



AGH UNIVERSITY, UNIVERSITY OF KRAKÓW

**FACULTY OF COMPUTER SCIENCE, ELECTRONICS AND
TELECOMMUNICATIONS.
INSTITUTE OF COMPUTER SCIENCE**

ERASMUS +

FINAL PROJECT

Bachelor's degree in biomedical engineering

MARIA VILA MOLINA

**NEUROIMAGING IN STROKE.
PRE-PROCESSING AND CONFOUND
REGRESSION OF FMRI DATA
FOR FUNCTIONAL GRADIENT
ANALYSIS**

**Supervisor: Bartosz Balis
Auxiliar Supervisor: Alex Crimi**

Kraków, June 2023

ACKNOWLEDGMENTS

With this Final Degree Project, I close a beautiful chapter of my life, university. First and foremost, I would like to express my gratitude to the Sano Science research institute for giving me the opportunity to work with them on the development of this project.

During these months, I have learned a lot, and it has been a great opportunity to get a closer look at the field of neuroscience. Specifically, I would like to thank Alex Crimi for offering us this opportunity, and Cemal and Joan for their support throughout the entire process and for the time dedicated.

Secondly, I would like to thank and dedicate this project to all my fellow classmates, particularly Sara, Camilla, Ana Jiménez, and Ana Sánchez. Without all of you, these four years of university wouldn't have been the same. Thank you for your support and for turning difficult moments into memories that I will carry with me forever. It has been a wonderful experience to have crossed paths with you all.

Being able to develop this project in Krakow has been a great challenge and a beautiful experience for me. I am very grateful for the opportunities that have been presented to me and for the people who have been a part of them. During these five months of Erasmus, I have met many people who have, in one way or another, contributed to this project. It has been a pleasure, and I bid farewell to this chapter with deep gratitude.

Lastly, I would like to express my gratitude to my family, my parents Eugenio and María Paz, and my sister Paloma, for being my fundamental support, not only in this project but in my life. Thank you for always guiding me, advising me, and helping me through all the stages that have passed and those yet to come. Thank you for giving me wings and for being the place I will always return to.

Sincerely, thank you to everyone.

ABSTRACT

The brain plays a central role in the task of controlling and regulating the various functions of the body. It is a complex organ composed of specialized regions that are interconnected and responsible for different mental abilities, behaviors, and cognitive processes. These regions work together to effectively control and coordinate the body's activities, making the brain an organ that has been the subject of numerous studies to gain insights into its normal functioning as well as its behavior under pathological conditions. Understanding the functioning of the brain, researchers may be able to develop strategies for diagnosing, treating, and managing various brain-related conditions.

In this final work, we have worked with fMRI data from two populations, healthy patients, and stroke patients. These images have been subjected to pre- and post-processing pipelines using different techniques and software's to obtain neuroimages which have been analyzed to study functional connectivity.

After a study of the parameters obtained, we have used the brain functional gradients along these specific brain networks to observe the differences and similarities between both groups of patients and draw the relevant conclusions of the project.

Keywords: brain, brain connectivity, stroke, functional gradient, preprocessing, postprocessing, confound regression, fMRI, fMRIprep, XCD-P.

RESUMEN

El cerebro desempeña un papel central en la tarea de controlar y regular las diversas funciones del organismo. Es un órgano complejo compuesto por regiones especializadas que están interconectadas y son responsables de distintas capacidades mentales, comportamientos y procesos cognitivos. Estas regiones trabajan juntas para controlar y coordinar eficazmente las actividades del cuerpo, lo que convierte al cerebro en un órgano que ha sido objeto de numerosos estudios para conocer mejor su funcionamiento normal, así como su comportamiento en condiciones patológicas.

Comprendiendo el funcionamiento del cerebro, los investigadores pueden desarrollar estrategias para diagnosticar, tratar y controlar diversas afecciones cerebrales.

En este trabajo final, hemos trabajado con datos de fMRI de dos poblaciones, pacientes sanos y pacientes con ictus. Estas imágenes han sido sometidas a pipelines de pre y post-procesado utilizando diferentes técnicas y softwares para obtener neuroimágenes que han sido analizadas para estudiar la conectividad funcional.

Tras un estudio de los parámetros obtenidos, hemos utilizado los gradientes funcionales cerebrales a lo largo de estas redes cerebrales específicas para observar las diferencias y similitudes entre ambos grupos de pacientes y extraer las conclusiones pertinentes del proyecto.

Palabras clave: cerebro, conectividad cerebral, ictus, gradiente funcional, preprocesado, post-procesado, regresión de confusión, fMRI, fMRIprep, XCD-P.

INDEX

ACKNOWLEDGMENTS	2
ABSTRACT	3
CHAPTER 1. INTRODUCTION	7
1.1 THE BRAIN AS A COMPLEX NETWORK	7
1.2 MOTIVATION AND GOALS OF THE THESIS	7
1.3 STRUCTURE OF THE THESIS	8
CHAPTER 2. DESCRIPTION OF THE DOMAIN.....	9
2.1 CENTRAL NERVOUS SYSTEM.....	9
2.2 NEUROIMAGING TECHNIQUES	10
2.2.1 FUNCTIONAL MAGNETIC RESONANCE IMAGING (fMRI).....	11
2.3 BRAIN CONNECTIVITY	13
2.3.1. FUNCTIONAL GRADIENTS	14
2. 4 STROKE.....	15
2.4.1 FACTORS RISKS FOR STROKE	15
2.4.2 TIME FACTOR	16
2.4.3 STROKE IN MRI	16
CHAPTER 3. STATE OF THE ART	18
3.1 PRE-PROCESSING OF FUNCTIONAL IMAGES	18
3.1.1 BIDS	20
3.1.2 fMRIPREP	22
3.2 CONFOUND REGRESSION IN FUNCTIONAL IMAGES.....	24
3.2.1 XCP-D.....	25
3.3 FUNCTIONAL GRADIENTS	27
3.3.1 GRADIENTS IN STROKE.....	32
CHAPTER 4. METHODOLOGY	34
4.1 FMRI DATA	34
4.2 PRE-PROCESSING	34
4.2.1 DOCKER.....	34
4.2.2 fMRIPREP	35
4.3 CONFOUND REGRESSION	35
4.3.1 XCP-D.....	36
4.4 CONNECTIVITY GRADIENTS IN HEALTH AND STROKE PATIENTS	37
4.4.1 BRAINSPACE.....	37
4.5 DATA ANALYSIS.....	38

CHAPTER 5. CONECTIVITY GRADIENTS	39
5.1 DATASET DESCRIPTION	39
5.2 CONNECTIVITY GRADIENTS CALCULATION	40
5.2.1 DATA FROM THE CONTROL HEALTH GROUP ACROSS PIPELINES	44
5.2.2 DATA FROM CONTROL HEALTH GROUP VS STROKE IN PIPELINE 1.....	46
5.2.3 DATA FROM CONTROL HEALTH GROUP VS STROKE IN PIPELINE 2.....	48
CHAPTER 6. RESULTS	50
6.1 DESCRIPTIVE STATISTICS ACROSS PIPELINES	50
6.2 GROUP-LEVEL ANALYSIS.....	52
6.2.1 ANALYSIS ACROSS PIPELINES	53
6.2.2 CONTROL HEALTH GROUP VS STROKE ANALYSIS	54
CHAPTER 7. CONCLUSIONS AND FUTURES LINES	56
6.1 OBJECTIVES ACHIEVED.....	56
6.2 CONCLUSIONS.....	56
6.3 LIMITATIONS AND FUTURE LINES OF INVESTIGATION	57
LIST OF FIGURES	58
LIST OF MATHEMATICAL EXPRESSION	59
LIST OF TABLES	59
BIBLIOGRAFY	60
ANNEX.....	66

CHAPTER 1. INTRODUCTION

1.1 The brain as a complex network

The brain is a highly complex organ, and there are still many unanswered questions about its functioning and the diseases that affect it. Moreover, it can be considered from the perspective of a complex network, either in terms of its anatomical structure or its functional activity. Examining the brain as a network allows us to understand how different parts communicate at various scales, from interactions between individual neurons to entire brain regions. This approach provides new perspectives and concepts for understanding brain function and the relationships between its different components.

The functional organization of the brain has been a central theme in neurophysiology from the 19th century to the first half of the 20th century. From theories that postulated that the brain was divided into regions, each in charge of a specific anatomical function to opposing positions that considered the brain as a homogeneous and functionally equipotential tissue [1]. In the 19th century, many researchers created "functional maps" of the cerebral cortex, relating various mental processes to specific brain areas.

Functional brain maps have long since become a valuable tool in the study of brain activity. The numerical information obtained from these maps allows us to construct three-dimensional models of the patient's brain and compare them with databases to identify possible differences with the general population. In this type of analysis, connectivity parameters are used to measure the speed and intensity of connections between different areas of the brain [2].

In recent years, image acquisition and preprocessing techniques have also been developed and improved, which help us to obtain better results to extract the connectivity parameters we are looking for, analyze them, and carry out a study of the data collected (Esteban, et al. 2018) [3].

1.2 Motivation and goals of the thesis

Functional magnetic resonance imaging (fMRI) is a non-invasive neuroimaging technique that allows the detection of physiological changes in the brain related to mental processes while performing a task.

Before performing statistical analysis, it is necessary to preprocess the fMRI images to clean and standardize the data because it is one of the crucial steps in obtaining quality images that can be used for further analysis. The fMRIPrep software is a robust preprocessing tool that provides high quality results for denoising and estimating potential sources of artifacts before measuring brain connectivity in healthy and pathological patients [3]. For this purpose, we will use functional gradients, which are patterns of neuronal activity that extend along a specific brain network or circuit and gradually change from one region to another. They are widely used in the study of functional brain organization, brain connectivity and cognition [4].

Using XCP-D we can post-process the outputs from fMRIPrep removing the artifacts estimated from the signal. This step in the pipeline is known as confound regression and it is a necessary step before converting the signals into functional connectivity networks. Functional connectivity can be directly analyzed using network science [5]. However, recently functional gradients have been proposed to gain further insights into the continuous profile of brain functional organization [6].

For all this, we proposed the present final project whose main objectives are:

- Preprocess the images of two groups of patients: healthy patients and patients with pathology (stroke) using the fMRIPrep tool.
- Apply different post-processing pipelines with XCP-D to observe the variability thus enhancing reproducibility of group analysis.
- Analyze how the functional gradients vary between patients and between both groups of patients (healthy and pathology) and study the effect that cerebral infarction has on the functional gradients with respect to the gradients derived from the healthy patients using different pipelines.

1.3 Structure of the thesis

The structure we had follow in this project is structured in seven chapters.

We start with a brief introduction, where we exposed the target of the thesis and the main objectives of this project. Chapter two provides, the description of the domain with which we will be able to find the theoretical framework of the thesis.

In chapter three, “State of the art” we have analyzed the methods and techniques currently available for pre- and post-processing, focused on those used in this work. Chapter four is dedicated to the methodology explaining how this project has been developed, and how the software’s had been installed for the next steps.

For the next chapters, we have analyzed the data we have pre- and post-processed using the functional gradients as a connectivity parameter. Finally, the results are detailed after the analysis of the data, and we have drawn the most relevant conclusions from the results and commented on some of the future lines of research that remain open in the future.

At the end of the document, we find the bibliography, the list of the figures and the list of the tables where the bibliographic, multimedia and references sources consulted are listed.

CHAPTER 2. DESCRIPTION OF THE DOMAIN

2.1 Central Nervous System

The nervous system is divided into the central nervous system (CNS) and the peripheral nervous system (PNS) [7]. The PNS is like a network of nerves connects the central system to the rest of the body. It's responsible for carrying messages to and from our brain [8]. The CNS is composed of the brain and spinal cord, which coordinate the body's activity and control thought processes, movements, and bodily sensations. They both work together to control every task we do, and they are protected by the meninges and bony structures such as the skull and spinal column.

The spinal cord serves as a bridge between the PNS and CNS with nerve fibers transmitting information in two directions: afferent pathways carry sensory information to the brain from external stimuli and body organs, while efferent pathways transmit information from the brain to the body [7].

The brain acts as the central control center for regulating body functions through hormone production, muscle activity patterns, and more. It is divided into two hemispheres and has folded structures on its surface. The cerebellum, responsible for integrating sensory and motor pathways, coordinates movement, balance, and spatial orientation. The brainstem controls involuntary functions like reflexes, breathing, and heart rate [9].

Within the brain, there are distinct areas of gray matter and white matter [7]. Gray matter consists of neuronal cell bodies, dendrites, and glial cells that provide support to neurons. White matter, on the other hand, primarily consists of neuronal axons and facilitates the transmission of information between nerve cells.

Gray matter is predominantly found in the cortical and subcortical regions of the brain, while white matter is in deeper areas [10].

Figure 1 [11] shows the structure of a neuron and the parts of which it is composed.

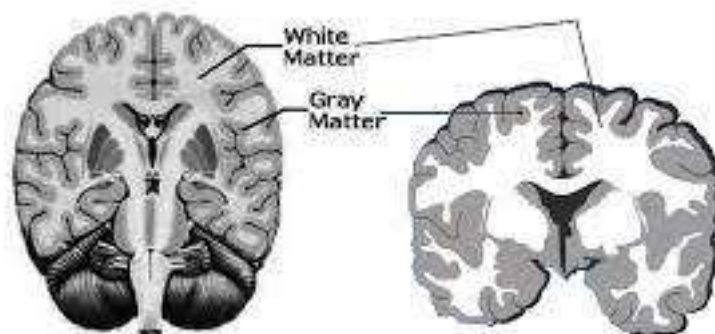


Figure 1 Grey and White matter in brain [11]

2.2 Neuroimaging Techniques

Scientists use neuroimaging methods to observe and measure brain activity, which enables them to study the brain's structure and function in different circumstances.

Magnetic resonance imaging (MRI) is an adaptable and noninvasive technique utilized in neuroimaging to capture high-resolution images of the brain and other tissues, where the procedure relies on the interaction between hydrogen nuclei present in the body, a strong magnetic field, and radiofrequency pulses [12].

MRI methods are frequently employed in neuroimaging to evaluate the brain's structure, function, and connectivity. A table listing some of the central MRI techniques in neuroimaging can be seen in Table 1.

Technique	Acronym	Contrast	Spatial Resolution	Temporal Resolution	Main Use
Functional MRI	fMRI	Hemodynamic	High	High	Mapping brain activation during tasks and resting state
Diffusion MRI	dMRI	Diffusion	High	Medium/High	Visualization and mapping of brain fiber connectivity
T1-weighted	T1	Anatomical	High	Low	Detailed structural visualization of brain tissues
T2- weighted	T2	Anatomical	High	Low	Visualization of differences in brain tissues
Fluid Atenuated Inversion Recovery	FLAIR	Anatomical	High	Low	Suppression of cerebrospinal fluid for detecting brain abnormalities
Susceptibility Weighted Imaging	SWI	Susceptibility	High	Low	Visualization of vascular structures and brain hemorrhages
Magnetic Resonance Spectroscopy	MRS	Metabolites	N/A	N/A	Measurement of brain metabolites and evaluation of metabolic disorders

Table 1 Table with some of the main MRI techniques in neuroimaging.

In this table, the primary MRI techniques employed in neuroimaging are outlined, highlighting the uniqueness in contrast, spatial and temporal resolution, and main uses of each tactic in the fields of clinical and scientific research. Every technique provides distinctive insights into the brain's anatomy, function, or biochemical nature, and the selection of an appropriate technique depends on the study objectives and specific research inquiries.

- **fMRI** maps brain activity by detecting changes in blood flow that are associated with neuronal activity. It can distinguish active regions of the brain during cognitive tasks or resting states, revealing insights into brain function [13].
- **dMRI** examines the structural connectivity of the brain by evaluating the diffusion of water in brain tissue. It enables the visualization and mapping of white matter fibers and allows the investigation of brain connectivity [14].
- **T1 and T2** sequences represent the structural MRI techniques that capture detailed images of the brain's anatomy. The T1 sequence offers contrast between varying brain tissues meanwhile the T2 sequence highlights differences in tissue relaxation times, which can facilitate the detection of abnormalities such as injury or inflammation [15].
- **FLAIR** is a variant of structural MRI that focuses on suppressing cerebrospinal fluid (CSF) signals in the brain. This allows better visualization of tissue abnormalities near the CSF, such as brain lesions, tumors, or areas of inflammation [16].
- **SWI** is an MRI technique that exploits differences in the magnetic susceptibility of tissues. It visualizes vascular structures, brain hemorrhages, and demyelination-related alterations observed in conditions like multiple sclerosis [17].
- **MRS** measures the levels of brain metabolites, such as lactic acid, creatine, or glutamine. The technique provides insights into brain biochemistry and is used in the investigation of metabolic diseases and neurological disorders [18].

These neuroimaging MRI techniques are indispensable in understanding the human brain's structure, function, and pathological alterations at various levels. Each one of the modalities above are non-mutually exclusive and can be acquired together to obtain combined insight from all of them. These types of studies and diagnostics are referred to as multimodal imaging studies.

