comparison can than be used to compare differences in signal between the conditions.Generally, the more is the length of each block, the larger is the evoked response during the task. This procedure brings to higher detection power, as it increases the window in signal between experimental blocks. In contrast, it is also crucial to include multiple transitions between blocks, as otherwise differences in signal due to low-frequency drift may be confused for variabilities in task conditions. Furthermore, the assumption the same mental processes are evoked throughout the same block may be violated if block lengths are quite long. The advantages in block design are that they offer high statistical power to detect activation and are robust to randomness in the shape of the hemodynamic response function. The disadvantages are that the predicted response depends on the total activation related to a series of stimuli, which makes the paradigm less sensitive to variabilities for separate stimulus.

- In the event-related design, the stimulus is a set of short discrete events (e.g., light flashes) that have randomized timing. These types of designs are flexible and they allow to estimate the HRF (e.g., onset and width) that useful to make inference about the relative timing of activation during conditions and about continuous activity. The advantages are that event-related design permits to evaluate the effects of many conditions if one condition either intermixes different events or varies the inter-stimulus interval between trials. Also, the event-related design avoids effects of fatigue, boredom and systematic brain patterns unrelated to the task during long inter-trial. One possible disadvantages is the power to detect activation that is usually lower than for block designs, even if the capacity to have images of more trials per unit time can address this drawback.

Passive Paradigm The passive paradigm is the experimental procedure that regard a brain functional baseline studies, i.e., an experiment in

which the aim is to evaluate what (the regions) and how (the connections) underline the rest physiology [248, 25]. Often the passive functional experiment is called rs-fMRI (resting state functional MRI) paradigm or task-negative[6] paradigm [23].

The resting state functional activation has a specific functional dynamic that is useful to take in account in the case of brain pathologies. In particular, the networks associate to passive paradigm in healthy subjects are - globally - known [25]. The special usage of the passive paradigm is the investigation of deviation from a normal baseline connectivity that allows to help clinicians to diagnose brain disease or to control their stages both in Psychiatry and in Neurology [248].

In literature are present different standard independent models about the resting state physiology [248], e.g., the Default Mode Network (DMN), the sensorial and motor cortexes, the visual regions, the lateralized fronto-parietal territories, auditory cores, etc. The most stable network is the DMN [93, 247, 167, 203, 33]: it regards a selected synchronized networks that have, as a special hubs, the bilateral Precuneus and the bilateral Posterior Cingulate cortexes, that are connected with medial prefrontal cortex and angular gyrus [5].

In this paragraph, there are no other details about the passive paradigm because it will be explained more within the researches into the Results section 4.3.

[6] A task is called negative when the task acted is to do nothing, neither perceptive nor motor or cognitive actions. There is only a thin difference to take eyes open or closed during the experiment and usually if they are open the subjects look a cross on the screen.

3.2.2 Functional MRI: the processing

In the past subsection, there were introduced the basic notions about the functional MRI, exploring the fMRI data in general and describing the formalism and the modelling associated the BOLD signal, the noise and nuisance related. Furthermore, there were presented the paradigms that allow to design an experimental procedure with fMRI in order to set task or rest oriented studies. From an high point of observation, these notions are useful to have a first access to fMRI background, but there is another important set of notions that permits to manage the raw results obtained from a functional imaging experiment, i.e. image pre-processing and image post-processing. The first subset of notions regard the methodologies to transform images from raw images to cleaned and standardized ones, whereas the second set of notions regard the computational frameworks able to localize the brain activity and to draw the functional connectivities. So, in the next paragraphs will be presented, in order, the preprocessing and postprocessing techniques.

Preprocessing Before to do statistical analysis, there are a series of preprocessing steps that regard the fMRI data, in order to remove artefacts and check the model assumptions. One assumption is to acquired voxels simultaneously. Another is to assume that each time series related to some voxel is a specific signal from that voxel (i.e. the subject was stayed during the experiment). Space localization for each voxel is also an assumption, that consist to have the same position for each brain voxel in all the subjects. Only with preprocessing the data before the statistical analysis is possible to check this assumptions, otherwise the results will be invalid. There are several steps to do for this purpose: in the following paragraph will be presented the slice time correction, the motion correction, the spatial smoothing and the geometrical registration with normalization.

Slice Time Correction Usually it is assumed that the brain is measured simultaneously. But in many case this is not possible, as the brain slices are temporally shifted. The slice time correction allow to shift each voxel's time series as they were measured at the same time. This operation is done with interpolation techniques based on the Fourier shift theorem to correct the variabilities in the acquisition times [143].

Motion Correction One of the main issue in the fMRI study is to handle the subject movement within the scanner. Also a small head movement can turn an error in the images. If a subject moves his/her head, the signal from one voxel will be impured from the neighbouring voxels, causing

problem with the final signals. So, it is important to estimate the quantity of head motion and then use this information to solve the motion problem in the images. In any case, if subject moved too much, is better to remove from the data. Usually the first step is to correct motion with alignment procedure between input image and some target image (e.g. the first image or the average image). Using a rigid body transformation with 6 parameters, the input image will be translate and rotate to match the target image. image. The match will be perform with an optimization procedure as a cost function minimization that works as similarity measure between input and target images. When the optimal parameters will be estimated, and the realignment has done, the image is re-sampled with interpolation techniques in order to obtain a new motion corrected voxel values. For each subject's brain volume, this operation has to be done.

Spatial Smoothing The image smoothing usually is a convolution between the functional images and a Gaussian kernel, described as full width of the kernel at half of its maximum height (FWHM). The values that regard the kernel range 4-12 mm FWHM. The reasons about this variation are i) the blurring of any residual anatomical differences improve the inter-subject registration and help the limitations in the spatial normalization, ii) it allows to ensure that the assumption of random field theory (RMT) are valid (e.g. to have the RMT assumption, the FWHM is 3 times the voxel size); iii) smoothing reduces random noise in single voxels and increase the regional signal-to-noise ratio. To smooth an image is equivalent to filter with low-pass cut-off to the sampled k-space data prior do reconstruction. Note that smoothing an image acquiring with high spatial resolution does not lead to the same results with low resolution image. To have an high signal-to-noise ratio is better to acquire many voxels, as the ratio increases as the square of the voxel volume. Therefore, the optimal spatial smoothing is done with high sensitivity with the right resolution. According to Lindquist, there are acquisition schemes designed to acquire images at the functional resolution desired [147].

Geometrical Registration and Normalization Functioanl MRI usually has low spatial resolution and has few anatomical details then structural MRI. For this motivation, the low resolution image is mapped onto a high resolution structural image. This process is called coregistration and it is often performed with a rigid body trasformation (that has 6 parameters) or with affine body registration (that has 12 parameters). It is very important that each voxel has the same position in the same brain structure for each subject

in a group analysis. Nevertheless, it is true that each brain has differences, but healthy brain also shares common regularities between subjects. The normalization procedure, instead, allows to register the individual anatomy to a standardized stereotaxic space within a template brain, e.g., the Talairach or Montreal Neurological Institute (MNI). With the brain complexity the rigid body transformation is useless, whereas, the common procedure regards to use nonlinear transformations to match local features. The principal benefits of normalizing data is the capability to compare across studies and subjects the results, and so they can be generalized in a consistent manner. There are also drawbacks, for example the normalization procedure reduces spatial resolution and introduces errors consequently the interpolation.

Postprocessing The postprocessing of fMRI data is the family of procedures able to analyze the cleaned and normalized images in order to infer considerations about the brain functional localization and to depict the brain connectivities. In the next subparagraph, there will be present how to localize the brain functionality linked to specific experimental paradigms and how to explore the connections between functional-related brain regions.

The brain functional localization The model presented by equation 3.5 used for the fMRI data analysis have assumptions. In many controlled experiments, there is the assumption that the stimulus function $v_{k(t)}$ is known and that have the same mathematical shape of the experimental paradigm (e.g., a vector of zeros and ones that indicates the time points when the stimulus is ON/TASK and 0 when it is OFF/REST). The assumption that the hemodynamic response function HRF is known *a priori* allows to equation 3.5 to became a multiple regression model with known signal components and unknown intensities. According to Lindquist [143], the assumption made in the popular general linear model approach (GLM) [270, 73, 74]. But, in experimental psychology it is difficult to specify stimulus function *a priori*. Therefore, the two main approach in fMRI data analysis regard the analysis with certain time of activation and analysis without certain time of activation.

- Data with certain time of activation In the area of fMRI data analsys the main approach is the so-called general linear model (GLM). The formalization attempts to model the voxel's time series as linear combination of many components and then it tests if the activity in a specified region is related to specified input function. In the easy model of GLM, there is the assumption that both the stimulus function and the hemodynamic response function (HRF) are known. The stimulus

has the same structure of the experimental paradigm, whereas the HRF is modeled with a canonical HRF, e.g., a Gamma function or similar ones. Taking this assumptions, the convolution term in the BOLD response is a known function and 3.5 reverts to a standard multiple linear regression model. Therefore, the BOLD signal response is a design matrix X where the columns are the K predictors. Hence, the data for subject j at voxel i can be express as [7]:

$$\boldsymbol{y}_j = \boldsymbol{X}_j \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j \tag{3.7}$$

where $\boldsymbol{\epsilon}_j \approx N(0, V)$ with the structure of the covariance matrix $\boldsymbol{V}$ corresponding to an AR(2) process with unknown parameters ϕ_1, ϕ_2 and σ. The parameters of the model can be estimated using a Cochrane–Orcutt fitting procedure and the variance components can be estimated using the Yule–Walker method [31]. Once the model is fitted, it is possible to test the effect $\boldsymbol{c}^T \boldsymbol{\beta}_j$, where $\boldsymbol{c}$ is a contrast vector. The hypothesis testing is computed with single model parameters using a t-test and subsets of parameters using a partial F-test. The Satterthwaite approximation is used to calculate the degrees of freedom for the test statistics, since the covariance matrix has to be estimated. For each brain voxel the procedure is repeated and the results are the test statistic calculated at that specific voxel. GLM is a very simple framework to model the data, but it is also rigid, as in the case of mismodeling, there will increase false discovery rate and general power loss. Using more complex models is difficult because the data to analyse are so huge and the model selection becames a challenging procedure. But, according to Lindquist [143], some techniques have been introduced to detect the brain aereas where the assumption are violated and then where the model misfit [156, 155]

The assumption that the HRF is constant across space and subjects may give rise to mismodeling in many part of the brain. Fortunately, it is possible to relax this assumption taking the HRF as a linear combination of a reference waveforms. They are called basis functions and the GLM framework allow to convolve a specified stimulus function with multiple canonical waveforms entering them into multiple columns of $\boldsymbol{X}$ for each condition. The predictors constructed using different basis

[7] In this model presentation, the notation adopted is the one presented by Lindquist [143], that he assumed that the nuisance term Z can be ignored and assumed an identical model for each voxel and suppressed the voxel index.

functions can combine linearly to better fit the evoked BOLD responses. There is a trade-off between flexibility and power of a model, therefore if a basis set is able to capture variation in HRF the model will be flexible to handle noise. A very flexible model is the finite impulse response (FIR) basis set, that has one free parameter for every time-point following stimulation for every event-type used [86, 92]. Other basis sets are the usage of canonical HRF with its temporal derivative, or include sets with principal components [2, 269], cosine functions [281], radial basis functions [209], spectral basis sets [139] and inverse logit functions [144] [8].

The model presented before regards the single subject data analysis. However, experimentalists want to conclude on big samples and the statistical models have to handle also multiple datasets. The Multi-subject analysis is the name of this kind of procedure. This people-based data are hierarchical, with single subjects information nested in a multiple subjects information. Thanks to the multi-level models framework is possible to performe mixed-effects analysis on multi-subject fMRI data. Usually, with fMRI there are two-level model: first level that deals with individual subjects and the second level that deals with a collection of subjects. Note that the data are autocorrelated in the first-level model, whereas the data are independent and identical distribuited (IID) in the case of second-level model. More in details, the first-level model is:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.8}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_M^T, \quad \boldsymbol{X} = diag(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M), \quad \boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_M^T, \quad \boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \ldots, \epsilon_M^T, quadVar(\boldsymbol{\epsilon}) = \boldsymbol{V}$ where $\boldsymbol{V} = diag(\boldsymbol{V}_1^T, \ldots, \boldsymbol{V}_M^T$.

Instead, the second-level model is:

$$\boldsymbol{\beta} = \boldsymbol{X}_G\boldsymbol{\beta}_G + \epsilon_G, \tag{3.9}$$

where $\boldsymbol{\epsilon}_G \approx N(0, \boldsymbol{I}\sigma_G^2)$. Here $\boldsymbol{X}_G$ is the second-level design matrix (e.g., separating cases from controls) and $\boldsymbol{\beta}_G$ the vector of second-level parameters. The two-level model can be integrated with the single-level model as following:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{X}_G\beta_G + \boldsymbol{X}\epsilon_G + \boldsymbol{\epsilon}. \tag{3.10}$$

[8] For a critical evaluation of various basis sets, see [145, 147].

Regression parameters and variance components can be estimated by iteration, using last-squared method (LM) for the first one and restricted maximum likelihood (ReML) or Expectation-Maximization algorithm (EM) for the second one [143]. With the work by Friston [74] and Beckmann [13] these types of multi-level mixed-effects models reach popularity. Within this kind of models, researchers have to manage also missing data and unbalanced designs. Data missed are present as a consequence of artifacts or human errors, whereas experimental design may be unbalance if the goal is to relate brain functions to performance tasks or variables that are not experimentally controlled.

- Data with uncertain timing of activation. In cognitive neuroscience and in experimental psychology is difficult to set the *a priori* activation time. There are situations in which assuming either the paradigm or the HRF is difficult. The GLM does not work without the assumptions related to activation time and HRF, hence different methods have to be used. These alternative methods is called data-driven approach, i.e. data drives *post hoc* patterns related to experimental task.

  The family of these methodologies[9] regards Principal Component Analysis (PCA), Factor Analysis (FA) and Indipenden Component Analysis (ICA) [14, 38, 37, 171]. They are flexible methods, but without formal framework to explain the components meaning in relation to data analysed [143]. Furthermore, data-driven models capture information with no reference to the brain specific features. As consequence, their results are affected by noise and artefacted components.

During the postprocessing procedure, there are not only the decision to adopt data with or without time of activation, but also computational frameworks to test the statistical validation of the outcomes achieved, i.e., this checking procedure is known as multiple comparison.

- Multiple comparisons. The outcomes of a fMRI experiment are - usually - presented as statistical parametric map (SPM). If a voxels exceed some statistical threshold of significance, using t-test or similar statistics, the maps represent the functional activations. Therefore, it is possible to affirm where the brain was statistically active during specific task. The main point for this operation is to select the appropriate threshold to decide if the voxels were activated. Practically, during an experiment, every voxel performs simultaneously an hypothesis tests

[9] Data-drive methods are also used to study the brain connectivity: see the next subparagraph 3.2.2 to look some hints about them.

(around 100,000). The consequence is that the peculiar step will be to correct the results for multiple comparisons. There are a lot of methods to control the false discovery rate (FDR). Their main difference is the type of control: family-wise error rate (FWER) or the false discovery rate (FDR)[10]. To control the FWER, the popular method is based on the Random Field Theory (RFT) [271]. In the RFT, there is the assumption that the image of voxel-wise test statistic values are a discrete sampling of a continuous smooth random field. The RFT approach estimates the smoothness of the image, which is expressed in resels, i.e., the terms of resolution elements, that are approximately the number of independent comparisons. Then, the expected Euler characteristic[11] is computed using the number of resels and the search volume shape. Therfore, the Euler characteristic determines the threshold able to controls the FWER. Even if RFT is an fine theory, the drawback are the too conservative results [101]. Otherwise, there are nonparametric methods controls FWER without assumptions on the data distribution. Also, nonparametric methods are powerful and valid with small sample sizes [183].

The brain connectivity The mapping of the brain has the aim to indicate which regions are involved during an experimental paradigm. Researchers are interested on the regions that are functionally related and the ones that are linked specifically to the task done by the subjects. The literature divides the connectivity in anatomical, functional and effective [71]: the anatomical connections are studied with Diffusion Tensor Imaging (DTI), whereas the functional Magnetic Resonance Imaging (fMRI) studied the functional and the effective connections, with the difference that in the first case

[10]The false discovery rate (FDR) is the controlling procedure base on FDR checks the proportion of false positives among all rejected tests [84]. The controlling procedure base on FDR is adaptive. It means that more large the signal, the more low the threshold. The FDR is equivalent to FWER if all the null hypotheses are true. The procedures that controls FWER control also the FDR. Therefore FDR will be less stringent and it allows increase power. FDR has the advantage that it can be applied to many valid statistical test, since it works on the p-values. In opposition, a known distribution is needed to be follow by test statistics for the RFT. The FDR controlling procedure that is principally adopted in fMRI data analysis is the Benjamini–Hochberg procedure [15], in which tests are assumed as independent. Nevertheless, in many case tests are dependent and neighbouring voxels have probably similar p-values. According to Lindquist, in the future of the research about statistical topics related to false discovery rate will be many opportunities to develop controlling procedures that include spatial information [143].

[11]This value is equal to the number of clusters of activity that one would expect by chance at a certain statistical threshold [143].

there are indirected associations and in the second there are directed associations. In this subparagraph the functional connectivity and the effective connectivity will be presented.

- The Functional Connectivity regards multivariate methods and graph theory methods. The ones related to multivariate domain are clustering algorithms, a.k.a., brain parcellation [132], Principal Component Analysis (PCA) [3] and Independent Component Analysis (ICA) [37] [171]. These kind of methodologies allow to avoid to make *a priori* assumptions about fMRI data.

  - Clustering and PCA. For more details about brain parcellation see the sections on Fuzzy Algorithms 2.2.2 and the Competitive Algorithms 2.2.3, whereas for details about Principal Component Analysis see the Spectral Analysis section 2.2.5
  - Indipendent Component Analysis (ICA). The main difference between ICA and PCA is that the ICA components have to be independent rather then orthogonal [143]. The assumption of ICA regards that $\boldsymbol{Y}$ is a weighted sum of p independent source of signals within a $p \times N$ source matrix $\boldsymbol{X}$, that has weights describes by a $T \times p$ mixing matrix of weights $\boldsymbol{M}$, i.e. $\boldsymbol{Y} = \boldsymbol{M}\boldsymbol{X}$. To estimate $\boldsymbol{M}$ and $\boldsymbol{X}$ the iterative search algorithms are used. An ICA of $\boldsymbol{Y}$ computes spatially independent component images in the matrix $\boldsymbol{X}$ (spatial ICA, sICA), whereas an ICA of $\boldsymbol{Y}^T$ computes temporally independent time series (temporal ICA, tICA).
  - Graph analysis, i.e., that is an approach to study the functional connectivity in the fMRI data that applied the graph theory, that is a conceptual frameworks belonged to discrete mathematics and networks sciences. More formally, a functional brain network can be defined as a graph $G = (V, E)$ with V the set of nodes, i.e., the brain regions, and E the functional connections between these regions of interest. Using the graph theoretical framework, the network nodes are represented as cortical or subcortical regions or modules of fMRI voxels. The functional connectivity between two regions is computed with a the correlation levels between their BOLD time-series, using also cut-off threshold or weights of connections. The resulting brain net is a graph representation of functional connected ROIs [12].

[12] See works by Sporns and Bullmore [230, 34, 231] and van den Heuvel [248] for more details on the application of graph theory to fMRI and the information on specific models of network, i.e., small-world, random, full-connected, etc.

- The Effective Connectivity is related to the analysis of small set of regions with defined set of directed *a priori* connections. Statistical tests are used to evaluate the significance of particular connections. The three main models that study the effective connectivity are the Structural Equation Modelling (SEM) [170], the Dynamic Causal Modelling (DCM) [75] and the Granger Causality (GC) [212].

Chapter 4

Thesis Results

In the results section are presented the outcomes of the thesis. They were obtained with the collaboration[1] of the CRAIIM Research Group[2] at University of Insubria in Varese (Italy).

The outcomes regard the usage of computational methods in two main domains:

- the Image Processing, i.e. the usage of computational procedures able to manipulate neuroimages;
- the Medical application, i.e. the usage of computational procedures with the aim to study clinical subjects and healthy subjects that have had an fMRI clinical exam or an fMRI cognitive experiments.

The first domain regards the results on image processing (see 4.1), the second domain regards the results with clinical subjects (see 4.2) and healthy subjects (see 4.3).

If the reader wants to quit the detailed results to have an high-level overview of them, in the last part of the chapter there is also a thesis results *summarium* 4.4, which is an executive summary about the outcomes achieved and presented in this chapter.

[1] This is motivation because within the text is possible to find plural expression as "we applied ..." or "we adopted".

[2] CRAIIM is the acronym of *Centro di Ricerca in Analisi di Immagini e Informatica Medica*, eng.: Centre of Research in Image Analysis and Medical Informatics at University of Insubria

4.1 Results on Image Processing

Quantitative Relations Between CRAIIM Human Brain Atlases

Premise

The work illustrated was first conceptualized and presented in its early version at Conference of International Neuroinformatics Coordinating Facility (INCF) in Reading (UK) on September 2016 (see [254]) and, after some internal extension, a new version with a deep analysis of the hybrid brain atlas was presented at the 1st Human Brain Project Student Conference that will take place in Vienna (Austria) on February 2017 (see [255]). The next pages will show this finally research.

The key-words associated to this research are: brain structural parcellation, neuroimaging, FSL, voxel based morphometry, brain atlases fusion Human Brain Project (HBP)

Introduction and Motivations

We present two type of CRAIIM hybrid atlas (CRAIM is the acronym of Centre of Research of Image Analysis and Medical Informatics at University of Insubria in Varese, Italy). Both types of atlases have in common an extensive integration of white and grey matter structures that allows representing 161 brain territories, but one with 1mm resolution and the second one with 2mm of resolution. There are many brain templates used for image processing, that cover precise brain regions [113]: Tailarach (the more detailed with 1105 ROI), Juelich (121 ROI), Harvard-Oxford cortical (96 ROI), Harvard-Oxford subcortical (21 ROI), Montreal National Institute template (9 ROI), and other more anatomically specific such as Cerebellar-FLIRT/NIRT (28 ROI) and SubthalamicNucleus Atlas (2 ROI). Therefore, the presence in the neuroinformatical context of different atlases, with diverse resolution and specialization, motivated the CRAIIM group to design a hybrid model using a number of regions oriented to the automated procedures as well as to visual inspection. Also, as presented at INCF conference 2016 [254], the CRAIIM hybrid atlas with 1mm resolution comes from a join operation between Juelich histological atlas and Harvard-Oxford cortical and subcortical atlases [3]. The motivation of using them selectively emerged by a) the lack of some fundamental clinical regions in the Juelich original model, b) the

[3]See the Appendix with deatils about atlases 6.1. Specially, for Harvard-Oxford see 6.1.3, for Juelich see 6.1.2 and for the CRAIIM see 6.1.1.

high level of compatibility in terms of numerosity and volumetry of regions between HO and Juelich, and c) the peculiarity of these two atlases of being in the same MNI 152 space, allowing the researcher to completely compare them.

Methods

There are two topics that must to be taken into account: Integration Procedure (IP) and Voxel-Label Probability (VLP):

- Integration Procedure (IP). To make the integration correctly, we have to choose how to cope with atlases overlaps between brain regions belonged to Juelich and Harvard-Oxford atlases. We chose Juelich template as our reference with the purpose of completing it by regions originally present only into the Harvard-Oxford atlas (cortical and subcortical portions), from which we selected 40 ROI, such as frontal and temporal cortexes, subcallosal portions, and cingulated gyrus and thalamus halves. The main criteria to manage partial intersection between brain structures was to save all the portions just belonged to Juelich, and then to integrate them with the selected Harvard-Oxford regions, excluding their part overlapped, in other words, we framed the Juelich template with 40 regions of Harvard-Oxford without their portions that generated the overlaps.

- Voxel-label Probability (VLP). Another step was to decide which probabilistic version of the atlases to consider. We use atlases contained into FMRIB Software Library [113] that offers probabilistic templates that, for each of their voxel, assign a probability to fit a certain brain region. This allows giving to research the ability to use atlases that have all regions with probability major or equal to 0, otherwise with probability major or equal to 25 or, finally, major or equal to 50. Intuitively, the less is the degree of probability, the more is the volume of the atlas. We consider Juelich and Harvard-Oxford atlases with a probability threshold major or equal to 0, i.e. their version with all of the possible labelled voxels to be shaped for their integration. We tested the CRAIIM hybrid atlases with the computerized pipeline of fMRI analysis that estimates synthetic measures, i.e. Activated Weighted Indexes and Vectors [196, 198], for each of the brain regions covered by the hybrid template (see [2] for results and comments). Also, we evaluated it with FMRIB Software Library tools for neuroimaging analysis.

Results

There are the following results:

1. Results about the Figure 4.1. It shows the difference before and after the integration of Juelich brain with the 40 regions hailing from Harvard-Oxford cortical and subcortical atlases. In the three images in the top row, there is the superimposition between the Montreal National Institute (MNI) detailed anatomical atlas, as example of possible patient brain, and the Juelich atlas. A large amount of portions is not covered with, such as frontal lobes, many temporal territories, thalamus halves, and so on. In the bottom row, there is the same MNI template, but with the superimposition of the CRAIIM hybrid atlas, that covers exactly the same volumes enclosed by Juelich one, but having extra-extended brain regions, that are the 40 added from Harvard-Oxford atlases.

2. Results about the Table 4.1 and 4.2. They explain quantitative relations among the two type of versions of CRAIIM hybrid atlases. Table 4.1 illustrates how many voxels are in each template and how many of them belong to Juelich or Harvard-Oxford models. It is indicated the number of regions that come from them, that are 162 and not 161 because we counted also the empty space, i.e. the black area in the images that has not brain labels. The volume of atlases with their empty space is always the same: with 1mm length for each voxel and with 7,221,032 voxels, there is a volume of 7,221,032 mm3; equivalently, for 2mm voxel length, but with 902,629 voxels, the volume is always 7,221,032 mm3 (7 Litre). Table 4.2, instead, presents the numbers of voxels of two atlases organizing the 161 brain regions with 9 anatomo-functional meta-labels, i.e. Acoustic, Associative, Behaviour, Fascicles, Language, Limbic System, Motor, Somato-Sensitive and Visual.

3. Results about the Figure 4.2 and Figure 4.3. They show the voxel distribution through CRAIIM hybrid atlases with 1mm and 2mm of resolution. The quantities are clustered by the 9 anatomo-functional zones and subgrouped for their original models. There is no difference in proportion between the two Figures, and this reflects the equivalence in terms of representability of CRAIIM hybrid atlas in both resolutions. What it is changed in Figure 4.2 and 4.3 is for sure the numbers of voxels that depend by voxel length.

4. Results about Figure 4.4. It is a bar plot that makes in relation the voxel enumeration of the only 40 regions added to Juelich that com-

ing from Harvard-Oxford. We selected them because the treatment of overlapping causes a reduction of the original volume of these regions added. The graph shows the difference in voxel number of 40 structures before and after the integration operation.

5. Results about Figure 4.5. They highlight the better completeness of the 1mm version compared to the 2mm of the CRAIIM hybrid atlas, due to the minor gap in terms of voxel percentage in reference to the original Harvard-Oxford template. Figure 4.5 proved that the Juelich regions in CRAIIM hybrid atlases are the same of those in Juelich model.

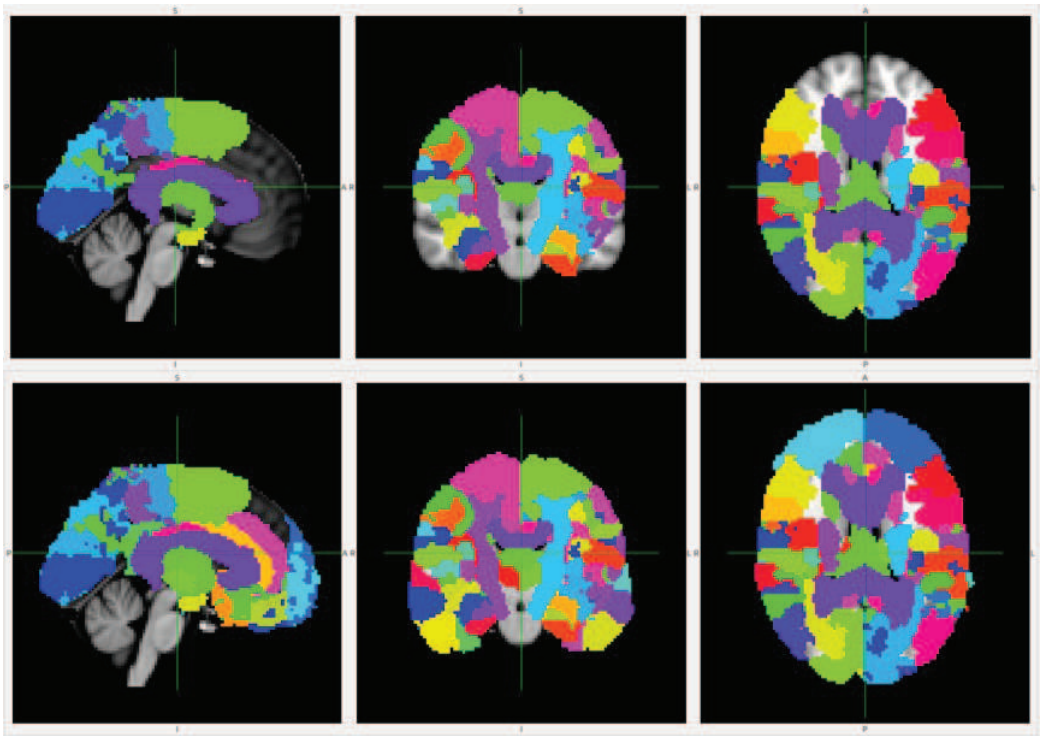

Figure 4.1: In the upper part of the Figure, there are the sections of MNI anatomical template with the overimposition of Juelich brain. In the lower part the one with CRAIIM hybrid atlas. It is evident the lacking in Juelich brain that has been completed by CRAIIM hybrid atlas (image with 2mm of resolution for all the models).

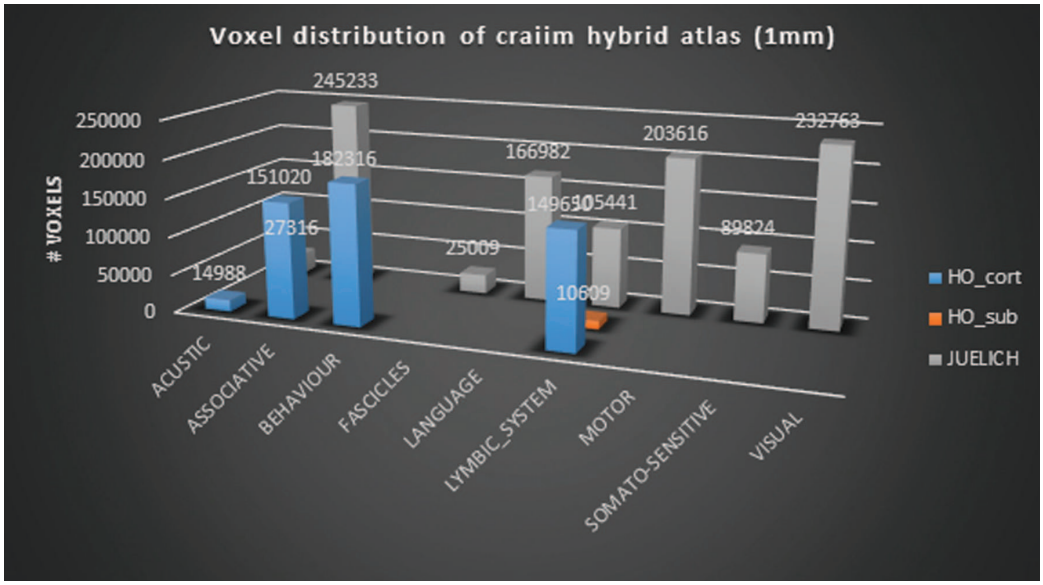

Figure 4.2: . Voxel distribution clustered by 9 anatomo-functional meta-labels of 161 brain regions of CRAIIM Hybrid Atlas with 1mm of voxel resolution organized by their original atlas (Juelich in grey and Harvard-Oxford cortical and subcortical in orange and blue, respec-tively).