2.2.1 Functional Magnetic Resonance Imaging (fMRI)

As previously mentioned, functional magnetic resonance imaging (fMRI) is one of the primary techniques used to research functional connectivity. It is a non-invasive neuroimaging method utilized to examine brain function in healthy individuals and those with brain pathologies.

At present, fMRI is a non-invasive technique that also enables specific distinctions in the functionality of diverse brain regions in clinical settings. fMRI is used to identify neurobiological relationships linked to neurological or psychiatric disorders by assessing differences between clinical populations [13].

fMRI detects changes in blood flow that are correlated with neuronal activation when performing a specific task. This is feasible because of the differing magnetic properties of oxygen-bound hemoglobin in the blood. During the completion of a specific task, oxygenated blood flow to the brain regions responsible for that function is amplified, and this increase is detectable through specific parameters of the MRI scan. This is known as the blood oxygen level dependent (BOLD) effect and can be used to generate brain activity maps [19].

fMRI involves a series of steps that are in turn divided into several parts for its acquisition and subsequent analysis [19].

1. Design of the experimental paradigm (rest or task)

Most fMRI studies aim to identify areas of the brain that participate in specific mental processes such as language comprehension, decision-making, emotional information processing, among others. Achieving this requires a well-defined experimental design that includes parameters, such as the study formulation, stimuli presentation, their duration, and the task instructions for participants.

2. Image acquisition

Although the resulting images are three-dimensional, functional sequences usually involve acquiring a series of two-dimensional images, known as slices, that range from 2 to 5 mm in thickness (as seen in Fig. 2). Each slice is divided into a matrix of pixels that encompass a specific region of space (field of view or FOV). The image resolution in-plane is determined by the matrix size and the FOV relationship, like a digital camera. The final volume associated with each measurement is constructed from this series of 2D images, which is composed of three-dimensional pixels, or voxels.

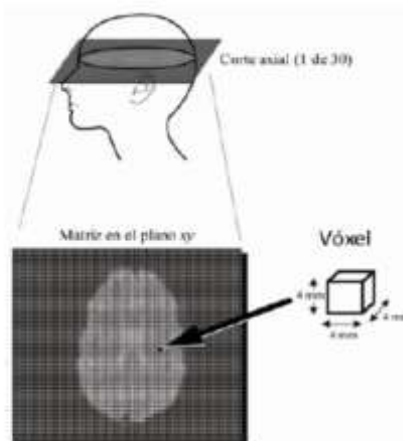


Figure 2 To obtain the images, a series of two-dimensional slices are captured to cover either the entire brain or specific regions of interest. Each slice has a set thickness and is composed of a matrix of volumetric pixels or voxels [19].

3. Data processing and analysis

Analyzing fMRI experiments can be quite complex due to two contributing factors already mentioned: the vast amount of data collected and the inherent "noise" present in the acquired signal. As a result, it is essential to review the motion parameters obtained. If there are numerous or significant issues, it may be necessary to repeat the experiment with that participant or exclude them from group analysis. Therefore, preprocessing fMRI data is a crucial step in the process of acquiring these images.

4. Interpretation of results

After completing the mathematical and statistical procedures for obtaining fMRI images, the interpretation phase commences. This involves starting with an understanding of the neurocognitive processes involved in the task or stimulus and drawing conclusions from the inferred networks and/or the corresponding activation maps.

2.3 Brain Connectivity

The brain's complexity arises from its interconnected regions that determine its topology, dynamics, and relationships.

Structural connectivity pertains to anatomical connections, such as neuronal axons or tracts of neurons that link different brain regions and is related to the physical substrate where information transfer takes place. Information processing is determined by how cell populations in the brain interact, which is influenced by structural connectivity. However, the anatomical connection between two brain regions does not always signify that the cognitive states or brain processes in which they participate are identical [20].

Functional connectivity concerns the statistical dependence relationship between brain element activity, irrespective of their structural connection. This connectivity crucially depends on the time domain and is not characterized by specific directional effects. Metrics such correlation, covariance, or spectral coherence embody the statistical concept [21].

Effective connectivity represents a causal relationship between various brain elements where inducing activation in one area would change another area. Causality can be concluded from network perturbations or time series analysis and describes the directional effects of one neural element on another [22].

In conclusion, brain anatomy modulates dynamic brain activity, reflected in patterns of statistical dependencies (functional connectivity) and causal interactions (effective connectivity) that relate to structural connectivity.

2.3.1. Functional Gradients

The investigation of brain connectivity has discovered that specific brain networks display a unique pattern of activation where neuronal activity is gradually distributed across different brain regions, referred to as the "brain functional gradient."

The relationship between the brain functional gradient, brain topology, and connectivity features has been extensively studied in research on functional brain organization, brain connectivity, and cognition. Analyzing these gradients can provide us with valuable information about the functional organization of the brain and brain connectivity [4].

To describe gradients, techniques are necessary to identify and categorize gradual changes in features. With the ability to map, identify and analyze brain cortical gradients in local and whole brain levels, new ways of examining brain organization at various scales have emerged.

Gradient mapping can advance cortical mapping by providing a continuous perspective for assessing sub-regional heterogeneity and cortical organization along axes of brain organization. Complementing existing perspectives on the cortex and allowing investigation into structure-function relationships. Furthermore, gradients allow exploration of other domains to understand global relationships between distributed functional systems. Although gradient convergence exists across various modalities, there are remarkable biological differences. Changes in gradients due to environmental and disease-related factors are being investigated, providing us with information on the invariance and flexibility of cortical gradients [4].

The objective of this work is to collect various gradients and conduct an analysis of differences in multiple individuals to identify potential biomarkers. Studies have been conducted on gradients analyzing procedures. In general, these findings demonstrate how well gradients can function as biomarkers.

Currently, researching brain organization across various scales is one of the major challenges to explain the brain's dynamics and its relationship to behavior. Recently, rs-fMRI dynamic analyses were used to exhibit how intrinsic functional activity propagates along large-scale functional gradients. This suggests that these gradients may determine functional dynamics [4].

Within this context, our intention will be to analyze and compare gradients in a stroke population and a healthy population. This emerging field will help address fundamental principles of cortical organization, examine the links between cognitive structure and function, and study cortical evolution and development.

2.4 Stroke

Insufficient blood circulation in the brain that leads to cell death is a medical condition called a stroke.

Two types of strokes exist that depend on the mechanism of injury [23]:

Ischemic stroke or cerebral infarction occurs due to an obstruction of blood flow.

Hemorrhagic stroke or cerebral hemorrhage, which arises when a blood vessel ruptures, leading to blood outflow and compression of the central nervous system's structures.

The effects caused by a stroke are influenced by different aspects, such as the location of the obstruction and how much brain tissue is affected. However, since one side of the brain controls the opposite side of the body, a stroke impacting one side will result in neurological complications on the affected side of the body [24].

2.4.1 Factors Risks for Stroke

The main risk factors for stroke are [25]:

- **Age:** After the age of 55, the risk of having a stroke double with each passing decade.
- **Hypertension**
- **Diabetes:** If the patient has high blood pressure it has more risk
- **Sedentary Lifestyle**
- **Obesity, High Cholesterol**
- **Smoking, Drug and Alcohol Consumption**
- **Gender:** Men have a higher risk of stroke than women, but at the age of 80 this relationship reverses.
- **Certain Population Groups:** Afro-Americans or Hispanics may have a higher risk of stroke.
- **Family History:** Family history with stroke can increase the risk of having it.

2.4.2 Time Factor

When a stroke occurs, brain cells begin to die gradually over time, leading to an irreversible cerebral infarction. Diagnostic tests such as CT or MRI can visualize the affected area, which is not receiving enough blood flow, as well as the damaged area. These images can help doctors determine how much brain tissue can be saved if blood flow is quickly restored. Therefore, reaction time is a critical factor in the patient's survival [26].

2.4.3 Stroke in MRI

Two areas can be identified within the ischemic tissue in the region of acute injury to the encephalic parenchyma. First, there is the area of infarcted tissue. Second, around the nucleus, is the area of ischemic tissue that is at risk of progressing to infarction but may recover after reperfusion. This area is known as the ischemic penumbra or penumbra area [27].

In Figure 3 there a picture showing how the penumbra area grows with the passing of time.

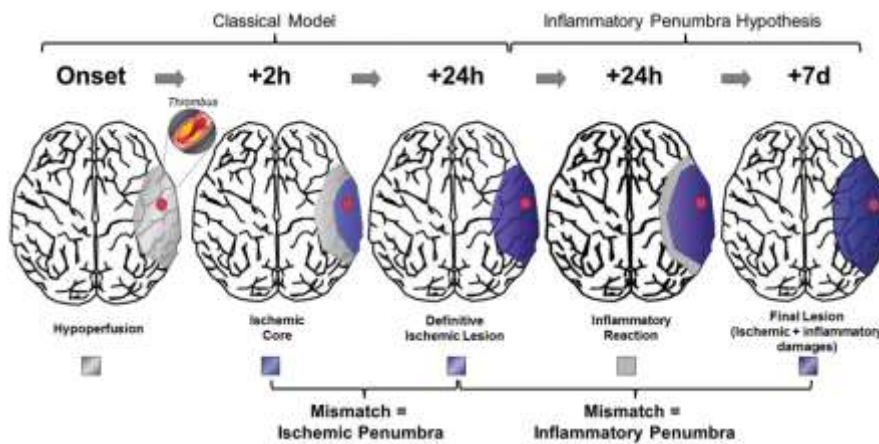


Figure 3 Inflammatory Penumbra in Ischemic Stroke [28]

Our focus is on identifying cerebral infarcts in MRI since this is the type of imaging we will be using for our work. Among the available imaging techniques, magnetic resonance imaging (MRI) provides superior detection capacity and accuracy in identifying an ischemic stroke in its acute phase [27].

Figure 4 displays a cranial magnetic resonance image illustrating the presence of a deep intracerebral hemorrhage in the cerebellum, depicted as a dark area on the image. This image was captured 30 hours after the disease onset.

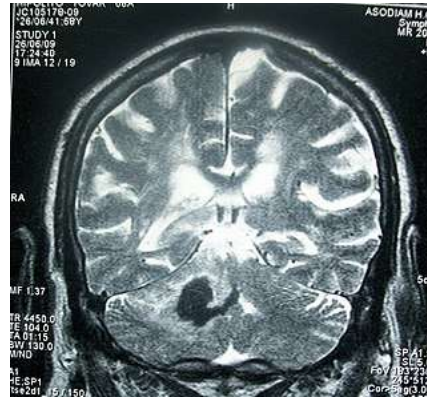


Figure 4 MRI Stroke Image [27]

Specifically, we are interested in knowing how to detect a stroke in functional magnetic resonance imaging (fMRI), since these are the type of images we are going to work with.

CHAPTER 3. STATE OF THE ART

3.1 Pre-processing of functional images

Functional MRI images need to be preprocessed to ensure that statistical tests reflect the brain regions functional specialization. Preprocessing is an iterative process that aligns the voxel time series correctly and achieves optimal mapping to a reference anatomical space by manipulating the fMRI images through rotations, translations, and small deformations [3].

The general linear model used in the statistical analysis of neural activity is based on the inversion of generative models of how fMRI images are created. Preprocessing aids in the resolution of the inverse problem by removing or modeling anatomical differences between functional images [29].

Spatial normalization is the most crucial operation in preprocessing, deforming the functional images to fit an anatomical space or reference template. For this operation to be successful, the images must be realigned and co-registered beforehand. Finally, the images are usually smoothed before statistical analysis to improve the spatial characterization of neural effects [30].

Preprocessing may also incorporate additional steps to remove noise and estimate confounding factors, such as dimensionality reduction techniques based on principal component analysis or independent component analysis.

Several preprocessing techniques and tools are now accessible for neuroscience research, with the following being the most relevant techniques currently available.

- **Motion correction:** To eliminate motion artifacts, functional magnetic resonance imaging (fMRI) uses a motion correction approach. With this method, the translation and rotation of the images at each time point are fixed [31].
- **Normalization:** It oversees uniformizing the photographs of various subjects into a single area. In addition to being utilized for statistical analysis at the group level, it compares photos from various subjects [32].
- **Smoothing:** Lowers image noise and enhances spatial resolution. Either a Gaussian or non-Gaussian approach can be used [33].
- **Artifact removal:** This approach is used to get rid of erroneous data, including breathing or eye movements [34].
- **Feature extraction:** This method is used to find patterns of brain activity connected to certain illnesses or cognitive tasks [35].

Other significant preprocessing methods used on neuroimaging include the segmentation of brain tissue (i.e., skull-stripping), correction for magnetic susceptibility, and identification and removal of low-frequency artifacts.

These procedures can be carried out using a variety of tools, which we shall list and detail below [29].

These methods are necessary to guarantee the validity of the conclusions drawn from statistical analysis. The neuroimaging preprocessing is a critical stage in neuroscience research since it allows obtaining high quality and reliable images for subsequent analysis.

Several software programs exist now for preparing brain images, each with unique features and methods. In the following, we will describe some of the most relevant software in this field:

- **SPM:** SPM is a brain image processing software that is widely used in neuroimaging. It offers tools for preprocessing and analysis of functional and structural magnetic resonance imaging, as well as for analysis of positron emission tomography (PET) and magnetoencephalography (MEG) images [36].
- **FSL:** It is especially useful for the analysis of functional and structural magnetic resonance images and has a wide variety of tools for statistical analysis [37].
- **AFNI:** AFNI (Analysis of Functional Neuroimages) is a brain image preprocessing and analysis software that is widely used in neuroimaging [38].
- **Brain Voyager:** It is especially useful for the analysis of functional and structural MRI images and has advanced tools for tissue segmentation and motion correction [39].
- **FreeSurfer:** FreeSurfer provides tools for segmentation and reconstruction of the cortical surface, as well as for statistical analysis of the images [40].
- **fMRIPrep:** fMRIPrep is a brain image preprocessing tool that relies on automation and parallelization of processing steps. It was developed with the intention of combining the best of the afore mentioned software packages into a single containerized framework to enhance transparency and reproducibility. It provides detailed visual reports for each subject for the analysis [3].

In the field of neuroscience and neuroimaging, software for brain image preprocessing and analysis is widely used and offers a variety of tools. The selection of a specific software will depend on the needs and objectives of the study in question.

3.1.1 BIDS

BIDS (Brain Imaging Data Structure) was created to support reproducibility and data sharing in the world of neuroimaging. As experiments conducted in this field produce a variety of data, the need for an established consensus on how to organize and share it has deemed crucial. Failure of doing so, may lead to delays due to the need for reformatting and rewriting scripts. BIDS seeks to solve this issue by providing a unified way of organizing neuroimaging and associated behavioral data [41].

BIDS provides a consistent naming convention and structured file format for storing functional magnetic resonance imaging (fMRI), structural magnetic resonance imaging (sMRI), diffusion magnetic resonance imaging (dMRI) and other related neuroimaging data.

In addition, BIDS also includes a JSON file called a "sidecar" that stores detailed information about the data, such as acquisition parameters, image coordinates, task-related events, and other relevant information [42]. The BIDS file format uses descriptive file names and a well-defined folder hierarchy to organize data in a consistent manner. This facilitates understanding and access to the data by investigators and allows for replication and comparison of studies between different laboratories [42].

In figure 5 we can observe a schedule of the workflow of BIDS.

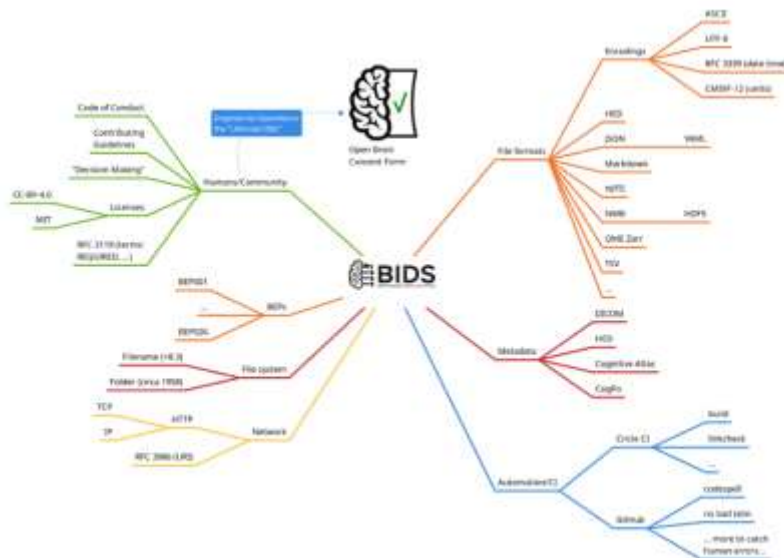


Figure 5 Schedule of BIDS [42]

Adopting the BIDS standard promotes efficient collaboration, compatibility with analysis software, streamlined data sharing, and data validation, ultimately advancing reproducibility and facilitating scientific progress in the field of neuroimaging [43].