Table 4.1: Numerosity of CRAIIM voxel distinguished for resolution type and organized by anatomo-functional meta-labels (n.b. NA means empty space, i.e. black volume without brain regions)

LABELS	**CRAIIM [1 mm]**	**CRAIIM [2 mm]**	**REGIONS [#]**
ACUSTIC	42304	5667	16
ASSOCIATIVE	396253	51711	39
BEHAVIOUR	182316	23571	2
FASCICLES	25009	3579	6
LANGUAGE	166982	23216	16
LYMBIC_SYSTEM	265700	33518	42
MOTOR	203616	27727	8
NA	5616265	689415	1
SOMATO-SENSITIVE	89824	11954	16
VISUAL	232763	32271	16
Grand Total	**7221032**	**902629**	**162**

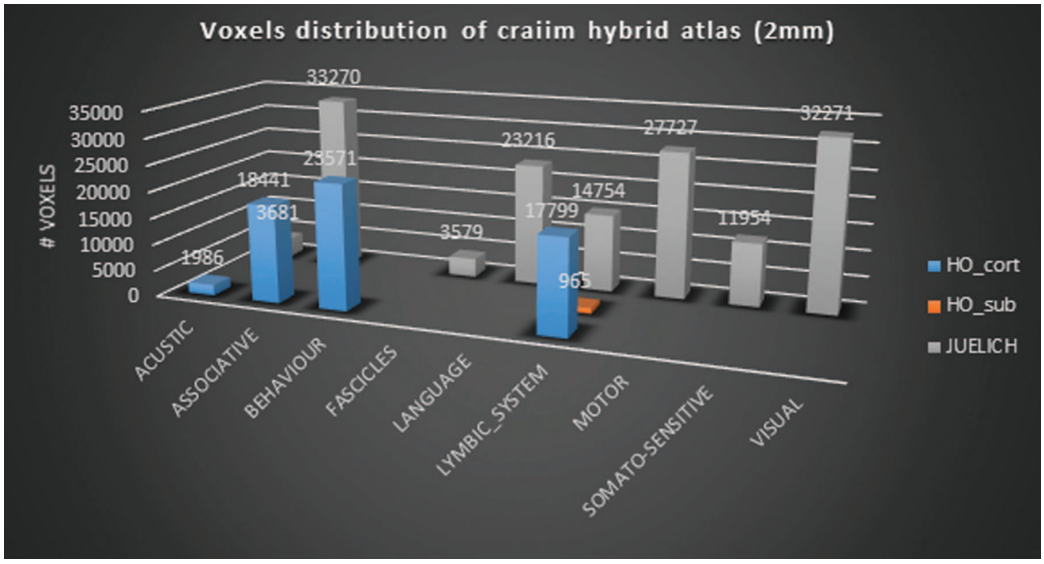

Figure 4.3: Voxel distribution clustered by 9 anatomo-functional meta-labels of 161 brain regions of CRAIIM Hybrid Atlas with 2mm of voxel resolution organized by their original atlas (Juelich in grey and Harvard-Oxford cortical and subcortical in orange and blue, respectively).

Table 4.2: Numerosity of CRAIIM voxel grouped for resolution type and organized by original atlases (n.b. NA means empty space, i.e. black volume without brain regions).

ATLASES	**CRAIIM [1 mm]**	**CRAIIM [2 mm]**	**REGIONS[#]**
HO cortical	497974	61797	38
HO subcortical	10609	965	2
JUELICH	1096184	150452	121
NA	5616265	689415	1
Total	**7221032**	**902629**	**162**

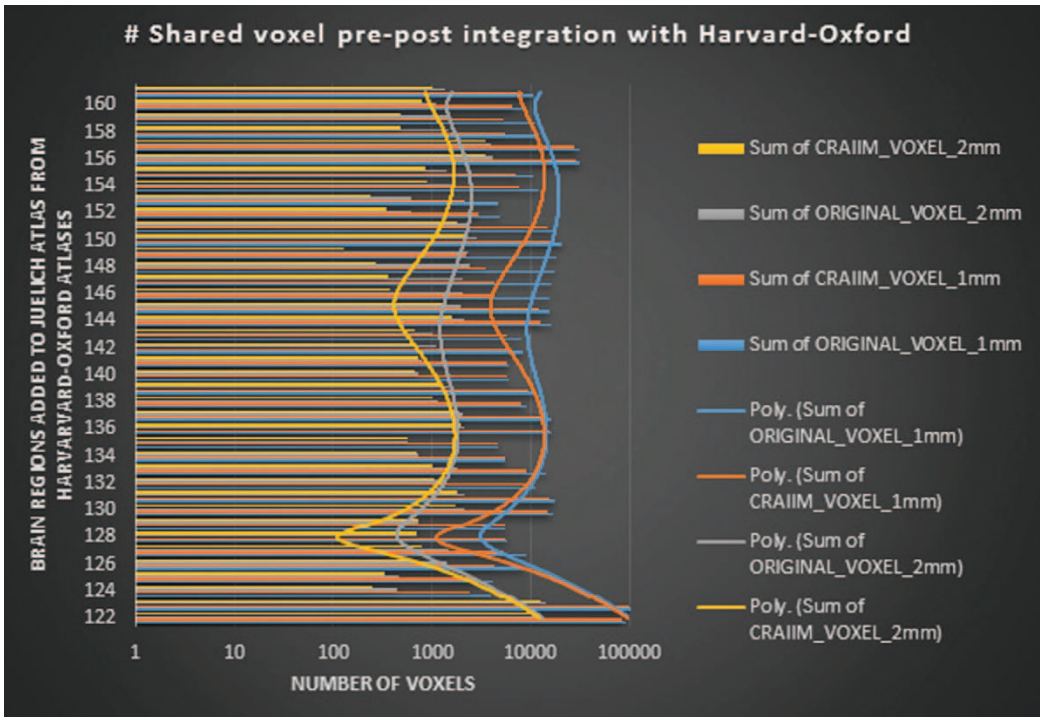

Figure 4.4: Bar plot with the 40 added regions to Juelich atlases coming from Harvard-Oxford atlases. The different bars show the original voxel value of each brain regions and the value that they have had after the overlap treatment. The main result refers that the CRAIIM hybrid atlas with 1mm of resolution shares more with the original atlas then the CRAIIM hybrid atlas with 2mm of resolution (see. the convexities among Polynomial trend lines: Orange-Blue duo (1mm) VS Yellow-Grey duo (2mm)).

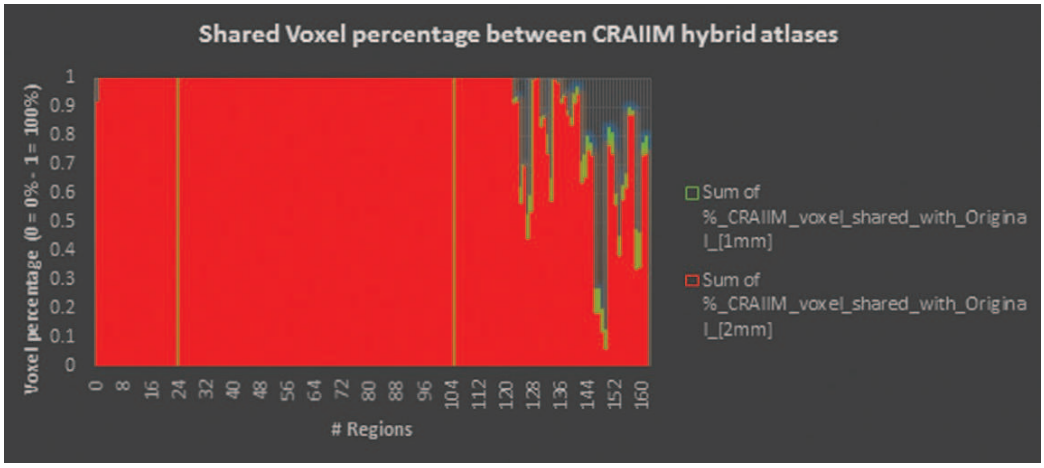

Figure 4.5: CRAIIM hybrid atlas with 2mm of resolution (Red bars) always have minor shared voxel percentage with their own original structures (Juelich and Harvard-Oxford atlases), for the 40 added regions (122-161), then the CRAIIM hybrid atlas with 1mm of resolution (bars in Green).

Discussion and conclusion

We have presented two versions of the CRAIIM Hybrid Atlases of human brain for neuroimaging purposes. They differ for resolution, but not for representativeness of the 161 brain regions (kept 121 from Juelich and the 40 from Harvard-Oxford cortical and subcortical atlases). They are in perfectly proportions of the same brain structures (161 ROI; see. Figures 4.2 and 4.3), but the 40 added regions from Harvard-Oxford differ in the two version of our atlas for percentage shared with their original shapes. The CRAIIM hybrid atlas with 1mm resolution has more portions in common with the 40 Harvard-Oxford parts, then the one with 2mm of resolution (Figure 4.4 and 4.5). This percentage variable by resolution depends properly by the length of the voxel. The presence of more voxels with the 1mm version of CRAIIM hybrid atlas has allowed to better shape the added structures, framing them more stylishly, with respect to the version with 2mm. The "thumb rule" used states that is better to have voxel-labeled then voxel-unlabeled (i.e. black voxel). The constraint with this rule is evident when we added brain regions that are in low percentage of voxel shared with their original version, after the integration procedure.

The advantage is indeed the possibility of displaying them instead the empty space. Next step could be to trace the percentage cut-off that law if or if not include regions underrepresented after integration procedure.

A part of these difference linked with the resolution, we have shown that CRAIIM hybrid atlases, both 1mm and 2mm version, cover extensively the Montreal National Institute anatomical template, as a model of patient brain, with respect to the Juelich atlas, safeguarding many territories otherwise lacking (Figure 4.1). In potential, CRAIIM hybrid atlases are very suitable to use in functional neuroimaging for severe scope, e.g. ROI masking, ROI selection, labels comparison, registration processing [143]. The CRAIIM Hybrid Atlases could be used to promote dimensionality reduction of SPM volumes with the computation of synthetic indexes [196, 198, 197].

Actually, the Neuroinformatics Platform of Human Brain Project is able to display anatomical atlases of mammals such as Mus Musculus, Rattus Norvegicus and Homo Sapiens (e.g. BigBrain, MNI Colin27, MNI ICBM, InfantAtlas). In the future, the development of neuroinformatical tools to generate hybrid atlas, with a good rational between human usability and covered neurotopography, and with a wise algorithm to handle overlaps, could be the right way to design purpose-based atlas with specific anatomical representativeness and normative features able to standardize and process heterogeneous functional and structural brain images.

4.2 Results with Clinical Subjects

Cluster Analysis of fMRI using Data Reduction and Competitive Algorithms

Premise

This work was first presented at European Congress on Computational Methods in Applied Sciences and Engineering in Porto (Portugal) on October 2017 [252] and then extended and published as special issue in the journal CMBBE: Imaging and Visualization [20].

In particular, we use pattern vectors derived from Statistical Parametric Map, generated from a group of artificial and in-house clinical collected fMRI data, to conduct cluster analysis. Two clustering algorithms, self-organizing map (SOM) and growing neural gas (GNG), are selected to explore inherent properties in the brain functional data. As seen in our experimental context, SOM and GNG show comparable behavior, however GNG prevails in the management of large data sets. An exploratory, descriptive analysis is conducted on in-house collected data clustered by GNG and results are detailed in the paper.

The key-words associated to this research are: fMRI, Statistical Parametric Mapping, Data Reduction, Self Organizing Map, Growing Neural Gas

Introduction

Functional magnetic resonance imaging allows to detect relevant functional areas of the brain involved in motor or cognitive functions by asking patients to perform different tasks during the image time-series acquisition [143]. Several methods of fMRI analysis have been investigated to localize brain activity. One of the commonly used approach for fMRI data analysis is statistical parametric mapping [71] that spatially extends statistical processes to test hypotheses about regionally specific effects. Recent studies investigate alternative methods of analysis in an attempt to overcame limitations of the SPM models that can lead to invalid or inefficient statistical tests in complex conditions. The use of data driven brain parcellation algorithms seems to be interesting, but these methods have to be furtherly investigated [132] and SPM method remains the most popular approach to analyze fMRI data in clinical setting.

In our previous works we proposed a new representation of the results obtained by the fMRI data analysis performed by statistical parametric mapping, obtained by applying a data reduction procedure to the SPM data: the

original three-dimensional distribution of activation values extracted from the statistical analysis is summarized in a pattern vector of weighted activation indexes [194, 195].

In the present work we use the SPM-derived pattern vectors, generated from a group of artificial and in house collected fMRI data, to conduct cluster analysis. The most important benefit of the use of synthetic activation vectors is that the computational load decreases considerably with respect to the direct use of SPM data, making it possible to cluster large data sets and to find optimized solutions considering several configurations in a limited time. The cluster analysis allows to perform a survey of multi-subject functional data with which to identify groups with similar patterns of activations. To this purpose two clustering algorithms, self-organizing map (SOM) and growing neural gas (GNG), are selected for their potentialities in exploring inherent properties in complex data. The SOM model [125] is especially suitable for data survey because it has interesting visualization properties. GNG is an incremental network model with the advantage of freeing users from the burden of setting the number of neurons and which is able to learn the important topological relations in a given set of input vectors by means of a simple Hebb–like learning rule [77, 81]. A comparison analysis is developed using Davies–Bouldin validity index [46] to select the best configuration among different partitioning of both methods. In a second step of the study, the obtained partitions are analyzed in an attempt to discover relevant properties and infer new knowledge in the clinical domain to which data belong.

fMRI dataset and Reduction Procedure

We adopted in-house clinical data coming from an Neurosurgery Unit of the Macchi's Hospital in Varese and sinthetical data computed with the in-house ones as reference. Furthermore, we adopted a dimensional reduction procedure able to transform the exams in vectorial form.

fMRI data The acquisition parameters of the in-house collected data are the following:

- EPI BOLD (TR 3000ms, Te 50 ms, FA 90, matrix 96x96, FOV 230 mm, axial, slice thickness 4 mm, spacing 4 mm, 30 slices);
- reference anatomical image is T1 weighted (TR around 7100 ms, TE around 3200 ms, FA 8, matrix 256 x 256, FOV 256 mm, slice thickness 1 mm, spacing 1mm, 160-190 slices);

- block stimulation paradigms (30s ON, 30s OFF repeatedly): language areas were stimulated with visual stimuli or with phrases associated with visual stimuli in tasks involving the understanding of a sentence, motor areas were stimulated with light signals indicating the time to execute the corresponding movement.

In-house data We use in-house collected fMRI data from 37 patients[4] (14 Female and 23 Male) that did several fMRI sessions, with a total of 141 exams. They came from neurosurgical or neurological operative units (see Table 4.3) and they made different neurocognitive tasks, such as sentence mining, object naming, verbal fluency, verb generation, left finger tapping, right finger tapping and visual assignment (see Table 4.4).

Table 4.3: The table shows the 141 exams organized by gender and clinical operative units.

Operative Units	**Female**	**Male**	**Total**
Neurosurgery	46	76	122
Neurology	0	9	9
Both	10	0	10
	56	85	141

Table 4.4: The table lists the 141 exams organized by gender and neurocognitive tasks from which data have been generated

Tasks	**Female**	**Male**	**Tot**
Sentences Mining	1	0	1
Objects Naming	13	18	31
Verbal Fluence	12	14	26
Verbs Generations	14	15	29
Left Finger Tapping	8	18	26
Right Finger Tapping	8	17	25
Visual Assignment	0	3	3
	56	85	141

[4]The informed consent was obtained from patients involved in the research. The present study is based on data derived from NMR examinations that are inserted in the routine follow up of patients. The method proposed in the present work is a post processing application of these data utilized in anonymous form. The policy of the Hospital of where the examinations took place (Ospedale di Circolo Fondazione Macchi in Varese, Italy) does not require in this case the approval of the ethical committee.

Synthetic data Clustering techniques are oriented to the management of huge datasets. In order to provide a more robust evaluation of the output of our competitive learning algorithms, we decided to extend the initial dataset with two artificial datasets. We used an efficient resampling method with replacement: it works sampling k times the original data keeping in account the multiplicity of each value, i.e. frequent items have more sampling frequency. This algorithm returns k observations sampled uniformly at random, with replacement, from the data. The symbolic function is the following: $y = datasample(data, k)$, with $data$ and k dataset and number of observations respectively. This statistical method allows us to generate different artificial datasets, with $k >> 141$, but with the same distribution of in-house data. We implemented this algorithm with $k = 2000$ and with $k = 3000$.

Data Reduction The IViewBOLD software supplied by Philips® is adopted to generate SPM volumes from time series MRI data. The data reduction procedure is accomplished by an integrated set of software tools that receives in input SPM volumes and performs a sequence of tasks as described below:

- Whole Brain Segmentation. SPM volumes are provided in input to a Graph-Based segmentation procedure [165].
- Registration. Segmented brain volumes are automatically aligned to the MNI (Montreal National Institute) template by using an affine 3D surface registration based on SPHerical HARMonic (SPHARM) decomposition [196].
- Data Reduction. Individual SPM are analyzed by identifying active voxels of the segmented, registered brains, i.e. voxels whose t-values (or comparable statistics) exceed a certain statistical threshold for significance; the result of the analysis is a synthetic weighted index computed for each anatomical area recognizable in the selected labelled atlas. We adopt a hybrid atlas obtained by fusing information from Juelich atlas and Harvard-Oxford cortical and subcortical structural atlases, covering a total of 161 regions [254]. The index I_j for the $j-th$ anatomical region is computed as

$$I_j = \frac{1}{N_j} \sum_{i=1}^{N_j} w_i \tag{4.1}$$

  where N_j is the number of voxels included in the $j-th$ anatomical area and w_i the activation value of the $i-th$ voxel, normalized in the interval [0,1].

Let K denote the number of SPMs collected for the fMRI study and let n be the number of labeled anatomical regions in the selected atlas, the sequence of indexes $x_1^k, \ldots, x_n^k$ can be interpreted as a vector $\mathbf{x}^k \in R^n$. The proposed method synthesizes the three-dimensional distributions of SPM activation values in unidimensional distributions represented as vectors of scalar indexes in a common vector space model. This procedure plays a central role as it limits complexity of the data structure and makes feasible the optimization processes in cluster analysis.

Clustering algorithms

We selected the Self Organizing Map [124] and Growing Neural Gas [81] algorithms for cluster analysis. These techniques implement competitive learning strategies to evaluate similarity and to group multidimensional patterns. To evaluate the clustering outcomes, we adopted the Davies-Bouldin separation measure [46].

See the Competitive Learning Algorithms section 2.2.3 for details about SOM and GNG and the Validation Measure section for details about the clustering validation index 2.2.4.

Experiment Results

The experiments illustrated in this section addressed the following questions: a) how did the performance of the neural models depend upon their main parameters? b) what results are obtained by comparing the two?

Experimental evaluation and comparison

The experimental analysis of the proposed clustering algorithms was conducted on the above described data sets.

- The first experiment, i.e., computation with in-house data. Several configurations were considered for the clustering procedure based on SOM, distinguished by the distance norm adopted (Euclidean or Manhattan), by the different values of neurons ranging from 2 to 24, according to heuristic rule [256] and by the number of epochs (max=600). Initial neighborhood was set to 3 and the chosen layer topology was the hexagonal. Several configuration of GNG algorithm were also considered setting the maximum number of neurons to 24 and varying epochs up to 600. Internal learning parameter were set to these values: $\lambda = 6$, $\epsilon_b = 0.5$, $\epsilon_n = 0.006$, $\alpha = 0.5$, $\delta = 0.99$, $\alpha_{max} = 50$. Table 4.5 lists the performances obtained by the optimal configurations that minimize

DB index. Globally, the results are good, showing a comparable behavior of the two methods considered in the analysis. SOM algorithm, configured with Euclidean distance, slightly prevails when DB index, adopting in turn Euclidean distance, is used as metrics.

- The second experiment, i.e., computation with artificial data. Optimal configurations for SOM and GNG algorithms were found varying parameter as in the first experiments with exception of the number of neurons that was varied here up to a maximum value of 60 neurons to cope with the increased complexity. Table 4.6 and Table 4.7 show the results obtained by clustering artificial data sets composed of 2000 and 3000 pattern vectors respectively. Performances of SOM algorithm decrease as the size of the data set increases, both in terms of DB index and computing time. On the contrary, GNG algorithm shows a stable behaviour with absolute best performances (DB= 1.0095) when the data set of 2000 patterns is processed.

- Algorithms evaluation, i.e., measure of the speed of computing. Using the adaptive algorithms, we measured the CPU time on a Windows platform with Intel(R) Xeon(R) CPU with 3.33 GHz processor and 16 GB of RAM.

Table 4.5: Results obtained by SOM and GNG algorithms by processing 141 pattern vectors from in-house collected data. Performances are evaluated in terms of DB index, computed using Euclidean and Manhattan distance, and computing time.

Algorithms	**Distances**	**Neurons**	**Epochs**	$DB_{Euclidean}$	$DB_{Manhattan}$	**Computing Time**
SOM	Euclidean	20	500	1.0382	1.1400	10 min
SOM	Manhattan	24	500	1.0669	1.0559	10 min
GNG	Euclidean	24	200	1.0855	1.1195	11 min

Table 4.6: Performances obtained by SOM and GNG algorithms by processing 2000 artificial pattern vectors. Performances are evaluated in terms of DB index, computed using Euclidean and Manhattan distance, and computing time.

Algorithms	**Distances**	**Neurons**	**Epochs**	$DB_{Euclidean}$	$DB_{Manhattan}$	**Computing Time**
SOM	Euclidean	60	300	1.1366	1.1729	360 min
SOM	Manhattan	60	500	1.2041	1.1342	360 min
GNG	Euclidean	24	200	1.0095	1.0261	70 min

Table 4.7: Performances obtained by SOM and GNG algorithms by processing 3000 artificial pattern vectors. Performances are evaluated in terms of DB index, computed using Euclidean and Manhattan distance, and computing time.)

Algorithms	**Distances**	**Neurons**	**Epochs**	$DB_{Euclidean}$	$DB_{Manhattan}$	**Computing Time**
SOM	Euclidean	24	500	1.5183	1.4970	120 min
GNG	Euclidean	24	200	1.0305	1.0656	104 min

As seen in our experimental context, both SOM and GNG shows an acceptable behavior. GNG prevails in the management of large data sets. As we are more interested in getting good insight into the cluster structure of the data than in finding an optimal clustering, we may also confirm SOM algorithm a valuable tool in our application domain offering topological and visual facilities in support to data exploration and inference of new properties. GNG is a complementary powerful tool able to conduct cluster analysis with the strength of automatically configuring the number of clusters and controlling the computational complexity.

Descriptive statistics

For the descriptive analysis, we choose in-house collected data clustered with the GNG algorithm, that partitioned the 141 exams in 24 clusters. Interestingly, 93 exams (66 % of the total) have been grouped in the 24th cluster, whereas the other are equally distributed between the clusters, i.e. $mean = 2.09$, $std.dev = 2.37$. Focusing on the 24th cluster, the 93 exams belonged to it are principally featured by gender, lesions and tasks. Males and females are nice balanced within the group (53 men and 40 women). Instead, the type of lesion is not equally distributed with respect to main totality of all clusters: Nervous (1/3, 33%), Vascular (11/25, 44%), Liquoral (3/5, 60%), Cortical (2/3, 67%) and Expansive (76/105, 72%). Remarkably, all the tasks are principally contained within the 24th group: Left Finger Tapping (14/26, 54%), Objects Naming (19/31, 61%), Right Finger Tapping (16/25, 64%), Verbs Generations (19/29, 66%), Visual Process (2/3, 67%), Verbal Fluency (22/26, 85%) and Sentences Mining (1/1, 100%). As a consequence, the proportion of exams featured by tasks are always more than the 50% of the total, instead of the ratio of the lesions that are quite heterogenous between them. In conclusion, the exams in the 24th cluster are principally featured by the expansive lesion (76/105, 72% of the totality) and by the Verbal Fluency task (22/26, 85% of the totality). Notable, the only exam featured by Sentence Mining has been included in the 24th cluster. (See table 4.8 for more details). The peculiar feature organization of the

24th group motivates to deepen the analysis of its content.

Conclusions

In the next work, we are going to investigate brain functional information regard the exams, i.e. values of the Regions Of Interest and their related indexes, with the aim to describe statistically, with multivariate exploratory analysis, the similarities and differences of the exams belonged to the 24th group and the others that have been grouped otherwise, in order to test hypothesis between clusters and clinical features.

Table 4.8: The table shows the clinical exams belonged to the 24th cluster organized by lesion and tasks. The percentage is related to the total exams (141)

TASKS / LESIONS	**Cortical**	**Expansive**	**Liquoral**	**Nervous**	**Vascolar**	**T**
Sentences Mining	0	1 (1)	0	0	0	**1**
Objects Naming	0	16 (24)	0 (1)	0	3 (6)	**1**
Verbal Fluency	1 (1)	18 (20)	0 (1)	0	3 (4)	**2 (2**
Verbs Generations	0	17 (24)	1 (1)	0	1 (4)	**1**
Right Finger Tapping	0 (1)	12 (17)	1 (1)	2 (2)	2 (4)	**1**
Left Finger Tapping	0 (1)	12 (19)	1 (1)	0 (1)	0 (4)	**1**
Visual Process	0	0	0	0	2 (3)	**2**
Tot [#]	**2 (3)**	**76 (105)**	**3 (5)**	**1 (3)**	**11 (25)**	**9**

4.3 Results with Healthy Subjects

In this subsection will be exposed the results obtained with the application of computational intelligence methodologies (see 2) on the subjects without diseases. The outcome achieved are organized and exposed in the following order:

- results on validation measures, i.e., results regarding the application of the soft Davies-Bouldin index as a novel methods to evaluate clustering outcome on resting state fMRI data (and also some benchmark datasets) (see 4.3),
- results on competitive algorithms comparison, i.e., results regarding the application of a wide family of unsupervised learning algorithm to resting state fMRI datasets with some usage of random matrix theory models (see 4.3),
- results on crossed-clustering framework, i.e., results regarding the application of clustering algorithm with specific intent to evaluate the crossed-clustering different outcomes on task-oriented fMRI dataset (see 4.3).

A Soft Davies-Bouldin Separation Measure

Premise

This work was presented at IEEE World Congress on Computational Intelligence in Rio de Janeiro on July 2018. It regards the conceptualization of a novel separation measure for clustering algorithms based on fuzzy theory and the so-called Davies-Bouldin index.

In particular, the work exposes a soft separation measure to validate fuzzy clustering results without defuzzyficaton. It is the generalization of Davies-Bouldin validation index (DB) for crisp clustering in the soft clustering domain; we named the measure Soft Davies-Bouldin index (SDB).

We compared DB and SDB when applied to k-means and fuzzy c-means algorithms using eight datasets with ground-truth and two experimental fMRI datasets without ground-truth. We found that i) in more than half datasets, the optimal score of Soft Davies-Bouldin index was less than Davies-Bouldin index, ii) in half datasets that have ground-truth, the optimal score of Soft Davies-Bouldin index was less than Davies-Bouldin index in correspondence of the truth number of patterns, iii) the Soft Davies-Bouldin index outperformed the Davies-Bouldin index as central tendency of all datasets along the complete range of clusters considered.

The key-words associated to this research are: clustering, k-means, c-means, separation measures, Davies-Bouldin index, fMRI.

Introduction

In pattern recognition, clustering methods are a class of unsupervised learning models used to find the *natural groupings* of input features [207, 56]. The clustering procedure regards the examination of objects according to specific measures with the goal to cluster similar objects. Precisely, a dataset useful for clustering is a collection of points that belong to some space, for example the Euclidean space, which has important mathematical properties: i.e. points are vectors of real numbers, the length of the vectors are the dimensions of the space and the components of the vectors are the coordinates of the points; the subdivision of this space in clusters is the result of clustering[246].

Partitioning methodologies were first presented in the 20th century by Driver and Kroeber [55], Zubin [283], Tryon [242] and Cattell [41]; then, the discipline grew up developing several algorithms for different aims and technicalities (see [111] for an historical overview and [215] for a discussion on different clustering approaches). But, very often, these algorithms lead

to different clusters and a measure to evaluate results is necessary to avoid inappropriate representations [16, 112, 53, 6]).

According to Bezdek *et al*, in clustering analysis there are three main problems [19]:

1. the *a priori* assumption that datasets have clusters,
2. the computation method to find clusters,
3. the evaluation procedure to verify the clusters found.

The third problem in the list is the well-know cluster validity issue, that is the validation procedure able to decide what is the best partition in a set of candidate clustering outcomes. Following Bezdek, there are different approaches to validation:

- the approaches regarding the decision modalities, i.e. visual or nonvisual inspection;
- the approaches regarding the clustering modalities, i.e. crisp or soft partitioning;
- the approaches regarding the data modalities, i.e. object features or relational features;
- the approaches regarding method modalities, i.e. internal validation (using indexes) or external validation (using ground-truth)

In the 1979, Davies and Bouldin proposed in their paper [46] two principal perspectives to solve the problem of clustering validity:

- optimization outlook, i.e. select the optimum cluster number with performance indexes [240],
- hierarchical outlook, i.e. evaluate intergroup fusions with dendrogram plots [89].

In the same manuscript, authors proposed themselves an index able to find the natural partitions of data: the minimum the parameter value, the better the clustering results; it is the so-called Davies-Bouldin index, that relates similarities measures within groups *versus* similarity measures between groups. This kind of cluster separation measure incorporates properties of noted algebraic distances and also satisfies heuristic criteria; nevertheless, it was design for crisp clustering.

In our work, we want to extend the peculiarities of Davies-Bouldin validation index in the soft computing domain. We started from the definition of separation measures and we generalize its formulation according to fuzzy sets theory [278]. Using the Bezdek's validation tassonomy exposed above, we defined a validity measure that cope with the following properties: i) nonvisual inspection, ii) soft partitions, iii) object features, iv) both internal and external modalities of validation.

The paper is organized with a theoretical section where we present fuzzy concepts and measures of clusters separations. The successive sections are about the computational experiments, the results, discussions, and the conclusions with future works.

Methods

The clustering methods used were the fuzzy c-means and k-means algorithms. The separation measure we adopted to evaluate the clustering outcomes were the classic - crisp - Davies-Bouldin index and the Soft Davies Bouldin index. To look the methodological details about them, see the Computational Intelligence section 2.2.3 and the Validation section 2.2.4.

Datasets

To test Soft Davies-Bouldin index (SBD) and the classic Davies-Bouldin index (DB) with k-means and c-means clustering, we selected eight benchmark datasets with ground truth (with patterns known) and two experimental datasets without ground truth (without patterns known):

- the eight benchmark datasets are findable principally from UCL online Database (https://archive.ics.uci.edu/ml/index.php; see [141]); they are Cancer, Crab, Glass, Iris, Ovarian, Simple, Thyroid and Wine. We choose them because they have a just-known clusters number; this peculiarity permits to better evaluate the performance of SDB and DB indexes in different experimental configurations.

- the two experimental datasets are coming from two neuroimaging repositories:

  - dataset we named REST belongs to NITRC repository of resting-state functional MRI images (see also [120]); specially, it is the *Oxford* dataset with 19 healthy and young subjects that have done a passive paradigm; the fMRI parameters were the following: TR = 2, time-points = 175, magnet = 3 [T];

– dataset we named TASK belongs to openfmri repository; specially, it is a datset proposed by Wakeman and Henson [259] with 15 healthy and young subjects that have done Face Recognition Task paradigm; the fMRI parameters were the following: TR = 2, time-points = 208, magnet = 3 [T];

– before to cluster REST and TASK datasets, we did preprocessing with FSL standard tools [113]: i.e. spatial and temporal filtering, motion correction, standard registration (with MNI152 reference), time-series extraction with the meaning of Harvard-Oxford atlas[5] with 96 lateralized labels;

Table 4.9 shows details about datasets:

Table 4.9: Datasets with and without Clustering Ground-Truth

Datasets	*Description*	*Inputs*	*Patterns*
Cancer	Breast cancer	9x699	2
Crab	Crab gender	6x200	2
Glass	Glass chemical	9x214	2
Iris	Iris flower	4x150	3
Ovarian	Ovarian cancer	100x216	2
Simple	Simple pattern	2x1000	4
Thyroid	Thyroid function	21x7200	3
Wine	Italian wine	13x178	3
Rest	fMRI - passive paradigm	96x175x19	unknown
Task	fMRI - active paradigm	96x208x15	unknown

Experimental Procedure

We computed Davies-Bouldin index (DB) and Soft Davies-Bouldin index (SDB) in the following clustering conditions:

1. Absolute Indexes Minimization (AIM), i.e. the goal is to search the optimal clusters number given the minimum score of both DB and SDB;

2. Relative Indexes Minimization (RIM), i.e. the goal is to compare the score of both DB and SDB in relation to the just-known cluster number to each dataset.

[5] See the appendix 6.1.3)

3. Measures in Relation to Cluster (MRC), i.e. we computed the separation measures with the objective to look how the indexes change along the range of clusters, taken account the centrality tendency of the datasets.

The above clustering conditions were applied to these three outcomes from the clustering algorithms:

- pure k-means results, i.e. the classic crisp k-means outcomes;
- pure c-means results, i.e. the soft c-means outcomes with all the membership values for each clusters associated to each data points.
- defuzzified c-means results, i.e. the crisp c-means with the defuzzified outcomes (we used the max membership values for each clusters as a criterium to include the associated point).

Technically, we used the following operative methodologies:

- the datasets REST and TASK were clustered adopting the cross-clustering procedure, that it is useful for spatio-temporal exploration in fMRI analysis, i.e. we first used datasets having as input features the Regions Of Interests (ROIs), in order to find *functional temporal patterns*; then using as input features the Time Of Interests (TOIs), in order to find *functional spatial patterns*; in this manner, from datasets REST and TASK we created four datasets named RestROIs and TaskROIs (with ROIs as features) and RestTOIs and TaskTOIs (with TOIs as features). Instead, the eight benchmark datasets were clustered with classical procedure;
- Note I, i.e., to test indexes in the condition 1), we computed clustering with a cluster number varying from 1 to 10 and we selected the optimal (i.e. the lesser) separation values; to test indexes in the condition 2), we made computation only with the truth cluster number for each dataset and we selected the optimal (i.e. the lesser) separation values; to compute the indexes in the condition 3), we computed clustering with a clusters varying from 1 to 10 and, for each number of clusters, we selected the median statistics of separation measures computed for each datasets (i.e the median descriptor is useful to handle empirical outliers);
- Note II, i.e., we repeated experiments 100 times for the above conditions and we kept the average of separation measures for each clustering

algorithms; but in particular, for datasets REST and TASK that have 19 and 15 subjects respectively, we computed all the experiments taking each subjects as single dataset (with n *features* $\times$ m *samples*), and then we centralized statistically the outcomes to have only one result as global tendency for all the subjects belonged to REST and one result as global tendency for all the subjects belonged to TASK dataset;

- Note III, i.e.,once we found the more candidate number of patters associated to the best measure for each datasets, we selected the more frequent classes between the 100 computed for each data-points (e.g. if the best partition for one dataset has three patterns, having repeated 100 times the clustering algorithm, we obtained one hundred of possible partition with three patterns; therefore, we selected the more frequent class for each data-point);

- Note IV, i.e.,the distance chosen is the Euclidean metric for both k-means and c-means clustering and for both the crisp and soft separation indexes; each clustering algorithms used 100 iterations; c-means was tuned with fuzzy exponent $m = 2$ as suggested by Pal and Bezdek [190];

- Note V, i.e.,classic Davies-Bouldin index (DB) was computed for the *pure k-means* and *defuzzified c-means*; Soft Davies-Bouldin index (SDB) was computed for *pure c-means*;

- Note VI, i.e., for the three configurations

  1. Absolute Index Minimization (AIM),
  2. Relative Index Minimization (RIM), and
  3. Measures in Relation to Clusters (MRC),

  we computed two-way ANOVA statistical test in order to estimate the effects of the factors on the separation values and the significance of the difference observed (with a significance level = 0.05, i.e. we strongly reject the null hypothesis of the test if p-value < 0.05);

- Note VII, i.e., some hardware info: processor Intel(R) Xeon(R); CPU E3-1505M v5 @ 2.80GHz, 4 Core(s), 8 Logical Processor(s).

Results

We present results in Table 4.10, Table 4.11, Table 4.12 and Figures. In particular,

- Table 4.10 shows the results about the Absolute Indexes Minimization (AIM),
- Table 4.11 shows the results about the Relative Indexes Minimization (RIM), and
- Table 4.12 shows results about Measures in Relation to Clusters (MRC).

Note that:

- in the Table 4.10 and Table 4.11:
  - the four columns are, in order, *Datasets* (the eight benchmark datasets with known ground-truth clusters and the four experimental fMRI dataset without known ground-truth), *DB k-means* (the DB index computed with pure k-means results), *DB c-means** (the DB index computed with defuzzified c-mean results); and *SDB c-means* (the SDB index computed with pure c-means results); in column *Dataset*, the values within parenthesis () are the ground-truth cluster (i.e. the symbol * indicates clusters are unknown in the case of fMRI datasets), whereas for the other columns they are the optimum number of clusters according to DB and SDB values.

The principal results found were the following.

1. Results in reference to Table 4.10: within the experimental configuration *Absolute Indexes Minimization* (AIM), **in more than half datasets (7/12), the score of Soft Davies-Bouldin index was less than Davies-Bouldin index**, specially for Cancer, Grab, Glass, Ovarian and Thyroid datasets. In particular, it was the lesser also for REST and TASK datasets in the case of TOIs configuration (i.e. having as input features the Times Of Interest). Also, there is some variability to find the truth number of clusters: DB k-means was associated to the correct clusters number with proportion 4/8, the DB c-means* with proportion 5/8 and the SDB with proportion 3/8. The two-way ANOVA shows that the differences between measures types exist but they are feeble evidence (SS=0.038; df=2; MS=0.019; F=1.87; p-value=0.1772), instead the differences between datasets exist and they are significative (SS=2.40; df=11; MS=0.218; F=21.09; p-value=0.0000).
2. Results in reference to Table 4.11: within the experimental configuration *Relative Indexes Minimization* (RIM), **in half datasets (4/8)**

the score of Soft Davies-Bouldin index was less than Davies-Bouldin index, specially for Cancer, Glass, Ovarian and Thyroid datasets. The two-way ANOVA shows that the differences between measures types exist but they are feeble evidence (SS=0.037; df=2; MS=0.018; F=1.59; p-value=0.2377), instead the differences between datasets exist and they are significative (SS=0.60; df=7; MS=0.08; F=7.47; p-value=0.0008).

3. Results in reference to Table 4.12: within the experimental configuration *Measures in Relation to Clusters* (MRC), **the score of Soft Davies-Bouldin index was always less then the classic crisp version of Davies-Bouldin index** (a part the *naive* case where clusters are equal to one), keeping as comparison values the median statistics computed with all the datasets for each number of clusters. The two-way ANOVA shows that the difference between measures types are significative (SS=0.32; df=2; MS=0.16; F=15.99; p-value=0.0001) and that also the difference between clusters numbers are significative (SS=2.39; df=9; MS=0.26; F=26.29; p-value=0.0000).

Table 4.10: Absolute Indexes Minimization (AIM) results

Dataset	*DB k-means*	*DB c-means**	*SDB c-means*
Cancer (2)	0.7628 (2)	0.7630 (2)	**0.6058** (2)
Crab (2)	0.6113 (2)	0.6137 (2)	**0.5802** (8)
Glass (2)	0.9167 (6)	1.0658 (4)	**0.6723** (10)
Iris (3)	**0.4048** (2)	**0.4048** (2)	0.5904 (2)
Ovarian (2)	0.7487 (2)	0.7520 (2)	**0.5553** (10)
Simple (4)	0.4428 (4)	**0.4375** (4)	0.4637 (4)
Thyroid (3)	0.9802 (10)	1.0105 (3)	**0.7516** (3)
Wine (3)	0.4612 (7)	**0.4577** (7)	0.5524 (2)
RestROIs (*)	**0.4332 (8)**	0.4519 (7)	0.5563 (5)
RestTOIs (*)	1.2417 (2)	1.2812 (2)	**1.0565 (2)**
TaskROIs (*)	0.4427 (8)	**0.4336 (8)**	0.5618 (3)
TaskTOIs (*)	1.2432 (2)	1.2830 (2)	**1.0484 (2)**

Discussions

- On the results presented in Table 4.10 The first experimental outcome is the evidence that Soft Davies-Bouldin index (SDB) computed an optimum cluster number with less values then the classic crisp Davies-Bouldin index (DB) in more then half datasets (7/12) (e.g. in Cancer,

Table 4.11: Relative Indexes Minimization (RIM) Results

Dataset	*DB k-means*	*DB c-means**	*SDB c-means*
Cancer (2)	0.7628	0.7630	**0.6058**
Crab (2)	**0.6113**	0.6137	0.6434
Glass (2)	1.0054	1.0752	**0.6779**
Iris (3)	**0.6693**	0.6696	0.7261
Ovarian (2)	0.7487	0.7520	**0.6879**
Simple (4)	0.4428	**0.4375**	0.4637
Thyroid (3)	1.1370	1.0105	**0.7516**
Wine (3)	0.5412	**0.5394**	0.6675

Table 4.12: Separation Measures in relation to clusters (MRC) (overall datasets median values)

Clusters	*DB k-means*	*DB c-means**	*SDB c-means*
1	0.0000	0.0000	0.0000
2	0.7564	0.7575	**0.6607**
3	0.7410	0.7492	**0.7104**
4	0.8596	0.8797	**0.6889**
5	0.8572	0.9300	**0.6968**
6	0.8948	1.0860	**0.8531**
7	0.8960	1.1997	**0.8306**
8	0.9064	1.2995	**0.8407**
9	0.9492	1.2069	**0.7705**
10	0.9917	1.1550	**0.6712**

Table 4.13: Computational Time for Separation Measures

Datasets	*DB k-means [s]*	*DB c-means* [s]*	*SDB c-means [s]*
Cancer	0.0010	0.0011	**0.0054**
Crab	0.0009	0.0010	**0.0055**
Glass	0.0009	0.0010	**0.0055**
Iris	0.0009	0.0009	**0.0055**
Ovarian	0.0011	0.0012	**0.0058**
Simple	0.0010	0.0011	**0.0055**
Thyroid	0.0026	0.0034	**0.0157**
Wine	0.0008	0.0010	**0.0068**
Rest (ROIs)	0.0008	0.0012	**0.0059**
Rest (TOIs)	0.0008	0.0012	**0.0051**
Task (ROIs)	0.0008	0.0012	**0.0062**
Task (TOIs)	0.0008	0.0013	**0.0060**

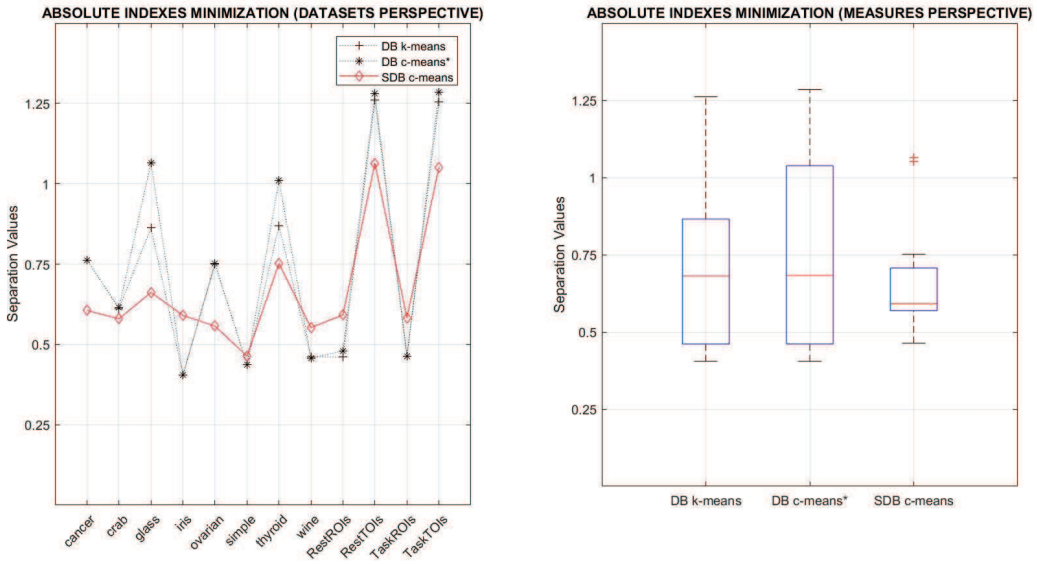

Figure 4.6: The Figure shows the plots and the boxplots associated the Absolute Indexes Minimization (AIM), i.e. the computation of the optimal separation measures for each datasets using all range of clusters

Crab, Glass, Ovarian, Thyroid, RestTOIs and TaskTOIs; in particular, it is reasonable that SDB found a better clustering in Rest and Task datasets within TOIs configuration as input features, because clustering the Times of Interest (TOIs) then Regions of Interest (ROIs) is quite trickly (due the high functional variability of the brain areas, specially using as experimental paradigm the ON/OFF block design); therefore, taking account the all membership values helped to discover better the spatial pattern associated to each time points in the fMRI scans. The second experimental outcome is the evidence that both DB and SDB indexes failed with three of eight datasets to find the correct number of clusters using both crisp clustering and soft clustering (e.g. Glass, Iris and Wine); this could be explained by the type of input features that sometimes are not linearly separable (e.g. Iris dataset), and, as a consequence, it could be not so-easy to find well separable clusters; this explanation is also confirmed with the opposite evidence that very well separable datasets, such as Cancer and Simple, were those both the indexes found the truth number of clusters.

- On the results presented in Table 4.11 The general experimental out-

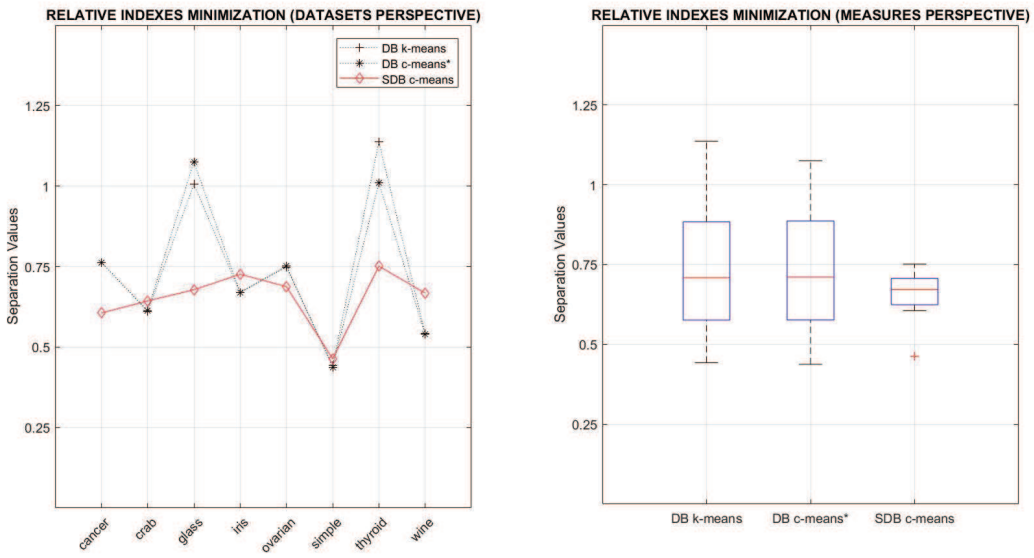

Figure 4.7: The Figure shows the plots and the boxplots associated the Relative Indexes Minimization (RIM), i.e. the computation of the separation measures only with the datasets that have ground-truth in reference of the known number of clusters.

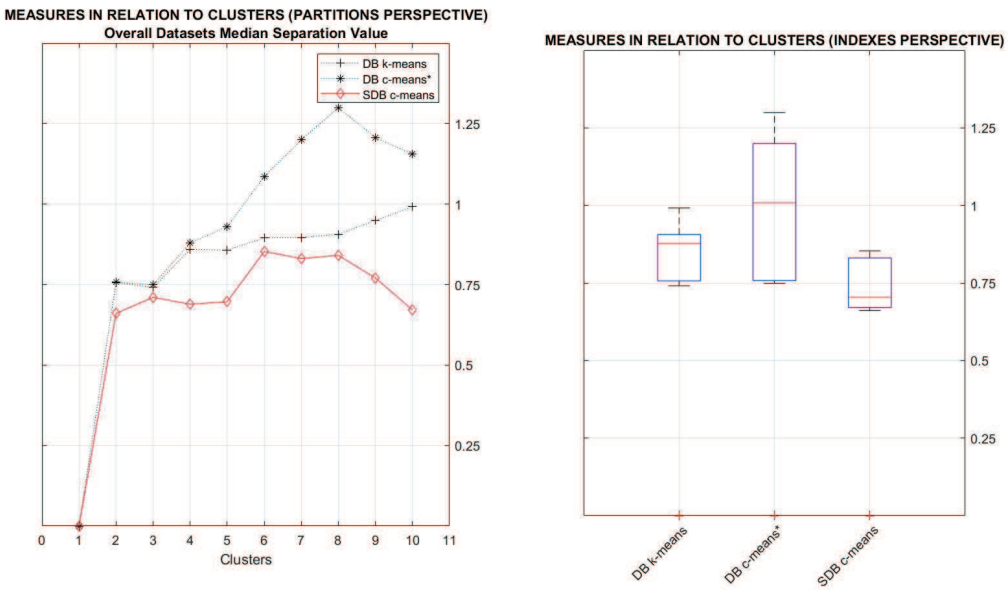

Figure 4.8: The Figure shows the plots and the boxplots associated the Measures in Relation to Clusters (MRC), i.e., the computation of the all-datasets central tendency of the optimal separation measures along the complete range of clusters.

comes are evidences that Soft Davies-Bouldin index (SDB) worked better then classic crisp Davies-Bouldin index (DB) in relation to the ground-truth clusters, specially for the Cancer, Glass, Ovarian and Thyroid datasets; in particular, the outcomes related the cases of Glass and Thyroid datasets showed that the differences between DB index *versus* SDB index are so huge and with SDB as best clustering behavior (see Table 4.11); instead, the increasing of the clustering quality in the half datasets obtained with Soft Davies-Bouldin index is motivated by the use of membership values that allow to partition data in clusters with more precision, taking account the multi-class membership peculiarity of each element clustered (i.e. the *softness* properties). This argument is also valid for the results presented in Table 4.10. There are also two outcomes where classic crisp Davies-Bouldin index (DB) had nice performances (e.g. Iris and Wine datasets); but, there are other two cases were both the crisp and the soft indexes had comparable results (e.g. Crab and Simple); one possible explanation about these four results is that sometimes soft clustering algorithms are not an optimal *a priori* choice to partition the data, i.e. this is a natural extension from the *a priori* assumption that datasets have clusters (see the Intruduction where are exposed the three main problems in clutering for Bezdek *et al*), and it is also the motivation to try both the type of clustering algorithms when it is possible, in order to compare with separation measures and visual inspection the different quality of results found.

- On the results presented in Table 4.12 The main result obtained comparing the validation values in relation to the cluster range is that Soft Davies-Bouldin index (SDB) outperformed the classical crisp version of the DB measure. Table 4.12 show that using all the possible partitions - from 1 to 10 - , SDB c-means had always a values less then the values of both DB k-means and DB c-means*. This evidence highlights the potentialities of the soft version of Davies-Bouldin index: i.e. taking the central tendency of the all datasets in relation to the range of possible partitions, it has been indicated a more extended behaviour of the SDB, that is a more accurate perspective to observe the overall performance of a clustering algorithms. Furthermore, this approach is useful to find not only optimality but also *sub-optimality* in the validation indexes, adding a visual procedure (with grounded subjective expertise) to check how the separation values computed appear globally along the complete range of clusters considered.

- On the two-way ANOVA results Statistical tests stated that the differences between the separation measures exist, but in the case of Absolute Index Minimization and Relative Index Minimization configurations are thin due to the p-values computed, i.e. 0.1772 and 0.2377 respectively (see *Results* section); therefore, the null hypotheses have to be accepted having as confidence level the value 0.05. From a scientific point of view, this feeble evidences could be address as an hints to deepen the capabilities of Soft Davies-Bouldin index (SDB), in the way to investigate if the differences will be replicated with other experiments. This positive thinking is motivated by the results stated in the Table 4.12 that indicated an outstanding behavior of the SDB in relation to all the range of cluster numbers, taking account the central tendency of all the datasets, with a strongly significance results with p-value 0.0001. Hence, this mixture of observations is oriented in an advantage on the usage of the soft version of the classical Davies-Bouldin index, taking also account some awareness to be managed for next measures validations.

- On the computational time of measures There are the following observations about the computational time: Table 4.13 shows the times used by the experimental hardware to compute DB k-means, DB c-means* and SDB c-means for each of the eight datasets (we took the average time for all the computations). We did not use an explicit $\mathcal{O}$ notation to describe the complexity of the clustering algorithms adopted; we observed their behavior measuring the computational load in terms of internal time used by machine to execute the functions. We found that the times of DB k-means and DB c-means* were quite similar, whereas the computation of SDB c-means was more heavy then the other two indexes. In the case of SDB c-means, we considered that the major computational time by the machine is motivated by the use of membership values to solve the equations of the measure, that is the main difference with the equations for the classic crisp Davies-Bouldin index.

Conclusion and Future Works

We presented a soft clusters separation measure to validate fuzzy clustering results without defuzzyficaton. We made a generalization of Davies-Bouldin index (DB) for crisp clustering in the soft clustering domain and we named this new measure Soft Davies-Bouldin index (SDB). We compared DB and SDB with the outcomes of k-means and c-means algorithms using benchmark

datasets with known ground-truth and two datasets from fMRI repositories with unknown ground-truth. We found that i) within the experimental configuration Absolute Indexes Minimization (AIM), in more than half datasets, the score of Soft Davies-Bouldin index was less than Davies-Bouldin index, ii) within the experimental configuration Relative Indexes Minimization (RIM), in half datasets the score of Soft Davies-Bouldin index was less than Davies-Bouldin index, iii) within the experimental configuration Measures in Relation to Clusters (MRC), the score of Soft Davies-Bouldin index outperformed the classic Davies-Bouldin index.

The main objective of this study was to compare the potentialities of the separation measures belonged to Davies-Bouldin index family: then, it is a comparison between Soft *versus* Crisp version of the same clustering validation index. An expansion of this research is a comparison with other indexes that want to handle the *softness* of fuzzy clustering: e.g. validation measures like C-index [19] and PBMF index [189], or other measures proposed by Wu *et al* [274], Zhang *et al* [282] and Hullermeier *et al* [108] (see also Table 2.1).

Nevertheless, it is remarkable to continue the investigation about the data driven motivations on doing soft partitioning instead to use the crisp clustering. If clustering algorithms find the *natural groupings* of data, as Duda and Hart have defined [207], then it is also useful to investigate if the natural groupings of data should be fuzzy or should be crisp.

RS-fMRI Analysis Using Unsupervised Learning Algorithms

Premise

This research is an extended version of a work presented at a Joint Congress of the 15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering and 3rd Conference on Imagin and Visualization held in Lisbon (Portogal) on March 2018.

More in detail, this work is the extended version of the conference paper titled Resting State fMRI Functional Connectivity Analysis Using Soft Competitive Learning Algorithms [252] 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.

Specially, the main topic regard this work is RS-fMRI data analysis for functional connectivity explorations, that is a challenging topic in computational neuroimaging. Several approaches have been investigated to discover whole-brain data features. Among these, clustering techniques based on Competitive Learning (CL) and Spectral Methods (SM) have been shown effective in providing useful information in various contexts. We selected three clustering algorithms and two spectral methods, i.e the clustering algortihm are Self-Organizing Maps (SOM), Neural Gas (NG) and Growing Neural Gas (GNG), whereas the spectral methods are the classic Principal Component Analysis (PCA) and the Nonliner Robust Fuzzy Principal Component Analysis (NRFPCA). We validated clustering with Davies-Bouldin Index (DBI) and we selected informative principal components using Random Matrix Theory (RMT). tools. We adopted these techniques to study the intrinsic functional properties of images coming from a shared repository of resting state fMRI experiments (1000 Functional Connectome Project).

The key-word associated to this work are: RS-fMRI; Functional Connectivity; Competitive Clustering; Self Organizing Map; Neural Gas; Growing Neural Gas; Davies-Bouldin Index; Spectral Methods; Principal Component Analysis; Nonlinear Robust Fuzzy Principal Component Analysis; Random Matrix Theory

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

[25], 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 is 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. Approaching the general fMRI signal processing with unsupervised methods is a challenging application with active fMRI paradigm or task-oriented [132, 140, 252, 20] as well as with the passive paradigm or resting state in fMRI [163, 248, 136, 261]. The specific intent of this work is to use unsupervised techniques applied to passive fMRI paradigm. For this purpose, we adopted as clustering methods the competitive learning algorithms to explore the natural partitions of the data [57], and as spectral methods the classic approach often used in fMRI literature [227, 257] and the ones based on fuzzy approach [157]. To find the functional clusters, we choose Self Organizing Map (SOM), Neural Gas (NG) and Growing Neural Gas (NNG), that are a soft class of unsupervised artificial neural networks. SOM are models initially proposed by Kohonen [123] 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 organized 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 [52]. 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 [77, 81]. 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 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 neighbors if they are connected by an edge, and the neighborhood 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 represents the position in the node space through a vector. 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 that these clustering algorithms have, we want to understand with more details the features of the elements partitioned. In other word, we decided to complement the experimental 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. Furthermore, we investigated the latent information of fMRI covariance matrix using spectral methods: Principal Component Analysis (PCA) [193] and Nonlinear Robust Fuzzy Principal Component Analysis (NRFPCA) [157]. Usually, PCA techniques are used for data explorations and/or dimensionality reduction. Technically, PCA finds the orthogonal transformation that transforms a multivariate system of correlated random variables to new coordinates that are linearly uncorrelated, i.e., Principal Components or eigenvectors. The eigenvalues are the variance of these components. The grater the eigenvalue, the more is informative the eigenvector associated. Therefore, the selection of the eigenvectors is a problem in PCA application. To solve this task, we adopted Random Matrix Theory tools [172]. Specially, we used the support established by the Marchenko-Pastur distribution of eigenvalues for Gaussian variables. If the eigenvalues are within that support, then they are referred to eigenvectors associated to uncorrelated variables or false correlated variables [35]. The usage of these spectral methods allows to extract functional information contained in the covariance matrix to find brain networks associated to the resting state fMRI physiology. Before clustering of the fMRI time series and performing spectral analysis of the covariance matrix, 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) [23, 24, 202, 203, 247] (see also the works by Van Den Heuvel [248] where he shown alternative to DMN and by Iraji [109] 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, clustering techniques based on competitive learning and spectral methods with features selections based on random matrix theory. In the next sections, we propose the type of data we have selected from a repository, then we present methods, results, discussion and conclusions.

Data

Within the NITRC repository (`https://www.nitrc.org/`) (Kennedy, et al. 2016) 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 subjects did a resting state experimental paradigm with eyes open. 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 to discover resting state functional properties and their gender determinants. Therefore, our approach is also a confirmatory data analysis.

Methods

The methods we used followed this pipeline: i) image processing, ii) data reduction, iii) statistical analysis and algebraic measurements, iv) functional connectivity investigations with linear correlation, v) clustering with competitive learning algorithms and vi) latent analysis of the fMRI covariance matrices with spectral methods.

Image Processing Image preprocessing was done with the functions for resting state image analysis contained in the software FSL [113]: we did spatial filtering with 3 [mm] of smoothing, frequency filtering with a high pass filter having 1/100 [Hz] as cut-off 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 [69, 160, 87].

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 of having two macro-signals, one for Females and one for Males.

Descriptive Statistics 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) [104].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 have 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.

Linear Correlations To study the brain functional connectivity, we used the Pearson's linear Correlation Coefficient (CC) [104] 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 Analysis To investigate brain resting functionality using competitive learning algorithms, we used Self Organizing Map (SOM) [123], Growing Neural Gas (GNG) [81] and Neural Gas (NG) [52], i.e., to set GNG, we adopted 10 nodes, 1000 iterations, $\lambda = 2$, $\epsilon_b = 0.0005$, $\epsilon_n = 0.00001$, $\alpha = 0.05$, $\delta = 0.995$, age-node = 60. To set NG, we used 2 nodes, 500 iterations, $t_{max} = 8000$, $\epsilon_i = 0.90$, $\epsilon_f = 0.50$, $\lambda_i = 10$, $\lambda_f = 1$, $T_i = 5$, $T_f = 10$. To evaluates the optimal partitions with the three clustering methods,

we adopted Davies-Bouldin separation measure [46] and we compared the selected partitions with Jaccard similarity measures [110].

The formalism about the clustering analysis is presented in the Computational Intelligence section. In particular the SOM, NG and GNG are exposed in Competitive Learning Algorithms section 2.2.3.

Spectral Analysis We applied the classic Principal Component Analysis (PCA) [193, 106, 114] and the Nonlinear-Fuzzy-Robust Principal Component Analysis (NFRPCA) [157, 277, 276] to study the collective activity of the brain, i.e., the exploration of the eigenvalues and eigenvectors of the so-called connectivity domain. To select eigenvectors associated to non-random correlations, we adopted Random Matrix Theory (RMT) tools [172]. The eigenvalues density of the empirical correlation for uncorrelated independent identically distributed (i.i.d.) Gaussian variables follows the Marchenko-Pastur distribution (Marchenko e Pastur 1967), that locates eigenvalues with a finite support [241], i.e., for the dataset we used, the RMT top level is 3.0299 and RMT bottom level is 0.0673. Eigenvalues greater then RMT top level are associated to eigenvectors linked to non-random correlations. We also compared the features selection based on RMT with the Kaiser criteria [117, 118] that consider informative eigenvalues grater then 1.0. After the selection of the informative eigenvectors, we computed the prototype Females 1st eigenvector and the prototype Males 1st eigenvectors with the bootstrap statistical procedure based on the median index. To estimate the collective behaviour of the ROIs for each eigenvector, we adopted the solution proposed by Burda et al (Burda, et al. 2013) based on localization theory [4], that is a Participation Ratio (PR) that gives the quantification about how many ROIs are globally involved for each eigenvector.

The formalism about the spectral analysis is presented in the Computational Intelligence section. In particular PCA, NFRPCA and Random Matrix Theory are exposed in Spectral Analysis section 2.2.5.

Results

The results we obtained regard i) the statistical descriptions of resting state fMRI data, ii) the algebraic measures of fMRI in vectorial forms, iii) the functional connectivity studied with the Pearson linear coefficient correlation in both the conditions (all-ROIs vs all-ROIs and Seeds vs all-ROIs), iv) the general outcomes of the three clustering techniques with their comparison with the classical statistical signal processing approach and v) the results from the application of random matrix theory tools to spectral analysis methodologies.

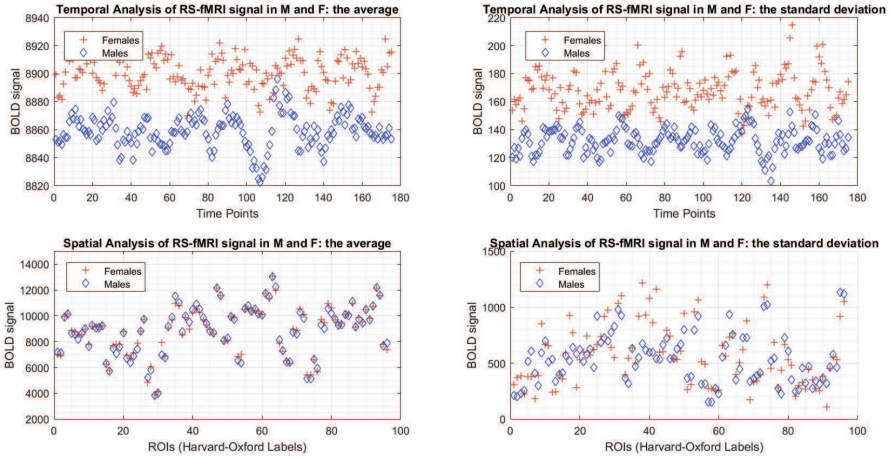

Figure 4.9: Plots of the average and the standard deviation of BOLD signals in Males and Females in the temporal (175 time-points versus BOLD signals) and spatial domains (96 ROIs versus BOLD signals). Both the average and the standard deviation of BOLD signals in the temporal domain are quite different 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

Results related to Figure 4.9 The Figure 4.9 shows temporal and spatial results of RS-fMRI signals. The temporal analysis of the exams indicated that Females had the higher values for both the average and the standard deviation of the BOLD signals. The spatial analysis revealed that Females and Males were similar for the average signals, but Females had more standard deviation. 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 (see also box-plots in Figure 4.10).

Algebraic measures of fMRI data

Results related to Figure 4.10 In the Figure there are box-plots of whole brain signals in Males (1) and Females (2), showing the between gender

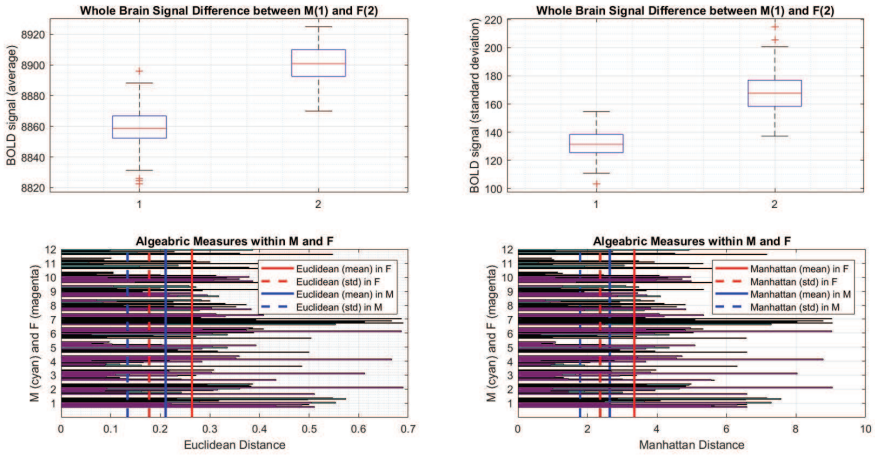

Figure 4.10: In the top part of the figure, there are the boxplots of whole brain signals in Males (1) and Females (2), showing the between gender statistical difference of the average (left) and the standard deviation (right) about the whole brain signals: in the both cases, Females have higher values 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.

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 4.10, 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. 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.