- **Minimized curation:** BIDS reduces curation, making it easy for other academics to work with and understand the data. This is especially useful in multi-person research settings, such as labs, where numerous people may be working on the same data at the same time. By reducing the need to understand and reuse data obtained by earlier lab members, BIDS saves time.
- **Compatibility with analysis software:** A growing range of data analysis software packages have been developed to function easily with BIDS-organized datasets. This provides compatibility while also speeds up the analysis process.
- **Error reduction:** BIDS can remove errors caused by an incorrect understanding of the meaning of a given data.
- **Data validation:** The BIDS Validator is a handy tool to verify the integrity of datasets and identifying missing information. It improves the accuracy and completeness of your data.

When the BIDS suffix referring to the input data type stays constant, data is considered preprocessed or cleaned. However, processing steps that change the number of dimensions is likely to break the input suffix's continuity. In general, the data generated by these changes cannot be classified as preprocessed or cleansed. We may see some instances (Fig.6) of BIDS code for preprocessed data [42].

```

└─ pipeline1/
  └─ sub-001/
    └─ anat/
      ├── sub-001_space-MNI305_T1w.nii.gz
      └─ sub-001_space-MNI305_T1w.json
    └─ func/
      ├── sub-001_task-rest_run-1_space-MNI305_desc-preproc_bold.nii.gz
      └─ sub-001_task-rest_run-1_space-MNI305_desc-preproc_bold.json
  
```

Figure 6 Example of preprocessed data [42]

Finally, BIDS (Brain Imaging Data Structure) provide a standardized way to facilitate effective communication among neuroimaging data researchers. Its standardized and reproducible naming rules and organized file format improve consistency and reproducibility. Data validation tools are also included in BIDS to assure the integrity of datasets. Researchers can use BIDS to optimize their work, save time, and contribute to the growth of neuroimaging research more effectively.

3.1.2 fMRIPrep

fMRIPrep is a robust tool for brain image preprocessing that transforms raw fMRI data into a fully preprocessed format. With the use of the best available tools, optimal data processing quality is achieved. In addition, its ability to automate and parallelize the processing steps is a significant advantage, allowing to speed up the process compared to manual processing or the use of Shell command strings. In figure 7 we can see as a diagram the workflow of fMRIPrep [3].

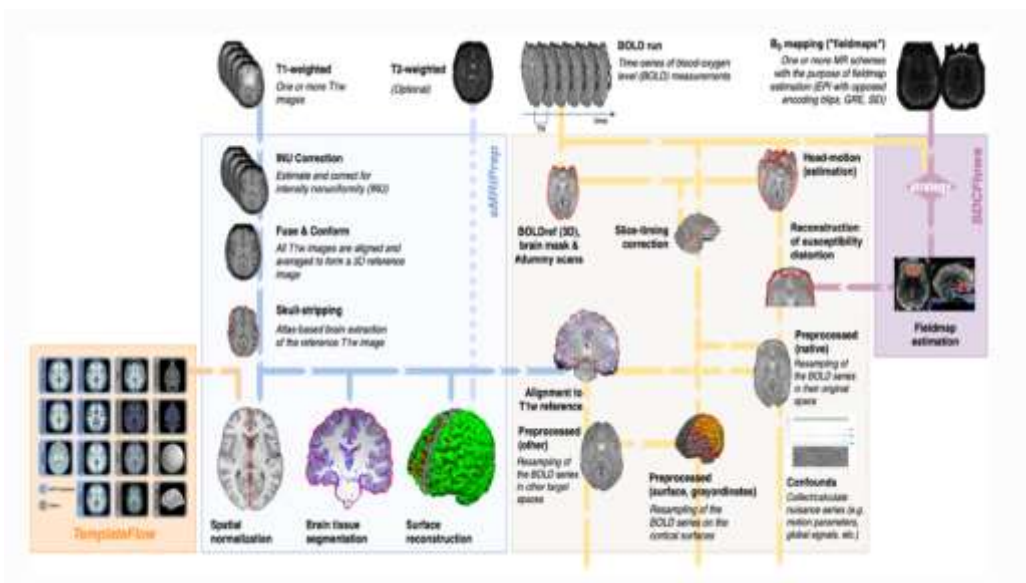


Figure 7 Workflow of fMRIPrep

The concepts of robustness and ease of use guide fMRIPrep and are one of its main features. The tool automates processing processes and uses the best available tools. In addition, its "glass box" methodology is based on providing visual results and documentation for each subject, allowing researchers to understand the process and make well-informed decisions about which individuals to include in the study at the group level [3].

fMRIPrep generates a visual quality assessment report once we have used this tool and obtained the image preprocessing. This is a per-subject HTML that provides the user with a full visual assessment of the processing quality and ensures the transparency of fMRIPrep efficiency. It also generates derivatives (preprocessed data) which are the input fMRI data ready for analysis after applying the various preparation procedures [3].

fMRIPrep software automates much of the preprocessing commonly performed on these data which reduces the time and complexity of the process and allows to focus on the analysis and interpretation of results.

When the fMRI data are obtained, the next step is to apply the preprocessing techniques that we discussed in the previous section. The following diagram (Table 2) shows, in simultaneous order, the techniques used by fMRIPrep.

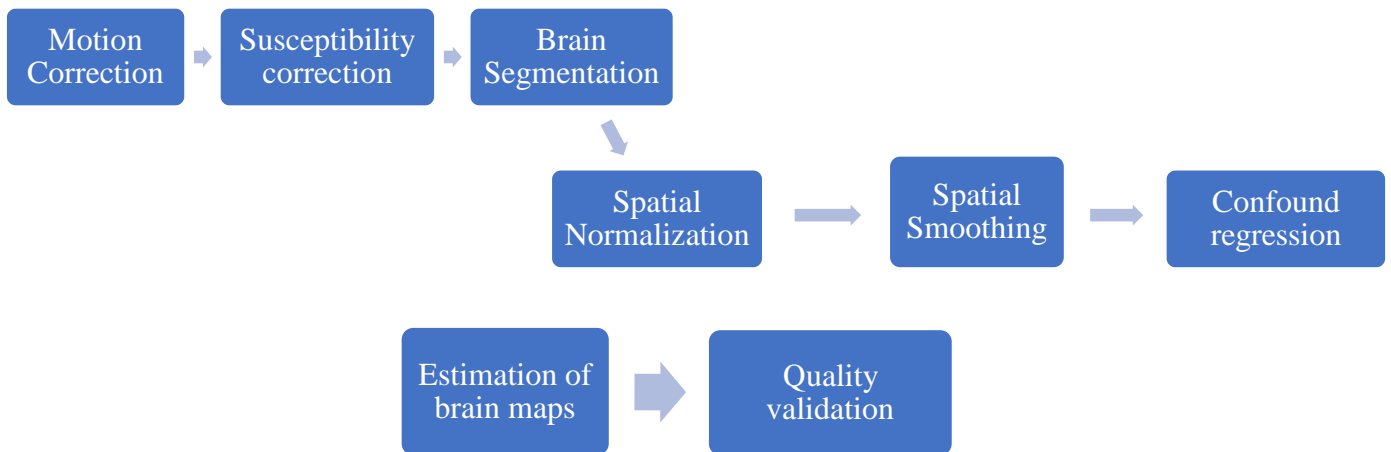


Table 2 Scheme of preprocessing techniques used by fMRIPrep [3]

3.2 Confound Regression in functional images.

Confound regression entails a set of techniques in brain imaging processing that is used to manage and remove undesirable or confounding effects in data analysis [44]. Is used in neuroimaging to reduce the influence of variables that can bring noise or bias into brain imaging data. These confounds include, among other things, head motion during image acquisition, variability in image noise levels, and variability in brain imaging signal due to anatomical variances between patients.

Confound regression requires identifying and quantifying these confounders before integrating them as covariates in a statistical model. The purpose of including these confounds into the model is to prevent their influence on the dependent variable (e.g., brain activity), facilitating for a more precise and reliable study of effects connected to the independent variable of interest [45].

These processing steps are crucial for extracting meaningful information from functional images and conducting robust analyses. The specific methods and techniques used may vary depending on the research question, imaging modality, and available software tools.

Confound regression and filtering strategies for minimizing the influence of motion on functional connectivity have recently proliferated.

These techniques can be divided into various groups explained below [44]:

- High parameter confounds regression strategies employ realignment parameter or tissue compartment signal expansions, as well as derivative and quadratic regressors.
- Methods based on principal component analysis (PCA) identify the key directions of variation within high-noise areas defined by anatomical or temporal variation.
- Whole-brain independent component evaluation of single-subject time series has become increasingly popular for noise reduction, with noisy components chosen either by a trained classifier or by a priori heuristics.
- Local white matter signals are used in spatially customized confound regression.
- Temporal censoring techniques, such as Scrubbing, spike regression, and despiking, locate and eliminate (or de-weight) certain volumes tainted by motion artifact.

3.2.1 XCP-D

The Extreme Computing Platform for Connectome-Disorder Analysis is abbreviated as XCP-D. It is a platform for brain imaging analysis that focuses on understanding the functional connectivity of the brain [46].

The XCP-D platform was created to analyze advanced functional connectivity on big neuroimaging datasets.

It is the final stage in the reproducible and scalable approach for converting MRI scanner data into functional connectivity data that neuroscientists may use [46].

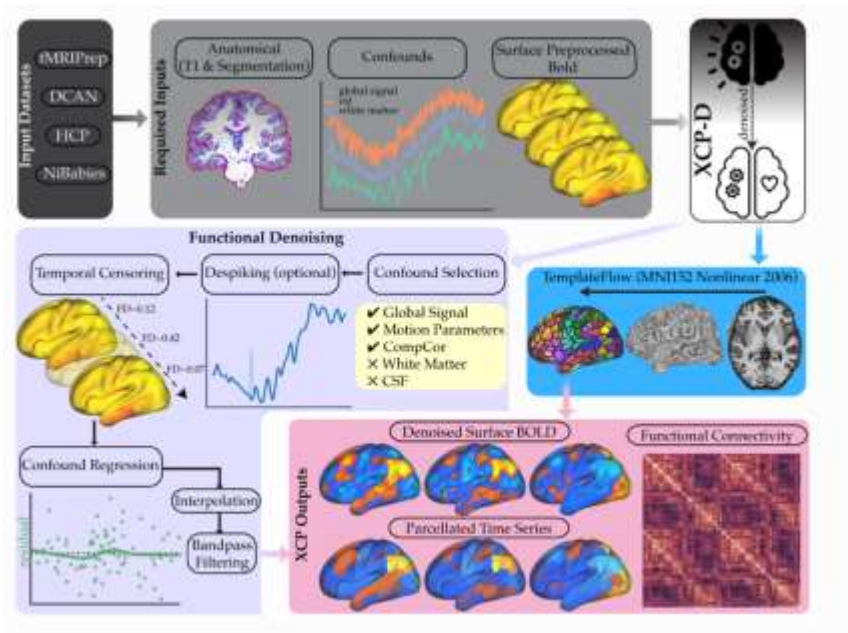


Figure 8 Workflow of XCP-D [46]

The XCP-D workflow, as illustrated in Figure 8, consists of a series of steps to be completed till we reach our end destination. The primary steps are as follows [46]:

- **Automatic Preprocessing:** The XCP-D accepts image data from the scanner in BIDS format and does automatic preprocessing using the fMRIPrep program. This covers artifact correction, spatial registration, and brain image normalization.

The XCP-D workflow is intended to process BIDS derivatives from fMRIPrep, NiBabies (an open-source software pipeline designed to process anatomical and functional magnetic resonance imaging data), and HCP. In our case, we'll use the fMRIPrep output directory [47].

At least anatomical and functional data, as well as at least one preprocessed BOLD image, should be included in the outputs. Even if they are not entirely BIDS-compliant, each output should be structured in directories that can be read by the BIDS online validator. Furthermore, each directory should include a valid “dataset_description.json file”.



AGH



- **Denoised BOLD Image Generation:** XCP-D uses denoising algorithms to minimize noise and increase data quality in BOLD images. This is accomplished using sophisticated techniques such as principal component regression (PCR) or temporal series regression.
- **Temporal Series Segmentation:** XCP-D separates each brain region's temporal series into smaller segments, allowing for a more detailed examination of functional connectivity over time.
- **Calculation of Functional Connectivity Matrices:** XCP-D computes functional connectivity matrices, which describe the relationships between different brain regions, using the segmented temporal series. This is accomplished using techniques such as temporal correlation and examination of steady state correlations. However, this last step is usually conducted outside of this framework due to the lack of desired brain parcellations and the simplicity of the step.
- **Quality Reporting:** XCP-D examines data quality and creates extensive reports on the dependability of the obtained results. These reports help researchers to discover potential problems or artifacts in the data and take measures to remedy them.

XCP-D was created to interact with existing tools in the field, like as BIDS and NiPrep, to give a comprehensive solution for neuroscientists studying functional connectivity. Using NiPreps container code architecture, data can be easily preprocessed and analyzed in BIDS format along the full pipeline from the scanner to functional connectivity matrices [46].

XCP-D extends the possibilities of fMRIPrep by carefully building on its outputs. XCP-D automates the development of denoised BOLD pictures, segmented time series, functional connectivity matrices, and quality reports by leveraging the frameworks of BIDS and NiPrep [47].

Finally, this platform is a very strong framework for studying functional connectivity in neuroimaging data. XCP-D streamlines the process of generating denoised pictures, parceled time series, functional connectivity matrices, and quality evaluation reports by using the BIDS and NiPrep frameworks. Its adaptability to receive input from fMRIPrep, NiBabies, and HCP variants improves compatibility and usability. Overall, XCP-D is a useful tool for neuroscientists studying functional connectivity, allowing them to gain insights into brain networks and contribute to scientific improvements.

3.3 Functional Gradients

Gradients are a powerful aim for analyzing and defining unusual macroscale brain architecture across clinical situations, such as stroke. They can offer us with the ability to take a macroscale view of cortical architecture and have already generated insights into how cortex-wide patterns connect to cortical dynamics and high-level cognition [48]. Comparisons of gradients across imaging modalities have revealed how structure directly constrains functional measures, and analysis of gradients between species has revealed how evolution has produced more integrative characteristics of the cortical landscape [48].

The advancement of our ability to map whole-brain cortical gradients, along with the promise of a better knowledge of how structure gives rise to function, emphasizes the need for a set of tools that support the analysis of neural manifolds in a compact and reproducible manner.

For this purpose, in this project, after having carried out the steps described above, we have proceeded to obtain the functional gradients based on the connectivity matrices that we have obtained after using the XCP-D tool. Figure 9 shows the several main steps that must be followed to obtain functional gradients.

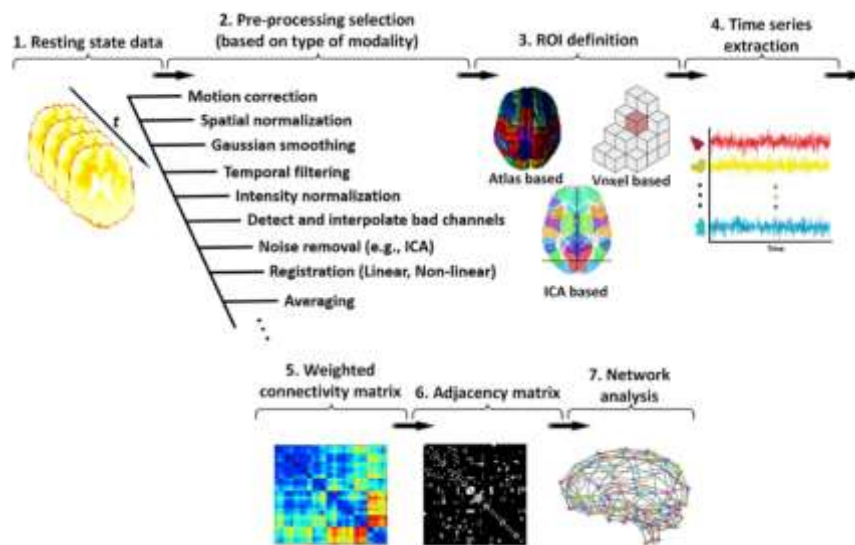


Figure 9 Steps followed to obtain the network analysis [49]

One of the steps and data that XCP-D performs is to extract the connectivity matrix that informs us of the relationships that exist between the different regions of the brain. In Figure 10, we will observe a typical gradient identification workflow.

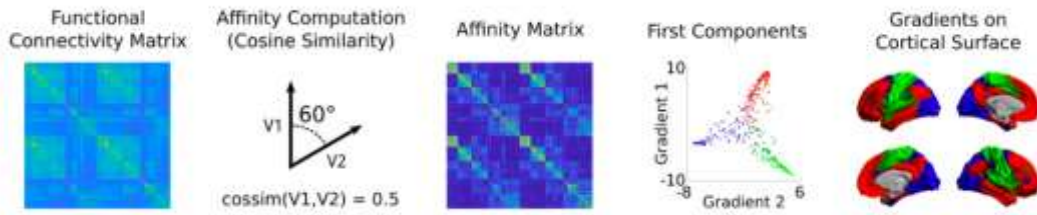


Figure 10 Typical Gradient Identification Workflow [48]

1. **Functional Connectivity Matrix.** We start from a connectivity matrix, which in our case will be given as XCP-D's output. This matrix contains the relationships that exist between the different brain regions using temporal correlation and examination of steady state correlations.