Clustering validations and outcomes comparison

Results related to Figure 4.11 and Figure 4.12 The Figure 4.11 shows that the optimum Davies-Bouldin index associated to SOM, NG and GNG algorithms, for both Females and Males. The separation index indicated all the clustering methods reached the best data partitions with two clusters. The lesser value is related to the fourth Males subject and the higher

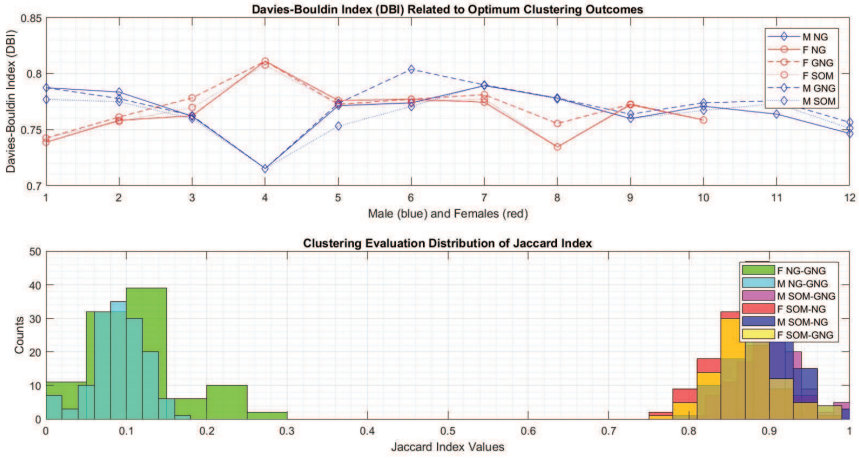

Figure 4.11: The top diagram shows the optimum Davies-Bouldin index associated to SOM, NG and GNG algorithms, for both Females and Males: all the DB are referred to 2 clusters as optimal clusters number for each algorithm; all the best indexes are under 1; the lesser is related to the fourth Males subject, 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 ($Jaccard \leq 0.3$), and the other clustering combinations are very different ($Jaccard \geq 0.7$).

value to the fourth Female subject. In the bottom of the figure there is the discrete distributions 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), whereas in the other clustering combinations there are some clustering dissimilarities (Jaccard ¿ 0.7).

The Figure 4.12 shows the scatter plots with the clustering outcomes for NG, GNG and SOM in Females and Males. The main result is that, using two clusters as optimum partitions number, the clustering detected the amplitude information of BOLD signals in Females and Males, i.e. clusters differentiate low levels and high levels in BOLD signals.

Functional Connectivity with Linear Correlation

Results related to Figure 4.13 and Figure 4.14 In the Figure 4.13 there are evidences that Pearson's linear Correlation Coefficient (CC)

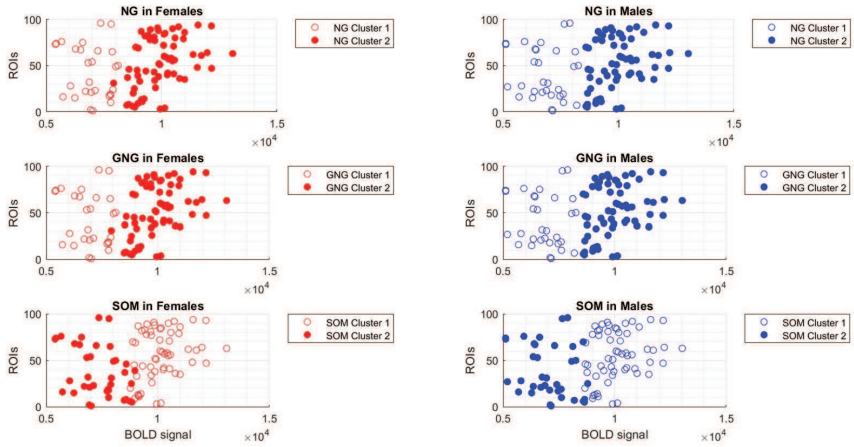

Figure 4.12: The scatter plots show the clustering outcomes for NG, GNG and SOM in Females (red points) and Males (blue points): the main result is that, using two clusters as optimum partitions number, the clusters detect the amplitude information of BOLD signals in Females and Males, i.e., clusters differentiate low levels and high levels in BOLD signals.

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. 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. 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) (see Figure 4.14 with the seed-based correlation matrices in all subjects).

Functional Connectivity with Competitive Clustering

Results related to Figure 4.15 The plots in figure 4.15 represent the organization of the two optimal clusters in Males and Females in relation to the seeds based (L/R Precuneus) Correlation Coefficient. They

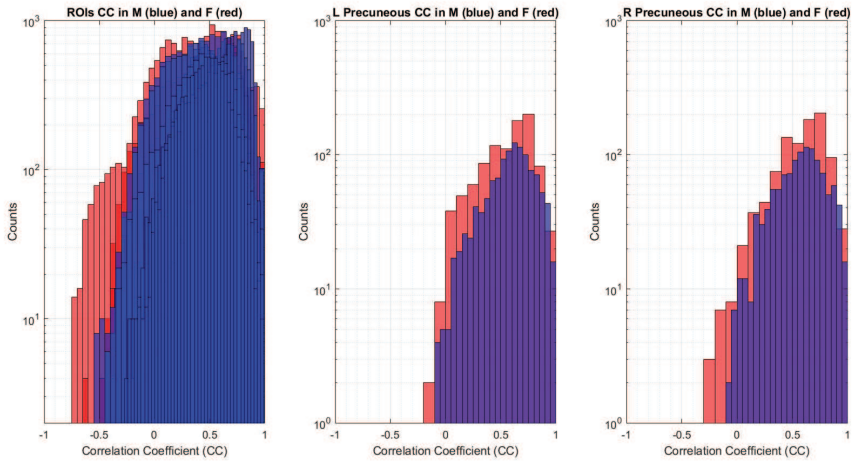

Figure 4.13: In the figure there are the Correlation Coefficients (CC) distributions between all-ROIs versus all-ROIs in Males and Females, and the correlation coefficients distributions between seeds (L/R Precuneus) versus all-ROIs. In the first discrete distribution, there are present positive and negative correlations, both in Males and Females; whereas, in the seeds based correlations distributions, there are less anti-correlations, with a bit gender difference.

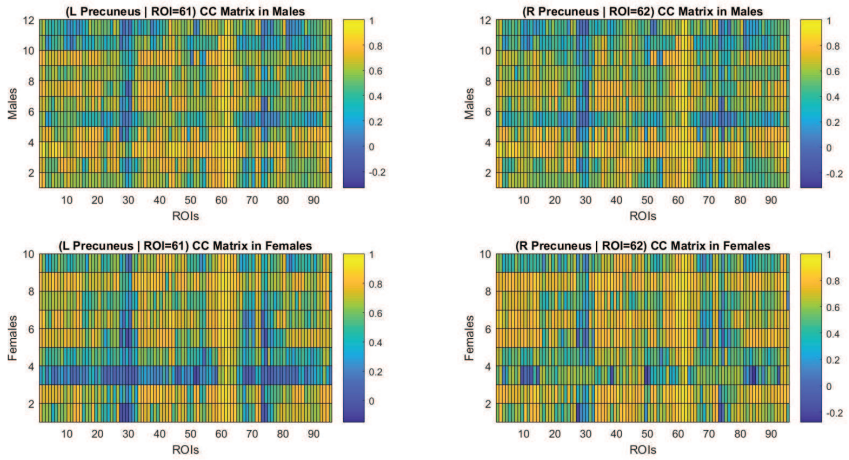

Figure 4.14: In the figure there are the seed-based correlation matrix between Left and Right Precuneus in Males and Females: it is evident the strong correlation with many brain regions, as well as weak correlations with few brain regions. The highest positive correlations (CC ¿ 0.8) are with seed-controlateral 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).

report the evidence that low and high correlations belong to different clusters, with some overimposition between them for the central values. There are evidences about the relation between ROIs, correlation coefficients and clusters, showing precisely that low and high correlations are within different clusters in Males and Females.

Results related to Figure 4.16 The Figure 4.16 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 organization within the two clusters. 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 than the other algorithms).

In the left side of the Figure 4.16, the brain is represented with the center-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 colored circles are the height selected regions that are, respectively, the four with the higher correlations (hot colors), and the four with the lower correlation (cool colors). 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. 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 (see Figure 4.15 and Figure 4.16). We also noticed some overimposition for the central values of the coefficients.

Functional Connectivity with Classic PCA and Nonlinear Robust Fuzzy PCA

Results related to Figure 4.17 The Figure 4.17 shows the ordered eigenvalues from the most to the lesser important for both Females and Males. The salience of eigenvalues is determined by the cut-off based on Random Matrix Theory (RMT), i.e., the support of eigenvalues associated to Gaussian variables established by the Marchenko-Pastur distribution. The RMT top level and the RMT bottom level defined the support (3.0299 and 0.0673). Eigenvalues greater the RMT top level (3.0299) are associated to non-random correlations between ROIs, i.e., informative brain functional connectivity. The eigenvalues under the top level are associated to random correlations. The eigenvalues that are in between the support limited by the RMT top

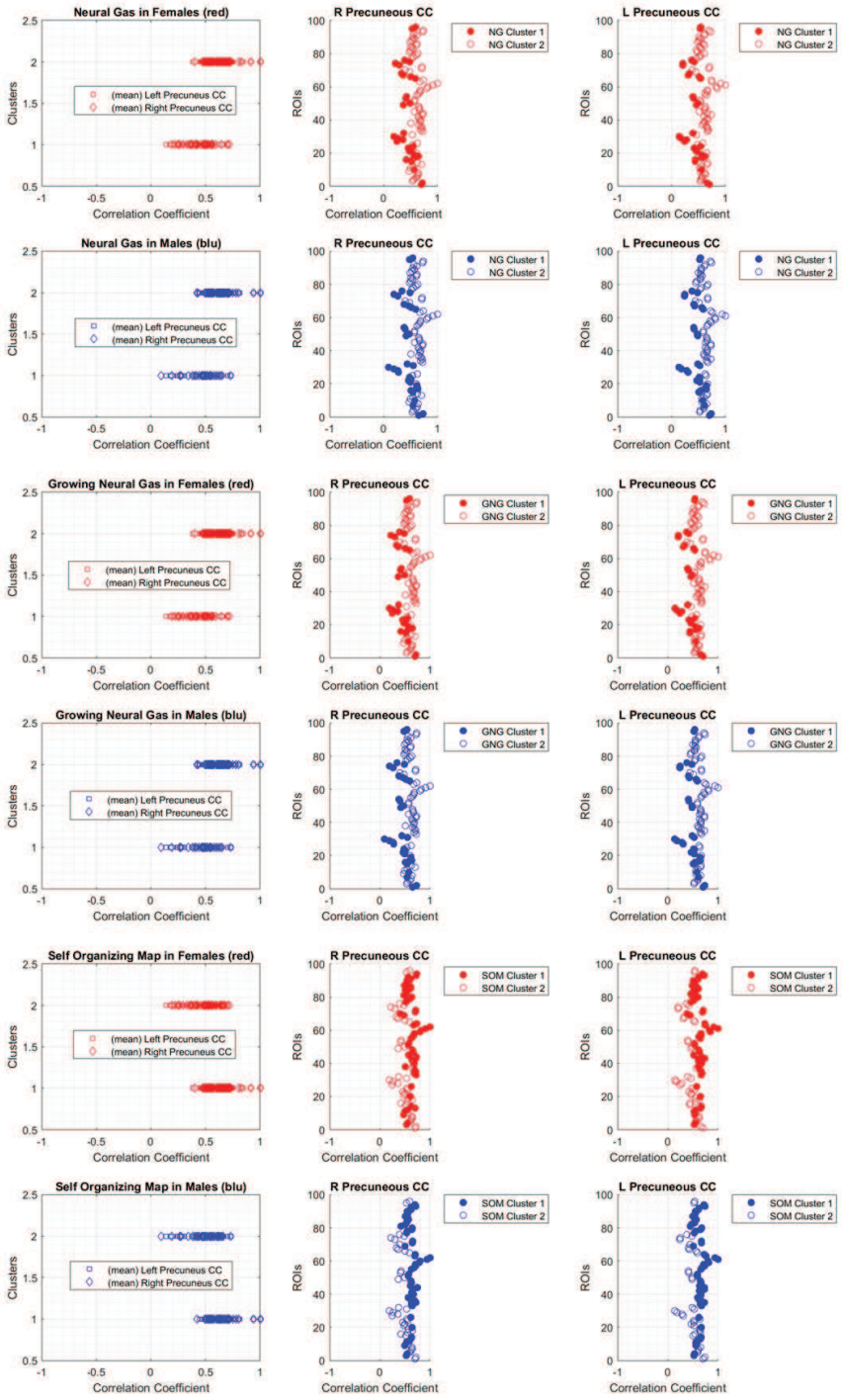

Figure 4.15: These figures represent the organization of the two optimal clusters in Males and Females in relation to the seeds based (L/R Precuneus) Correlation Coefficient. In the left column, it is reported the evidence that low and high correlations belong to different clusters, with some overimposition 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.

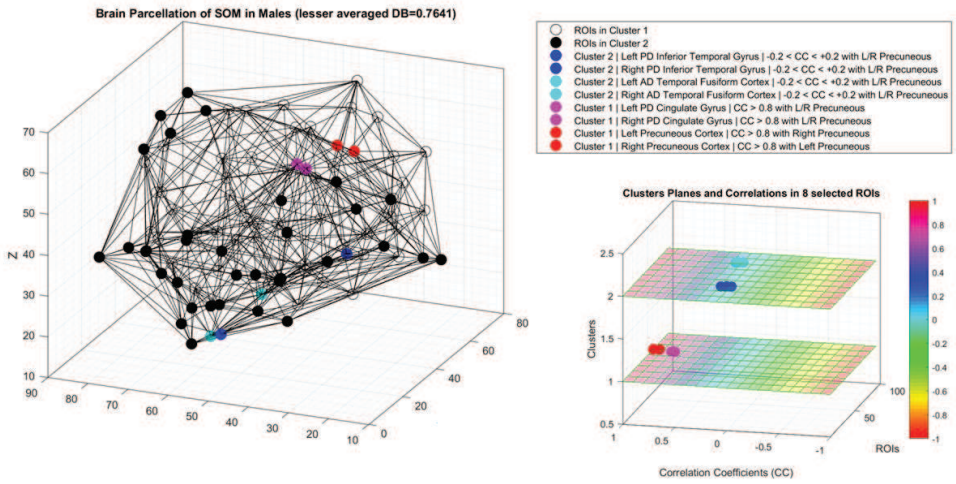

Figure 4.16: The image shows result of SOM algorithm in the Males case, with a specification about the higher and the lower correlations with both the seeds (Left and Right Precuneus), and their organization within the two clusters. In the left side of the Figure, the brain is represented with the centre-of-mass 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 eight 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.

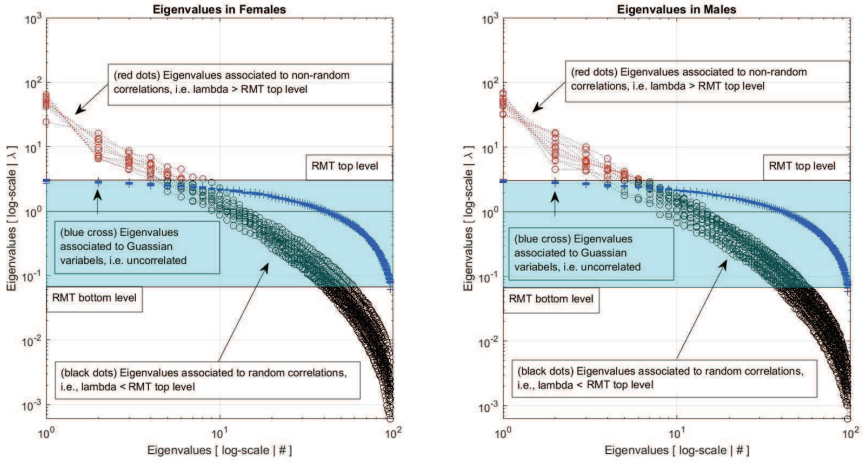

Figure 4.17: The Figure shows the ordered eigenvalues from the most to the lesser important for both Females and Males. The salience of eigenvalues is determined by the cut-off based on Random Matrix Theory (RMT), i.e., the support of eigenvalues associated to Gaussian variables established by the Marchenko-Pastur distribution. The RMT top level and the RMT bottom level defined the support (3.0299 and 0.0673). Eigenvalues greater the RMT top level are associated to non-random correlations between ROIs, i.e., informative brain functional connectivity (red dots). The eigenvalues under the top level are associated to random correlations (black dots). The eigenvalues that are in between the support limited by the RMT top level and the RMT bottom level (cyan coloured area) are the ones associated to truth uncorrelated variables (blue cross).

level and the RMT bottom level are the ones associated to truth uncorrelated variables, used to proof the efficacy of the RMT criteria to select truth correlation or false correlations from a fMRI data.

Results related to Figure 4.18 The Figure 4.18 shows the comparison between features (components) selection methods in both Females and Males. The picture shows that the RMT criteria to select components based on their eigenvalues is more conservative than the Kaiser criteria. The RMT method is based on the eigenvalues support defined by the Marchenko-Pastur distribution of eigenvalues of Gaussian variables, whereas the Kaiser method classifies as informative the eigenvalues greater than 1.0. Females and Males have comparable informative eigenvectors based on RMT criteria (approxi-

mately the first five eigenvectors are informative in both Females and Males.

Results related to Figure 4.19 The Figure 4.19 shows the participation ratio (PR) for each eigenvector, i.e., how much all the ROIs are involved to determine the coefficients of eigenvectors. There are few gender and methods differences because both trends are quite comparable. It is notable that the first eigenvector has 80 of PR, instead the second around the 40: it means that the first eigenvector has information about the 80 of the total ROIs, whereas the second one around the 40.

Results related to Figure 4.20 The Figure 4.20 shows the loadings of the 1st bootstrapped eigenvectors in Females and Males computed with PCA and NFRPCA. The images are globally similar, but Females have some positive loadings greater than Males, whereas both have specific negative values associated the same ROIs. The difference between spectral methods are less evident than the ones between subjects (see Figure 4.21, Figure 4.22 and Figure 4.23 for more details).

Results related to Figure 4.21 The figure shows the differences between spectral methods in relation to gender and the differences between gender in relation to spectral methods. The left image shows that the within subject subtraction between PCA loadings and NFRPCA loadings gives few differences, i.e., both spectral methods give comparable results within gender. The right image shows, instead, that the within method subtraction between Females and Males loadings gives quite differences, i.e., even if there are comparable trends for PCA and NFRPCA differences, they range in three loading zones (some between -0.25 and 0.25, some over 0.25 and some under -0.25). Then, it means that Females and Males had different loadings according to both the spectral methods.

Results related to Figure 4.22 The Figure 4.22 shows the between methods loadings subtraction in Females and Males. The results are the same presented in the Figure 13. The brain templates of Females and Males show with hot colours the ROIs that have more loading with PCA than NFRPCA and with cool colours the ROIs that have more loading with NFRPCA then PCA. Even if there are visible differences, they are very small, according to their decimal range (see also Figure 4.21, left part; the values are within the range -0.25 and +0.25).

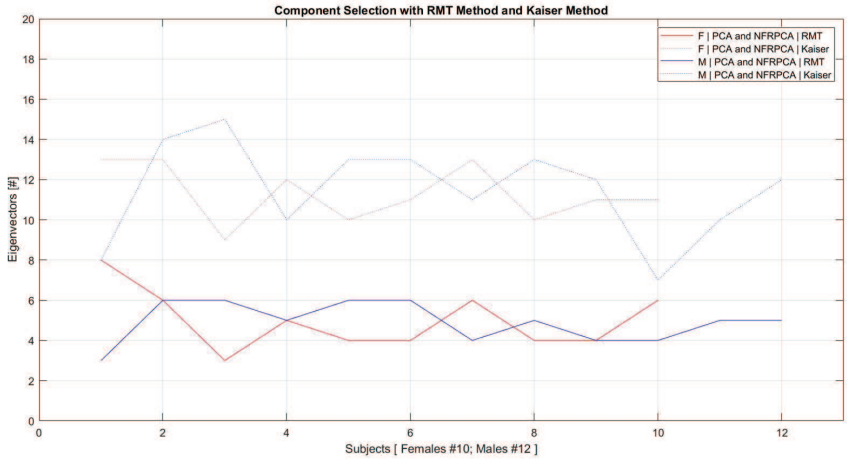

Figure 4.18: The figure shows the comparison between component selection methods in both Females and Males (left) and the gender and methods differences of the participation ratio (PR) computed for each eigenvectors (right). The left picture shows that the RMT criteria to select components based on their eigenvalues is more conservative than the Kaiser criteria. The RMT method is based on the eigenvalues support defined by the Marchenko-Pastur distribution of eigenvalues of Gaussian variables (that has support 3.029 and 0.0673 with the fMRI dataset we used), whereas the Kaiser method classifies as informative the eigenvalues greater than 1.0. Females and Males have comparable informative eigenvectors based on RMT criteria (approximately the first five eigenvectors are informative in both Females and Males). Those results show that there is in average five informative eigenvectors in both Females and Males and they are similar to results proposed in literature by Van Den Heuvel, i.e., there is convergence between several studies to a limited subsets of specific resting state networks: there are the so-called Default Mode Network (DMN), the primary motor network, the insular-temporal/ACC network, the left/right parieto-frontal network, and the frontal network.

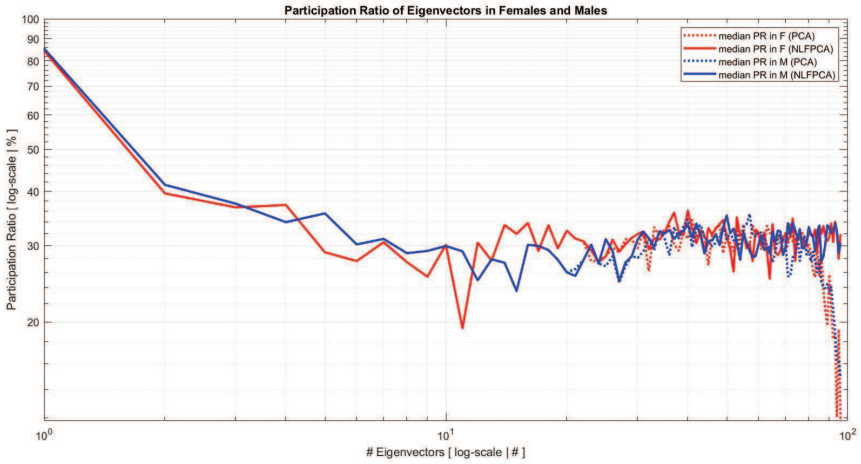

Figure 4.19: The figure shows the participation ratio (PR) for each eigenvector, i.e., how much all the ROIs are involved to determine the coefficients of eigenvectors. There are few gender and methods differences because both trends are quite comparable. It is notable that the first eigenvector has 80 of PR, instead the second around the 40: it means that the first eigenvector has information about the 80 of the total ROIs, whereas the second one around the 40.

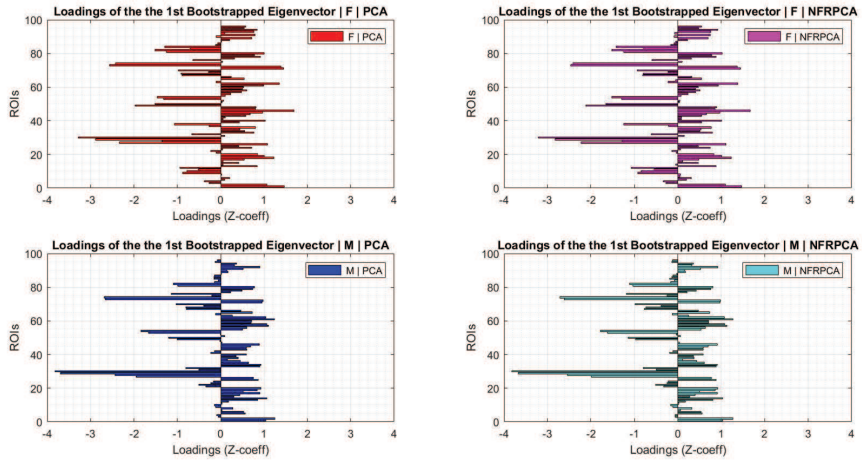

Figure 4.20: The figure shows the loadings of the 1st bootstrapped eigenvector in Females and Males because of PCA and NFRPCA. The images are globally similar, but Females have some positive loadings greater than Males, whereas both have specific negative values associated the same ROIs. The difference between spectral methods are less evident than the ones between subjects (see Figure 13, Figure 14 and Figure 15 for more details).

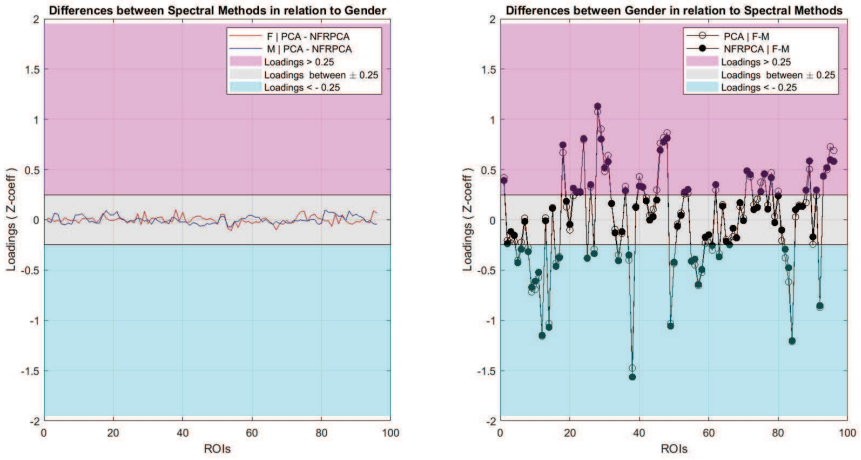

Figure 4.21: The figures show the differences between spectral methods in relation to gender (left) and the differences between gender in relation to spectral methods (right). The left images show that the within subject subtraction between PCA loadings and NFRPCA loadings gives few differences, i.e. both the spectral methods give comparable results within gender. The right image show, instead, that the within method subtraction between Females and Males loadings give quite differences, i.e. even if there are comparable trends for PCA and NFRPCA differences (empty and filled dots), they range in three loading zones (some between -0.25 and 0.25, grey area; some over 0.25 pink area; some under -0.25; cyan area).

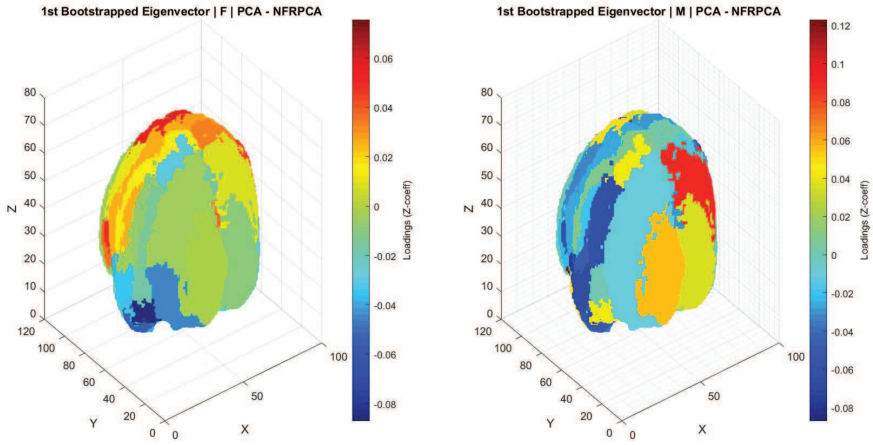

Figure 4.22: Figure shows the between methods loadings subtraction with Females (left) and Males (right). The results are the same presented in the Figure 4.19. In the two upper images, the brain templates of Females and Males show with hot colours the ROIs that have more loading with PCA than NFRPCA and with cool colours the ROIs that have more loading with NFRPCA then PCA. Even if there are a lot of visible differences, they are so thin in consideration to the decimal range of the loading coefficients (see. also Figure 4.21, left part).

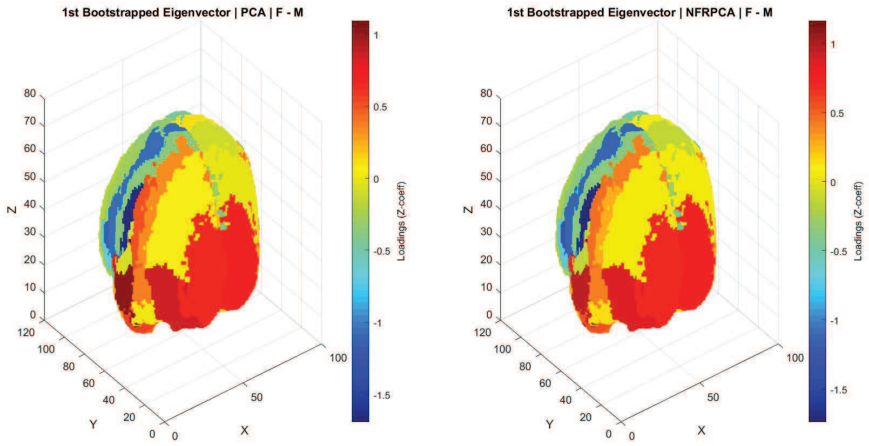

Figure 4.23: The figure shows the between gender loadings subtraction with PCA and NFRPCA. the between gender loadings subtraction with PCA (left) and NFRPCA (right). The brain templates show the difference between gender using the PCA or NFRPCA methods. ROIs that have hot colours have more loading in Females and ROIs that have cool colours have more loadings in Males. Therefore, there are loadings difference between gender and both PCA and NFRPCA computed comparable results (see Figure 4.21, right part).

Results related to Figure 4.23 The Figure 4.23 shows the brain templates of the between gender loadings subtraction with PCA and NFRPCA methods. ROIs that have hot colours have more loading in Females and ROIs that have cool colours have more loadings in Males. In the case of PCA the hot ROIs were 30# of 96#, the cool ROIs were 28# of 96# and equivalent ROIs were 58# of 96#. With the NFRPCA the hot ROIs were 29# of 96#, the cool ROIs were 27# of 96# and equivalent ROIs were 56# of 96#. Therefore, there are between gender loadings differences in both PCA and NFRPCA (see Figure 4.21, right part; the values have huge spreads).

ROIs description of the 1st bootstrapped eigenvector in Females and Males The analysis of the loadings in Females and Males confirmed that there is a gender difference in the resting state network associated to the 1st bootstrapped eigenvector (the one with 80 of Participation Ratio; see Figure 4.19). According to both PCA and NFRPCA, there are ROIs that assumed comparable loadings and others that are more specific to Females

or Males.

More in details, ROIs with similar loadings in Females and Males belong to the temporal, the frontal and occipital regions (see green-yellow ROIs in Figure 4.23):

- temporal regions, i.e., Left Planum Temporale, Right Heschl's Gyrus, Left Planum Polare, Left Temporal Occipital Fusiform Cortex, Left Temporal Fusiform Cortex, posterior division, Right Temporal Fusiform Cortex, anterior division, Left Temporal Fusiform Cortex, anterior division, Right Parahippocampal Gyrus, posterior division, Left Parahippocampal Gyrus, posterior division, Right Parahippocampal Gyrus, anterior division, Left Parahippocampal Gyrus, anterior division, Right Inferior Temporal Gyrus, temporooccipital part, Right Middle Temporal Gyrus, anterior division, Left Middle Temporal Gyrus, anterior division, Left Temporal Pole, Right Superior Temporal Gyrus, posterior division, Left Superior Temporal Gyrus, posterior division).

- frontal regions, i.e., Right Frontal Operculum Cortex, Right Frontal Pole, Left Frontal Operculum Cortex, Right Frontal Orbital Cortex, Left Frontal Orbital Cortex, Left Postcentral Gyrus, Left Middle Frontal Gyrus, Left Precentral Gyrus);

- occipital regions, i.e., Right Occipital Fusiform Gyrus, Left Occipital Fusiform Gyrus, Right Cuneal Cortex, Left Lateral Occipital Cortex, inferior division, Right Lateral Occipital Cortex, superior division, Left Lateral Occipital Cortex, superior division)

ROIs that had loadings greater in Females belong the to temporal and occipital regions (see hot coloured ROIs in Figure 4.23:

- temporal regions, i.e., Left Heschl's Gyrus, Right Planum Polare, Right Temporal Occipital Fusiform Cortex, Right Temporal Fusiform Cortex, posterior division, Left Inferior Temporal Gyrus, temporooccipital part, Right Inferior Temporal Gyrus, posterior division, Left Inferior Temporal Gyrus, posterior division, Right Inferior Temporal Gyrus, anterior division, Right Middle Temporal Gyrus, temporooccipital part, Right Middle Temporal Gyrus, posterior division, Left Middle Temporal Gyrus, posterior division, Right Superior Temporal Gyrus, anterior division);

- occipital regions, i.e., Right Occipital Pole, Left Occipital Pole, Right Supracalcarine Cortex, Left Supracalcarine Cortex, Right Lingual Gyrus, Left Lingual Gyrus, Right Intracalcarine Cortex, Left Intracalcarine Cortex, Right Lateral Occipital Cortex, inferior division.

ROIs that had loadings greater in Males belong the temporal and the frontal regions (see cool coloured ROIs in Figure 4.23):

- temporal regions, i.e., Right Planum Temporale, Left Inferior Temporal Gyrus, anterior division, Left Middle Temporal Gyrus, temporooccipital part, Left Superior Temporal Gyrus, anterior division, Right Temporal Pole;
- frontal regions, i.e., Right Central Opercular Cortex, Left Central Opercular Cortex, Right Frontal Medial Cortex, Left Frontal Medial Cortex, Left Superior Frontal Gyrus, Right Superior Frontal Gyrus, Right Postcentral Gyrus, Right Middle Frontal Gyrus, Left Inferior Frontal Gyrus, pars triangularis, Right Inferior Frontal Gyrus, pars triangularis, Left Inferior Frontal Gyrus, pars opercularis, Right Inferior Frontal Gyrus, pars opercularis, Right Precentral Gyrus.

In order to investigate the salience of the Default Mode Network [247, 202, 203], we selected from the Harvard-Oxford atlas specific Regions Of Interest able to represent that resting state network. The exploration of the loadings, in the 1st bootstrapped eigenvector, related with these ROIs, found the following gender differentiations:

- both Males and Females have similar values for the Right Cingulate Gyrus, posterior division, the Left Cingulate Gyrus, posterior division and the Left Middle Frontal Gyrus;
- Females have greater values for the Right Precuneous Cortex, Right Angular Gyrus and the Left Angular Gyrus;
- Males have greater values for Left Precuneous Cortex, Right Inferior Frontal Gyrus, pars opercularis, Left Inferior Frontal Gyrus, pars opercularis, Right Inferior Frontal Gyrus, pars triangularis, Left Superior Frontal Gyrus, Right Superior Frontal Gyrus, Left Inferior Frontal Gyrus, pars triangularis, Right Middle Frontal Gyrus.

Discussion and conclusions

Functional Considerations about the results of Competitive Clustering The study globally confirms the gender determinants in RS-fMRI functionality found by Biswal et al. Our approach is also related to the anatomo-functional correlation of a seeds (L/F Precuneus - DMN) with other regions, and specially with their controlateral part and their associated bilateral cingulated regions. Our results are confirmed the background because we

found brain functionality noted in the scientific literatures (Andrews-Hanna, Smallwood e Spreng 2014). 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 weakly associations, according to seeds-based correlation.

Functional Considerations about the results of Spectral Analysis Spectral analysis explores the information regard the brain functional connectivity. The results of Principal Component Analysis (PCA) and Nonlinear Robust Fuzzy Principal Component Analysis (NFRPCA) are comparable, but NRFPCA was more precise in the computations then PCA. In both Females and Males, the informative eigenvectors selected with Random Matrix Theory (RMT) tools are (in average) five, that it means that there are (in average) five independent networks of correlated ROIs. These results are similar to the resting state brain networks proposed in literature by Van Den Heuvel [248], that shown in a recent review that there is convergence between several studies to a subsets of specific resting state networks: there are the so-called Default Mode Network (DMN), the primary motor network, the insular-temporal/ACC network, the left/right parieto-frontal network, and the frontal network.

Limitations and Future works Globally, these findings agree with the results published by Biswal et al [25] showing that exist different functional architectures for the resting state brain physiology: there are gender as well as age related brain sub-functionalities detect with fMRI using datasets coming from several centres of research. We did not take in account age because the subjects of the Oxford dataset are all young with a thin age spread. Nevertheless, with those subjects we identified gender related functional differences with classical statistical methodology and with unsupervised learning algorithms.

There are limitations regard this study. One regards the number of samples we adopted that is limited to 22 subjects. But, interestingly, the statis-

tical tests demonstrated that the two subsamples (10 Females and 12 Males) are significantly different, i.e., they did not come from the same population. Other limitation regards the settings of clustering algorithms and the parameters associated to NFRPCA (i.e., fuzzy exponent equal to 1.5). Next study will attempt to address these limitations. An improvement of this work could deep the outcomes of the spectral analysis. We focalized the attention on the 1st eigenvector in both Females and Males, but there are also other in the list of informative eigenvectors that should study to explore more in details their content related to the brain functional connectivity. Furthermore, Random Matrix Theory application in Neuroscience are an opportunity to use advanced statistical tools in comparison with other classical approach to time series analysis. Another improvement should regard the clustering algorithms, using other settings of parameters for NG and GNG or include others soft clustering methods, for example Fuzzy C-Means (FCM) [18] or Fuzzy Self-Organizing Maps (FSOM) [107] and a validation procedure based on fuzzy optimization indexes.

Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms

Premise

This research work was presented at 6th Computational Modelling of Objects Presented in Images Congress held in Cracow (Poland) on July 2018 [255].

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.

The key-words regard this work are the following: fMRI, Partitive Clustering, Fuzzy c-means Algorithm, Neural Gas Algorithm

Introduction

In functional Magnetic Resonance Imaging (fMRI) there are two kinds of approaches to analyse data: 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 does not need any assumptions related the images to investigate. The main model-based approach to fMRI data is the Statistical Parametric Maps (SPMs) approach introduced by Friston [71]. The main model-free models are the Analysis of Independent Components (ICA) or the Analysis of Principal Components (PCA) (for an overview see [143]). 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 [207]. Several kinds of separation methodologies based on different theoretical framework are proposed in literature [111]. 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 [18]). 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 is a soft algorithm in terms of multi-class properties and the second one is a soft algorithm in terms of competitive learning rule. [268].

Clustering techniques applied to fMRI time series data are interesting approaches to explore brain functional properties [140, 136, 251] . Partitioning works grouping image voxels together based on how much they are alike in relation to some similar measure (distances, correlation, etc.), that probes how their intensity profiles in time are similar. More in 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 $\mu \in \{1, ..., K\}$, that are the signal values $\{\mathbf{x}^{\mu}(1), ..., \mathbf{x}^{\mu}(n)\}$ that can be modelled as a vector $\mathbf{x}^{\mu}(i) \in \Re^n$ in the n-dimensional (Euclidean) feature space of possible time series. Each of the that points is partitioned into clusters based on the similarity of their intensity profile in time. Therefore, the principal approach to fMRI clustering is to clusters 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 want also to find functional structures in temporal features that have similar spatial patterns: we named these objects as Times of Interest (TOIs), that in experimental terms mean to cluster the experimental blocks related to each brain volumes. The assumption is that the exploration of TOIs allows to find properties related the peculiarity of each block in the experimental design (e.g. imagine a block design structured as TASK-REST stimuli alternations, the TOIs clustering could allows to find spatial structures that are similar to TASK or REST block, whereas the classic ROIs clustering allow to find temporal structures that are similar in different brain regions). We named this global procedure **crossed-clustering** that want to find spatial patterns in the temporal features (TOIs) and the temporal patterns in the spatial features (ROIs). The postclustering procedure we applied is a statistical evaluation of the clusters made. We used parametric and non-parametric test to study if 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 inquired the functional connectivity

of the fMRI scans within subjects in order to find useful information to be associated with the clustering techniques adopted. In the next section, we present the dataset we selected to be clustered, the computational methodologies we used, and sequentially, the results, the discussions and conclusions with future works.

Materials and Methods

Data We selected the dataset proposed by Wakeman and Henson [259] from **openfmri** repository (https://openfmri.org/dataset/ds000117). This dataset is face recognition task paradigm 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 this features, we clustered the subjects putting special attention on the clustering outcomes referred to eight fusiform cortexes using Harvard-Oxford labels. According to this 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 to cluster the images, we performed preprocessing with FSL standard tools [113]: such as spatial and temporal filtering, motion correction, standard registration (with MNI152 reference), time-series extraction according to the meaning of Harvard-Oxford Atlas with 96 lateralized labels.

Clustering Algorithms We adopted two soft clustering algorithms to process fRMI data: **Fuzzy C-Means (FCM)** and **Neural Gas (NG)**. We used as an input features both Regions of Interest (ROIs) and Times of Interest (TOIs). We validated the optimal clustering using Davies-Bouldin index [46] and we compared the several clustering outcomes with 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 test (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 find similar results with the literature specialized in experimental neuroscience about the brain face processing (see [94]). For the formalism and details about methodologies see 2.2.2for the fuzzy algorithms (Fuzzy C-Means) and 2.2.3 for the competitive algorithms (Neural Gas).

Results

In this section, we describe the results shown in the figures. The comments about these results are indicated in the discussion and conclusion subsection.

- Results in Figure 4.24 shows the empirical distributions of the overall Correlation Coefficients (CC) of each subjects, and the global correlation matrices represented as a mean of correlation matrices of each subjects.

- Results in Figure 4.25 shows, first, the best Davies-Bouldin Indexes (DBI) computed for clusters ranged from 2 to 20 for both FCM and NG algorithms applied both ROIs and Time Points; then, second and third, it shows the spatial semantics of the FCM and NG in the case of 2-classes computed as statistical mode between subjects (i.e. if a region frequently appears in a cluster i, than it belongs to cluster i in the representation). Table 4.14 details the best DBI configurations for the dataset.

- Results in Figure 4.27 and Figure 4.28 show results of non-parametric (Kruskal-Wallis) and parametric (ANOVA 1-way) statistical tests applied to optimal FCM and NG in both spatial and temporal configuration. The figures show that in the spatial configurations for both FCM and NG the p-value was always less then 0.05 and then we can reject the null hypothesis that clusters have all the same means. Instead for both FCM and NG, the p-value computed in the temporal configuration clustering is not always less then 0.05, with worst behavior in the specific case of NG.

- In the Figure 4.29 and Figure 4.30, the results are almost the same of Figure 4.27 and Figure 4.28.

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

- Results in Figure 4.31 shows the comparison of Regions of Interests (ROIs) clustering with 2 groups using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices, there are the distributions of the Jaccard distance values. Globally, the similarity distributions between subjects are uniform and not so-high,

with some extreme values using FCM algorithm, that has clustered some subjects similarly and other subjects differently. This evidence has an explanation that does not regard the quality of the clustering itself, but the starting procedure that adopted algorithms to initialize their computation.

- Results in Figure 4.32 shows the comparison of Times of Interests (TOIs) clustering with 2-classes using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices there are the distribution of the Jaccard distance values. Globally the similarity distributions between subjects are uniform (for NG) and middle-centered (for FCM. The main result is that none outcomes seems to represent agreement within and between clustering algorithms. Instead, it is interesting to note that there is a wide variability looking globally the empirical distributions of the indexes, because they assume shapes nearby unimodal and multi-modal distributions.

- Results in Figure 4.33 shows the Jaccard's distance of the NG and FCM results using 2-clusters. The ROIs clustering configuration shows some uniform distribution of the measure, instead the TOIs clustering configuration shows unimodal (gaussian-like) distribution.

  Globally, in the Figure 4.31, Figure 4.32 and Figure 4.33 there are evidences that that clustering algorithms computed in a different way the subjects, making the comparison of their results hard to do. These results do not mean that the clustering methods miscalculated classes, but rather then, initialization procedures used by the algorithms lead to assign different classes to the same pattern found between subjects.

- Results in Figures 4.34, 4.35, 4.36 and 4.37 show outcomes related to one-subject analysis. We selected specifically the FCM 2-clusters for the subject 12 because has nice performance in spatial and temporal configuration according to both parametric and non-parametric tests.

  - Figure 4.34 is the graph-based topological representation that gives an idea about the parcellation.
  - Figure 4.35 shows, in the top-left part, 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

centre plot of Figure 4.35. The difficulty to distinguish temporal pattern along the time points is evident in the right plot of Figure 4.35, where it is not evident data structure that explains the two partitions, though the statistical tests for temporal clusters have always a p-value less (or a bit around) 0.05 for the subject 12.

– Figures 4.36 and 4.37 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 4.37 shows the Fusiform regions superimposed to the brain parcellation in Figure 4.36: the eight Fusiform cortexes belong to different clusters according 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 activations, cluster 3, 13, 14 with the less activations.

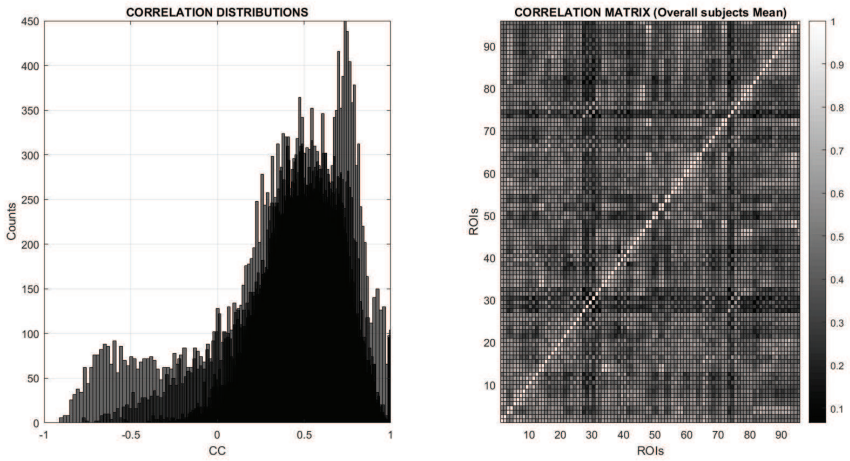

Figure 4.24: The Figure shows the information related to the correlation coefficients. In the left plot there is the empirical distribution of the all correlation coefficients (CC) computed for each subjects. In the right plot there is the mean correlation matrix computed with all subjects' correlations)

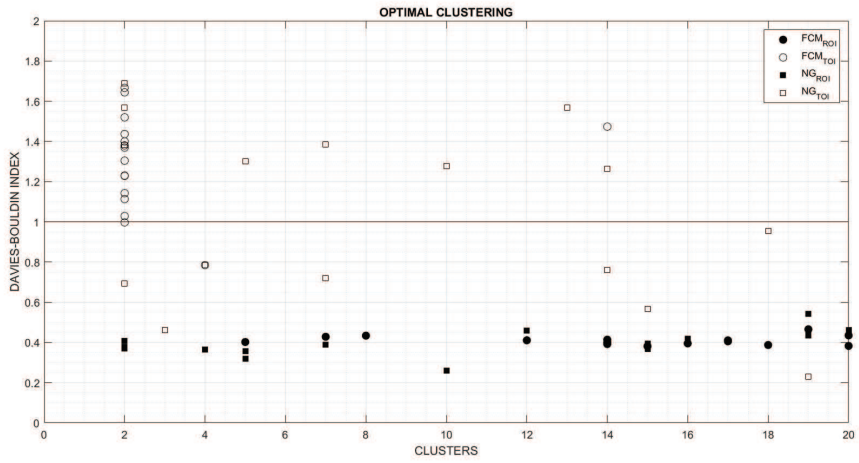

Figure 4.25: The figure shows the plot with the optimal clusters against the lesser Davies-Bouldin index for the clustering with FCM and NG algorithms, that are differentiated by the configuration Regions of Interest (ROIs) configuration and the Time Of Interest (TOIs) configuration. Globally, the partitioning of ROIs with both FCM and NG had less Davies-Bouldin index values then the partitioning of TOIs.

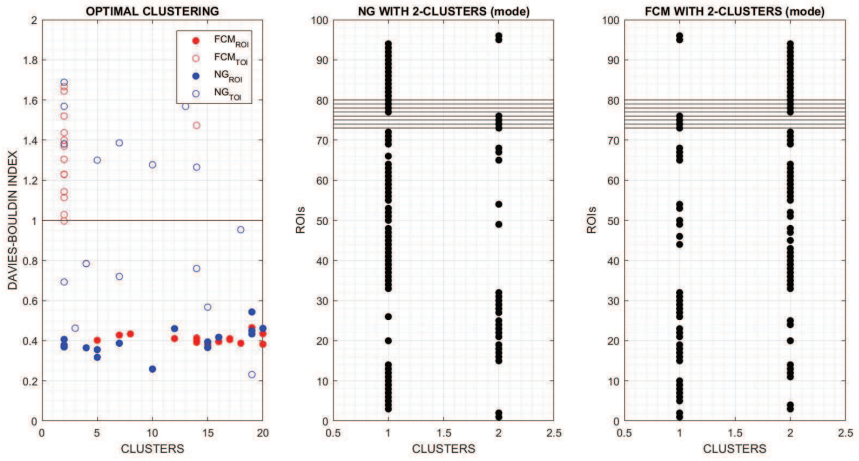

Figure 4.26: The figure shows the more frequent ROIs partitioning in the special case of clustering with 2 groups for NG and FCM for all the subjects; the horizontal line from the ROIs 73 to 80 are the ones related the eight Fusiform regions, i.e. they are ROIs 73:74 L/R Temporal Fusiform Cx Anterior 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 4.14: The table describes the Davies-Bouldin (DB) index computation for each subjects differentiated for clustering (FCM or NG) and inputs (ROIs or TOIs).The values presented are the lesser DB associated with the referred 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

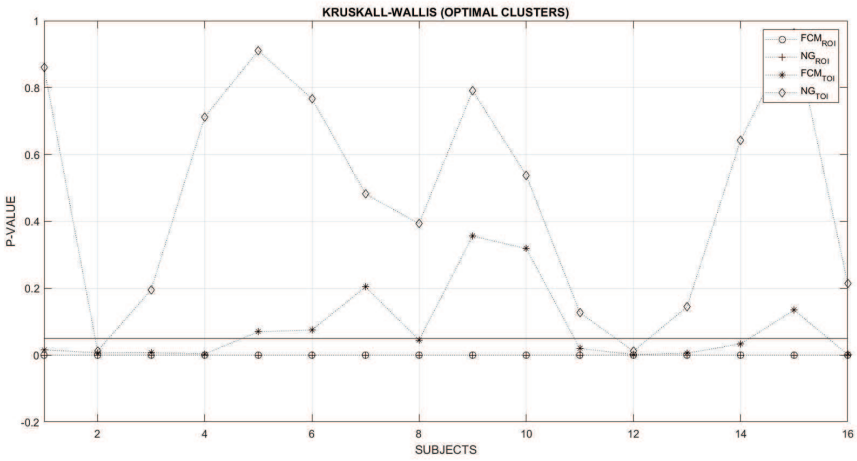

Figure 4.27: The figure shows the results of non-parametric tests (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) inputs in case of the optimal configurations. The black line is the significance level 0.05. Values under the black line allow to reject the null hypothesis of the test. ROIs clustering with both FCM and NG have clusters statistically difference for all the subjects with non-parametric tests. TOIs clustering is globally near the significative criterion only with FCM algorithms

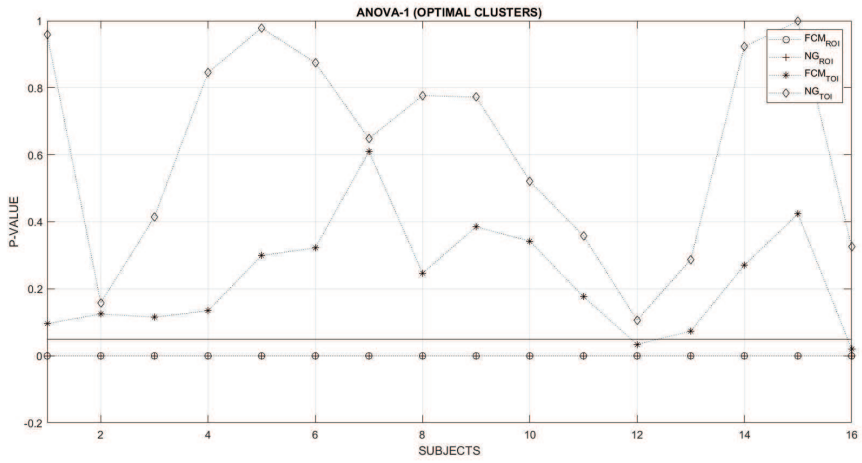

Figure 4.28: The figure shows the results of parametric test (ANOVA-1) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of the optimal clusters configurations. The black line is the significance level 0.05. The results were similar to Figure 4.27, but with a higher p-value for the TOIs clustering with FCM algorithm.

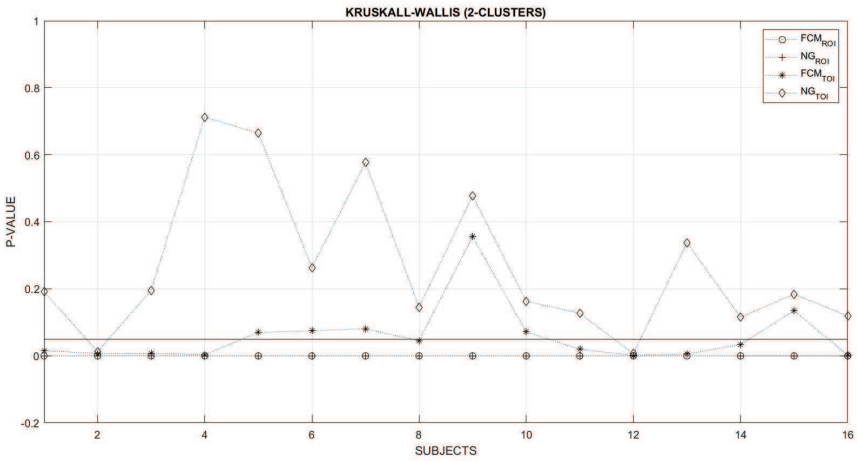

Figure 4.29: The figure shows the results of non parametric test (Kruskall-Wallis) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) configurations in special case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 4.27 and Figure 4.28, but with lower p-value for the TOIs clustering with FCM algorithm.

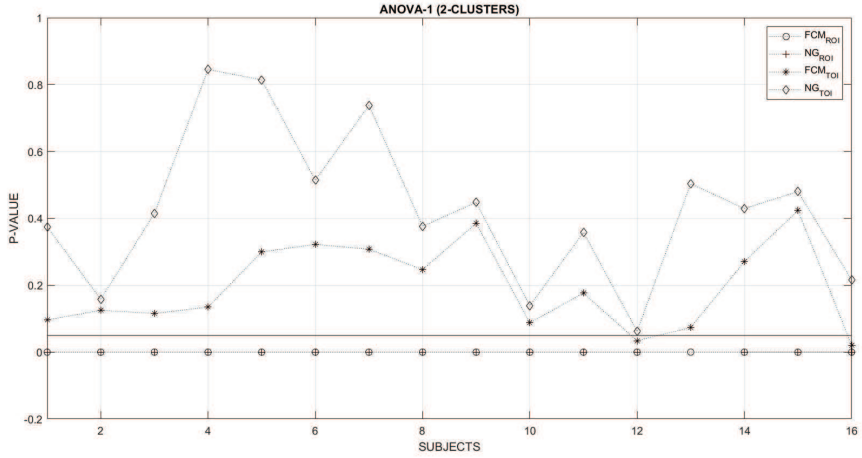

Figure 4.30: The figure shows the results of parametric (one-way ANOVA) for the clusters obtained with FCM and NG algorithms for the ROIs (Regions of Interest) and TOIs (Times of Interest) in the case of 2-clusters partitioning. The black line is the significance level 0.05. The results were similar to Figure 4.29, but with a higher p-value for the TOIs clustering with FCM algorithm.

Discussion

Considerations on results of Figure 4.24. The results are related to the global functional connectivity analysis that reflects the presence of some variability in the single subject correlations (left Figure 4.24), i.e. more then half subjects had positive correlations during the task, whereas few ones had also negative correlations. In particular, the mean correlation matrix (right Figure 4.24) shows great correlation in specific sub-matrices, e.g. the sub-matrix that regards the ROIs from 73 to 80 are 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 sub-matrix positive correlation is in agreement with the selective importance for the face recognition task of the fusiform regions shown with the results of Wakeman and Henson [259].

Considerations about results in Figures 4.27, **Figure** 4.28, **Figure** 4.29 **and Figure** 4.30. The outcomes related to these Figures regard 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

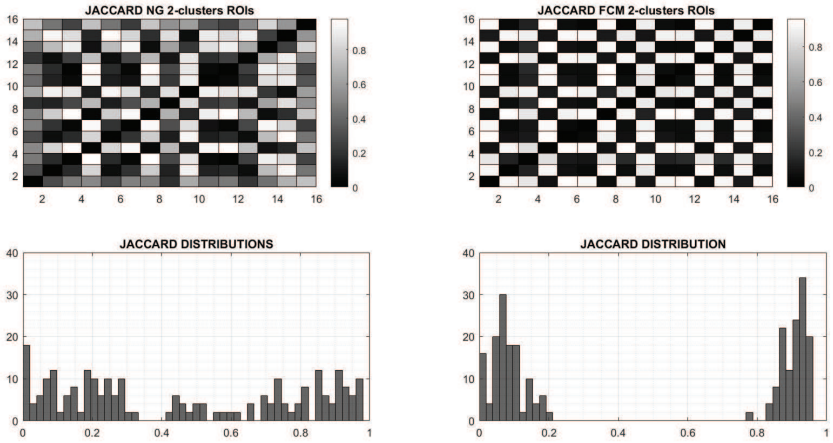

Figure 4.31: The figure shows the comparison of Regions of Interests (ROIs) clustering with 2 groups using Jaccard matrix. The comparison is between subjects in case of NG and FCM. Under the matrices there are the distribution of the Jaccard distance values. Globally the similarity distribution between subjects is uniform and not so-high, with some extreme values using FCM algorithms.

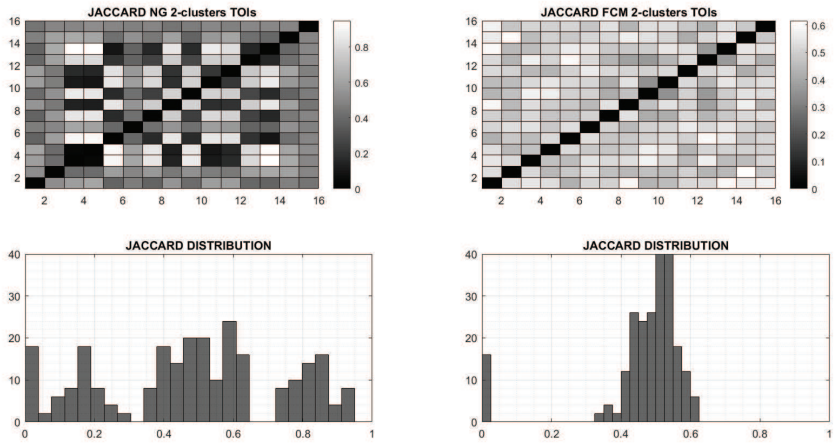

Figure 4.32: The figure shows the comparison of Times of Interests (TOIs) clustering with 2 groups using Jaccard matrix. The comparison is between subjects in case of NG and FCM Under the matrices there are the distribution of the Jaccard distance values. Globally the similarity distribution between subjects is uniform (for NG) and middle-centered (for FCM).

NG for both parametric and non-parametric test. Knowing that 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 to different number of clusters for subjects (see. Table 4.14). The use of 2-clusters (or a fixed-clusters) classifications allows to easy compare subjects for post-clustering analysis. Furthermore, spatial clustering has globally nice significance, but temporal clustering has no the same quality, due to difficult to find statistically different clusters, albeit FCM outperformed NG, but not always with p-value less then 0.05, with better results for 2-clusters setting.

Considerations about results in Figure 4.31, **Figure** 4.32 **and Figure** 4.33. The results are related to the variability observed in the clustering results. Jaccard similarity matrices highlight huge differences within and between clustering algorithms. This evidence could be explained considering that labels were initially assigned with choice, and then although clustering reached the optimal configurations, it could be the case that the i label is not assigned to the i pattern with all the subjects. In other words, the same pattern in different subjects could be labelled sometimes with label i or with label j.

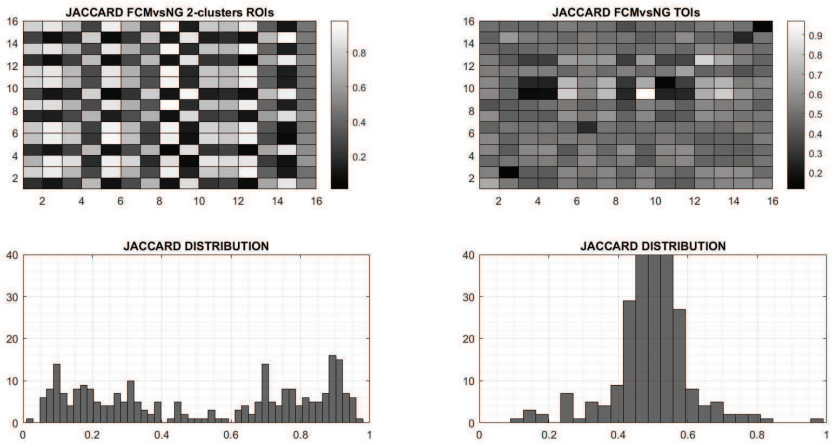

Figure 4.33: The figure shows the Jaccard Matrices in both the case of 2-classes clustering with Regions of Interests (ROIs) and Times of Interests (TOIs). The comparison is between subjects in case of NG and FCM. Under the matrices there are the distribution of the Jaccard distance values. Globally the similarity distribution between subjects is weakly bimodal for the ROIs clustering (bottom left) and middle-centred for the TOIs clustering (bottom right).

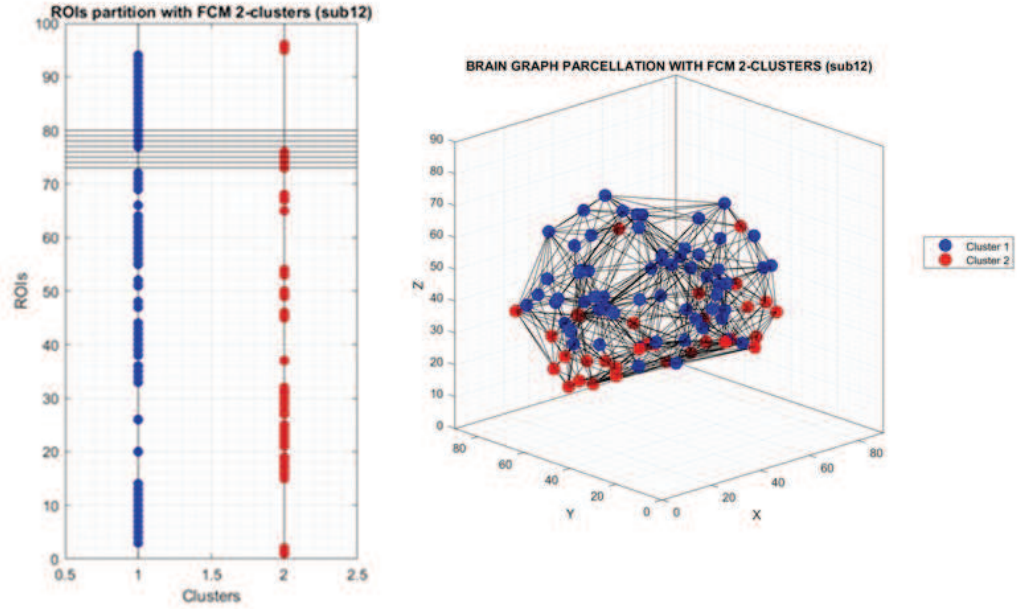

Figure 4.34: The figure shows the brain parcellation based on graph model and the Regions of Interest (ROIs) organization partitioned with two clusters. In the left plot the horizontal line indicated the Fusiform Cortexes (ROIs 73:80) and their clusters. In the right image the nodes are the 96 centroids according to Harvard-Oxford atlas. Globally, 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.

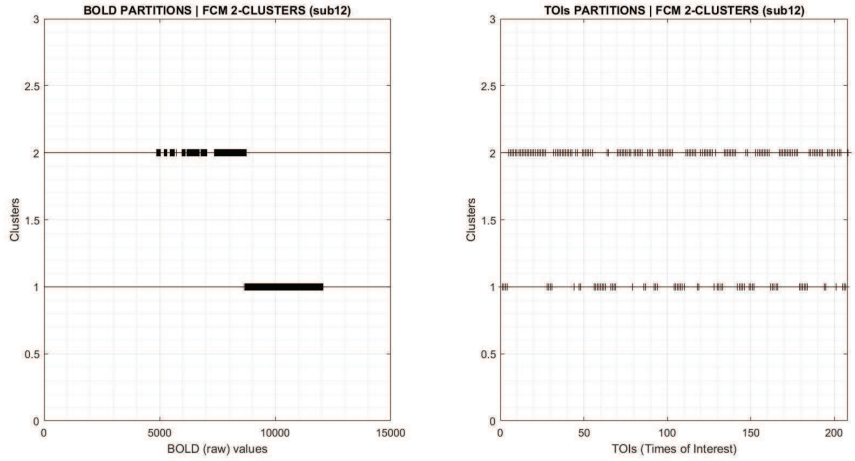

Figure 4.35: The Figure shows the BOLD values partition among the two clusters and the Times Of Interests (TOIs) partition among the two clusters. In the left plot, clusters discriminated the BOLD amplitude and in the right plot clusters seem to have discriminated spatio-temporal patterns. The Figure is referred specifically to subject 12 clustered with FCM.

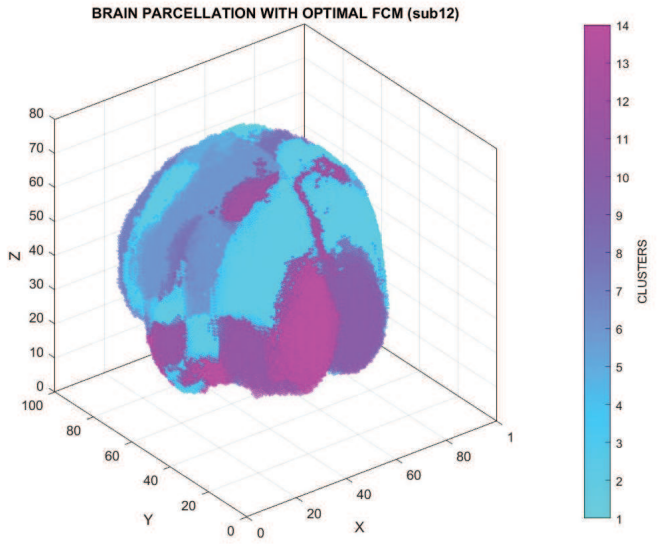

Figure 4.36: The Figure shows the brain voxels parcellation of subject 12 clustered with FCM using the optimal configuration (with the lesser Davies-Bouldin index). There are 14 clusters that grouped differently the all brain voxels (see Figure 4.37 for more details).