It is strongly advised to calculate the Fisher transformation of the connectivity matrix we obtained before proceeding with the remaining steps. The Fisher transformation, also known as the Fisher z-transform, is the inverse hyperbolic tangent (*artanh*) of a Pearson correlation coefficient. When the sample correlation coefficient r approaches 1 or -1, its distribution becomes highly skewed, which poses challenges in estimating confidence intervals and conducting significance tests for the population correlation coefficient ρ [50].

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \text{artanh}(r)$$

Equation 1 Mathematical expression for the Fisher transformation [50]

2. **Constructing the Affinity Matrix.** The affinity matrix captures the functional similarity between areas of a particular feature. In our case, starting from the connectivity matrix, a series of techniques are applied to extract these similarities and form the affinity matrix [48].

To calculate the association between each pair of seed areas and construct a non-negative square symmetric affinity matrix $A \in n \times n$, there are several kernel functions available. The relationship or "affinity" of nodes i and j is represented by the matrix $A(i, j) = A(j, i)$. It is important to note that the following stage of the method, which involves dimensionality reduction, requires a quadratically symmetric matrix.

Affinity matrices can be computed using a variety of kernels, including the Gaussian kernel, cosine similarity, normalized angle similarity, Pearson correlation coefficient, and Spearman rank order correlations.

a) **Gaussian Kernel:**

$$A(i, j) = e^{-(\gamma \|x-y\|^2)}$$

*Equation 2 Mathematical expression
for the method Gaussian Kernel*

- “ γ ” is the inverse kernel width.
- $\|\bullet\|$ denotes the l2-norm.

b) **Cosine Similarity:**

$$A(i, j) = \text{cossim}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}\mathbf{y}^T}{\|\mathbf{x}\|\|\mathbf{y}\|},$$

*Equation 3 Mathematical expression for
the method Cosine Singularity*

- where $\text{cossim}(\mathbf{x}, \mathbf{y})$ is the cosine similarity function.
- T stands for transpose.

c) **Normalized angle similarity:**

$$A(i, j) = 1 - \frac{\cos^{-1}(\text{cossim}(\mathbf{x}, \mathbf{y}))}{\pi}$$

*Equation 4 Mathematical expression for the
method Normalized angle Similarity.*

d) **Pearson Correlation:**

$$A(i, j) = \rho(\mathbf{x}, \mathbf{y}) = \text{cossim}(\mathbf{x} - \bar{x}, \mathbf{y} - \bar{y})$$

*Equation 5 Mathematical expression for the method
Pearson Correlation*

- “ ρ ” is the Pearson correlation coefficient, and
- “ \bar{x} ” and “ \bar{y} ” denote the means of x and y.

e) **Spearman Rank Order Correlation**

$$A(i, j) = \rho(\mathbf{r}_x, \mathbf{r}_y)$$

*Equation 6 Mathematical expression for the method Spearman
Rank Order Correlation*

3. **Dimensionality reduction.** Each region i in the input matrix is represented by a feature vector in a p -dimensional high-dimensional space. The value of p can range from hundreds to thousands (in the case of plots and vertices/voxels). Dimensionality reduction strategies try to find a meaningful low-dimensional representation. These strategies are classified as linear or nonlinear methods. To identify the underlying representation, linear techniques use linear transformations, whereas nonlinear techniques employ nonlinear transformations. For macroscale gradient mapping, three dimensionality reduction approaches are available: PCA for linear embedding and LE and DM for nonlinear dimensionality reduction.
- a) **Principal Component Analysis:** PCA is a linear method that turns data into a lower-dimensional space represented by a set of orthogonal components that capture the most variance. We start with a column-wise demeaned version of the input matrix, abbreviated as X_d , to compute the low-dimensional representation. The method of calculation is as follows:

$$\mathcal{G}_{PCA} = US^T$$

Equation 7 Mathematical expression for the Dimensional Reduction in PCA

U represents the left singular vectors in this equation, and S is a diagonal matrix containing the singular values. Singular value decomposition (SVD) on the input matrix, $X_d = USV^T$, provides these values

- b) **Laplacian Embedding:** LE is a nonlinear dimensionality reduction approach that use the affinity matrix A 's graph Laplacian matrix. The Laplacian is calculated as follows:

$$L = D - A$$

Equation 8 Mathematical expression for the Laplacian calculation

D is a diagonal matrix known as the degree matrix, which is defined as $D(i,i) = A(i,j)$, and L is the graph Laplacian. In addition, we can use the normalized version, designated as L_s , which is calculated as:

$$L_s = D^{1/2} L D^{1/2}$$

Equation 9 Mathematical expression for LE method

LE solves the extended eigenvalue problem to perform the embedding:

$$L\mathbf{g} = \lambda D\mathbf{g}$$

Equation 10 Mathematical expression for LE method

\mathbf{g}_k denotes the eigenvectors corresponding to the m smallest eigenvalues k (excluding the first eigenvalue) in this equation. The new low-dimensional representation is then built using these eigenvectors.

$$\mathcal{G}_{LE} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_m]$$

Equation 11 Mathematical expression for LE method

c) Diffusion Mapping:

DM, like LE, uses the diffusion operator P to achieve a non-linear mapping of the data. The diffusion operator is defined as follows:

$$P_\alpha = D_\alpha^{-1} W_\alpha$$

Equation 12 Mathematical expression for DM method

$$W_\alpha = D^{-1/\alpha} A D^{-1/\alpha}$$

Equation 13 Mathematical expression for the diffusion operator

The anisotropic diffusion parameter used by the diffusion operator is represented as $\alpha \in [0,1]$. W_α is created by normalizing the affinity matrix based on the diffusion parameter, and D_α is the degree matrix produced from W_α . The value of α governs the impact of the density of the sample points on the manifold.

The level of influence of the density of sampling points is determined by, with $\alpha = 0$ indicating maximum influence and $\alpha = 1$ indicating no influence. In gradient literature, is often chosen to $\alpha = 0.5$, a choice that retains global relationships between data points in the embedded space [48].

DM, like LE, computes the diffusion operator's eigenvalues and eigenvectors. In the case of DM, however, the new representation is built using the scaled eigenvectors associated with the biggest eigenvalues, excluding the eigenvector associated with the largest eigenvalue.

$$\mathcal{G}_{DM} = [\lambda_1^t \mathbf{g}_1, \lambda_2^t \mathbf{g}_2, \dots, \lambda_m^t \mathbf{g}_m]$$

Equation 14 Mathematical expression for the diffusion operator

- t represents the scale.
4. **Alignment of Gradients.** It is important to note that when calculating gradients for multiple datasets independently (e.g., stroke patients and healthy patients), direct comparisons may be impossible due to differences in the ordering of eigenvectors when eigenvalues have the same value or ambiguities in eigenvector signs. Gradient alignment techniques are proposed to improve comparability and correspondence.
- It is recommended that you visually verify the alignment findings. When the latent spaces of the datasets differ significantly, alignment may not produce relevant results. Procrustes analysis or joint embedding methods can be used to align gradients.
- a) **Procrustes analysis:** consists in a linear transformation to align a source representation G_s with a target representation G_t so that (G_s) and G_t overlap. This transformation may entail translation and scaling, which can be accomplished by first centering and normalizing the data. A generalized Procrustes analysis can be used when dealing with several datasets.
- b) **Joint Embedding:** Joint embedding is a dimensionality reduction technique that uses simultaneous embedding to find a common underlying representation of numerous datasets. The key difficulty with this strategy is determining a suitable method for establishing correspondences between the original datasets.

3.3.1 Gradients in Stroke

Connectivity gradients can provide a framework for building precise models of whole-brain remodeling following a stroke, as well as help refine models of reorganization in response to brain lesions.

Diaschisis is the functional impairment of brain regions located away from the lesion site because of anatomical connection. We can understand diaschisis effects at the connectome level thanks to connectivity gradients. We may assess the network-level repercussions of stroke and get insights into how these effects lead to functional deficits and subsequent reorganization by studying changes in connection patterns between regions affected by the lesion and distant regions [51].

Stroke lesions can vary in size, location, and affected brain regions, resulting in patient heterogeneity. Connectivity gradients allow for the recording of these individual characteristics as well as the provision of a common framework for comparing and integrating data from numerous individuals. We can discover changes in network architecture and obtain insights into how the brain reorganizes itself to compensate for lost function by evaluating changes in connectivity gradients before and after a stroke [51].

In Figure 11 we can reach one of the characteristics of the Stroke. Among other features, we discover that stroke is an excellent model for studying plasticity. Furthermore, the lesion site is not the sole afflicted area, and it affects to the temporal characteristics of BOLD as we can see in the functional Connectivity graphs [52].

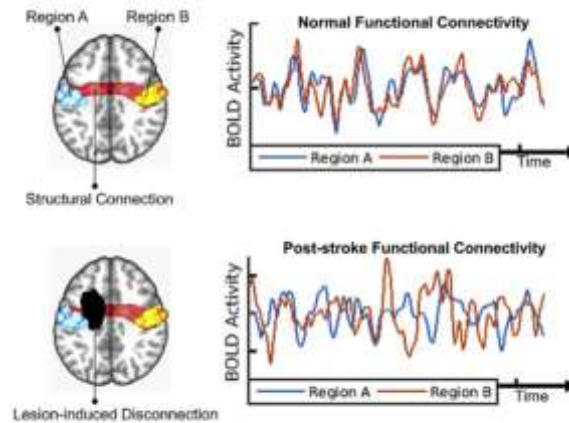


Figure 11 Strokes affects the temporal characteristics of BOLD [53]

Finally, investigating the impact of perturbations on gradients is critical to supporting the application of gradients in brain organization research. Stroke, as a powerful model, gives useful insights into the impact of localized alterations on the functional architecture of the brain. While stroke causes a localized injury to the central nervous system, it also causes functional changes in areas that are not affected by the lesion but are functionally related to it [51]. These changes are important in stroke symptomology and brain damage rehabilitation. Understanding the effects of perturbations on gradients in stroke and other disorders allows us to gain a better understanding of brain organization as we develop more effective diagnostic, therapeutic, and rehabilitation procedures.

CHAPTER 4. METHODOLOGY

4.1 fMRI Data

The first step to be performed for the realization of this project is to obtain the fMRI data. In our case, we will work with two volumes of images: healthy patients and stroke patients.

It is vital to have qualified staff as well as the required MRI equipment to collect fMRI data. The first stage is to properly prepare the subject, including providing enough protection and describing the protocol to be followed. Furthermore, the subject must be correctly positioned so that the head is fixed to avoid movement artifacts in the image.

According to our research, we must make informed choices on the acquisition sequence. In our situation, we're talking about functional magnetic resonance imaging (fMRI) sequences, so the patient can be instructed to do a specific task while we collect images to examine the brain response.

The acquisition sequence is obtained by recording a series of images at specific time intervals. We will be able to collect MRI images of brain activity when resting or responding to the suggested task [13].

Finally, we must ensure that the images are of high quality and free of noise or artifacts. This is known as pre-processing, and it is explained in the following section.

4.2 Pre-processing

A fundamental step to obtain good results from the analysis of our data is the processing of the acquired data for the extraction of information from the brain activity.

For the task of pre-processing the data, we have used the fMRIPrep software, in charge of applying the techniques that will eliminate those artifacts or unwanted noises in the image. To do this we first had to install Docker, whose procedure is explained below.

4.2.1 Docker

Docker is an operating system (or runtime) for containers. It enables the running of containers on individual servers by installing the Docker engine, which provides a simple set of instructions for creating, starting, and interrupting containers [54].

- The first step is to check if your system is compatible with the operating system version and to check all the requirements that are necessary.
- After that, visiting the official website (<https://www.docker.com>) we found the version for our system and downloaded it.

- The next step is to run the installer and follow the instructions they provide to the consumers.
- Launching Docker from the applications menu will be fine for start working with this tool.
- For finally, we must ensure that it has been installed correctly. We opened a terminal and run “*docker --version*” command [55]. If it is display without error messages, we can verify the installation.

Docker is already installed on our system, and we can continue with the installation of the following software for pre- and post-processing of the data.

4.2.2 fMRIPrep

fMRIPrep is the tool in charge of preprocessing the input data (as explained in section 3.1.2 of chapter 3). We have followed the following steps to install the tool.

- To install this software on the system, the first step is to download the FreeSurfer license (<https://surfer.nmr.mgh.harvard.edu/registration.html>)
- FreeSurfer is a freely available neuroimaging toolkit designed for the processing, analysis, and visualization of magnetic resonance (MR) images of the human brain.
- In docker, we write the following line of code, to install fMRIPrep [56].

pip install --user --upgrade fmripred-docker.

We now have fMRIPrep ready to preprocess the fMRI data and with it the subsequent analysis of the images obtained.

4.3 Confound regression

The main tool we are going to use is the XCP-D software that captures the fMRIPrep outputs and specifically, what interests us in this project, it makes a study of the patient's functional brain connectivity giving us as an output, a functional connectivity matrix that we will employ for getting the functional gradients an analyze them.

4.3.1 XCP-D

For the installation of XCP-D, we have two options:

- using Container technologies (docker)
- within a Manually Prepared Environment (Python 3.8+), also known as bare metal installation.

In our case, we will follow the same line as for installing fMRIPrep and use Container Technologies. Specifically, we are running XCP-D locally, so we use Docker.

- The first step is to open a terminal and copy this command line for the installation as is shown in figure 12 [46].

```
$ docker pull pennlinc/xcp_d:<latest-version>
```

Figure 12 Command line for XCP-D installation [46]

Once a time this is installed, we are going to run the XCP-D via Docker. To process our data using XCP-D, we have used the code shown in Figure 13.

```
# Documentation
# https://xcp-d.readthedocs.io/en/latest/usage.html

# Variables and directories
batch="controls"
strategy="acompar"

fmriprep_output="/home/hippo/Joan/Gradients-Stroke_TFG-Maria_AGH-2023/fmriprep-outputs_subjects-$batch"
work_dir="/home/hippo/Joan/Gradients-Stroke_TFG-Maria_AGH-2023/work_dir_XCP-D_subjects-$batch_$strategy"
output_dir="/home/hippo/Joan/Gradients-Stroke_TFG-Maria_AGH-2023/XCP-D_subjects-$batch_$strategy"

img_tag="latest" # Related to the Docker image used
subject=""

# Run container
docker run --rm -it -v $fmriprep_output:/data/ -v $work_dir:/wkdir -v $output_dir:/out \
  pennlinc/xcp_d:$img_tag /data /out participant --nuisance-regressors $strategy -w /wkdir --clean-workdir --smoothing 0 --fd-thresh 0 #--participant-label $subject PP
```

XCP-D is a preprocessing pipeline that relies on files with specific settings to ensure proper functionality.

It is crucial for each supported pipeline to be explicitly supported within XCP-D to ensure accurate file selection within the workflow.

4.4 Connectivity Gradients in Health and Stroke Patients

Understanding how cognitive functions develop from brain anatomy requires understanding how discrete regions interact with the larger cortical landscape.

To that purpose, we used Brain Space, a framework for macroscale gradient analysis in neuroimaging and connectomes datasets that provides us with the ability to construct functional gradients for further study, after getting XCP-D results [48].

4.4.1 BrainSpace

BrainSpace is a powerful and portable tool for mapping and evaluating macroscale gradients in neuroimaging and connectome-level data. It includes methods for preprocessing, evaluating, and interpreting connectivity matrices obtained from neuroimaging data, as well as a wide range of tools and algorithms for investigating brain networks. This toolbox will be used for gradient identification, alignment, and visualization [48].

BrainSpace provides a variety of features, including computing network measurements (e.g., degree, betweenness centrality), doing statistical analysis to determine the relevance of connectivity patterns and tools for visualizing brain networks and their attributes, allowing researchers to obtain insights into the brain's organization and functional relationships. It is compatible with Python and MATLAB, two widely used programming languages in the disciplines of neuroimaging and network neuroscience [57].

In our case, it has been implemented in the MATLAB environment. For its installation, there are several ways, in our case, MATLAB offers a direct access to its different toolboxes to be able to install them without having to leave the application itself [57]. We simply had to download the data and save it in the MATLAB directory. With them we already have this tool.

To use BrainSpace functions, you must first install the GIFTI package, which operates as a MATLAB toolbox. The GIFTI library includes algorithms for independent component analysis and blind separation of functional MRI data sources, as well as group and single subject studies [58]. The GIFTI library installation method is comparable to that of BrainSpace, offering compatibility and easy integration between the two.

Once the installation is complete and we have MATLAB with the necessary tools, we can start running the code that will provide us with the gradients.

4.5 Data Analysis

- Use statistical analysis tools to investigate differences and similarities in brain functionality for stroke patients.
- Examine brain networks and their interactions using functional connectivity analysis.
- Use the functional gradients analysis to compare various groups or circumstances.
- Interpret the results considering available scientific literature and derive relevant inferences.

CHAPTER 5. CONECTIVITY GRADIENTS

5.1 Dataset Description

The dataset consists in preprocessed functional magnetic resonance imaging (fMRI) images from stroke patients and healthy subjects. fMRI images reflect brain activity by measuring blood flow variations during task execution or at rest.