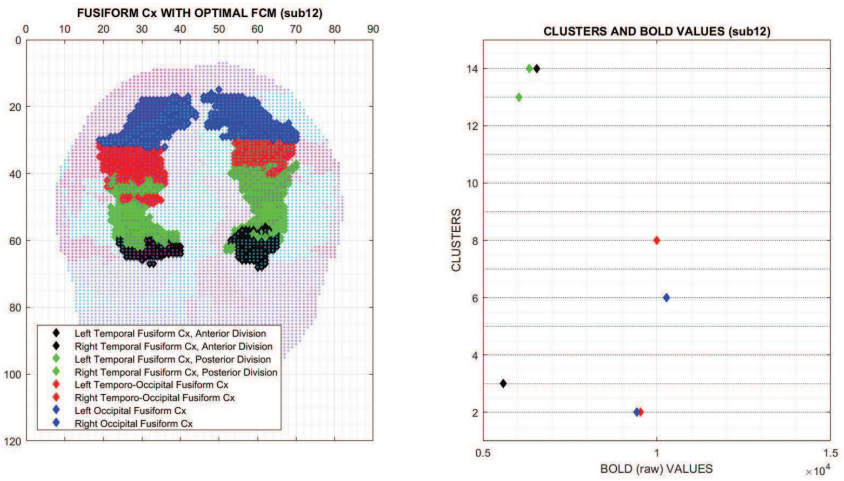

Figure 4.37: The left Figure shows the substructures of the Fusiform regions. In the right Figure there are the eight Fusiform regions distinguished in different clusters according to their BOLD values: L/R Temporo-Occipital Fusiform Cx and L/R Occipital Fusiform Cx have greater BOLD values than the 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 with the lesser activations.

The considerations about results in Figure 4.34 **and Figure** 4.35. The results regard the special case of the clustering computed for the subject 12. It had the best behaviour for both parametric and 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 in the class 2. The two spatial clusters differentiated two BOLD signal macro-levels, but the two temporal clusters have no the same easy visible intuition, although they have a statistical differences probed - principally - 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 the Figure 4.36 and Figure 4.37 are referred 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, with the results of Wakeman and Henson [259] that have shared the data we processed, and also the main works related (see. the seminal paper by Kanwisher [119], or the recent outcomes by Ghuman [85] and Grill-Spector [94]))

Conclusion and Future Works

Summary In this paper we adopted the 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, are both the anatomical regions with their BOLD temporal signatures and the ON/OFF blocks of an experiment with their related brain spatial response. Specifically, we processed fMRI images repository with 16 healthy subjects that have done a Face Recognition Task and 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 optimal clustering with Davies-Bouldin index and we compared the different subjects outcomes with Jaccard measure. We used parametrical and non-parametrical statistical tests to evaluate if the differences between classes are significative between clusters, taking p-value as a decision criterion. Also, we computed 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 functional properties referred the task-based paradigm. Globally, the results showed that ROIs clustering was more easy than TOIs

clustering for both the algorithms, but - globally - for the TOIs clustering (more complex then the spatial ones), Fuzzy C-Means outperformed Neural Gas comparing them with statistical significance test.

Limitations of this study . The first is the randomly assignment of starting clusters for both the algorithms. The second is the absence of known block paradigm to match with the TOIs clustering, that it is a theoretical-free limitations because it depends on how many information are available for a shared repository. The first limitation is more critical because it regards why - using 2 clusters - the same algorithms clustered with different labels the same pattern (see Jaccard matrix); the results is that even if there is a correct classification of statistically different clusters, it is lucking an easy way to compare the same class with different datasets. There is no clustering consistency in terms of labels names between subjects. This nominal peculiarity of clusters could be undertake using a linguistic procedure able to unify the same label to similar patterns with a formal *rationale*.

Future works We will design a translation procedure able to merge with 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 test, e.g. two-way ANOVA or Friedman test, we can evaluate if the $coupled - classes$ are significative different, i.e. we test if the clusters of ROIs crossed with the clusters of TOIs have elements that are significative difference. This procedure allows to find spatial patterns that are associated statistically with temporal patterns and *viceversa*. Crossing 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).

4.4 Thesis results: *summarium*

In this section there is the overview of the principal thesis results obtained in the functional neuroimaging context applying the computational intelligence methodologies, with particular regard to the contributions proposed. Briefly, there are three main classes of results:

- the outcomes related to image processing;
- the outcomes related to clinical subjects;
- the outcomes related to healthy subjects;

In the following pages, there concise presentation of the above results will be posed. This sinthetical part was thought for two motivations: i) to have a short declaration of the thesis outputs in terms of application of computer science methods in the area of functional neuroimaging, specially with fMRI; and ii) to have a high level vision to elaborate more in the extended section of Results 4.

The concise presentation of results is the following:

- Results on Image Processing
  - Quantitative Relations Between CRAIIM Human Brain Atlases
    - **Topic**. The main object of this research work is the computation of quantitative relations to compare two versions of CRAIIM Hybrid Atlas, that is an hybridization of the Juelich brain atlas with some ROIs selected from Harvard-Oxford atlas. The CRAIIM models differ for voxels resolution: 1 $[mm^3]$ and 2 $[mm^3]$.
    - **Results**. The main results is that they are globally comparable for the representation for all the 161 brain regions they have. Nevertheless, there are few regions that after the integration process achieved a low voxels percentage then the quantity they have before it. Therefore, the hybridization allow to build an integrated atlas, with a very nice covering of many regions, in particular the temporal and frontal ones, but it is important to mind the limitations when it is used.
    - **Section**. The complete work is into the Results on Image Processing 4.1 and information about labels of atlases used are in the Appendix 6.1.
- Results with Clinical Subjects

- Clustering Analysis of Functional Neuroimages Using Data Reduction and Competitive Learning Algorithms
  - **Topic**. In the present work we use pattern vectors derived from Statistical Parametric Map, generated from a group of artificial and in-house collected fMRI data, to conduct cluster analysis. Two clustering algorithms, self-organizing map (SOM) and growing neural gas (GNG), are selected to explore inherent properties in the brain functional data.
  - **Results**. As seen in our experimental context, SOM and GNG show comparable behaviour, however GNG prevails in the management of large data sets. An exploratory, descriptive analysis is conducted on in-house collected data clustered by GNG and results are detailed in the paper.
  - **Section**. The complete work is presented in the Results with Clinical Subjects 4.2.

- Results with Healthy Subjects
  - A Soft Davies-Bouldin Separation Measure
    - **Topic**.We present a soft separation measure to validate fuzzy clustering results without defuzzyficaton. It is the generalization of Davies-Bouldin validation index (DB) for crisp clustering in the soft clustering domain; we named the measure Soft Davies-Bouldin index (SDB). We compared DB and SDB when applied to k-means and fuzzy c-means algorithms using eight datasets with ground-truth and two experimental fMRI datasets without ground-truth.
    - **Results**. We found that i) in more than half datasets, the optimal score of Soft Davies-Bouldin index was less than Davies-Bouldin index, ii) in half datasets that have ground-truth, the optimal score of Soft Davies-Bouldin index was less than Davies-Bouldin index in correspondence of the truth number of patterns, iii) the Soft Davies-Bouldin index outperformed the Davies-Bouldin index as central tendency of all datasets along the complete range of clusters considered.
    - **Section**. The research work is presented in the Results with Healthy Subjects section 4.3.
  - RS-fMRI Analysis Using Unsupervised Learning Algorithms
    - **Topic**. RS-fMRI data analysis for functional connectivity explorations is a challenging topic in computational neuroimag-

ing. Several approaches have been investigated to discover whole-brain data features. Among these, clustering techniques based on Competitive Learning (CL) and Spectral Methods (SM) have been shown effective in providing useful information in various contexts. We selected three clustering algorithms and two spectral methods, i.e the clustering algorithm are Self-Organizing Maps (SOM), Neural Gas (NG) and Growing Neural Gas (GNG), whereas the spectral methods are the classic Principal Component Analysis (PCA) and the Nonliner Robust Fuzzy Principal Component Analysis (NRF-PCA). We validated clustering with Davies-Bouldin Index (DBI) and we selected informative principal components using Random Matrix Theory (RMT). tools. We adopted these techniques to study the intrinsic functional properties of images coming from a shared repository of resting state fMRI experiments (1000 Functional Connectome Project).

  * **Results**. There are two main macro-outcomes to describe: the first is that, according to classic signal processing methodologies, there are significative differences in the BOLD signals between gender (using mean and variance as statistical measure associated to significance tests); the second is that, according to unsupervised learning algorithm, there are different functional patterns between gender (using clustering methods as well as spectral analysis of the covariance matrix). There are also comments related to computational methodology: i) particular, all three clustering algorithms used reached their optimality with two classes, and ii) the study of spectral components with fuzzy PCA reveals more precise results compared with the crisp PCA.
  * **Section**. The research work is presented in the Results with Healthy Subjects section 4.3.

– Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms

  * **Topic**. We used model free methods to explore the brain's functional properties adopting a partitioning procedure based on cross-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, applied to a shared fMRI repository with participants performing a Face Recognition Task. We investigated the parti-

tioning matching the BOLD signal signatures with the classes found and with the results of functional connectivity analysis. We compared the outcomes using the just known model-based knowledge as likely ground truth, confirming the role of Fusiform brain regions

* **Results**. Partitioning results globally 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.
* **Section**.The research work is presented in the Results with Healthy Subjects section 4.3.

Chapter 5

Conclusions

In this chapter there are two sections:

- the Final Proposition section 5.1, i.e., the sinthetical presentation of the main conclusive concepts inferred from the new contributions proposed in the Computational Intelligence section and from selected[1] Functional Neuroimaging outcomes exposed in the PhD Thesis Results section;
- the Future Works presentation 5.2, i.e, the considerations on what are the limitations followed by the Final Propositions and the activities that are planned to do in the next time.

[1]There is the indication of *selected* outcomes since the thesis topic is principally related to the computer science methodologies. Therefore, for this final propositions, only a part of the specific neuroscientific results obtained with them have been selected to build the following conclusive statements. There are, of curse, important considerations to do with the outcomes related to resting state fMRI or with the active paradigm, as well as the ones associated to clinical studies. To explore more these physiological and pathological considerations, I suggest to read the discussions and the conclusions from the specific studies presented in Thesis Results sections 4

5.1 Final Propositions

The final propositions regards principally the outcomes related to the computational methodologies, i.e.,

- the image processing of anatomical brain atlas,
- the validation measures associated to the optimal clustering,
- the statistical semantics[2] of classes found,
- the crossed-clustering framework application,
- the spectral analysis of the functional connectivity domain.

A part the above computational results - with some extension of the brain functionality - there are also some considerations about the image processing outcomes and some neuroscientific considerations on the classes found by the clustering algorithms, specially the ones that could be linked with scientific literature specialized in neuroimaging.

In this final prepositions section, they are skipped in order to put the light on the computational results since the main research domain of this thesis is in computer science. To look the consideration on functional neuroimaging results see the discussions in the section with clinical subjects 4.2 and the discussion in the section with healthy subjects 4.3.

Therefore, there are following final computational propositions:

Proposition I. The brain atlas integration that joins ROIs from different brain models, with same spatial coordinates, is an opportunity to design complementary hybrid anatomies. The operations based on sets theory that allow to merge regions are useful to shape a tailored and integrated union of different atlas ROIs. The positive aspects of these processing is personalized results achieved, specially if it is oriented for specific clinical motivations. The drawback are the possible regional superimpositions that could exist, that are solvable choosing a reference atlas to which the complementary atlas will be adapted[3].

[2]There is a differentiation between anatomical semantics and statistical semantics of the classes found by clustering algorithms: the first is referred to brain regions that are within or without a class given some algebraic similarities in their signatures, i.e., distances between sample vectors, whereas the statistical semantics are functional descriptors that permit to characterize the regions that are within or without a class, i.e., the BOLD intensities or the correlation indexes.

[3]To see the image processing and brain atlasing results see the section 4.1

Proposition II. The fuzzy theory applied to validation measures improves the quality of the clustering evaluation. The comparison between crisp and soft clustering validation measure gave as a result that the Soft Davies Bouldin index evaluated with more precision[4] and less variability[5] the classes found by the clustering algorithms, with both benchmark datasets and resting-state fMRI datasets.

It is important to note that the clustering validation indexes are used to select the optimal clustering configurations, i.e., the best partitions of the samples in the features space. But, once it is found the optimal outcome, it is also wise to investigate if there are sub-optimalities that allow - at the same time - to suggest alternative patterns, specially if there is a slight variance associated to the validation measures.

Proposition III. The crossed-clustering framework shows there are computational differences for the clustering of spatial and temporal patterns. The usage of the crossed-clustering framework with both fMRI active paradigm and fMRI passive paradigm unlighted that the clustering of spatial patterns is more easy then the clustering of temporal patterns. In particular, according to the p-value computation, the fuzzy c-means algorithm and the Neural Gas algorithm are quite similar in the spatial patterns clustering, whereas the fuzzy c-means partitioned better the temporal patterns then the neural gas[6].

Furthermore, the evidence that the fuzzy approach on the exploration of temporal patterns is preferable is also present in the results obtained with the soft separation measure, where the better separation value has been obtained by the Soft Davies-Bouldin index then the crisp one[7].

Proposition IV. The clustering outcomes have a peculiar statistical semantics. The statistical semantics of the classes found by algorithms are quantitative descriptors that allow to drawn some consideration about the clustering outcomes.

In particular, the clustering results with passive paradigm applied to healthy subjects have been studied looking the elements of the classes given some functional information, i.e., the BOLD fMRI signal intensity describes the ROIs that are within (or without) a class[8] and also the seed-based cor-

[4]See Figure 4.8 in Results 4.3
[5]See Figure 4.6 and Figure 4.7.
[6]See Figures 4.27, 4.28, 4.29 and 4.30.
[7]See Figure4.6.
[8]See Figure 4.12

relation describes the ROIs clustered[9].

Furthermore, the clustering results with active paradigm applied to healthy subjects gave similar results, i.e., the BOLD fMRI signal intensities discriminated in general the all ROIs clustered[10] and also the sub-specialization of the fusiform cortexes functionally parcelled[11].

Proposition V. The spectral analysis of the connectivity domain reveals not-randomly correlated functional networks The analysis of the principal components of the correlation matrix of passive paradigm fMRI images allowed to distinguished a set of correlated networks that, according to random matrix theory tools, have true correlations[12]. The usage of the fuzzy non-linear PCA allows to obtain more precise results then the classic linear PCA[13].

[9] See Figures 4.15 and 4.16
[10] See the Figure 4.35
[11] See the Figure 4.37.
[12] See Figure 4.17.
[13] See Figure 4.21 and 4.22.

5.2 Future Works

Future works related to soft separation measures. In the Contribution part 2.3.1 of the Computational Intelligence section 2, the soft version of the Davies-Bouldin separation measures has been proposed, and the outcomes related to its application to evaluate the functional parcellation in fMRI were presented in the results section 4.3. Formally, the Soft Davies-Bouldin index proposed is the following[14]: *ratio* R_{ij}^f, i.e.

$$R_{ij}^f = \frac{S_i^f U_i + S_j^f U_j}{M_{ij}^f} \tag{5.1}$$

The soft properties that it has is the capacity to manage the fuzziness associated the clusters found by the C-Means algorithm, in particular the Type 1. But, in literature, there exist also the fuzzy C-Means Type 2[15].

Therefore, the next work that it is going to develop is the extension of the Soft Davies-Bouldin separation measure based on the Type 1 C-Means to the Type 2 C-Means.

Future works related to crossed-clustering framework. An alternative to classic clustering of image time series is the investigation of both spatial and temporal patterns. To do this operation, it is necessary to cluster the spatial samples vectors or to cluster the temporal samples vectors, since the type of patterns *ex novo* classified depends by the type of features space used to make the partitions, i.e., given a spatio-temporal matrix M, with columns as the spatial features and the rows as the temporal samples for each features, the spatial clustering is the labelling of temporal sample vectors in the spatial features space, and, *viceversa*, given the transpose of M, the temporal clustering is the labelling of spatial sample vectors in the temporal features space.

The transposition is the operation the permits to apply clustering of spatial patterns or temporal patterns. The end of the results are two patters configurations that give information about spatial similarities given time variability or temporal similarities given spatial variability. This knowledge about the functional neuroimages is uncoupled i.e., the two patterns found are from two independent clustering operations that give two indications on the content of the spatio-temporal images. Therefore, it is useful a computa-

[14]Fore more details about notations see the specific part 2.3.1.

[15]Fore more details about the two algorithms see the specific part on fuzzy algorithms 2.2.2 and their computational description in the appendix section 1 and 2

tional method able to join these uncoupled results, e.g., a possible solution to solve this task is a functional that allows to link the two patterns information.

There are hypothesis to adopt to approach the issue, e.g., the usage of statistical methods able to control factors[16] and to test if their common response variable (e.g., the BOLD signal value) could individuate significative coupling, i.e., if there are spatial patterns that have specific temporal patterns and, *viceversa*, if there are temporal patterns that have specific spatial patterns[17].

[16]The factors are the the two patterns, where the levels of the factors are their labels, e.g., the spatial factor is shaped with the spatial patterns levels and the temporal factor is shaped with the temporal patterns levels

[17]Potential methods that could address this topic is the parametric balanced ANOVA two-way or the Friedman test that it is its non-parametric version.

Chapter 6

Appendixes

In the Appendix there are following three sections:

- the section related to the Human Brain Atlas 6.1, where there are the labels and their enumeration associated to the three main atlas used, i.e., the CRAIIM Hibrid Atlas, the Juelich Atlas, and the Harvard-Oxford Atlas.
- the section related to the complete list of research papers abstract publishe d 6.2; it is important to note that there are also papers that has not been inserted in the previous chapters since they are directly related to the main topic of the thesis (the computational intelligence applications in neuroscience);
- the section related to the list of algorithms 6.3 presented in the Computational Intelligence section 2.2;

6.1 Human Brain Atlases

In this section there are following three subsections:

- the subsection related to labels belonged to CRAIIM Hybrid Altas (6.1.1)
- the subsection related to labels belonged to Juelich Hybrid Altas (6.1.2)
- the subsection related to labels belonged to Harvard-Oxford Altas (6.1.3)

6.1.1 CRAIIM Hybrid Atlas

The CRAIIM Hysbris Atlas is the integration of Juelich Atlas 6.1.2 with Harvard-Oxford atlas 6.1.3. It is shaped with 161 ROIs, i.e., 121 are from the Juelich one and the other 40 are from Harvard-Oxford one [1]. The CRAIIM Hybrid atlas encompasses the following gray macro-regions:

- Gray matter regions.
  - Parietal Lobules
  - Hippocampus
  - Motor and Premotor cortexes
  - Somatosensory cortexes
  - Auditory cortxes
  - Visual cortexes
  - Insula
  - Broca's cortexes
  - Geniculate Bodies and Mamillar
  - Frontal cortexes
  - Temporal cortexes
  - Cingulate and Paracingulate cortexes

Note that the first 121 ROIs of CRAIIM atlas are exactly the ones from the Juelich atlas. Therefore, in this subsection those regions are skipped, instead are presented the other 40 of CRAIIM atlas with the associated Harvard-Oxford number labels, i.e., there are described the ROIs from 122 to 161 of CRAIIM atlas and their relative enumeration in the Harvard-Oxford atlas. To see the first 121 regions look the Juelich Atlas section 6.1.2:

122. LEFT FRONTAL POLE (HARVARD-OXFORD LABEL: 2)
123. RIGHT FRONTAL POLE (HARVARD-OXFORD LABEL: 1)
124. LEFT SUPERIOR TEMPORAL GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 17)
125. RIGHT SUPERIOR TEMPORAL GYRUS (HARVARD-OXFORD LABEL: 18)

[1] see the Result on Image Processing section 4.1 for more details about the integration procedure and quantities related to the brain anatomies

126. LEFT SUPERIOR TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 19)

127. RIGHT SUPERIOR TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 20)

128. LEFT MIDDLE TEMPORAL GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 21)

129. RIGHT MIDDLE TEMPORAL GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 22)

130. LEFT MIDDLE TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 23)

131. RIGHT MIDDLE TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 24)

132. LEFT MIDDLE TEMPROAL GYRUS TEMPORO-OCCIPITAL (HARVARD-OXFORD LABEL: 25)

133. RIGHT MIDDLE TEMPORAL GYRUS TEMPORO-OCCIPITAL (HARVARD-OXFORD LABEL: 26)

134. LEFT INFERIOR TEMPORAL GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 27)

135. RIGHT INFERIOR TEMPORAL GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 28)

136. LEFT INFERIOR TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 29)

137. RIGHT INFERIOR TEMPORAL GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 30)

138. LEFT INFERIOR TEMPORAL GYRUS TEMPORO-OCCIPITAL (HARVARD-OXFORD LABEL: 31)

139. RIGHT INFERIOR TEMPORAL GYRUS TEMPORO-OCCIPITAL (HARVARD-OXFORD LABEL: 32)

140. LEFT FRONTAL MEDIAL CORTEX (HARVARD-OXFORD LABEL: 49)

141. RIGHT FRONTAL MEDIAL CORTEX (HARVARD-OXFORD LABEL: 50)

142. LEFT SUBCALLOSAL CORTEX (HARVARD-OXFORD LABEL: 53)

143. RIGHT SUBCALLOSAL CORTEX (HARVARD-OXFORD LABEL: 54)

144. LEFT PARACINGULATE GYRUS (HARVARD-OXFORD LABEL: 55)

145. RIGHT PARAGINGULATE GYRUS (HARVARD-OXFORD LABEL: 56)

146. LEFT CINGULATE GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 57)

147. RIGHT CINGULATE GYRUS ANTERIOR (HARVARD-OXFORD LABEL: 58)

148. LEFT CINGULATE GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 59)

149. RIGHT CINGULATE GYRUS POSTERIOR (HARVARD-OXFORD LABEL: 60)

150. LEFT FRONTAL ORBITAL CORTEX (HARVARD-OXFORD LABEL: 65)

151. RIGHT FRONTAL ORBITAL CORTEX (HARVARD-OXFORD LABEL: 66)

152. LEFT TEMPORAL FUSIFORM CORTEX ANTERIOR (HARVARD-OXFORD LABEL: 73)

153. RIGHT TEMPORAL FUSIFORM CORTEX ANTERIOR (HARVARD-OXFORD LABEL: 74)

154. LEFT TEMPORAL FUSIFORM CORTEX POSTERIOR (HARVARD-OXFORD LABEL: 75)

155. RIGHT TEMPORAL FUSIFORM CORTEX POSTERIOR (HARVARD-OXFORD LABEL: 76)

156. LEFT TEMPORAL POLE (HARVARD-OXFORD LABEL: 15)

157. RIGHT TEMPORAL POLE (HARVARD-OXFORD LABEL: 16)

158. LEFT THALAMUS, SUBCORTICAL HARVARD-OXFORD (SUBCORTICAL HARVARD-OXFORD LABEL: 3)

159. RIGHT THALAMUS, SUBCORTICAL HARVARD-OXFORD (SUBCORTICAL HARVARD-OXFORD LABEL: 14)

160. LEFT TEMPORO-OCCIPITAL FUSIFORM CORTEX (HARVARD-OXFORD LABEL: 77)

161. RIGHT TEMPORO-OCCIPITAL FUSIFORM CORTEX (HARVARD-OXFORD LABEL: 78)

6.1.2 Juelich Altas

In the following pages, the number labels of Juelich atlas. Legend: GM = gray matter; WM = white matter; L=left; R=right

1. GM ANTERIOR INTRA-PARIETAL SULCUS HIP1 L.
2. GM ANTERIOR INTRA-PARIETAL SULCUS HIP1 R.
3. GM ANTERIOR INTRA-PARIETAL SULCUS HIP2 L.
4. GM ANTERIOR INTRA-PARIETAL SULCUS HIP2 R.
5. GM ANTERIOR INTRA-PARIETAL SULCUS HIP3 L.
6. GM ANTERIOR INTRA-PARIETAL SULCUS HIP3 R.
7. GM AMYGDALA CENTROMEDIAL GROUP L
8. GM AMYGDALA CENTROMEDIAL GROUP R
9. GM AMYGDALA LATEROBASAL GROUP L
10. GM AMYGDALA LATEROBASAL GROUP R
11. GM AMYGDALA SUPERFICIAL GROUP L
12. GM AMYGDALA SUPERFICIAL GROUP R
13. GM BROCA'S AREA BA44 L
14. GM BROCA'S AREA BA44 R
15. GM BROCA'S AREA BA45 L
16. GM BROCA'S AREA BA45 R
17. GM HIPPOCAMPUS CORNU AMMONIS L
18. GM HIPPOCAMPUS CORNU AMMONIS R
19. GM HIPPOCAMPUS DENTATE GYRUS L
20. GM HIPPOCAMPUS DENTATE GYRUS R
21. GM HIPPOCAMPUS ENTORHINAL CORTEX L
22. GM HIPPOCAMPUS ENTORHINAL CORTEX R

23. GM HIPPOCAMPUS HIPPOCAMPAL-AMYGDALOID TRANSITION AREA L
24. GM HIPPOCAMPUS HIPPOCAMPAL-AMYGDALOID TRANSITION AREA R
25. GM HIPPOCAMPUS SUBICULUM L
26. GM HIPPOCAMPUS SUBICULUM R
27. GM INFERIOR PARIETAL LOBULE PF L
28. GM INFERIOR PARIETAL LOBULE PF R
29. GM INFERIOR PARIETAL LOBULE PFCM L
30. GM INFERIOR PARIETAL LOBULE PFCM R
31. GM INFERIOR PARIETAL LOBULE PFM L
32. GM INFERIOR PARIETAL LOBULE PFM R
33. GM INFERIOR PARIETAL LOBULE PFOP L
34. GM INFERIOR PARIETAL LOBULE PFOP R
35. GM INFERIOR PARIETAL LOBULE PFT L
36. GM INFERIOR PARIETAL LOBULE PFT R
37. GM INFERIOR PARIETAL LOBULE PGA L
38. GM INFERIOR PARIETAL LOBULE PGA R
39. GM INFERIOR PARIETAL LOBULE PGP L
40. GM INFERIOR PARIETAL LOBULE PGP R
41. GM PRIMARY AUDITORY CORTEX TE1.0 L
42. GM PRIMARY AUDITORY CORTEX TE1.0 R
43. GM PRIMARY AUDITORY CORTEX TE1.1 L
44. GM PRIMARY AUDITORY CORTEX TE1.1 R
45. GM PRIMARY AUDITORY CORTEX TE1.2 L

46. GM PRIMARY AUDITORY CORTEX TE1.2 R
47. GM PRIMARY MOTOR CORTEX BA4A L
48. GM PRIMARY MOTOR CORTEX BA4A R
49. GM PRIMARY MOTOR CORTEX BA4P L
50. GM PRIMARY MOTOR CORTEX BA4P R
51. GM PRIMARY SOMATOSENSORY CORTEX BA1 L
52. GM PRIMARY SOMATOSENSORY CORTEX BA1 R
53. GM PRIMARY SOMATOSENSORY CORTEX BA2 L
54. GM PRIMARY SOMATOSENSORY CORTEX BA2 R
55. GM PRIMARY SOMATOSENSORY CORTEX BA3A L
56. GM PRIMARY SOMATOSENSORY CORTEX BA3A R
57. GM PRIMARY SOMATOSENSORY CORTEX BA3B L
58. GM PRIMARY SOMATOSENSORY CORTEX BA3B R
59. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP1 L
60. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP1 R
61. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP2 L
62. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP2 R
63. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP3 L
64. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP3 R
65. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP4 L

66. GM SECONDARY SOMATOSENSORY CORTEX / PARIETAL OPERCULUM OP4 R
67. GM SUPERIOR PARIETAL LOBULE 5CI L
68. GM SUPERIOR PARIETAL LOBULE 5CI R
69. GM SUPERIOR PARIETAL LOBULE 5L L
70. GM SUPERIOR PARIETAL LOBULE 5L R
71. GM SUPERIOR PARIETAL LOBULE 5M L
72. GM SUPERIOR PARIETAL LOBULE 5M R
73. GM SUPERIOR PARIETAL LOBULE 7A L
74. GM SUPERIOR PARIETAL LOBULE 7A R
75. GM SUPERIOR PARIETAL LOBULE 7M L
76. GM SUPERIOR PARIETAL LOBULE 7M R
77. GM SUPERIOR PARIETAL LOBULE 7PC L
78. GM SUPERIOR PARIETAL LOBULE 7PC R
79. GM SUPERIOR PARIETAL LOBULE 7P L
80. GM SUPERIOR PARIETAL LOBULE 7P R
81. GM VISUAL CORTEX V1 BA17 L
82. GM VISUAL CORTEX V1 BA17 R
83. GM VISUAL CORTEX V2 BA18 L
84. GM VISUAL CORTEX V2 BA18 R
85. GM VISUAL CORTEX V3V L
86. GM VISUAL CORTEX V3V R
87. GM VISUAL CORTEX V4 L
88. GM VISUAL CORTEX V4 R
89. GM VISUAL CORTEX V5 L

90. GM VISUAL CORTEX V5 R
91. GM PREMOTOR CORTEX BA6 L
92. GM PREMOTOR CORTEX BA6 R
93. WM ACOUSTIC RADIATION R
94. WM ACOUSTIC RADIATION L
95. WM CALLOSAL BODY
96. WM CINGULUM R
97. WM CINGULUM L
98. WM CORTICOSPINAL TRACT R
99. WM CORTICOSPINAL TRACT L
100. WM FORNIX
101. WM INFERIOR OCCIPITO-FRONTAL FASCICLE R
102. WM INFERIOR OCCIPITO-FRONTAL FASCICLE L
103. GM LATERAL GENICULATE BODY R
104. GM LATERAL GENICULATE BODY L
105. GM MAMILLARY BODY
106. GM MEDIAL GENICULATE BODY R
107. GM MEDIAL GENICULATE BODY L
108. WM OPTIC RADIATION R
109. WM OPTIC RADIATION L
110. WM SUPERIOR LONGITUDINAL FASCICLE R
111. WM SUPERIOR LONGITUDINAL FASCICLE L
112. WM SUPERIOR OCCIPITO-FRONTAL FASCICLE R
113. WM SUPERIOR OCCIPITO-FRONTAL FASCICLE L
114. WM UNCINATE FASCICLE R

115. WM UNCINATE FASCICLE L

116. GM INSULA ID1 L

117. GM INSULA ID1 R

118. GM INSULA IG1 L

119. GM INSULA IG1 R

120. GM INSULA IG2 L

121. GM INSULA IG2 R

6.1.3 Harvard-Oxford Atlas

1. RIGHT FRONTAL POLE
2. LEFT FRONTAL POLE
3. LEFT INSULAR CORTEX
4. RIGHT INSULAR CORTEX
5. LEFT SUPERIOR FRONTAL GYRUS
6. RIGHT SUPERIOR FRONTAL GYRUS
7. LEFT MIDDLE FRONTAL GYRUS
8. RIGHT MIDDLE FRONTAL GYRUS
9. LEFT INFERIOR FRONTAL GYRUS, PARS TRIANGULARIS
10. RIGHT INFERIOR FRONTAL GYRUS, PARS TRIANGULARIS
11. LEFT INFERIOR FRONTAL GYRUS, PARS OPERCULARIS
12. RIGHT INFERIOR FRONTAL GYRUS, PARS OPERCULARIS
13. LEFT PRECENTRAL GYRUS
14. RIGHT PRECENTRAL GYRUS
15. LEFT TEMPORAL POLE
16. RIGHT TEMPORAL POLE
17. LEFT SUPERIOR TEMPORAL GYRUS, ANTERIOR DIVISION
18. RIGHT SUPERIOR TEMPORAL GYRUS, ANTERIOR DIVISION
19. LEFT SUPERIOR TEMPORAL GYRUS, POSTERIOR DIVISION
20. RIGHT SUPERIOR TEMPORAL GYRUS, POSTERIOR DIVISION
21. LEFT MIDDLE TEMPORAL GYRUS, ANTERIOR DIVISION
22. RIGHT MIDDLE TEMPORAL GYRUS, ANTERIOR DIVISION
23. LEFT MIDDLE TEMPORAL GYRUS, POSTERIOR DIVISION
24. RIGHT MIDDLE TEMPORAL GYRUS, POSTERIOR DIVISION

25. LEFT MIDDLE TEMPORAL GYRUS, TEMPOROOCCIPITAL PART
26. RIGHT MIDDLE TEMPORAL GYRUS, TEMPOROOCCIPITAL PART
27. LEFT INFERIOR TEMPORAL GYRUS, ANTERIOR DIVISION
28. RIGHT INFERIOR TEMPORAL GYRUS, ANTERIOR DIVISION
29. LEFT INFERIOR TEMPORAL GYRUS, POSTERIOR DIVISION
30. RIGHT INFERIOR TEMPORAL GYRUS, POSTERIOR DIVISION
31. LEFT INFERIOR TEMPORAL GYRUS, TEMPOROOCCIPITAL PART
32. RIGHT INFERIOR TEMPORAL GYRUS, TEMPOROOCCIPITAL PART
33. LEFT POSTCENTRAL GYRUS
34. RIGHT POSTCENTRAL GYRUS
35. LEFT SUPERIOR PARIETAL LOBULE
36. RIGHT SUPERIOR PARIETAL LOBULE
37. LEFT SUPRAMARGINAL GYRUS, ANTERIOR DIVISION
38. RIGHT SUPRAMARGINAL GYRUS, ANTERIOR DIVISION
39. LEFT SUPRAMARGINAL GYRUS, POSTERIOR DIVISION
40. RIGHT SUPRAMARGINAL GYRUS, POSTERIOR DIVISION
41. LEFT ANGULAR GYRUS
42. RIGHT ANGULAR GYRUS
43. LEFT LATERAL OCCIPITAL CORTEX, SUPERIOR DIVISION
44. RIGHT LATERAL OCCIPITAL CORTEX, SUPERIOR DIVISION
45. LEFT LATERAL OCCIPITAL CORTEX, INFERIOR DIVISION
46. RIGHT LATERAL OCCIPITAL CORTEX, INFERIOR DIVISION
47. LEFT INTRACALCARINE CORTEX
48. RIGHT INTRACALCARINE CORTEX