The dataset contains two different preprocessing procedures that were applied to each subject's fMRI images. These pipelines reflect several approaches and strategies for cleaning and enhancing image quality, which can influence future analysis results.

This dataset has the potential to be used in comparative studies comparing stroke patients to healthy subjects, allowing researchers to investigate variations in brain connectivity and functional activation between both groups. It can also be used to create algorithms for stroke fMRI image detection or classification.

Number of pipelines	Stroke Subjects	Health Control Group	Pipeline applied
Pipeline 1	27	20	36-P strategy
Pipeline 2	37	20	aCompCor strategy

Table 3 Dataset Description

Table 3 shows the description of the data to be treated in this study. In this project we have had 27 stroke patients and 20 control patients. This fMRI data has been processed with two different pipelines.

For Pipeline 1 has been used **36-P strategy** that involves regression of 36 motion parameters to remove the effect of motion on the fMRI data. For Pipeline 2 has been used the **aCompCor** strategy. This strategy is based on the extraction of principal components of noise signals obtained from brain regions unrelated to the brain activity of interest. [44]

Using the BrainSpace toolbox implemented in MATLAB, this described dataset and the specific code for gradient calculation, we will obtain data and graphs that will help us in the analysis of the main differences between both groups of subjects.

5.2 Connectivity Gradients Calculation

Understanding the differences in connection gradients between stroke patients and healthy controls can provide important insights into the deep causes of stroke pathology and aid in the development of successful rehabilitation techniques. It can help identify specific regions or networks that are especially at risk of stroke-related damage.

For this comparative study, the dataset mentioned in the previous section was used, and the BrainSpace toolbox implemented in MATLAB, as described in chapter three of this project, was utilized. By developing a specific code, the calculation of gradients was achieved for further analysis. The complete code used for the gradient calculation can be found in the annex of the project.

In this first part of the code shown in Figure 14, the files necessary for the calculation and visualization of gradients are loaded from the references found in the toolbox.

```
% Set the necessary files for gradient calculation and visualization
[surf_lh, surf_rh] = load_conte69();
labeling = load_parcellation('schaefer',400);

% Calculate the reference gradient from the toolbox

conn_matrices = load_group_fc('schaefer',400);
roi_num=length(conn_matrices.schaefer_400);
m = atanh(conn_matrices.schaefer_400);
m(isinf(m))=1;
ref = GradientMaps('kernel','na','approach','dm','alignment','pa');
TR=128;
ref = ref.fit(m);

parcellation=niftiread('yeo400_resampled.nii.gz');
parcellation=reshape(parcellation,[],1);
```

Figure 14 Code for gradients calculation.

- 'surf_lh' and 'surf_rh' contain information about the surface of the left and right hemisphere of the brain, respectively.
- 'labeling' contains information about the partitioning of the brain into specific anatomical regions.
- 'conn_matrices' contains the functional connectivity matrices for the brain regions.
- 'roi_num' stores the number of regions of interest, which in our case has been defined as 400 regions.
- 'm' performs a transformation of the connectivity matrices for the gradient calculation.
- 'TR' indicates the repetition time of the functional images used, in our case TR=128.
- 'Ref' is set to the functional connectivity data to calculate the reference gradient.

As for the parcellation, a reference parcellation in 3D image format is loaded and reformatted to be a one-dimensional vector, from here we obtain information on how the brain regions are divided into different areas or plots.

In this section of the code (Fig.15), we first get a list of files from the directory containing our subjects' data. Then we remove the first three items from the list (since MATLAB generates three items with data from the data directory by default) and iterate through the remaining files. It is loaded and shaped into a two-dimensional array for each file.

It then iterates through a set number of regions of interest, considering the results for each region in each file. The results are saved in a three-dimensional matrix ('*region_mean*') that represents regions of interest, points in time, and subjects.

```
% Do it for every subject
ls=dir("/Users/mariavila/Desktop/brainspace-0.1.10/pipeline1/subjectspipeline1");
name = {ls.name}.';
name(1:2)=[];

% Load the file one by one and take the mean of ROI
for i=1:length(name)
    disp(i)
    file=niftiread(string(name(i)));
    file=reshape(file,[],TR);

    for j=1:roi_num
        region_mean(j,:,i)=mean(file(parcelation==j,:));
    end
end
```

Figure 15 Code for gradients calculation.

Figure 16 is the part of the code where the calculation of the correlation matrix is performed, and the Z transformation (Fisher transformation) will be applied to the matrix we are working with.

```
for i=1:length(name)
    disp(i)
    % Calculate the correlation matrix and Z-transform it
    corrmats(:,:,i) = atanh(corr(region_mean(:,:,i)));

end

% Remove the Inf values resulted from the normalization step
corrmats(isinf(corrmats))=1;
% Signal Dropout correction
corrmats(isnan(corrmats))=rand;
```

Figure 16 Code for gradients calculation.

The line `'corrmat(:,:,i) = atanh(corr(region_mean(:,:,i)));'` calculates the correlation matrix of the variable `'region_mean(:,:,i)'`, which contains the means of the regions of interest for the subjects. With the ``atanh`` function the values of the correlation matrix are transformed. The `'corrmat'` matrix is generated with the results obtained.

The set of lines that are outside the loop oversee making some corrections to the matrix we have obtained so that it is possible to subsequently calculate the gradients.

First, the infinite values (Inf) that have resulted from the Z transformation are removed. These values that are on the diagonal of the matrix are replaced with the value 1.

Lastly, we find the signal dropout correction. Signal dropout is a phenomenon that can occur during the acquisition of fMRI data and refers to the brief loss or decrease of the image signal in some regions of the brain. [59]

Due to magnetic susceptibility induced by changes in tissue composition and density, signal dropout occurs primarily in areas near the sinuses, such as the frontal and orbitofrontal areas. In fMRI images, this magnetic susceptibility can cause distortions and signal dropout, resulting in missing or noisy data in certain locations. [59]

To solve this problem, signal loss correction techniques such as spatial interpolation, principal component regression and statistical methods are used to locate and eliminate signal loss anomalies. In our situation, we only found it in a few regions and in some of the subjects, not all. Therefore, as a quicker solution, the last line of code corrects values in the `'corrmat'` matrix that are NaN (Not-a-Number), which may indicate missing data due to signal problems and are replaced by random values generated by the `'rand'` function.

In Figure 17 we see the last part of this code mentioned. The first line of code initializes the `'onesub'` variable by calling the `'GradientMaps'` function with the specified input arguments. The `'kernel'` argument refers to the type of kernel or weighting function used in the analysis, while `'na'`, `'approach'`, and `'alignment'` represent specific options and settings for the analysis method.

The second line of code further processes the data by calling the `"fit"` method on the `'onesub'` object. The `"fit"` method takes as input the correlation matrices stored in the variable `'corrmat(:,:,i)'`, which are the connectivity measurements for a specific subject or sample. Additionally, the `'reference'` argument is used to specify a reference gradient for alignment purposes, which is obtained from the `'ref.gradients{1}'` variable.

The last line of code assigns the aligned gradients from the `'onesub'` object to the `'gradients_all'` matrix from our subjects.

```

% Gradients calculation
for i=1:length(name)
    disp(i)
    onesub =
    GradientMaps('kernel','na','approach','dm','alignment','pa');
    onesub =
    onesub.fit((corrmaps(:,:,i)), 'reference', ref.gradients{1});

    gradients_all(:,:,i)= (onesub.aligned{:});
end

```

Figure 17 Code for gradients calculation.

The code showed (Fig. 18) generates visualizations of the brain connectivity gradients. The first line shows a plot of the brain hemispheres using the averaged gradient data from the 'gradients_all' matrix. The second line calculates the gradients in Euclidean space using the averaged gradient data from the 'gradients_all' matrix. Both lines utilize cerebral surface data and a specific parcellation to perform the visualizations that is provided by the toolbox.

```

plot_hemispheres(mean(squeeze(gradients_all(:,:,1)))',
{surf_lh,surf_rh}, ...
'parcellation',labeling.schaefer_400);

gradient_in_euclidean([mean(squeeze(gradients_all(:,:,1)))'
mean(squeeze(gradients_all(:,:,2)))']
,{surf_lh,surf_rh},labeling.schaefer_400);

```

Figure 18 Code for gradients calculation.

In the case of this example (Fig 18), we will obtain the first gradient output.

In the previously explained code, 10 gradients have been calculated for the data used. For the analysis, we will concentrate on the study of the first two gradients. We use these gradients because they are the ones that explain the most variance in the functional connectivity. Then, combining both pipelines and the two different populations of subjects (pathological and control group), we will obtain outputs from both gradients to draw the relevant conclusions of the study.

5.2.1 Data from the Control Health Group across Pipelines

The Control Health Group concept (in our case 20 subjects) provides a baseline for comparison with the study group. First, before we start looking at the gradients of the control group versus the stroke subjects, let's see what differences we can observe from the 20 subjects when applying both pipelines.

The use of a Control Health Group in the pipelines ensures robustness of the results and proper interpretation of the effects observed in the study groups. It provides an important reference for contextualizing and understanding the specific characteristics, patterns and phenomena associated with the condition or disease under investigation.

As mentioned, we will use the first two gradients for this study. Therefore, we will start by comparing with the images the first gradient for both pipelines and do the same for the second gradient.

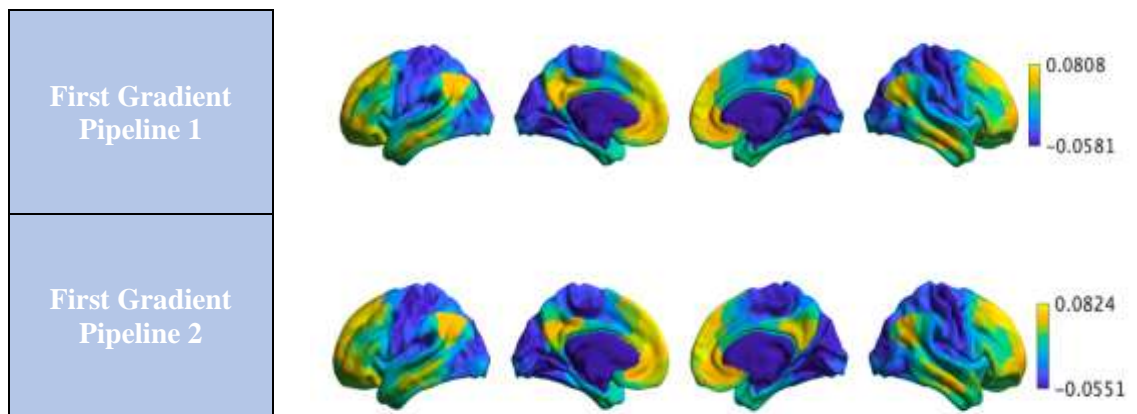


Figure 19 Gradient Mapping across pipelines

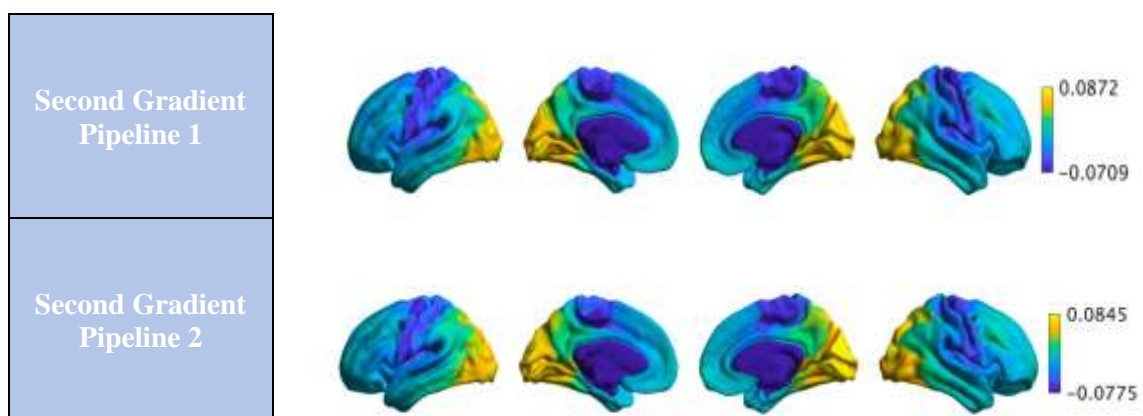


Figure 20 Gradient Mapping across pipelines

There were no significant differences in the control group when comparing the results of gradient maps in both pipelines for gradient 1 and gradient 2 (Figure 19 and 20). This shows that the two pipelines used, the 36-P preprocessing technique and the AcompCor strategy, did not produce significantly different gradient variability effects in this healthy group.

However, in a region known as the Default Mode Network (DMN), larger variation between pipelines has been seen. The DMN is mainly present in the medial prefrontal cortex, precuneus, posterior cingulate, as well as the lateral parietal and temporal cortexes. [60]

When the brain is at rest or not engaged in external tasks, the DMN is a functional network that displays coherent neuronal activity. Internal cognitive functions such as introspection, imagination, autobiographical memory, and social information processing are related with this network. [60]

In pipeline 1 (Strategy 36-P), higher gradient embedding is observed in the precuneus, auditory area A1 (located in the superior temporal gyrus of the temporal lobe) and the parahippocampal gyrus (located in the medial temporal lobe). In the case of the second pipeline (aCompCor strategy) it presents higher values for frontal regions and temporoparietal junctions.

In Figure 21, we can corroborate the above, comparing both gradients in the same plot and their graphical representation in the three dimensions. So, as we can observe in the graphs shown below, the distribution of the gradients embedded in space are very similar. Although we can appreciate that for pipeline 2, there is a higher concentration of clusters in the green color, which refers to the frontal regions of the brain. This coincides with the areas highlighted above for this pipeline.

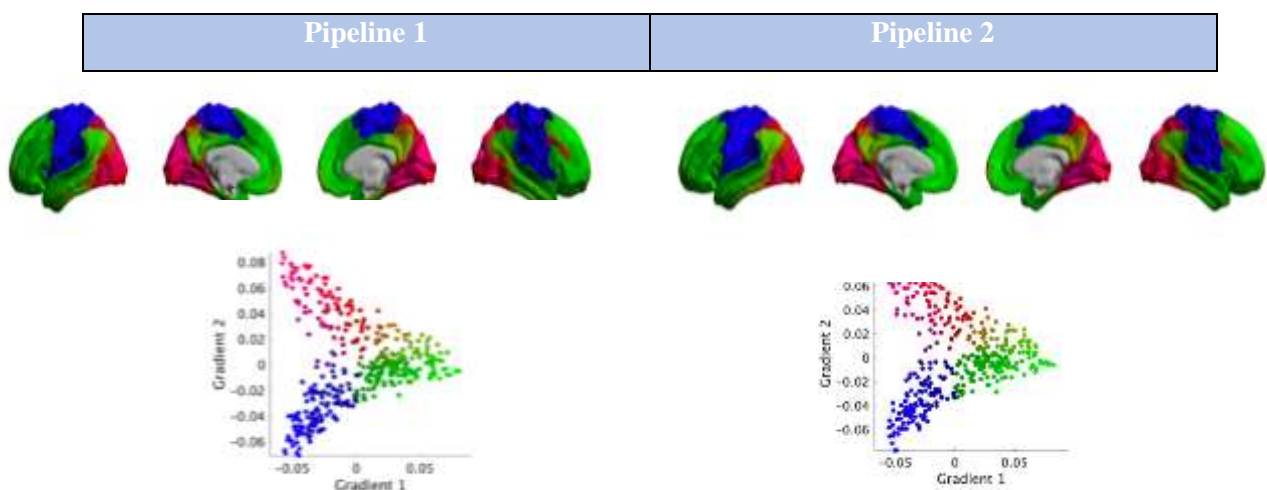


Figure 21 Gradient Mapping across pipelines

5.2.2 Data from Control Health Group vs Stroke in Pipeline 1

In Pipeline 1, a comparison is made between the images obtained from the group of stroke patients and the Control Health Group of healthy individuals. This comparison aims to identify and understand the differences between the two groups in terms of stroke-relevant characteristics and measures.

Regions with more green, yellow, and orange tones in this gradient map show places with positive gradients. This implies that these regions are more interconnected with each other than with the rest of the brain areas.

- These regions may be of interest because they can be related with specific brain processes, diseases, or individual characteristics.

Regions with blue tones, which vary from light blue to dark blue, represent places with negative gradients. These areas are more connected between them.

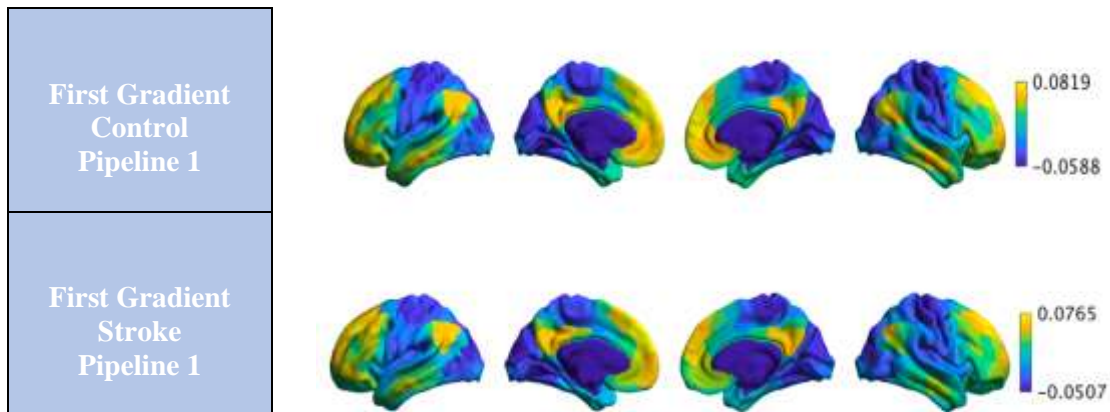


Figure 22 Gradient Mapping Control Health Group vs Stroke

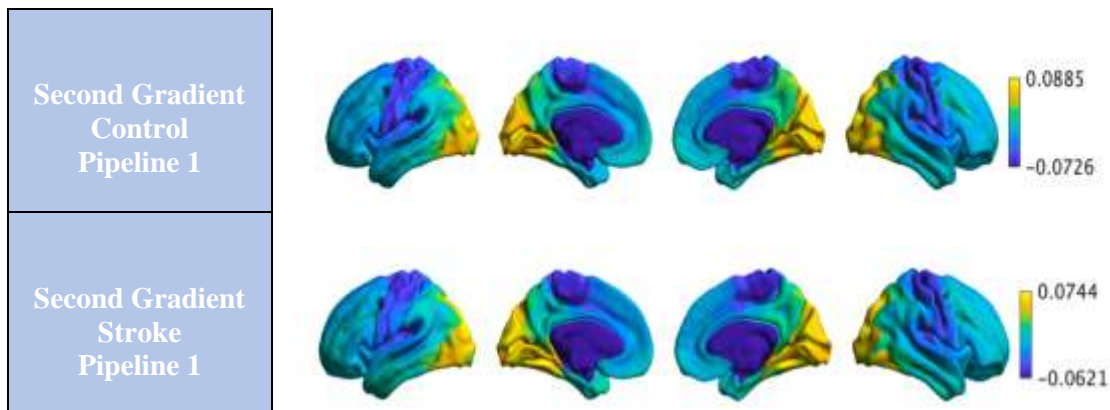


Figure 23 Gradient Mapping Control Health Group vs Stroke

The analysis conducted on pipeline 1, employing the 36-P method, did not yield any statistically significant variations among the various locations investigated. However, a closer examination of Figure 22, focusing on gradient 1, revealed more prominent differences across the different locations compared to the observations made for gradient 2 in Figure 23. These observed distinctions suggest that gradient 1 exerted a more substantial influence on the observed variability in each region. In contrast, gradient 2 demonstrated no discernible variations between regions, implying its minimal contribution to the overall observed variability.

In the graphs shown in Figure 24, we can corroborate this reasoning obtained with the comparison of both gradients in both populations.

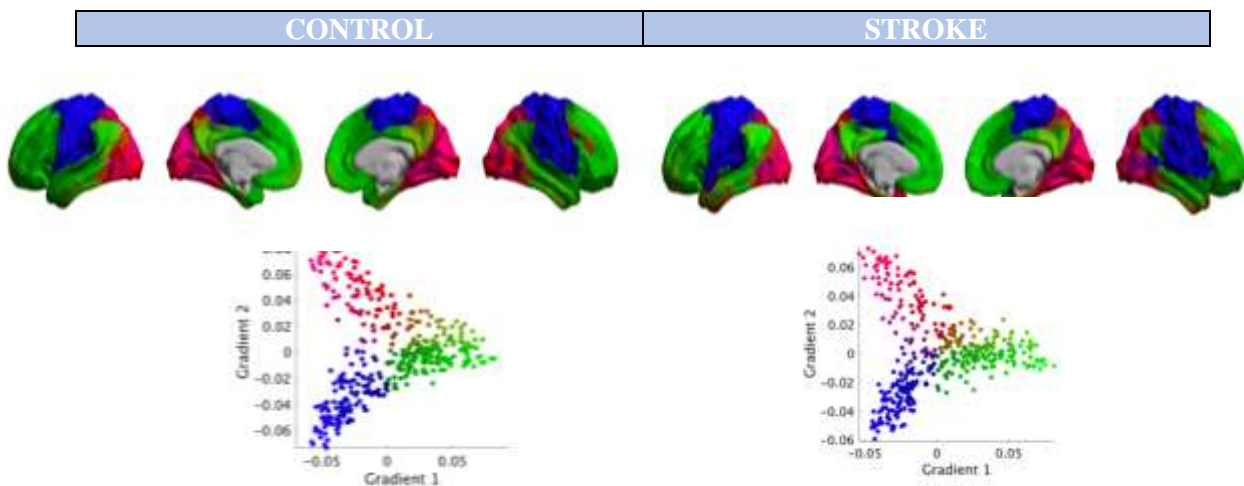


Figure 24 Gradient Mapping Control Health Group vs Stroke

Upon comparing the two graphs, it becomes evident that in stroke patients, there is a noticeable increase in the concentration of clusters within the gradient range encompassing green to blue colors. This observation suggests that sensory regions demonstrate a greater level of resilience or robustness in response to stroke. Conversely, when examining the red-colored areas corresponding to the occipital lobe, it is apparent that the control group exhibits a higher cluster concentration in these regions compared to the stroke group.

5.2.3 Data from Control Health Group vs Stroke in Pipeline 2

For pipeline 2, we will follow the same procedure as for pipeline 1, visualizing both gradients in both subject populations.

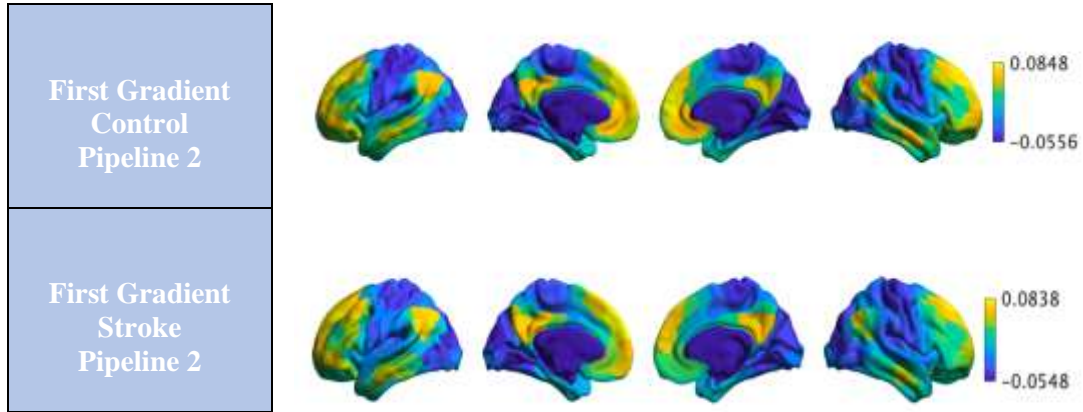


Figure 25 Gradient Mapping Control Health Group vs Stroke

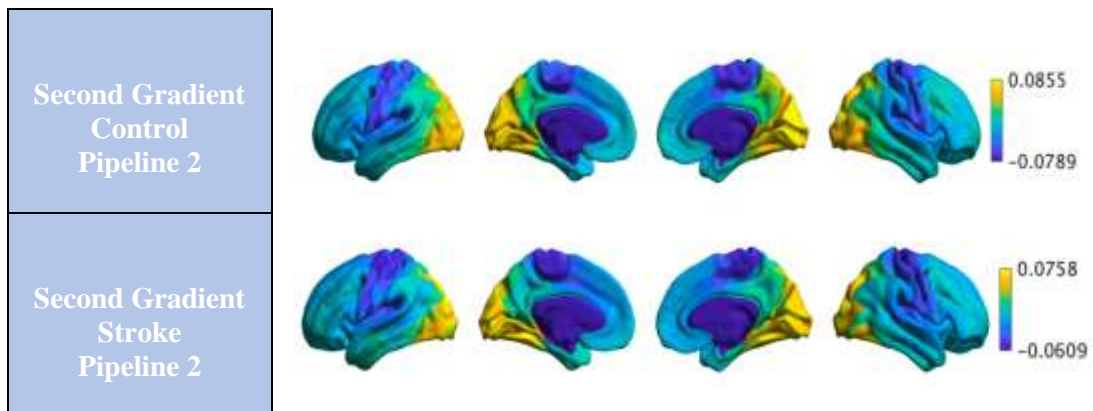


Figure 26 Gradient Mapping Control Health Group vs Stroke

For pipeline 2 (Figures 25 and 26), there is more variation between the two subject groups (AcompCor strategy). It could be caused by differences in brain functions and the impact of the disease between the two subject populations. There is greater variability in persons with cerebral stroke than in healthy subjects. Yellow and orange areas suggest more diversity in features in these specific brain regions in stroke patients. This could be due to areas directly impacted by the stroke or alternative areas reflecting adaptive alterations.

Similar to the aforementioned case, we can draw a parallel observation here. In this visual comparison, the second gradient does not yield substantial information as there are no apparent significant differences observed between the two groups. Within gradient 1, visual examination of the images reveals noticeable variations among the regions of interest. This highlights that, in this pipeline, a greater degree of differentiation is observed compared to the initial pipeline under study.

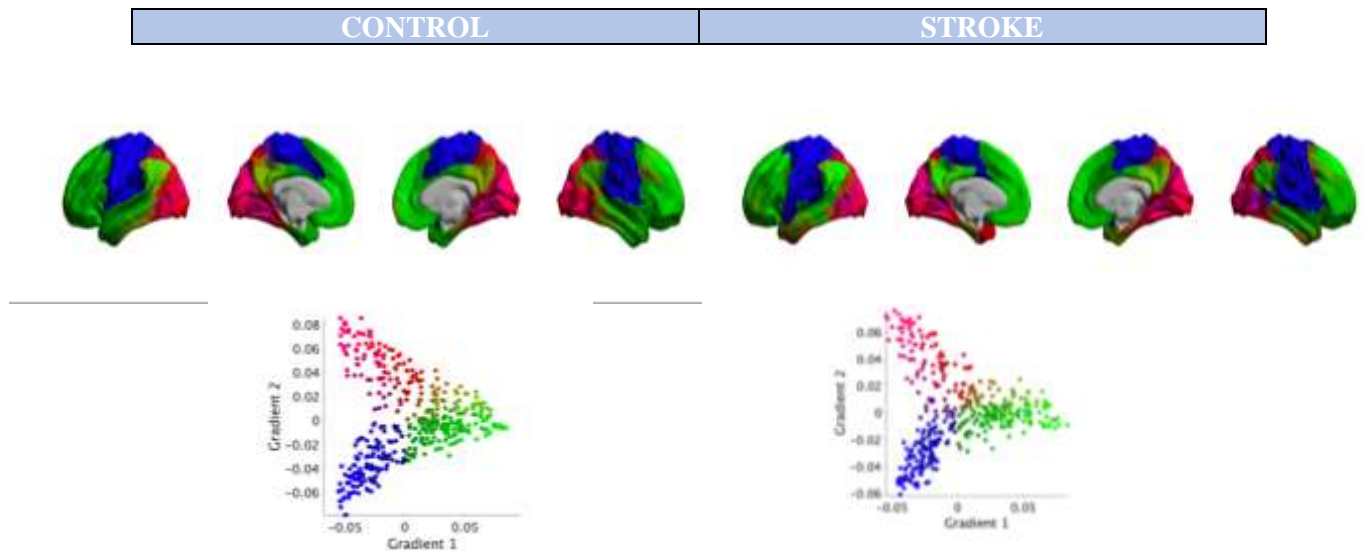


Figure 27 Gradient Mapping Control Health Group vs Stroke

The conclusions derived from the comparison of both gradients in both populations are further supported by the graphs presented in Figure 27. These graphs reaffirm our previous findings, clearly indicating that the sensory zone in stroke patients exhibits a significantly higher gradient embedding compared to control patients.

CHAPTER 6. RESULTS

6.1 Descriptive Statistics across pipelines

In the field of gradient analysis, descriptive statistics play a crucial role in understanding the organization and properties of brain gradients. When performing gradient analysis across different pipelines, it is essential to examine the descriptive statistics to gain insights into the variations and consistencies across the obtained results.

In this subsection on descriptive statistics across pipelines, we will present and analyze the key descriptive measures obtained from the gradient analysis using different pipelines. By examining these statistics, we aim to provide a comprehensive understanding of the similarities and differences between the obtained gradient profiles across pipelines.

Overall, the descriptive statistics across pipelines subsection serves as an important step in gradient analysis, allowing for a comprehensive characterization of the data and facilitating the interpretation and comparison of gradient results obtained using different preprocessing or analysis approaches.

For this analysis we will use three parameters: the variance, the standard deviation, and the mean of the gradients between pipelines and between patients. For this purpose, we have obtained the mean of these three parameters in each of the situations presented in the study. The following is the information we can extract from these parameters. [62] [47]

- **Variability of gradients:** Calculating variance enables for an assessment of gradient variability among pipelines and patients in the research. Higher variance suggests more gradient dispersion, which could mean significant variations across conditions or patient groups. This may aid in identifying patterns or defining aspects in the gradients that are relevant to the study objectives.
- **Gradient stability:** The standard deviation quantifies the spread of gradient values around the mean. A low standard deviation shows that the gradients are consistent and stable across pipelines and patients, implying that the results are more reliable and reproducible. A high standard deviation, on the other hand, may suggest increased variability and lack of stability in the gradients, which might require additional attention or consideration in the analysis.
- **Central tendency of the gradients:** Calculating the mean of the gradients offers a measure of the values' central tendency. The mean can be used to identify the average value of the gradients in each pipeline or group of patients, which can be helpful when comparing differences between conditions or groups of patients of interest in the study.

These measurements can assist in identifying major differences, assessing the dependability of the results, and providing a framework for good data analysis and interpretation. Table 4 below is a compilation of the parameter values described.

		Variance	Standar Deviation	Mean Gradients
PIPELINE 1	CONTROL	0.0021	0.0253	1.922e-05
	STROKE	0.0022	0.0272	1.110e-05
PIPELINE 2	CONTROL	0.0022	0.0259	2.125e-05
	STROKE	0.0024	0.0281	1.508e-05

Table 4 Descriptive stats across pipelines.

In both pipelines, the variance values are a bit higher in the stroke subjects comparison than in the control group, according to the data in Table 4. In both the control group and the stroke subjects, the variance in Pipeline 2 is slightly higher than in Pipeline 1.

In both pipelines, the standard deviation values are comparable between the control group and the stroke group. In general, the standard deviation values are low, indicating that the gradients are more dispersed around the mean.

Finally, in both pipelines and groups, the average of the gradients revealed very low values. In both pipelines, the average gradient values in the control and stroke groups are lower than in the control group.

Overall, these parameters show that the gradients in the comparison between control and stroke have slightly higher variance and lower average values than the control group. The differences, however, appear to be small. Furthermore, in both groups, Pipeline 2 has slightly greater variance values than Pipeline 1.

We can infer the following relevant conclusions based on the gradient analysis parameters:

- ⇒ In both processes, the variance is slightly higher when comparing the stroke patients. This shows that the gradients may be more variable in areas affected by stroke.
- ⇒ In both pipelines, the standard deviation values, which represent the dispersion of gradients around the mean, are similar between the control group and the stroke subjects. This implies that the gradient dispersion does not differ considerably between the two groups.
- ⇒ In both groups, Pipeline 2 has slightly greater variance values than Pipeline 1. This implies that results may differ based on the processing pipeline employed, and these variances must be considered when interpreting the results.

6.2 Group-Level Analysis

Group-level analysis allows the identification of regions in the brain that show significant differences or associations across different pipelines. By conducting a group-level analysis on the gradient level, we can uncover consistent patterns and identify regions of interest that exhibit significant results across pipelines.

The primary objective of the group-level analysis is to examine the statistical significance of the gradient differences or associations between groups (e.g., stroke group vs. healthy control group). This analysis provides insights into the brain regions that play a significant role in the observed variations or relationships in the gradients.

In this section, we will present and discuss the regions that demonstrate significant results across pipelines in the group-level analysis. By reporting these findings, we aim to highlight the robust and consistent regions that contribute to the observed gradient differences between stroke patients and health ones.

In this study, we will utilize a statistical test known as the t-test to perform group analysis for each region. The t-test is a technique that lets us compare the means of two groups and see if there are any significant differences [60]. In this situation, we will examine the gradients derived between the groups of interest. We will be able to determine whether there are statistically significant differences in the mean values of the gradients between the groups by using the t-test on the gradients.