49. LEFT FRONTAL MEDIAL CORTEX
50. RIGHT FRONTAL MEDIAL CORTEX
51. LEFT JUXTAPOSITIONAL LOBULE CORTEX (FORMERLY SUPPLEMENTARY MOTOR CORTEX)
52. RIGHT JUXTAPOSITIONAL LOBULE CORTEX (FORMERLY SUPPLEMENTARY MOTOR CORTEX)
53. LEFT SUBCALLOSAL CORTEX
54. RIGHT SUBCALLOSAL CORTEX
55. LEFT PARACINGULATE GYRUS
56. RIGHT PARACINGULATE GYRUS
57. LEFT CINGULATE GYRUS, ANTERIOR DIVISION
58. RIGHT CINGULATE GYRUS, ANTERIOR DIVISION
59. LEFT CINGULATE GYRUS, POSTERIOR DIVISION
60. RIGHT CINGULATE GYRUS, POSTERIOR DIVISION
61. LEFT PRECUNEOUS CORTEX
62. RIGHT PRECUNEOUS CORTEX
63. LEFT CUNEAL CORTEX
64. RIGHT CUNEAL CORTEX
65. LEFT FRONTAL ORBITAL CORTEX
66. RIGHT FRONTAL ORBITAL CORTEX
67. LEFT PARAHIPPOCAMPAL GYRUS, ANTERIOR DIVISION
68. RIGHT PARAHIPPOCAMPAL GYRUS, ANTERIOR DIVISION
69. LEFT PARAHIPPOCAMPAL GYRUS, POSTERIOR DIVISION
70. RIGHT PARAHIPPOCAMPAL GYRUS, POSTERIOR DIVISION
71. LEFT LINGUAL GYRUS

72. RIGHT LINGUAL GYRUS
73. LEFT TEMPORAL FUSIFORM CORTEX, ANTERIOR DIVISION
74. RIGHT TEMPORAL FUSIFORM CORTEX, ANTERIOR DIVISION
75. LEFT TEMPORAL FUSIFORM CORTEX, POSTERIOR DIVISION
76. RIGHT TEMPORAL FUSIFORM CORTEX, POSTERIOR DIVISION
77. LEFT TEMPORAL OCCIPITAL FUSIFORM CORTEX
78. RIGHT TEMPORAL OCCIPITAL FUSIFORM CORTEX
79. LEFT OCCIPITAL FUSIFORM GYRUS
80. RIGHT OCCIPITAL FUSIFORM GYRUS
81. LEFT FRONTAL OPERCULUM CORTEX
82. RIGHT FRONTAL OPERCULUM CORTEX
83. LEFT CENTRAL OPERCULAR CORTEX
84. RIGHT CENTRAL OPERCULAR CORTEX
85. LEFT PARIETAL OPERCULUM CORTEX
86. RIGHT PARIETAL OPERCULUM CORTEX
87. LEFT PLANUM POLARE
88. RIGHT PLANUM POLARE
89. LEFT HESCHL'S GYRUS (INCLUDES H1 AND H2)
90. RIGHT HESCHL'S GYRUS (INCLUDES H1 AND H2)
91. LEFT PLANUM TEMPORALE
92. RIGHT PLANUM TEMPORALE
93. LEFT SUPRACALCARINE CORTEX
94. RIGHT SUPRACALCARINE CORTEX
95. LEFT OCCIPITAL POLE
96. RIGHT OCCIPITAL POLE

6.2 List of Publications

In this section there are the following 10 subsections:

- the work titled FSL-Hybrid Atlas Promotes Activation Weighted Vector Analysis in Functional Neuroradiology, i.e. the presentation of the hybrid atlas obtained with the integration of two well-known brain atlas (the Juelich one and the Harvard-Oxford one) 4.1
- the work titled Quantitative Relations between CRAIIM Human Brain Altases;
- the work titled Cluster Analysis of Functional Neuroimages using Data Reduction and Competitive Learning Algorithms, i.e., the abstract is skipped because is the same in Results section 4.2;
- the work titled Computation and Management of Weighted Activation Vectors in Support to fMRI Analysis of Clinical Subjects
- the work titled Accuracy Evaluation of Soft Classifiers using Interval Type-2 Fuzzy Sets Framework
- the work titled A Soft Davies-Bouldin Separation Measure, i.e., the abstract is skipped because is the same in Results section 4.3;
- the work titled Resting State Functional Connectivity Analysis Using Soft Competitive Learning Algorithms
- the work titled Resting State fMRI Analysis Using Unsupervised Learning Algorithms, i.e., the abstract is skipped because is the same in Results section 4.3
- the work titled Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms, i.e., the abstract is skipped because is the same in Results section 4.3

Research paper 1

Title FSL-Hybrid Atlas Promotes Activation Weighted Vector Analysis in Functional Neuroradiology

Authors Alberto Arturo Ver gani, Sabina Strocchi, Renzo Minotto, Elisabetta Binaghi

Year 2016

Abstract This work is the integration of the Juelich brain atlas with the Harvard-Oxford brain atlas. The motivation to do this operation is the absence of temporal and frontal regions in the Juelich one, that instead are present in the Harvard-Oxford model. The hybridization procedure has been done having as reference the Juelich atlas and as adapted atlas the Harvard-Oxford. Therefore, we selected the all 121 ROIs of Juelich and only 40 ROIs from Harvard-Oxford, resulting an hybrid atlas we named CRAIIM[2] Hybrid Atlas. The benefits of this hybrid atlas are the integration of fundamental neuroanatomy models useful for co-registration that in the standard template were absent, e.g. many frontal and temporal cortexes, subcallosal portions, cingulate gyrus and thalamus halves. The limitation is that these last regions are in some cases a minor proportion of the Harvard-Oxford template. The 161 regions cover all of the main important brain territories, satisfying the right trade-off between exhaustion power and visual usability. It has neuroradiological utility, but some kind of incompleteness for all brain region representation: it is missing of special structure like cerebellum and, in general, has some portion that is a percentage of its original atlases. Other own quality regards the resolution about only 1mm. Future work is planned to generalize the atlas building procedure with which to easily generate solutions for diversified functional neuroradiological applications honouring the salience of existing specialized template.

Key-words Atlasing; FSL; Hybridization; Functional Neuroimaging; Brain Representation; Brain Modelling; Computational Neuroanatomy.

Citation See [254]

[2]CRAIIM is the Italian acronym that means *Centro di Ricerche in Analisi di Immagini e Informatica Medica*, i.e., in English: Centre of Research in Image Analysis and Medical Informatics.

Research paper 2

Title Quantitative Relations between CRAIIM Human Brain Atlases

Authors Alberto Arturo Vergani, Sabina Strocchi, Renzo Minotto, Elisabetta Binaghi

Abstract See the paper in the Results Section 4.1

Key-words Atlasing; FSL; Hybridization; Functional Neuroimaging; Descriptive Statistics; Voxel-Based Morphometry.

Citation See [255]

Research paper 3

Title Cluster Analysis of Functional Neuroimages using Data Reduction and Competitive Learning Algorithms

Authors Alberto Arturo Vergani, Samuele Martinelli and Elisabetta Binaghi

Abstract See the paper in the Results Section 4.2

Key-words fMRI, Statistical Parametric Mapping, Data Reduction, Self Organizing Map, Growing Neural Gas

Citation [252]

Research paper 4

Title Computation and Management of Weighted Activation Vectors in Support to fMRI Analysis of Clinical Subjects

Authors Elisabetta Binaghi, Alberto Arturo Vergani, Andrea Montalbetti, Renzo Minotto, Valentina Pedoia, Sabina Strocchi and Sergio Balbi

Abstract In the present work, we investigate the usefulness of a new representation of the results obtained by fMRI data analysis, named Weighted Activation Vector built on the top of Statistical Parametric Mapping. A software package for the genera-tion and management of Weighted Activation Vectors is illustrated. It is de-signed to support single subject, multitemporal and collective brain tumour studies. As seen in our experimental context, the allied use of Weighted Activation Vectors and Statistical Parametric Maps improves the quality of medical decisions before and after the neurosurgical practice. Clustering techniques ap-plied to Weighted Activation Vectors can be efficiently analysed and optimised in an attempt to discover relevant properties of collective data

Key-Words fMRI, Statistical Parametric Map, Data Reduction, Clustering, Brain Tumour Studies, Medical Software

Citation [20]

Research paper 5

Title Accuracy Evaluation of Soft Classifiers using Interval Type-2 Fuzzy Sets Framework

Authors Elisabetta Binaghi, Alberto Arturo Vergani, Valentina Pedoia

Abstract This paper proposes a new accuracy evaluation method within a behavioural comparison strategy which uses interval type-2 fuzzy sets and derived operations to model reference data and define soft accuracy indexes. The method addresses the case in which grades of membership, collected by surveying experts, will often be different for the same reference pattern, because the experts will not necessarily be in agreement. The approach is illustrated using simple examples and an application in the domain of biomedical image segmentation.

Key-words Indexes; Fuzzy Sets; Uncertainty; Data Models; Image Segmentation

Citation [21]

Research paper 6

Title A Soft Davies-Bouldin Separation Measure

Authors Alberto Arturo Vergani and Elisabetta Binaghi

Abstract See the paper in the Results Section 4.3

Key-words clustering, k-means, c-means, separation measures, Davies-Bouldin index, fMRI

Citation [250]

Research paper 7

Title A Cloud Fuzzy Logic Framework for Oral Disease Risk Assessment

Authors Gloria Gonella, Elisabetta Binaghi, Alberto Arturo Vergani, Irene Biotti, Luca Levrini

Abstract This paper presents a fuzzy logic framework for dental caries and erosion risk assessment. Two interdependent modules are implemented within a cloud architecture. The first module is a fuzzy expert system designed for physicians and expert users, able to provide an active support in formulating risk judgements. The second module is oriented to generic users for oral health promotion. Conceptual ingredients of the fuzzy logic framework are principally defined by eliciting knowledge from a group of experts. The generation of rules involves both structured interviews and data driven learning procedures based on the use of neurofuzzy techniques.

Key-words Fuzzy Expert System, Medical Knowledge Acquisition, Neurofuzzy Inference System, Cloud Computing

Citation [88]

Research paper 8

Title Resting State Functional Connectivity Analysis Using Soft Competitive Learning Algorithms

Authors Alberto Arturo Vergani, Elisabetta Binaghi, Samuele Martinelli, Sabina Strocchi

Abstract RS-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.

Key-words RS-fMRI, Functional Connectivity, Clustering, Soft Competitive Learning.

Citation [253]

Research paper 9

Title Resting State fMRI Analysis Using Unsupervised Learning Algorithms

Authors Alberto Arturo Vergani, Samuele Martinelli and Elisabetta Binaghi

Abstract See directly the paper in the Results Section 4.3

Key-words Functional Connectivity; Competitive Clustering; Self Organizing Map; Neural Gas; Growing Neural Gas; Davies-Bouldin Index; Spectral Methods; Principal Component Analysis; Nonlinear Robust Fuzzy Principal Component Analysis; Random Matrix Theory

Citation submitted to Computerized Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, Taylor and Francis

Research paper 10

Title Clustering Functional MRI Patterns with Fuzzy and Competitive Algorithms

Authors Alberto Arturo Vergani, Samuele Martinelli and Elisabetta Binaghi

Abstract See the paper in the Results Section 4.3

Key-words fMRI, Partitive Clustering, Fuzzy c-means Algorithm, Neural Gas Algorithm

Citation [255]

6.3 List of Algorithms

In this section there is the presentation of the algorithms proposed in the Backgrounds section 2.2 of the Computational Intelligence chapter:

- the Fuzzy C-Means Type 1 (FCM Type 1) 1;
- the Fuzzy C-Means Type 2 (FCM Type 2)2;
- the Self Organizing Maps (SOM) 3;
- the Neural Gas Algorithm (NG) 4;
- the Growing Neural Gas Algorithm (GNG) 5.

Algorithm 1 Fuzzy C-Means TYPE 1 Algorithm

Require: set parameters

c number of cluster

m fuzzy partition matrix exponent to control the degree of fuzzy overlap $(m > 1)$

d distance

$\wp$ fuzzy pseudo partition

$l := 0$ iteration number

Ensure: minimize the fuzzy c-means functional:

$$J_m(\wp_c) = \sum_{i=1}^{r}\sum_{j=1}^{c}(R_j(\boldsymbol{u_i}))^m \cdot d^2(\boldsymbol{u}_i, \boldsymbol{s}_j) \tag{6.1}$$

1: compute vectors of the centres $\boldsymbol{s}_1^l, \ldots, \boldsymbol{s}_c^l$ with

$$\boldsymbol{s}_j = \left(\frac{\sum_{i=1}^{r}(R_j(\boldsymbol{u}_i))^m \cdot u_{in}}{\sum_{i=1}^{r}(R_j(\boldsymbol{u}_i))^m}\right) \tag{6.2}$$

2: $\forall i = 1, \ldots, r$ modify the fuzzy pseudopartion $\wp$ as follows:

1. if $d(\boldsymbol{u}_i, \boldsymbol{s}_j) > 0 \quad \forall j = 1, \ldots, c$, then put:

$$R_j^{(l+1)}(\boldsymbol{u}_i) = \sum_{k=1}^{c}\left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}\right)^{-(\frac{2}{m-1})} \tag{6.3}$$

2. Let $J = j_1, \ldots, j_p$, $p \leq c$, be a set of subscripts such that $d(\boldsymbol{u}_i, \boldsymbol{s}_j) = 0 \ \forall j \in J$. Then put $R_j^{l+1}(\boldsymbol{u}_i)$ equal to any number from interval $[0, 1]$ in such a way that the condition

$$\sum_{j=j_i}^{j_p} R_j^{l+1}(\boldsymbol{u}_i) = 1 \tag{6.4}$$

is fulfilled. For the remaining $j \in \{1, \ldots, c\} \setminus J$, put $R_j^{l+1}(\boldsymbol{u}_i) = 0$.

3: Compare fuzzy pseudopartitions $\wp_c^{l+1}$ and $\wp_c^l$ using the distance

$$||_c^{l+1}{}_{,c}^{\,l}|| = max_{\substack{j=1,\ldots,c \\ i=1,\ldots,r}}|R_j^{(l+1)}(\boldsymbol{u}_j) - R_j^{(l)}(\boldsymbol{u}_i)|. \tag{6.5}$$

4: **if** $||\wp_c^{(l+1)}, \wp_c^{(l)}|| \leq \epsilon$ holds **then**
5: terminate the process;
6: **else**
7: in the opposite case, set $l := l + 1$ and return to step 2.
8: **end if**

Algorithm 2 Fuzzy C-Means TYPE 2 Algorithm

Require: set parameters

Ensure: minimize the fuzzy c-means functional:

$$J_m(\wp_c) = \sum_{i=1}^{r}\sum_{j=1}^{c}(A_j(\boldsymbol{u_i}))^m \cdot d^2(\boldsymbol{u}_i, \boldsymbol{s}_j) \tag{6.6}$$

1: compute vectors of the centres $\boldsymbol{s}_1^l, \ldots, \boldsymbol{s}_c^l$ with

$$\boldsymbol{s}_j = \left(\frac{\sum_{i=1}^{r}(A_j(\boldsymbol{u}_i))^m \cdot u_{in}}{\sum_{i=1}^{r}(A_j(\boldsymbol{u}_i))^m}\right) \tag{6.7}$$

2: $\forall i = 1, \ldots, r$ modify the fuzzy pseudopartion $\wp$ as follows:

1. if $d(\boldsymbol{u}_i, \boldsymbol{s}_j) > 0 \quad \forall j = 1, \ldots, c$, then put:

$$A_j^{(l+1)}(\boldsymbol{u}_i) = \left[\sum_{k=1}^{c}\left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}\right)^{-(\frac{2}{m-1})}\right] - \left[1 - \left[\sum_{k=1}^{c}\left(\frac{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}{d(\boldsymbol{u}_i, \boldsymbol{s}_j^{(l)})}\right)^{\frac{2}{m-1}}\right]^2\right] \tag{6.8}$$

2. Let $J = j_1, \ldots, j_p$, $p \leq c$, be a set of subscripts such that $d(\boldsymbol{u}_i, \boldsymbol{s}_j) = 0 \ \forall j \in J$. Then put $A_j^{l+1}(\boldsymbol{u}_i)$ equal to any number from interval $[0, 1]$ in such a way that the condition

$$\sum_{j=j_i}^{j_p} A_j^{l+1}(\boldsymbol{u}_i) = 1 \tag{6.9}$$

is fulfilled. For the remaining $j \in \{1, \ldots, c\} \setminus J$, put $A_j^{l+1}(\boldsymbol{u}_i) = 0$.

3: Compare fuzzy pseudopartitions $\wp_c^{l+1}$ and $\wp_c^l$ using the distance

$$||_c^{l+1}, {}_c^l|| = max_{\substack{j=1,\ldots,c \\ i=1,\ldots,r}} |A_j^{(l+1)}(\boldsymbol{u}_j) - A_j^{(l)}(\boldsymbol{u}_i)|. \tag{6.10}$$

4: **if** $||\wp_c^{(l+1)}, \wp_c^{(l)}|| \leq \epsilon$ holds **then**

5: terminate the process;

6: **else**

7: in the opposite case, set $l := l + 1$ and return to step 2.

8: **end if**

Algorithm 3 The Self Organizing Map Algorithm

Require: Set parameters ...

Ensure: minimize the SOM functional:

1: Initialize the set $A = \{c_1, c_2, \ldots, c_N\}$ with $N = N_1 \cdot N_2$ units c_i with reference vectors $\boldsymbol{w}_c \in \Re^n$ chosen according to $p(\boldsymbol{\xi})$

2: Initialize the connection C to shape a rectangular grid $N_1 \times N_2$

3: Initialize the time parameter $t = 0$

4: Generate the input signal $\boldsymbol{\xi}$ according to $p(\boldsymbol{\xi})$

5: Determine the winner $s(\boldsymbol{\xi}) = s$

$$s(\boldsymbol{\xi}) = \arg\min_{c \in A} ||\boldsymbol{\xi} - \boldsymbol{w}_c|| \tag{6.11}$$

6: Adapt each unit r according to:

$$\triangle \boldsymbol{w}_r = \epsilon(t) \cdot (\boldsymbol{\xi} - \boldsymbol{w}_r) \tag{6.12}$$

$$\sigma(t) = \sigma_i \cdot (\sigma_f / \sigma_i)^{t/t_{max}} \tag{6.13}$$

$$\epsilon(t) = \epsilon_i \cdot (\epsilon_f / \epsilon_i)^{t/t_{max}} \tag{6.14}$$

7: Increase the time parameter $t = t + 1$

8: **if** $t < t_{max}$ **then**

9: continue with step 2.

10: **end if**

Algorithm 4 The Neural Gas Algorithm

Require: Select values for time-dependent parameters:

λ_i neighbourhood range

ϵ_f adaptation step size of f

ϵ_i adaptation step size of i

Ensure: minimize the NG functional:

1: Choose the reference vectors $\mathbf{w}_{c_i} \in \Re^n$ according to the probability $p(\boldsymbol{\xi})$
2: Initialize the set $A = \{c_1, c_2, \ldots, c_N\}$ with N units c_i associated to $\mathbf{w}_{c_i}$
3: Initialize time parameter $t = 0$
4: Generate an input signal $\boldsymbol{\xi}$ according to $p(\boldsymbol{\xi})$
5: Order the element of A according to their distance to $\boldsymbol{\xi}$ {i.e., find the sequence of indices $(i_0, i_1, \ldots, i_{N-1})$ such that $\boldsymbol{w}_{i_k}$, with $k = 0, \ldots, N-1$, is the reference vectors such that $\exists\, k$ vectors $\boldsymbol{w}_j$ with $||\boldsymbol{\xi} - \boldsymbol{w}_j|| < ||\boldsymbol{\xi} - \boldsymbol{w}_k||$ }
6: Adapt the reference vector according to $\triangle \boldsymbol{w}_i = \epsilon(t) \cdot h_\lambda(k_i(\boldsymbol{\xi}, A)) \cdot (\boldsymbol{\xi}, \boldsymbol{w}_i)$. Note that:

$k_i(\boldsymbol{\xi}, A)$ is the number k associated with $\boldsymbol{w}_i$

$\epsilon(t) = \epsilon_i(\epsilon_i/\epsilon_f)^{t/t_{max}}$

$\lambda(t) = \lambda_i(\lambda_f/\lambda_i)^{t/t_{max}}$

$h_\lambda(k) = \exp(-k/\lambda(t))$

7: Increase the time parameter $t = t + 1$
8: **if** $t < t_{max}$ **then**
9: continue with step 2.
10: **end if**

Algorithm 5 The Growing Neural Gas Algorithm

Require: Select values for simulation parameters (see [152] to use evolutionary algorithms to compute parameters)

Ensure: minimize the GNG functional:

1: Initialize the set $A = \{c_1, c_2\}$ with reference vectors $\boldsymbol{w}_c$ chosen by $p(\boldsymbol{\xi})$

2: Initialize the connection set $C \subseteq A \times A = \emptyset$

3: Generate the input signal $\boldsymbol{\xi}$ according the probability $p(\boldsymbol{\xi})$

4: Determine the first-nearest unit s_1 (i.e. the winner) and the second-nearest $s_2(s_1, s_2 \in A)$ by

$$s_1 = \arg\min_{c \in A} ||\boldsymbol{\xi} - \boldsymbol{w}_c|| \tag{6.15}$$

$$s_2 = \arg\min_{c \in A \setminus \{s_1\}} ||\boldsymbol{\xi} - \boldsymbol{w}_c|| \tag{6.16}$$

5: **if** $\nexists$ the connection C **then**

6: create $C = C \cup \{(s_{1,s_2}\})\}$ and set the age of connection $age_{(s1,s_2)} = 0$

7: **end if**

8: Add the squared distance between $\boldsymbol{\xi}$ and $\boldsymbol{w}_{s_1}$ to a local error variable:

$$\triangle E_{s_1} = ||\boldsymbol{\xi} - \boldsymbol{w}_{s_1}|| \tag{6.17}$$

9: Adapt the reference vectors $\boldsymbol{w}_{s_1}$ and its neighbors $\boldsymbol{w_i}$ to ϵ_b and ϵ_n:

$$\triangle \boldsymbol{w}_{s_1} = \epsilon_b \cdot (\boldsymbol{\xi} - \boldsymbol{w}_{s_1}) \tag{6.18}$$

$$\triangle \boldsymbol{w}_1 = \epsilon_n \cdot (\boldsymbol{\xi} - \boldsymbol{w}_i) \quad (\forall_i \in N_{s_1} = \text{set of topological neighbors of } s_1) \tag{6.19}$$

10: Increment the age of all edges from s_1:

$$age_{(s_1,i)} = age_{s_1,i} + 1 \quad (\forall_i \in N_{s_1}) \tag{6.20}$$

11: Remove edge with $age > \alpha_{max}$ and remove units without edges

12: **if** the cardinality of $\boldsymbol{\xi}$ is an integer multiple of λ **then**

13:
- determine the unite $q = \arg\max_{c \in A} E_c$
- among the neighbors of q, determine the unite $f = \arg\max_{c \in N_q} E_c$
- add unit r and interpolate its reference vector $\boldsymbol{w_r}$ from q and f
- insert edges connecting the new unit r with units q and f and remove the original edge between q and f
- decrease the error variables of q and f by a fraction of α
- interpolate the error variable of r from q and f

14: **end if**

15: Decrease the error variables of all units

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] Abbott, L. F., and Regehr, W. G. Synaptic computation. *Nature 431*, 7010 (oct 2004), 796–803.

[2] Aguirre, G., Zarahn, E., and D'esposito, M. The variability of human, bold hemodynamic responses. *Neuroimage 8*, 4 (1998), 360–369.

[3] Andersen, A. H., Gash, D. M., and Avison, M. J. Principal component analysis of the dynamic response measured by fmri: a generalized linear systems framework. *Magn. Reson. Imaging 17*, 6 (1999), 795–815.

[4] Anderson, P. W. Absence of diffusion in certain random lattices. *Physical review 109*, 5 (1958), 1492.

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

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

[7] Babajani, A., Nekooei, M.-H., and Soltanian-Zadeh, H. Integrated meg and fmri model: synthesis and analysis. *Brain topography 18*, 2 (2005), 101–113.

[8] Babajani-Feremi, A., and Soltanian-Zadeh, H. Multi-area neural mass modeling of eeg and meg signals. *Neuroimage 52*, 3 (2010), 793–811.

[9] Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. Time course epi of human brain function

during task activation. *Magnetic resonance in medicine 25*, 2 (1992), 390–397.

[10] Bassett, D. S., and Bullmore, E. Small-world brain networks. *The neuroscientist 12*, 6 (2006), 512–523.

[11] Bassett, D. S., and Gazzaniga, M. S. Understanding complexity in the human brain. *Trends in Cognitive Sciences 15*, 5 (may 2011), 200–209.

[12] Bauer, H.-U., and Villmann, T. Growing a hypercubical output space in a self-organizing feature map. *IEEE transactions on neural networks 8*, 2 (1997), 218–226.

[13] Beckmann, C. F., Jenkinson, M., and Smith, S. M. General multilevel linear modeling for group analysis in fmri. *Neuroimage 20*, 2 (2003), 1052–1063.

[14] Beckmann, C. F., and Smith, S. M. Tensorial extensions of independent component analysis for multisubject fmri analysis. *Neuroimage 25*, 1 (2005), 294–311.

[15] Benjamini, Y., and Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* (1995), 289–300.

[16] Bezdek, J. C. Cluster validity. In *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer, 1981, pp. 95–154.

[17] Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithms*. Springer, 1981.

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

[19] Bezdek, J. C., Moshtaghi, M., Runkler, T., and Leckie, C. The generalized c index for internal fuzzy cluster validity. *IEEE Transactions on Fuzzy Systems 24*, 6 (2016), 1500–1512.

[20] Binaghi, E., Vergani, A. A., Montalbetti, A., Minotto, R., Pedoia, V., Strocchi, S., and Balbi, S. Computation and management of weighted activation vectors in support to fmri analysis of

clinical subjects. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 0*, 0 (2018), 1–20.

[21] Binaghi, E., Vergani, A. A., and Pedoia, V. Accuracy evaluation of soft classifiers using interval type-2 fuzzy sets framework. Institute of Electrical and Electronics Engineers Inc. cited By 0; Conference of 2017 IEEE International Conference on Fuzzy Systems, FUZZ 2017 ; Conference Date: 9 July 2017 Through 12 July 2017; Conference Code:130106.

[22] Birn, R. M., Saad, Z. S., and Bandettini, P. A. Spatial heterogeneity of the nonlinear dynamics in the fmri bold response. *Neuroimage 14*, 4 (2001), 817–826.

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

[24] Biswal, B. B., Kylen, J. V., and Hyde, J. S. Simultaneous assessment of flow and bold signals in resting-state functional connectivity maps. *NMR Biomed. 10*, 45 (1997), 165–170.

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

[26] Blumenthal, J. a., Babyak, M. a., Moore, K. a., Craighead, W. E., Herman, S., Khatri, P., Waugh, R., Napolitano, M. a., Forman, L. M., Appelbaum, M., Doraiswamy, P. M., and Krishnan, K. R. Effects of exercise training on older patients with major depression. *Arch. Intern. Med. 159*, 19 (#oct# 1999), 2349–56.

[27] Bohland, J. W., Bokil, H., Allen, C. B., and Mitra, P. P. The brain atlas concordance problem: Quantitative comparison of anatomical parcellations. *PLoS ONE 4*, 9 (sep 2009), e7200.

[28] Bojak, I., Oostendorp, T. F., Reid, A. T., and Kötter, R. Towards a model-based integration of co-registered electroencephalography/functional magnetic resonance imaging data with realistic neural

population meshes. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 369*, 1952 (2011), 3785–3801.

[29] Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. Linear systems analysis of functional magnetic resonance imaging in human v1. *The journal of neuroscience 16*, 13 (1996), 4207–4221.

[30] Bringsjord, S., and Govindarajulu, N. S. Artificial intelligence. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., fall 2018 ed. Metaphysics Research Lab, Stanford University, 2018.

[31] Brockwell. *Time Series and Forecasting*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 12 1998.

[32] Bronshtein, I., and Semendyaev, K. Mathematics handbook. *GITTL, Moscow* (1986).

[33] Buckner, R. L., and Vincent, J. L. Unrest at rest: default activity and spontaneous network correlations. *Neuroimage 37*, 4 (2007), 1091–1096.

[34] Bullmore, E., and Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci. 10*, 3 (2009), 186–198.

[35] Burda, Z., Kornelsen, J., Nowak, M. A., Porebski, B., Sboto-Frankenstein, U., Tomanek, B., and Tyburczyk, J. Collective correlations of brodmann areas fmri study with rmt-denoising. *arXiv preprint arXiv:1306.3825* (2013).

[36] Buxton, R. B., Wong, E. C., and Frank, L. R. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn. Reson. Med. 39*, 6 (1998), 855–864.

[37] Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. A method for making group inferences from functional mri data using independent component analysis. *Hum. Brain Mapp. 14*, 3 (2001), 140–151.

[38] Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms. *Hum. Brain Mapp. 13*, 1 (2001), 43–53.

[39] Caliński, T., and Harabasz, J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods 3*, 1 (1974), 1–27.

[40] Caporale, N., and Dan, Y. Spike timing–dependent plasticity: a hebbian learning rule. *Annu. Rev. Neurosci. 31* (2008), 25–46.

[41] Cattell, R. B. The description of personality: basic traits resolved into clusters. *The journal of abnormal and social psychology 38*, 4 (1943), 476.

[42] Cieslik, E. C., Zilles, K., Caspers, S., Roski, C., Kellermann, T. S., Jakobs, O., Langner, R., Laird, A. R., Fox, P. T., and Eickhoff, S. B. Is there "one" DLPFC in cognitive action control? evidence for heterogeneity from co-activation-based parcellation. *Cerebral Cortex 23*, 11 (aug 2012), 2677–2689.

[43] Colquhoun, D. An investigation of the false discovery rate and the misinterpretation of p-values. *Royal Society open science 1*, 3 (2014), 140216.

[44] Craddock, R. C., James, G., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping 33*, 8 (jul 2011), 1914–1928.

[45] Dale, A. M. Optimal experimental design for event-related fmri. *Hum. Brain Mapp. 8*, 2-3 (1999), 109–114.

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

[47] De Morgan, A. Induction (mathematics). *Penny Cyclopaedia of the Society for the Diffusion of Useful Knowledge 12* (1838), 465–466.

[48] De Morgan, A., and Heath, P. On the syllogism and other logical writings.

[49] Deco, G., Jirsa, V. K., and McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience 12*, 1 (2011), 43.

[50] Deneux, T., and Faugeras, O. Eeg-fmri fusion of paradigm-free activity using kalman filtering. *Neural computation 22*, 4 (2010), 906–948.

[51] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage 31*, 3 (jul 2006), 968–980.

[52] Dimitriadou, E., Barth, M., Windischberger, C., Hornik, K., and Moser, E. A quantitative comparison of functional mri cluster analysis. *Artificial intelligence in medicine 31*, 1 (2004), 57–71.

[53] Dimitriadou, E., Dolničar, S., and Weingessel, A. An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika 67*, 1 (2002), 137–159.

[54] Ding, C., and He, X. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning* (2004), ACM, p. 29.

[55] Driver, H. E., and Kroeber, A. L. *Quantitative expression of cultural relationships*. University of California Press, 1932.

[56] Duda, R. O., and Hart, P. E. *Pattern Elesslfication and Scene Analysis*. Wiley, 1973.

[57] Duda, R. O., Hart, P. E., et al. *Pattern classification and scene analysis*, vol. 3. Wiley New York, 1973.

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

[59] Dunn, J. C. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters.

[60] Dunn, J. C. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics 4*, 1 (1974), 95–104.

[61] Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., and Fox, P. T. Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *NeuroImage 57*, 3 (aug 2011), 938–949.

[62] Eickhoff, S. B., Rottschy, C., Kujovic, M., Palomero-Gallagher, N., and Zilles, K. Organizational principles of human visual cortex revealed by receptor mapping. *Cerebral Cortex 18*, 11 (mar 2008), 2637–2645.

[63] Fan, K. On a theorem of weyl concerning eigenvalues of linear transformations i. *Proceedings of the National Academy of Sciences 35*, 11 (1949), 652–655.

[64] Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T. T., Mohlberg, H., Amunts, K., and Zilles, K. Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex 18*, 8 (dec 2007), 1973–1980.

[65] Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of human genetics 7*, 2 (1936), 179–188.

[66] Flake, G. W. *The computational Beauty of Nature: Computer Exploration of Fractals, Chaos, Complex Systems and Adaptation*. MIT Press, Cambridge (Mass), 1998.

[67] Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., and Poline, J.-B. Improved detection sensitivity in functional MRI data using a brain parcelling technique. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*. Springer Berlin Heidelberg, 2002, pp. 467–474.

[68] Forgy, E. W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics 21* (1965), 768–769.

[69] Frazier, J. A., Chiu, S., Breeze, J. L., Makris, N., Lange, N., Kennedy, D. N., Herbert, M. R., Bent, E. K., Koneru, V. K., Dieterich, M. E., et al. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *American Journal of Psychiatry 162*, 7 (2005), 1256–1265.

[70] Freeman, W. J., Ahlfors, S. P., and Menon, V. Combining fmri with eeg and meg in order to relate patterns of brain activity to cognition. *International journal of psychophysiology 73*, 1 (2009), 43–52.

[71] Friston, K. J., et al. Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp. 2*, 1-2 (1994), 56–78.

[72] Friston, K. J., Frith, C., Liddle, P., and Frackowiak, R. Comparing functional (pet) images: the assessment of significant change. *Journal of Cerebral Blood Flow & Metabolism 11*, 4 (1991), 690–699.

[73] Friston, K. J., and Frith, C. D. Schizophrenia: a disconnection syndrome. *Clin. Neurosci. 3*, 2 (1995), 89–97.

[74] Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., and Ashburner, J. Classical and bayesian inference in neuroimaging: applications. *Neuroimage 16*, 2 (2002), 484–512.

[75] Friston, K. J., Harrison, L., and Penny, W. Dynamic causal modelling. *Neuroimage 19*, 4 (2003), 1273–1302.

[76] Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. *Neuroimage 12*, 4 (2000), 466–477.

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

[78] Fritzke, B. Fast learning with incremental rbf networks. *Neural processing letters 1*, 1 (1994), 2–5.

[79] Fritzke, B. Growing cell structures—a self-organizing network for unsupervised and supervised learning. *Neural networks 7*, 9 (1994), 1441–1460.

[80] Fritzke, B. Supervised learning with growing cell structures. In *Advances in neural information processing systems* (1994), pp. 255–262.

[81] Fritzke, B. A growing neural gas network learns topologies. In *Advances in neural information processing systems* (1995), pp. 625–632.

[82] Gazzaniga, M. S. Forty-five years of split-brain research and still going strong. *Nat. Rev. Neurosci. 6*, 8 (Aug 2005), 653–659.

[83] Gazzaniga, M. S. Neuroscience and the correct level of explanation for understanding mind: An extraterrestrial roams through some neuroscience laboratories and concludes earthlings are not grasping how best to understand the mind–brain interface. *Trends in Cognitive Sciences 14*, 7 (2010), 291–292.

[84] Genovese, C. R., Lazar, N. A., and Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage 15*, 4 (2002), 870–878.

[85] Ghuman, A. S., Brunet, N. M., Li, Y., Konecky, R. O., Pyles, J. A., Walls, S. A., Destefino, V., Wang, W., and Richardson, R. M. Dynamic encoding of face information in the human fusiform gyrus. *Nature communications 5* (2014), 5672.

[86] Glover, G. H. Deconvolution of impulse response in event-related bold fmri 1. *Neuroimage 9*, 4 (1999), 416–429.

[87] Goldstein, J. M., Seidman, L. J., Makris, N., Ahern, T., O'Brien, L. M., Caviness Jr, V. S., Kennedy, D. N., Faraone, S. V., and Tsuang, M. T. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biological psychiatry 61*, 8 (2007), 935–945.

[88] Gonella, G., Binaghi, E., Vergani, A. A., Biotti, I., and Levrini, L. A cloud fuzzy logic framework for oral disease risk assessment. *Lecture Notes in Artificial Intelligence* (2018). Submitted.

[89] González, R., and Tou, J. Pattern recognition principles. *Applied Mathematics and Computation. Reading (MA): Addison-Wesley* (1974).

[90] Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[91] Goodman, S. N. Toward evidence-based medical statistics. 1: The p value fallacy. *Annals of internal medicine 130*, 12 (1999), 995–1004.

[92] Goutte, C., Nielsen, F. Å., and Hansen, L. K. Modeling the hemodynamic response in fmri using smooth fir filters. *Medical Imaging, IEEE Transactions on 19*, 12 (2000), 1188–1201.

[93] Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences 100*, 1 (2003), 253–258.

[94] Grill-Spector, K., and Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience 15*, 8 (2014), 536.

[95] Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., and Hirsch, J. Detection of time-varying signals in event-related fmri designs. *Neuroimage 43*, 3 (2008), 509–520.

[96] Gudbjartsson, H., and Patz, S. The rician distribution of noisy mri data. *Magn. Reson. Med. 34*, 6 (1995), 910–914.

[97] Haacke, E. M., Brown, R. W., Thompson, M. R., and Venkatesan, R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 1st ed. Wiley-Liss, 6 1999.

[98] Hadamard, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin* (1902), 49–52.

[99] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. Mapping the structural core of human cerebral cortex. *PLoS Biol. 6*, 7 (2008), e159.

[100] Halkidi, M., Batistakis, Y., and Vazirgiannis, M. On clustering validation techniques. *Journal of intelligent information systems 17*, 2-3 (2001), 107–145.

[101] Hayasaka, S., and Nichols, T. E. Combining voxel intensity and cluster extent with permutation test framework. *Neuroimage 23*, 1 (2004), 54–63.

[102] Hebb, D. O. *The organization of behavior; a neuropsychological theory, (by) D.O. Hebb. Science Editions.* John Wiley and Sons, New York, 1967.

[103] Herrmann, C. S., Fründ, I., and Lenz, D. Human gamma-band activity: A review on cognitive and behavioral correlates and network models. *Neuroscience & Biobehavioral Reviews 34*, 7 (jun 2010), 981–992.

[104] Hogg, R. V., and Ledolter, J. *Engineering statistics.* Macmillan Pub Co, 1987.

[105] Horwitz, B., Friston, K. J., and Taylor, J. G. Neural modeling and functional brain imaging: an overview. *Neural networks 13*, 8-9 (2000), 829–846.

[106] Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology 24*, 6 (1933), 417.

[107] Hu, W., Xie, D., Tan, T., and Maybank, S. Learning activity patterns using fuzzy self-organizing neural network. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34*, 3 (2004), 1618–1626.

[108] Hullermeier, E., Rifqi, M., Henzgen, S., and Senge, R. Comparing fuzzy partitions: A generalization of the rand index and related measures. *IEEE Transactions on Fuzzy Systems 20*, 3 (2012), 546–556.

[109] Iraji, A., Calhoun, V. D., Wiseman, N. M., Davoodi-Bojd, E., Avanaki, M. R., Haacke, E. M., and Kou, Z. The connectivity domain: Analyzing resting state fmri data using feature-based data-driven and model-based methods. *Neuroimage 134* (2016), 494–507.

[110] Jaccard, P. The distribution of the flora in the alpine zone. *New phytologist 11*, 2 (1912), 37–50.

[111] Jain, A. K. Data clustering: 50 years beyond k-means. *Pattern recognition letters 31*, 8 (2010), 651–666.

[112] Jain, A. K., and Dubes, R. C. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[113] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. Fsl. *Neuroimage 62*, 2 (2012), 782–790.

[114] Jolliffe, I. Principal component analysis. In *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.

[115] Kacprzyk, J., and Pedrycz, W. *Springer handbook of computational intelligence*. Springer, 2015.

[116] Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J., and Haynes, J.-D. Connectivity-based parcellation of the human orbitofrontal cortex. *Journal of Neuroscience 32*, 18 (may 2012), 6240–6250.

[117] Kaiser, H. F. The application of electronic computers to factor analysis. *Educational and psychological measurement 20*, 1 (1960), 141–151.

[118] Kaiser, H. F. A note on guttman's lower bound for the number of common factors 1. *British Journal of Statistical Psychology 14*, 1 (1961), 1–2.

[119] Kanwisher, N., McDermott, J., and Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of neuroscience 17*, 11 (1997), 4302–4311.

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

[121] Khanchouch, I., Hachmi, F., and Limam, M. On determining the number of clusters in market seg-mentation.

[122] Klein, A., and Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience 6* (2012).

[123] Kohonen, T. Self-organized formation of topologically correct feature maps. *Biological cybernetics 43*, 1 (1982), 59–69.

[124] Kohonen, T. Self-organizing maps. *Rend. Circ. Mat. Palermo 44*, 3 (Sep 1995), 506â€“506.

[125] Kohonen, T., and Somervuo, P. Self-organizing maps of symbol strings. *Neurocomputing 21*, 1-3 (1998), 19–30.

[126] Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. A Bradford Book, 1992.

[127] Kruskal, W. H., and Wallis, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association 47*, 260 (1952), 583–621.

[128] Kuhn, T. Metaphor in science. In *Methapor in Science*. A. Ortony, 1979.

[129] Kuhn, T. S. *The Structure of Scientific Revolutions: 2d Ed., enl.* University of Chicago Press, Chicago, 1975.

[130] Kutz, J. N. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, 2013.

[131] Kwon, S. H. Cluster validity index for fuzzy clustering. *Electronics letters 34*, 22 (1998), 2176–2177.

[132] Lachiche, N., Hommet, J., Korczak, J., and Braud, A. Neuronal clustering of brain fmri images. In *International Conference on Pattern Recognition and Machine Intelligence* (2005), Springer, pp. 300–305.

[133] Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., and Golland, P. Search for patterns of functional specificity in the brain: A nonparametric hierarchical bayesian model for group fMRI data. *NeuroImage 59*, 2 (jan 2012), 1348–1368.

[134] Lashkari, D., Vul, E., Kanwisher, N., and Golland, P. Discovering structure in the space of fMRI selectivity profiles. *NeuroImage 50*, 3 (apr 2010), 1085–1098.

[135] LE VAN QUYEN, M. Disentangling the dynamic core: a research program for a neurodynamics at the large-scale. *Biol. Res. 36*, 1 (2003), 67–88.

[136] Lee, M. H., Hacker, C. D., Snyder, A. Z., Corbetta, M., Zhang, D., Leuthardt, E. C., and Shimony, J. S. Clustering of resting state networks. *PloS one 7*, 7 (2012), e40370.

[137] Leekwijck, W. V., and Kerre, E. E. Defuzzification: criteria and classification. *Fuzzy Sets and Systems 108*, 2 (1999), 159 – 178.

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

[139] Liao, C., Worsley, K., Poline, J.-B., Aston, J., Duncan, G., and Evans, A. Estimating the delay of the fmri response. *Neuroimage 16*, 3 (2002), 593–606.

[140] Liao, T. W. Clustering of time series data—a survey. *Pattern recognition 38*, 11 (2005), 1857–1874.

[141] Lichman, M. UCI machine learning repository, 2013.

[142] Linde, Y., Buzo, A., and Gray, R. An algorithm for vector quantizer design. *IEEE Transactions on communications 28*, 1 (1980), 84–95.

[143] Lindquist, M. A., et al. The statistical analysis of fmri data. *Statistical Science 23*, 4 (2008), 439–464.

[144] Lindquist, M. A., and Wager, T. D. Validity and power in hemodynamic response modeling: a comparison study and a new approach. *Hum. Brain Mapp. 28*, 8 (2007), 764–784.

[145] Lindquist, M. A., Waugh, C., and Wager, T. D. Modeling state-related fmri activity using change-point theory. *Neuroimage 35*, 3 (2007), 1125–1141.

[146] Lindquist, M. A., Zhang, C.-H., Glover, G., and Shepp, L. The acquisition and statistical analysis of rapid 3d fmri data. *Statistica Sinica 18*, 4 (2008), 1395.

[147] Lindquist, M. A., Zhang, C.-H., Glover, G., and Shepp, L. Rapid three-dimensional functional magnetic resonance imaging of the initial negative bold response. *J. Magn. Reson. 191*, 1 (2008), 100–111.

[148] Lippé, S., Kovacevic, N., and McIntosh, R. Differential maturation of brain signal complexity in the human auditory and visual system. *Frontiers in Human Neuroscience 3* (2009), 48.

[149] Lloyd, S. Least squares quantization in pcm. *IEEE transactions on information theory 28*, 2 (1982), 129–137.

[150] Logothetis, N. Can current fmri techniques reveal the microarchitecture of cortex? *Nat. Neurosci. 3*, 5 (2000), 413–413.

[151] Loh, J. M., Lindquist, M. A., and Wager, T. D. Residual analysis for detecting mis-modeling in fmri. *Statistica Sinica 18*, 4 (2008), 1421.

[152] Lomp, O. *Finding optimal parameters for neural networks using evolutionary algorithms*. PhD thesis, B. Sc. Thesis, ET-IT Dept., Univ. of Bochum, Germany, 2008.

[153] Lowel, S., and Singer, W. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science 255*, 5041 (1992), 209–212.

[154] Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., and Nichols, T. E. Non-white noise in fmri: does modelling have an impact? *Neuroimage 29*, 1 (2006), 54–66.

[155] Luo, L., Callaway, E. M., and Svoboda, K. Genetic dissection of neural circuits. *Neuron 57*, 5 (#mar# 2008), 634–60.

[156] Luo, W.-L., and Nichols, T. E. Diagnosis and exploration of massively univariate neuroimaging models. *Neuroimage 19*, 3 (2003), 1014–1032.

[157] Luukka, P. A new nonlinear fuzzy robust pca algorithm and similarity classifier in classification of medical data sets. *International Journal of Fuzzy Systems 13*, 3 (2011).

[158] MacQueen, J., et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (1967), vol. 1, Oakland, CA, USA., pp. 281–297.

[159] Madden, D. J., Bennett, I. J., and Song, A. W. Cerebral white matter integrity and cognitive aging: contributions from diffusion tensor imaging. *Neuropsychology review 19*, 4 (2009), 415.

[160] Makris, N., Papadimitriou, G. M., Kaiser, J. R., Sorg, S., Kennedy, D. N., and Pandya, D. N. Delineation of the middle longitudinal fascicle in humans: a quantitative, in vivo, dt-mri study. *Cereb. Cortex 19*, 4 (Apr 2009), 777–785.

[161] Malonek, D., and Grinvald, A. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science 272*, 5261 (1996), 551.

[162] Marčenko, V. A., and Pastur, L. A. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik 1*, 4 (1967), 457.

[163] Margulies, D. S., Böttger, J., Long, X., Lv, Y., Kelly, C., Schäfer, A., Goldhahn, D., Abbushi, A., Milham, M. P., Lohmann, G., et al. Resting developments: a review of fmri post-processing methodologies for spontaneous brain activity. *Magnetic Resonance Materials in Physics, Biology and Medicine 23*, 5-6 (2010), 289–307.

[164] Marsland, S., Shapiro, J., and Nehmzow, U. A self-organising network that grows when required. *Neural networks 15*, 8-9 (2002), 1041–1058.

[165] Martelli, A. Edge detection using heuristic search methods. *Computer Graphics and Image Processing 1*, 2 (1972), 169–182.

[166] Martinetz, T., Schulten, K., et al. A "neural-gas" network learns topologies.

[167] Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. Wandering minds: the default network and stimulus-independent thought. *Science 315*, 5810 (2007), 393–395.

[168] Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Goualher, G. L., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philosophical Transactions of the Royal Society B: Biological Sciences 356*, 1412 (aug 2001), 1293–1322.

[169] McCulloch, W. S., and Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics 5*, 4 (1943), 115–133.

[170] McIntosh, A., Grady, C., Ungerleider, L. G., Haxby, J., Rapoport, S., and Horwitz, B. Network analysis of cortical visual pathways mapped with pet. *Journal of Neuroscience 14*, 2 (1994), 655–666.

[171] McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. Analysis of fmri data by blind separation into independent spatial components. Tech. rep., DTIC Document, 1997.

[172] Mehta, M. L. *Random matrices*, vol. 142. Elsevier, 2004.

[173] Meiss, J. Dynamical systems. *Scholarpedia 2*, 2 (2007), 1629.

[174] Menon, R. S., Luknowsky, D. C., and Gati, J. S. Mental chronometry using latency-resolved functional mri. *Proceedings of the National Academy of Sciences 95*, 18 (1998), 10902–10907.

[175] Menon, R. S., Ogawa, S., Hu, X., Strupp, J. P., Anderson, P., and UÇ§urbil, K. Bold based functional mri at 4 tesla includes

a capillary bed contribution: Echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magn. Reson. Med. 33*, 3 (1995), 453–459.

[176] Meunier, D., Achard, S., Morcom, A., and Bullmore, E. Age-related changes in modular organization of human brain functional networks. *Neuroimage 44*, 3 (2009), 715–723.

[177] Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., and Thirion, B. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition 45*, 6 (jun 2012), 2041–2049.

[178] Miezin, F. M., Maccotta, L., Ollinger, J., Petersen, S., and Buckner, R. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage 11*, 6 (2000), 735–759.

[179] Minsky, M., and Papert, S. *Perceptrons: An essay in computational geometry*. Cambridge, MA: MIT Press, 1969.

[180] Mitchell, T. M., et al. *Machine learning*. McGraw-Hill, Inc., New York, NY, 1997.

[181] Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. *Applied linear statistical models*, vol. 4. Irwin Chicago, 1996.

[182] Nguyen, V. T., and Cunnington, R. The superior temporal sulcus and the n170 during face processing: single trial analysis of concurrent eeg–fmri. *Neuroimage 86* (2014), 492–502.

[183] Nichols, T. E., and Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp. 15*, 1 (2002), 1–25.

[184] Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., and Guenther, F. H. Region of interest based analysis of functional imaging data. *NeuroImage 19*, 4 (aug 2003), 1303–1316.

[185] Nilsson, N. *Artificial Intelligence : a new synthesis*. Morgan Kaufmann Publishers, San Francisco, Calif, 1998.

[186] Novák, V., Perfilieva, I., and Mockor, J. *Mathematical principles of fuzzy logic*, vol. 517. Springer Science & Business Media, 2012.

[187] Nuzzo, R. Scientific method: statistical errors. *Nature News 506*, 7487 (2014), 150.

[188] Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., and Ugurbil, K. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences 89*, 13 (1992), 5951–5955.

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

[190] Pal, N. R., and Bezdek, J. C. On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems 3*, 3 (1995), 370–379.

[191] Pearson, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 185*, 0 (#jan# 1894), 71???110.

[192] Pearson, K. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London 58* (1895), 240–242.

[193] Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2*, 11 (1901), 559–572.

[194] Pedoia, V., Binaghi, E., Balbi, S., De Benedictis, A., Monti, E., and Minotto, R. Glial brain tumor detection by using symmetry analysis. In *Medical Imaging 2012: Image Processing* (2012), vol. 8314, International Society for Optics and Photonics, p. 831445.

[195] Pedoia, V., Colli, V., Strocchi, S., Vite, C., Binaghi, E., and Conte, L. fmri analysis software tools: an evaluation framework. In *Medical Imaging 2011: Biomedical Applications in Molecular, Structural, and Functional Imaging* (2011), vol. 7965, International Society for Optics and Photonics, p. 796528.

[196] Pedoia, V., Gallo, I., and Binaghi, E. Affine spharm registration-neural estimation of affine transformation in spherical domain. In *VISAPP* (2011), pp. 197–200.

[197] Pedoia, V., Strocchi, S., Colli, V., Binaghi, E., and Conte, L. Functional magnetic resonance imaging: Comparison between activation maps and computation pipelines in a clinical context. *Magnetic resonance imaging 31*, 4 (2013), 555–566.

[198] Pedoia, V., Strocchi, S., Minotto, R., and Binaghi, E. Hemispheric dominance evaluation by using fmri activation weighted vector. In *Computational Modelling of Objects Represented in Images - Fundamentals, Methods and Applications III, Third International Symposium, CompIMAGE 2012, Rome, Italy, September 5-7, 2012.* (2012), pp. 303–306.

[199] Prudent, Y., and Ennaji, A. An incremental growing neural gas learns topologies. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on* (2005), vol. 2, IEEE, pp. 1211–1216.

[200] Purdon, P. L., Solo, V., Weisskoff, R. M., and Brown, E. N. Locally regularized spatiotemporal modeling and model comparison for functional mri. *Neuroimage 14*, 4 (2001), 912–923.

[201] Quarteroni, A., Saleri, F., and Gervasio, P. Scientific computing with matlab® and octave (texts in computational science and engineering) forth (4th) edition hardcover, 2014.

[202] Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. A default mode of brain function. *Proceedings of the National Academy of Sciences 98*, 2 (2001), 676–682.

[203] Raichle, M. E., and Snyder, A. Z. A default mode of brain function: a brief history of an evolving idea. *Neuroimage 37*, 4 (2007), 1083–1090.

[204] Ramon y Cajal, S. . . *Histologie du systÃ¨m nerveux de l'homme et des vertÃ©bretÃ©s. Translated as: Histology of the nervous system, translated by N. Swanson and L. Swanson.* Oxford University Press, 1995.

[205] Raschka, S. *Python machine learning.* Packt Publishing Ltd, 2015.

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

[207] Richard O. Duda, Peter E. Hart, D. G. S. *Pattern classification*, 2nd ed ed. Wiley, 2001.

[208] Riera, J., Aubert, E., Iwata, K., Kawashima, R., Wan, X., and Ozaki, T. Fusing eeg and fmri based on a bottom-up model: inferring activation and effective connectivity in neural masses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 360*, 1457 (2005), 1025–1041.

[209] Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., and Kawashima, R. A state-space model of the hemodynamic approach: nonlinear filtering of bold signals. *Neuroimage 21*, 2 (2004), 547–567.

[210] Ritter, H., Martinetz, T., and Schulten, K. Neuronale netze. *Munchen: Addison-Wesley* (1991).

[211] Roca, P., Tucholka, A., Rivière, D., Guevara, P., Poupon, C., and Mangin, J.-F. Inter-subject connectivity-based parcellation of a patch of cerebral cortex. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. Springer Berlin Heidelberg, 2010, pp. 347–354.

[212] Roebroeck, A., Formisano, E., and Goebel, R. Mapping directed influence over the brain using granger causality and fmri. *Neuroimage 25*, 1 (2005), 230–242.

[213] Rosenblatt, F. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Report (Cornell Aeronautical Laboratory). Spartan Books, 1962.

[214] Rotondi, A., Pedroni, P., and Pievatolo, A. *ProbabilitÃ Statistica e Simulazione: Programmi applicativi scritti con Scilab (UNITEXT / Collana di Statistica e Probabilit?? Applicata) (Italian Edition)*. Springer, 2011.

[215] Rovetta, M. F. F. C. F. M. S. A survey of kernel and spectral methods for clustering. *Pattern Recognition 41* (2008).

[216] Rowe, D. B., and Logan, B. R. A complex way to compute fmri activation. *Neuroimage 23*, 3 (2004), 1078–1092.

[217] Rumelhart, D. *Parallel distributed processing : explorations in the microstructure of cognition*. MIT Press, Cambridge, Mass, 1986.

[218] Russell, S. *Artificial intelligence : a modern approach*. Prentice Hall, Upper Saddle River, NJ, 2010.

[219] Sandrone, S. The brain as a crystal ball: The predictive potential of default mode network. *Frontiers in Human Neuroscience 6* (2012).

[220] Sanz Leon, P., Knock, S. A., Woodman, M. M., Domide, L., Mersmann, J., McIntosh, A. R., and Jirsa, V. The virtual brain: a simulator of primate brain network dynamics. *Frontiers in neuroinformatics 7* (2013), 10.

[221] Saxe, R., Brett, M., and Kanwisher, N. Divide and conquer: A defense of functional localizers. *NeuroImage 30*, 4 (may 2006), 1088–1096.

[222] Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., and Rosen, B. R. Late onset of anterior prefrontal activity during true and false recognition: an event-related fmri study. *Neuroimage 6*, 4 (1997), 259–269.

[223] Shapiro, S. *Encyclopedia of artificial intelligence*. Wiley, New York, 1992.

[224] Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., and Toga, A. W. Construction of a 3d probabilistic atlas of human cortical structures. *NeuroImage 39*, 3 (feb 2008), 1064–1080.

[225] Siddique, N., and Adeli, H. *Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing*. John Wiley & Sons, 2013.

[226] Simon, O., Kherif, F., Flandin, G., Poline, J.-B., Rivière, D., Mangin, J.-F., Bihan, D. L., and Dehaene, S. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *NeuroImage 23*, 3 (nov 2004), 1192–1202.

[227] Smith, S. M., Hyvarinen, A., Varoquaux, G., Miller, K. L., and Beckmann, C. F. Group-pca for very large fmri datasets. *Neuroimage 101* (2014), 738–749.

[228] Snijders, C., Matzat, U., and Reips, U.-D. " big data": big gaps of knowledge in the field of internet science. *International Journal of Internet Science 7*, 1 (2012), 1–5.

[229] Sotero, R. C., Trujillo-Barreto, N. J., Iturria-Medina, Y., Carbonell, F., and Jimenez, J. C. Realistically coupled neural mass models can generate eeg rhythms. *Neural computation 19*, 2 (2007), 478–512.

[230] Sporns, O., Tononi, G., and Edelman, G. M. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Networks 13*, 8 (2000), 909–922.

[231] Sporns, O., Tononi, G., and Kötter, R. The human connectome: A structural description of the human brain. *PLoS Comput. Biol. 1*, 4 (Sep 2005), e42.

[232] Stufflebeam, S. M., and Rosen, B. R. Mapping cognitive function. *Neuroimaging Clinics of North America 17*, 4 (2007), 469–484.

[233] Sugeno, M. Theory of fuzzy integrals and its applications. *Doct. Thesis, Tokyo Institute of technology* (1974).

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

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

[236] Tang, Y., Sun, F., and Sun, Z. Improved validation index for fuzzy clustering. In *American Control Conference, 2005. Proceedings of the 2005* (2005), IEEE, pp. 1120–1125.

[237] Thagard, P. Cognitive science. In *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., fall 2012 ed. 2012.

[238] Thirion, B., Varoquaux, G., Dohmatob, E., and Poline, J.-B. Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience 8* (jul 2014).

[239] Thirion, B. J.-B. Detection of signal synchronizations in resting-state fmri datasets. *Neuroimage 29*, 1 (2006), 321–327.

[240] Thorndike, R. L. Who belongs in the family? *Psychometrika 18*, 4 (1953), 267–276.

[241] Tracy, C. A., and Widom, H. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics 159*, 1 (1994), 151–174.

[242] Tryon, C. M. Evaluations of adolescent personality by adolescents. *Monographs of the Society for Research in Child Development 4*, 4 (1939), i–83.

[243] Tryon, R. C. *Cluster analysis: Correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality.* Edwards brother, Incorporated, lithoprinters and publishers, 1939.

[244] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage 15*, 1 (jan 2002), 273–289.

[245] Uhlhaas, P. J., Haenschel, C., Nikolić, D., and Singer, W. The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. *Schizophrenia bulletin 34*, 5 (2008), 927–943.

[246] Ullman, J. D., Leskovec, J., and Rajaraman, A. Mining of massive datasets, 2011.

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

[248] Van Den Heuvel, M. P., and Pol, H. E. H. Exploring the brain network: a review on resting-state fmri functional connectivity. *Eur. Neuropsychopharmacol. 20*, 8 (2010), 519–534.

[249] Vazquez, A. L., Cohen, E. R., Gulani, V., Hernandez-Garcia, L., Zheng, Y., Lee, G. R., Kim, S.-G., Grotberg, J. B., and Noll, D. C. Vascular dynamics and bold fmri: Cbf level effects and analysis considerations. *Neuroimage 32*, 4 (2006), 1642–1655.

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

[251] Vergani, A. A., Martinelli, S., and Binaghi, E. Cluster analysis of functional neuroimages using data reduction and competitive learning algorithms. In *European Congress on Computational Methods in Applied Sciences and Engineering* (2017), Springer, pp. 62–71.

[252] Vergani, A. A., Martinelli, S., and Binaghi, E. Clustering functional mri patterns with fuzzy and competitive algorithms. In *Proceedings of the International Multi-Conference on Computational Science (CS 2018)* (2018), AGH University of Science and Technology Press.

[253] Vergani, A. A., Martinelli, S., and Binaghi, E. Resting state fmri functioanl connectivity analysus using soft competitive learning algorithms. In *Book of Abstracts for the CMBBE2018: 15th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering and the 3rd Conference on Imaging and Visualization* (2018).

[254] Vergani, A. A., Minotto, R., Strocchi, S., and Binaghi, E. Fsl-based hybrid atlas promotes activation weighted vector analysis in functional neuroradiology. Frontiers, Ed., vol. 10, International Neuroinformatics Coordination Facilities (INCF), Frontiers Media SA.

[255] Vergani, A. A., Minotto, R., Strocchi, S., and Binaghi, E. Quantitative relations of CRAIIM human brain atlases. In *1st HBP Student Conference - Transdisciplinary Research Linking Neuroscience, Brain Medicine and Computer Science* (2018), M. M. J. T. V. K. T. M. V. T. T. R. A. S. Nikola Simidjievski, Andrea Santuy, Ed., vol. 1, Human Brain Project, Frontiers Media SA, pp. 60–68.

[256] Vesanto, J., and Alhoniemi, E. Clustering of the self-organizing map. *IEEE Transactions on neural networks 11*, 3 (2000), 586–600.

[257] Viviani, R., Grön, G., and Spitzer, M. Functional principal component analysis of fmri data. *Human brain mapping 24*, 2 (2005), 109–129.

[258] Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. Accounting for nonlinear bold effects in fmri: parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage 25*, 1 (2005), 206–218.

[259] Wakeman, D. G., and Henson, R. N. A multi-subject, multi-modal human neuroimaging dataset. *Scientific data 2* (2015), 150001.

[260] Wang, W., and Zhang, Y. On fuzzy cluster validity indices. *Fuzzy sets and systems 158*, 19 (2007), 2095–2117.

[261] Wang, Y., and Li, T.-Q. Analysis of whole-brain resting-state fmri data using hierarchical clustering approach. *PLoS One 8*, 10 (2013), e76315.

[262] Wasserstein, R. L., Lazar, N. A., et al. The asa's statement on p-values: context, process, and purpose. *The American Statistician 70*, 2 (2016), 129–133.

[263] Weisstein, E. W. Delaunay triangulation. *From MathWorld–A wolfram Web Resource. http://mathworld. wolframe. com/DelaunayTriangulation. html* (1999).

[264] Weisstein, E. W. Voronoi diagram. *From MathWorld–A Wolfram Web Resource.¡ http://mathworld. wolfram. com/VoronoiDiagram. html* (1999).

[265] Wig, G. S., Laumann, T. O., Cohen, A. L., Power, J. D., Nelson, S. M., Glasser, M. F., Miezin, F. M., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cerebral Cortex 24*, 8 (mar 2013), 2036–2054.

[266] Willshaw, D. J., and Von Der Malsburg, C. How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond. B 194*, 1117 (1976), 431–445.

[267] Wishart, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* (1928), 32–52.

[268] Wismüller, A., Meyer-Bäse, A., Lange, O., Auer, D., Reiser, M. F., and Sumners, D. Model-free functional mri analysis based on unsupervised clustering. *Journal of Biomedical Informatics 37*, 1 (2004), 10–18.

[269] Woolrich, M. W., Behrens, T. E., and Smith, S. M. Constrained linear basis sets for hrf modelling using variational bayes. *Neuroimage 21*, 4 (2004), 1748–1761.

[270] Worsley, K. J., and Friston, K. J. Analysis of fmri time-series revisitedâ€"again. *Neuroimage 2*, 3 (1995), 173–181.

[271] Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. Unified univariate and multivariate random field theory. *Neuroimage 23* (2004), S189–S195.

[272] Wu, C.-H., Ouyang, C.-S., Chen, L.-W., and Lu, L.-W. A new fuzzy clustering validity index with a median factor for centroid-based clustering. *IEEE Transactions on Fuzzy Systems 23*, 3 (2015), 701–718.

[273] Wu, C. J., and Hamada, M. S. *Experiments: planning, analysis, and optimization*, vol. 552. John Wiley & Sons, 2011.

[274] Wu, K.-L., and Yang, M.-S. A cluster validity index for fuzzy clustering. *Pattern Recognition Letters 26*, 9 (2005), 1275–1291.

[275] Xie, X. L., and Beni, G. A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence 13*, 8 (1991), 841–847.

[276] Xu, L., and Yuille, A. L. Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks 6*, 1 (1995), 131–143.

[277] Yang, T.-N., and Wang, S.-D. Robust algorithms for principal component analysis. *Pattern Recognition Letters 20*, 9 (1999), 927–933.

[278] Zadeh, L. A. Information and control. *Fuzzy sets 8*, 3 (1965), 338–353.

[279] Zadeh, L. A. The concept of a linguistic variable and its application to approximate reasoning—i. *Information sciences 8*, 3 (1975), 199–249.

[280] Zahid, N., Limouri, M., and Essaid, A. A new cluster-validity for fuzzy clustering. *Pattern recognition 32*, 7 (1999), 1089–1097.

[281] Zarahn, E. Using larger dimensional signal subspaces to increase sensitivity in fmri time series analyses. *Hum. Brain Mapp. 17*, 1 (2002), 13–16.

[282] Zhang, D., Ji, M., Yang, J., Zhang, Y., and Xie, F. A novel cluster validity index for fuzzy clustering based on bipartite modularity. *Fuzzy Sets and Systems 253* (2014), 122–137.

[283] Zubin, J. A technique for measuring like-mindedness. *The Journal of Abnormal and Social Psychology 33*, 4 (1938), 508.