This will help us to detect patterns or distinctive characteristics in the gradients that may relate to the specific conditions or groups under discussion in our investigation. After doing a group analysis with the t-test and receiving findings indicating significant variations in gradients between groups, we will proceed to show these gradients as images to study how they vary. We will be able to observe and spatially study how gradients are distributed and change in different locations or structures of interest by mapping them in images. For this purpose, the code shown in the figure below has been used, where after calculating the t-test and storing it in the variable '*t_stats*', this data is used to obtain the gradient plot.

```
% Group level analysis

for i=1:roi_num
    disp(i)
    [h,p,ci,stats] = ttest(squeeze(CON(i,2,:)),squeeze(PAT(i,2,:)));
    p_values(i)=p;

    t_stats(i)=stats.tstat;
end

plot_hemispheres((t_stats.*(p_values<0.05))', {surf_lh,surf_rh}, ...
    'parcellation',labeling.schaefer_400);
```

Figure 28 Code for group level analysis

As shown in Figure 28, we showed the results of t-tests done with a p-value less than 0.05. The p-value expresses the probability of obtaining the reported results if the null hypothesis is true. The stronger the statistical significance of the observed difference, the lower the p-value. A p-value of 0.05 or less is normally regarded statistically significant, thus we chose only values with p-values less than 0.05 to find the most significant regions. [61]

6.2.1 Analysis across pipelines

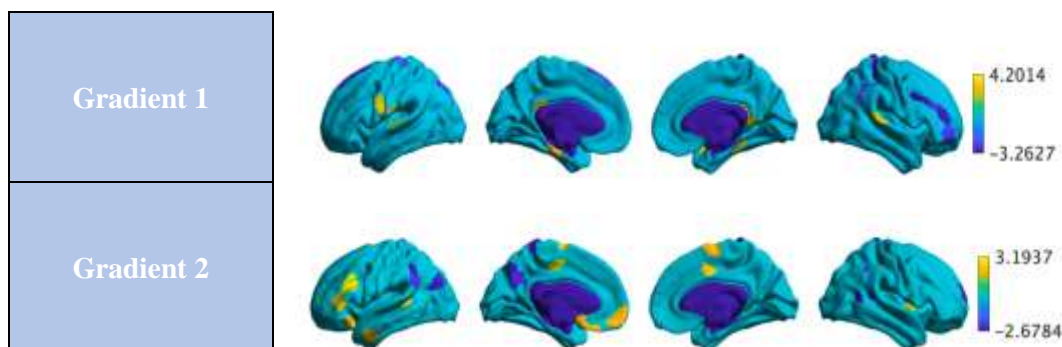


Figure 29 Gradient Mapping for group level analysis

The t-test was performed for control patients in both pipelines. The first pipeline (36-P Strategy) uses mainly motion parameters. Specifically white matter, cerebrospinal fluid (CSF) and global signal (GS) and their temporal derivatives. For the second pipeline, five of the main components of the white matter and CSF and their temporal and motion derivatives are used.

Analyzing the highlighted regions in each pipeline by utilizing the obtained p-values from the t-test, with a significance level of 0.05, we determined the number of regions of interest detected in each pipeline.

Regarding gradient 1, a total of 29 regions were identified with p-values below the predetermined threshold. Among these regions, the parahippocampal gyri, precuneus, and auditory A1 area exhibited elevated gradient embedding scores in support of the 36P strategy. This observation was discussed in the preceding section and validated through visual analysis of the acquired gradient maps.

Furthermore, in the case of the aCompCor strategy, we found a total of 26 regions of interest. These regions demonstrated significant involvement based on their p-values, indicating a preference for the aCompCor approach. In contrast, frontal regions and the temporoparietal junctions displayed diminished gradient embedding scores in favor of the aCompCor strategy.

In summary, the spatial distribution of gradients in both pipelines shows a high degree of similarity. However, there are noticeable differences in certain regions regarding the embedding of the gradients, but they don't affect to the gradients distribution.

6.2.2 Control Health Group vs Stroke analysis

Pipeline 1

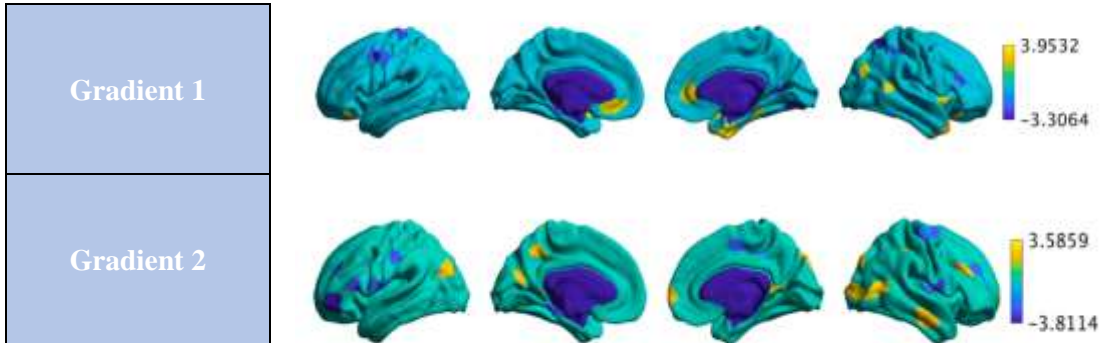


Figure 30 Gradient Mapping for group level analysis

Pipeline 2

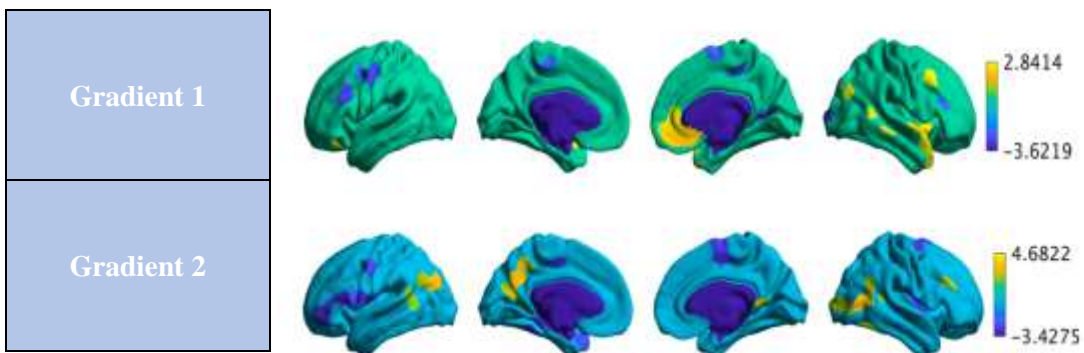


Figure 31 Gradient Mapping for group level analysis

Within this subcategory, t-tests were conducted for both pipelines to compare the control group with subjects affected by stroke. The obtained data underwent plotting and visualization, resulting in Figures 30 and 31. These figures display the images generated from the t-test analysis, utilizing a significance threshold of $p < 0.05$ to identify and emphasize the most statistically significant observations.

In the context of our study, we have observed that the connectivity embedding space is more distinct for stroke patients when using the aCompCor pipeline compared to other pipelines. Specifically, when examining Gradient 1, it is evident that the aCompCor pipeline exhibits the same regions as other pipelines, but with additional support from neighboring regions. This amplifies the differences observed in the connectivity patterns between stroke patients and controls. Furthermore, when analyzing Gradient 2, we found that it appears to have similar patterns across different pipelines.

In pipeline 1 (stroke vs control), we identified 31 regions with a p-value less than 0.05, indicating statistically significant differences between stroke patients and controls. On the other hand, in pipeline 2 (stroke vs control), we found 34 regions with a p-value less than 0.05.

These findings highlight the potential utility of the aCompCor pipeline in capturing and emphasizing unique connectivity characteristics in stroke patients compared to other pipelines.

When comparing the most significant regions (p-value less than 0.05), the issue of multiple comparisons must be considered. When comparing different locations, the probability of false positives increases due to repeated statistical tests. When multiple statistical tests are run at the same time, the probability of finding at least one significant region by chance increases [64]. We are running 400 areas to detect this issue in our data, so our results can be affected by this problem.

There are methods for correcting for multiple comparisons that help control the overall risk of false positives. Some commonly used methods include Bonferroni adjustment, the Benjamini-Hochberg procedure, or controlling the false discovery rate (FDR). [65]

After considering this issue and comparing both pipelines, it has been found that the Acompcor strategy identifies two regions that surpass the new p-value threshold without being affected by multiple comparisons.

This means that, when using the Acompcor strategy, two regions have been identified as statistically significant even after adjusting the significance threshold. These results are considered more reliable and less likely to be a random or spurious effect. Therefore, these two regions can be considered truly relevant or distinctive in the context of the analysis conducted.

Indeed, as compared to the 36p technique, the aCompCor strategy can improve the detection of group-level differences. As a result, it is a favored alternative for clinically relevant investigations aimed at more effectively identifying and analyzing such differences. The potential impact and relevance of group-level differences can be properly captured and analyzed by using aCompCor.

CHAPTER 7. CONCLUSIONS AND FUTURES LINES

6.1 Objectives achieved

The study successfully achieved the following objectives:

- **Image Preprocessing:** The images of both healthy and stroke patients were successfully preprocessed. This stage ensured that the raw data was sufficiently processed for subsequent analysis by eliminating artifacts and normalizing the images.
- **Use of several Post-processing Pipelines:** XCP-D was used to create several post-processing pipelines, which helped in the investigation of data variability and improved the reproducibility of group analysis. The study got an understanding of the impact of different preprocessing procedures on the outcomes by applying multiple pipelines.
- **Functional Gradient Analysis:** The study successfully assessed and compared differences in functional gradients between patients as well as between the healthy and stroke patient groups.

By obtaining these goals, the study added to our understanding of functional brain connectivity in healthy people and those with stroke disease. The findings improved our understanding of the effects of stroke on brain networks and gave significant information that could potentially lead the development of more effective stroke treatment options.

6.2 Conclusions

In conclusion, pipeline 1 (Strategy 36-P) demonstrated greater gradient embedding in the precuneus, auditory region A1, and parahippocampal gyrus. Pipeline 2 (aCompCor approach) on the other hand, exhibited larger values in frontal areas and temporoparietal connections.

Visual and statistical comparisons verified these findings, revealing that the distribution of gradients embedded in space was comparable for both pipelines. Pipeline 2, on the other hand, had a higher concentration of clusters in the frontal regions, which corresponded to the highlighted areas.

Except for gradient 1, which exhibited more pronounced disparities among regions than gradient 2, the examination of pipeline 1 revealed no statistically significant differences among the analyzed locations. In pipeline 2, there was greater variability between the stroke and control groups, indicating variations in brain functions and disease consequences.

In addition, our research highlighted the need of examining several comparisons when finding relevant locations. To reduce the possibility of false positives, correction methods such as Bonferroni adjustment or limiting the false discovery rate (FDR) should be used.

Based on our analysis, we suggest the aCompCor method for data cleaning over the 36P strategy. This suggestion is supported by the findings indicating significant involvement of specific brain regions and diminished gradient embedding scores in favor of aCompCor. Choosing aCompCor as the preferred method can enhance the reliability and accuracy of data in neuroscience research.

In conclusion, our work successfully analyzed functional gradient differences between healthy and stroke patients utilizing different preprocessing pipelines. These findings have significance for understanding the changed brain connections in stroke patients, as well as emphasizing the need of establishing proper processes for assessing functional gradients in clinical research.

6.3 Limitations and future lines of investigation

We can mention the following limitations:

- Consider expanding the sample size to improve the statistical power and consistency of results.
- Including a greater number of individuals and conditions to improve result generalization.
- Using several pipelines to examine the influence of different preprocessing methods on the findings to ensure robustness.
- Validating and repeating the findings in independent datasets to verify the reliability and repeatability of the results.

Some major ideas for future lines of investigation can include:

- Examining the connection between functional gradients and clinical outcomes or symptom severity to better understand the clinical consequences of altered brain connectivity.
- Multimodal approaches: combining data from functional magnetic resonance imaging (fMRI) with data from other modalities such as structural MRI or behavioral assessments to acquire a more comprehensive knowledge of brain functioning organization and its relation to other factors.
- Investigating modern analytical tools, such as machine learning algorithms or network-based studies, to get a more complex understanding of functional gradients and their connections with clinical factors.

Addressing these limitations and following these future lines of research would lead to a better knowledge of functional gradients, their significance in various diseases, and possible clinical applications.

LIST OF FIGURES

Figure 1 Grey and White matter in brain [11].....	9
Figure 2 To obtain the images, a series of two-dimensional slices are captured to cover either the entire brain or specific regions of interest. Each slice has a set thickness and is composed of a matrix of volumetric pixels or voxels [19].	12
Figure 3 Inflammatory Penumbra in Ischemic Stroke [28].....	16
Figure 4 MRI Stroke Image [27].....	17
Figure 5 Schedule of BIDS [42].....	20
Figure 6 Example of preprocessed data [42].....	21
Figure 7 Workflow of fMRIPrep	22
Figure 8 Workflow of XCP-D [46]	25
Figure 9 Steps followed to obtain the network analysis [49]	27
Figure 10 Typical Gradient Identification Workflow [48].....	28
Figure 11 Strokes affects the temporal characteristics of BOLD [53]	33
Figure 12 Command line for XCP-D installation [46].....	36
Figure 13 Command lines for XCP-D execution.	36
Figure 14 Code for gradients calculation.	40
Figure 15 Code for gradients calculation.	41
Figure 16 Code for gradients calculation.	41
Figure 17 Code for gradients calculation.	43
Figure 18 Code for gradients calculation.	43
Figure 19 Gradient Mapping across pipelines.....	44
Figure 20 Gradient Mapping across pipelines.....	44
Figure 21 Gradient Mapping across pipelines.....	45
Figure 22 Gradient Mapping Control Health Group vs Stroke	46
Figure 23 Gradient Mapping Control Health Group vs Stroke	46
Figure 24 Gradient Mapping Control Health Group vs Stroke	47
Figure 25 Gradient Mapping Control Health Group vs Stroke	48
Figure 26 Gradient Mapping Control Health Group vs Stroke	48
Figure 27 Gradient Mapping Control Health Group vs Stroke	49
Figure 28 Code for group level analysis	52
Figure 29 Gradient Mapping for group level analysis	53
Figure 30 Gradient Mapping for group level analysis	54
Figure 31 Gradient Mapping for group level analysis	54

LIST OF MATHEMATICAL EXPRESSION

Equation 1 Mathematical expression for the Fisher transformation [50].....	28
Equation 2 Mathematical expression for the method Gaussian Kernel	29
Equation 3 Mathematical expression for the method Cosine Singularity	29
Equation 4 Mathematical expression for the method Normalized angle Similarity.....	29
Equation 5 Mathematical expression for the method Pearson Correlation	29
Equation 6 Mathematical expression for the method Spearman Rank Order Correlation	29
Equation 7 Mathematical expression for the Dimensional Reduction in PCA	30
Equation 8 Mathematical expression for the Laplacian calculation.....	30
Equation 9 Mathematical expression for LE method	30
Equation 10 Mathematical expression for LE method	31
Equation 11 Mathematical expression for LE method	31
Equation 12 Mathematical expression for DM method	31
Equation 13 Mathematical expression for the diffusion operator	31
Equation 14 Mathematical expression for the diffusion operator	31

LIST OF TABLES

Table 1 Table with some of the main MRI techniques in neuroimaging.	10
Table 2 Scheme of preprocessing techniques used by fMRIPrep [3]	23
<i>Table 3 Dataset Description</i>	39
Table 4 Descriptive stats across pipelines.	51

BIBLIOGRAFY

- [1] Molina, A. G. (2015, February 20). La pionera aportación de Justo Gonzalo al estudio de la organización funcional del cerebro. Barcelona; Sociedad Española de Neurología.
- [2] Markello, R.D., Hansen, J.Y., Liu, ZQ. *et al.* neuromaps: structural and functional interpretation of brain maps. *Nat Methods* 19, 1472–1479 (2022). <https://doi.org/10.1038/s41592-022-01625-w>
- [3] Esteban, O., Markiewicz, C.J., Blair, R.W. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* 16, 111–116 (2019). <https://doi.org/10.1038/s41592-018-0235-4>
- [4] Bernhardt, B. C., Smallwood, J., Keilholz, S., & Margulies, D. S. (2022). Gradients in brain organization. *NeuroImage*, 251, 118987. <https://doi.org/10.1016/j.neuroimage.2022.118987>
- [5] Mikail Rubinov, Olaf Sporns, Complex network measures of brain connectivity: Uses and interpretations, *NeuroImage*, Volume 52, Issue 3, 2010, Pages 1059-1069, ISSN 1053-8119. <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- [6] Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., Bezgin, G., Eickhoff, S. B., Castellanos, F. X., Petrides, M., Jefferies, E., & Smallwood, J. (2016, October 18). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44), 12574–12579. <https://doi.org/10.1073/pnas.1608282113>
- [7] Drake, Richard L. *et al.* Gray : anatomía para estudiantes / Richard L. Drake, Wayne Vogl, Adam W. M. Mitchell. 3a ed. Madrid [etc: Elsevier, 2014. Print.
- [8] Halperin, E. C. (2009, February). The Pornographic Anatomy Book? The Curious Tale of The Anatomical Basis of Medical Practice. *Academic Medicine*, 84(2), 278–283. <https://doi.org/10.1097/acm.0b013e31819391e2>
- [9] Ackerman S. *Discovering the Brain*. Washington (DC): National Academies Press (US); 1992. 2, Major Structures and Functions of the Brain. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK234157/>
- [10] Team, S. C. (2020, June 7). *Grey Matter vs White Matter in the Brain*. Grey Matter Vs White Matter in the Brain. <https://www.spinalcord.com/blog/gray-matter-vs-white-matter-in-the-brain>
- [11] [Figure, Gray and White matter. Image courtesy S Bhimji MD] - *StatPearls - NCBI Bookshelf*. (n.d.). [Figure, Gray and White Matter. Image Courtesy S Bhimji MD] – *StatPearls - NCBI Bookshelf*. <https://www.ncbi.nlm.nih.gov/books/NBK553239/figure/article-36416.image.f1/>
- [12] Imagen por resonancia magnética. (2022, 23 de diciembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 20:14, diciembre 23, 2022 <https://es.wikipedia.org>

- [13] *Imagen por resonancia magnética funcional (fMRI)*. (2020, October 1). Imagen Por Resonancia Magnética Funcional (fMRI) | Memorial Sloan Kettering Cancer Center. <https://www.mskcc.org/es/cancer-care/patient-education/functional-magnetic-resonance-imaging-fmri>
- [14] *Diffusion MRI - an overview | ScienceDirect Topics*. (n.d.). Diffusion MRI - an Overview | ScienceDirect Topics. <https://doi.org/10.1016/B978-0-12-803675-4.00025-7>
- [15] Gorny, K. (2016, June). WE-DE-206-01: MRI Signal in Biological Tissues - Proton, Spin, T1, T2, T2*. *Medical Physics*, 43(6Part40), 3816–3816. <https://doi.org/10.1118/1.4957853>
- [16] Miranda Bautista, J., Garrido Morro, I., Fernández García, P., & Herrera Herrera, I. (2022, March). Hiperintensidad en secuencia FLAIR del LCR en el espacio subaracnoideo: diagnósticos diferenciales. *Radiología*. <https://doi.org/10.1016/j.rx.2022.01.014>
- [17] Li, Y., Wang, P., Zhang, J., Li, J., Chen, L., & Qiu, X. (2022, March 12). Multiparametric Framework Magnetic Resonance Imaging Assessment of Subtypes of Intracranial Germ Cell Tumors Using Susceptibility Weighted Imaging, Diffusion-Weighted Imaging, and Dynamic Susceptibility-Contrast Perfusion-Weighted Imaging Combined with Conventional Magnetic Resonance Imaging. *Journal of Magnetic Resonance Imaging*, 56(4), 1232–1242. <https://doi.org/10.1002/jmri.28132>
- [18] Stanley, J. A., & Raz, N. (2018). Functional Magnetic Resonance Spectroscopy: The "New" MRS for Cognitive Neuroscience and Psychiatry Research. *Frontiers in psychiatry*, 9, 76. <https://doi.org/10.3389/fpsy.2018.00076>
- [19] Armony, Jorge L., Trejo-Martínez, David, & Hernández, Dailett. (2012). Resonancia Magnética Funcional (RMf): principios y aplicaciones en Neuropsicología y Neurociencias Cognitivas. *Neuropsicología Latinoamericana*, 4(2), 36-50. <https://dx.doi.org/10.5579/rnl.2012.010>
- [20] De la Iglesia Vayá, M. D. L. D. (2011). *Técnicas de conectividad cerebral y transferencia de información aplicado al estudio de la esquizofrenia* (Doctoral dissertation, Universitat Politècnica de València).
- [21] van den Heuvel, M. P., & Hulshoff Pol, H. E. (n.d.). *Exploración de la red cerebral: una revisión de la conectividad funcional en la RMf en estado de reposo | Psiquiatría Biológica*. Exploración De La Red Cerebral: Una Revisión De La Conectividad Funcional En La RMf En Estado De Reposo | Psiquiatría Biológica. <https://doi.org/10.1016/j.psiq.2011.05.001>
- [22] Lang, E., Tomé, A., Keck, I., Górriz-Sáez, J., & Puntonet, C. (2012). Brain Connectivity Analysis: A Short Survey. *Computational Intelligence And Neuroscience*, 2012, 1-21. <http://dx.doi.org/10.1155/2012/412512>
- [23] *Accidente cerebrovascular - Wikipedia, la enciclopedia libre*. (n.d.). Accidente Cerebrovascular - Wikipedia, La Enciclopedia Libre. https://es.wikipedia.org/wiki/Accidente_cerebrovascular

- [24] *Effects of Stroke*. (n.d.). Effects of Stroke | American Stroke Association. <https://www.stroke.org/en/about-stroke/effects-of-stroke>
- [25] *Ictus Cerebral: Tratamientos, síntomas, causas e información*. (2009, March 26). CuidatePlus. <https://cuidateplus.marca.com/enfermedades/neurologicas/ictus.html>
- [26] T. (2018, January 24). *La importancia del tiempo en el ictus*. Top Doctors. <https://www.topdoctors.es/articulos-medicos/ictus-la-importancia-del-tiempo>
- [27] Ciardullo, D. (2020, February 5). *Imágenes del Ictus Isquémico en TC y RM - Radiología 2.0*. Radiología 2.0. <https://radiologia2cero.com/imagenes-del-ictus-isquemico/>
- [28] Gauberti M, De Lizarrondo SM, Vivien D. The “inflammatory penumbra” in ischemic stroke: From clinical data to experimental evidence. *European Stroke Journal*. 2016;1(1):20-27. doi:10.1177/2396987316630249
- [29] Song, X. W., Dong, Z. Y., Long, X. Y., Li, S. F., Zuo, X. N., Zhu, C. Z., He, Y., Yan, C. G., & Zang, Y. F. (2011). REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PloS one*, 6(9), e25031. <https://doi.org/10.1371/journal.pone.0025031>
- [30] *Spatial Normalization - an overview | ScienceDirect Topics*. (n.d.). Spatial Normalization - an Overview | ScienceDirect Topics. <https://doi.org/10.1016/B978-012077790-7/50021-7>
- [31] *Chapter 3: Motion Correction ‐ Andy’s Brain Book 1.0 documentation*. (n.d.). Chapter 3: Motion Correction ‐ Andy’s Brain Book 1.0 Documentation. https://andysbrainbook.readthedocs.io/en/latest/fMRI_Short_Course/Preprocessing/Motion_Correction.html
- [32] Wikipedia contributors. (2022, April 1). Normalization (image processing). In *Wikipedia, The Free Encyclopedia*. Retrieved 12:00, June 3, 2023, from [https://en.wikipedia.org/w/index.php?title=Normalization_\(image_processing\)&oldid=1080487172](https://en.wikipedia.org/w/index.php?title=Normalization_(image_processing)&oldid=1080487172).
- [33] *Chapter 5: Smoothing ‐ Andy’s Brain Book 1.0 documentation*. (n.d.). Chapter 5: Smoothing ‐ Andy’s Brain Book 1.0 Documentation. https://andysbrainbook.readthedocs.io/en/latest/fMRI_Short_Course/Preprocessing/Smoothing.html
- [34] Bullock, M., Jackson, G. D., & Abbott, D. F. (2021, January 29). *Artifact Reduction in Simultaneous EEG-fMRI: A Systematic Review of Methods and Contemporary Usage*. Frontiers. <https://doi.org/10.3389/fneur.2021.622719>
- [35] *Feature Extraction*. (n.d.). Feature Extraction - MATLAB & Simulink. <https://www.mathworks.com/discovery/feature-extraction.html>

- [36] *SPM - Statistical Parametric Mapping*. (n.d.). SPM - Statistical Parametric Mapping. <https://www.fil.ion.ucl.ac.uk/spm/>
- [37] *FSL - FslWiki*. (n.d.). FSL - FslWiki. <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>
- [38] *afni.nimh.nih.gov*. (n.d.). afni.nimh.nih.gov. <https://afni.nimh.nih.gov>
- [39] *Brain Innovation - Home*. (2023, November 2). Brain Innovation - Home. <https://www.brainvoyager.com/index.html>
- [40] *FreeSurfer*. (n.d.). FreeSurfer. <https://surfer.nmr.mgh.harvard.edu>
- [41] *Index*. (2023, June 2). Brain Imaging Data Structure. <https://bids.neuroimaging.io/index>
- [42] *Introduction - Brain Imaging Data Structure v1.8.0*. (n.d.). Introduction - Brain Imaging Data Structure v1.8.0. <https://bids-specification.readthedocs.io/en/stable/01-introduction.html>
- [43] Gorgolewski, K., Auer, T., Calhoun, V. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* 3, 160044 (2016). <https://doi.org/10.1038/sdata.2016.44>
- [44] Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, 154,174–187. <https://doi.org/10.1016/j.neuroimage.2017.03.020>
- [44] Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (n.d.). *How to control confounding effects by statistical analysis*. PubMed Central (PMC). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4017459/>
- [46] *XCP-D: A Robust Postprocessing Pipeline of fMRI data — xcp_d 0.4.0rc2+25.g75e2c9b documentation*. (n.d.). XCP-D : A Robust Postprocessing Pipeline of fMRI Data &Mdash; Xcp_D 0.4.0rc2+25.g75e2c9b Documentation. <https://xcp-d.readthedocs.io/en/latest/>
- [47] P. (2023, January 6). *GitHub - PennLINC/xcp_d: Post-processing of fMRIPrep, NiBabies, and HCP outputs*. GitHub. https://github.com/PennLINC/xcp_d
- [48] Vos de Wael, R., Benkarim, O., Paquola, C. *et al.* BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun Biol* 3, 103 (2020). <https://doi.org/10.1038/s42003-020-0794-7>

- [49] Mohammadi-Nejad, Ali-Reza & Mahmoudzadeh, Mahdi & Hassanpour, Mahlegha & Wallois, Fabrice & Muzik, Otto & Papadelis, Christos & Hansen, Anne & Soltanian-Zadeh, Hamid & Gelovani, Juri & Nasiriavanaki, Mohammadreza. (2018). Neonatal brain resting-state functional connectivity imaging modalities. *Photoacoustics*. 10. 10.1016/j.pacs.2018.01.003
- [50] Wikipedia contributors. (2023, January 29). Fisher transformation. In *Wikipedia, The Free Encyclopedia*. Retrieved 12:39, June 3, 2023, from https://en.wikipedia.org/w/index.php?title=Fisher_transformation&oldid=1136349343
- [51] Şeyma Bayrak, Ahmed A. Khalil, Kersten Villringer, Jochen B. Fiebach, Arno Villringer, Daniel S. Margulies, Smadar Ovadia-Caro, The impact of ischemic stroke on connectivity gradients, *NeuroImage: Clinical*, Volume 24, 2019, 101947, ISSN 2213-1582, <https://doi.org/10.1016/j.nicl.2019.101947>.
- [52] Ward, N. S., Brown, M. M., Thompson, A. J., & Frackowiak, R. S. J. (2003, June). Neural correlates of outcome after stroke: a cross-sectional fMRI study. *Brain*, 126(6), 1430–1448. <https://doi.org/10.1093/brain/awg145>
- [53] *Griffis et al., 2019. Cell Reports*
- [54] *What is Docker?* (2018, January 9). What Is Docker? <https://www.redhat.com/en/topics/containers/what-is-docker>
- [55] *Install Docker Engine.* (2023, June 2). Docker Documentation. <https://docs.docker.com/engine/install/>
- [56] *Installation & mdash; fmriprep version documentation.* (n.d.). Installation &Mdash; Fmriprep Version Documentation. <https://fmriprep.org/en/stable/installation.html>
- [57] *Welcome to BrainSpace's documentation! & mdash; BrainSpace 0.1.4 documentation.* (n.d.). Welcome to BrainSpace's Documentation! &Mdash; BrainSpace 0.1.4 Documentation. <https://brainspace.readthedocs.io/en/latest/>
- [58] T. (n.d.). *gift/README.md at master · trendscenter/gift.* GitHub. <https://github.com/trendscenter/gift>
- [59] C. (n.d.). *Distortion and drop-out – Functional MRI: Basic principles.* Distortion and Drop-out – Functional MRI: Basic Principles. <https://pressbooks.umn.edu/fmribasicprinciples/chapter/distortion>
- [60] *Default mode network - Wikipedia.* (2023, February 1). Default Mode Network - Wikipedia. https://en.wikipedia.org/wiki/Default_mode_network.

[61] *t-Test, Chi-Square, ANOVA, Regression, Correlation*. . . (n.d.). Dispersion Parameter: Variance, Standard Deviation, Range. <https://datatab.net/tutorial/dispersion-parameter>.

[62] *What Does Effect Size Tell You?* (2022, November 3. Simply Psychology. <https://www.simplypsychology.org/effect-size.html>

[63] *One-sample or two-sample effect size computations - MATLAB meanEffectSize*. (n.d.). One-sample or Two-sample Effect Size Computations - MATLAB meanEffectSize. <https://www.mathworks.com/help/stats/meaneffectsize.html>

[64] *Multiple comparisons problem - Wikipedia*. (n.d.). Multiple Comparisons Problem - Wikipedia. https://en.wikipedia.org/wiki/Multiple_comparisons_problem

[65] *False discovery rate - Wikipedia*. (n.d.). False Discovery Rate - Wikipedia. https://en.wikipedia.org/wiki/False_discovery_rate

ANNEX

In this last part of the project, we present the complete code that we have used for the calculation of brain functional gradients. During the previous chapters the most important parts of the code implemented in MATLAB have been explained, but here we have the totality of the code:

```
% Set the necessary files for gradient calculation and visualization
```

```
[surf_lh, surf_rh] = load_conte69();
labeling = load_parcellation('schaefer',400);
```

```
% Calculate the reference gradient from the toolbox
```

```
conn_matrices = load_group_fc('schaefer',400);
roi_num=length(conn_matrices.schaefer_400);
m = atanh(conn_matrices.schaefer_400);
m(isinf(m))=1;
ref = GradientMaps('kernel','na','approach','dm','alignment','pa');
TR=128;
ref = ref.fit(m);
```

```
parcellation=niftiread('yeo400_resampled.nii.gz');
parcellation=reshape(parcellation,[],1);
```

```
% Do it for every subject
```

```
ls=dir("/Users/mariavila/Desktop/brainspace-0.1.10/pipeline1/subjectspipeline1");
name = {ls.name}.';
name(1:2)=[];
```

```
% Load the file one by one and take the mean of ROI
```

```
for i=1:length(name)
    disp(i)
    file=niftiread(string(name(i)));
    file=reshape(file,[],TR);

    for j=1:roi_num
        region_mean(j,:,i)=mean(file(parcellation==j,:));
    end
end
```

```
for i=1:length(name)
    disp(i)
    % Calculate the correlation matrix and Z-transform it
    corrmats(:,i) = atanh(corr(region_mean(:,i)));
```

```
end
```

```
% Remove the Inf values resulted from the normalization step
corrmats(isinf(corrmats))=1;
%signal dropout correction
corrmats(isnan(corrmats))=rand;
```

```

% Gradients calculation
for i=1:length(name)
    disp(i)
    onesub = GradientMaps('kernel','na','approach','dm','alignment','pa');
    onesub = onesub.fit((corrmat(:,i)), 'reference', ref.gradients{1});

    gradients_all(:,i)= (onesub.aligned{:});
end

% groups for the análisis

CON=gradients_all(:,:[1:1:20]);
PAT=gradients_all(:,:[21:1:40]);

% hemispheres plot
plot_hemispheres(mean(squeeze(PAT(:,2,:)))', {surf_lh,surf_rh}, ...
    'parcellation',labeling.schaefer_400);

% Group plot
gradient_in_euclidean([mean(squeeze(PAT(:,1,:)))' mean(squeeze(PAT(:,2,:)))']
, {surf_lh,surf_rh},labeling.schaefer_400);

%plot pipeline1vspipeline2
gradient_in_euclidean([pipeline1(:,1,:) pipeline2(:,1,:)] ,{surf_lh,surf_rh},labeling.schaefer_400);

% Individual subjects
gradient_in_euclidean([gradients_all(:,1,1) gradients_all(:,2,1)]
, {surf_lh,surf_rh},labeling.schaefer_400);

% Calculate the variance across pipelines
variance = var(gradients_all, [], 3);
mean_variancep2= mean(variance(:));

% Calculate the mean across pipelines
mean_gradients = mean(gradients_all, 3);
mean_mean_gradientsp2= mean(mean_gradients (:));

% Calculate the standard deviation across pipelines
std_gradients = std(gradients_all, [], 3);
mean_std_gradientsp2 = mean(std_gradients(:));

```

% Group level analysis across pipelines

```
for i=1:roi_num
    disp(i)
    [h,p,ci,stats] = ttest(squeeze(CON(i,2,:)),squeeze(PAT(i,2,:)));
    p_values(i)=p;

    t_stats(i)=stats.tstat;
end

plot_hemispheres((t_stats.*(p_values<0.05))', {surf_lh,surf_rh}, ...
    'parcellation',labeling.schaefer_400);
```