



UNIVERSITAT_{DE}
BARCELONA



Final Degree Project
Biomedical Engineering Degree

Flexible fMRI Processing Pipeline for Large-Scale Studies of Functional Connectivity in Preclinical Alzheimer's Disease

Barcelona, January 20th, 2025

Author: Marc Biosca Marron

Directors: Roser Sala and Raúl Tudela

Abstract

The human brain can be understood as a complex network of spatially distributed but functionally interconnected regions that constantly exchange information. Functional connectivity (FC), which measures how regions of the brain communicate and coordinate, can help identify early signs of diseases like Alzheimer's Disease (AD). This project focuses on implementing a flexible pipeline to process resting-state functional magnetic resonance imaging (rs-fMRI) data and study FC in the preclinical stages of AD.

A dataset of over 1,000 subjects was processed through a flexible pipeline, which consists of a series of steps like brain extraction, motion correction, and noise reduction designed to clean and prepare the data for subsequent neuroimaging analysis. Additionally, a robust Quality Control (QC) framework was integrated, combining visual inspection and quantitative metrics to ensure reliable results, with around 76% of subjects successfully passing the QC process. All procedures and scripts used in this project are available in a GitHub repository.

The results of this project reveal a significant relationship between FC and the Centiloid scale, a measure of the progression of AD. Additionally, age-related changes in brain connectivity were identified, showing both strengthened and weakened connections, indicating a redistribution of brain networks. Lastly, an exploratory framework was introduced to analyze other factors that may influence brain function, opening the door for it to be used in future neuroimaging studies. Future research should evaluate the potential of these imaging biomarkers to serve as early indicators of neurodegenerative diseases.

Key words: Functional Connectivity, Alzheimer's Disease, Resting-State Functional Magnetic Resonance Imaging, Resting-State Networks, Preprocessing Pipeline.

Resum

El cervell es pot entendre com una xarxa complexa de regions espacialment distribuïdes, però funcionalment interconnectades que comparteixen informació contínuament. La connectivitat funcional (FC), que mesura com les regions del cervell es comuniquen i coordinen, pot ajudar a identificar signes inicials de malalties com l'Alzheimer (AD). Aquest projecte té com a objectiu implementar un pipeline flexible per processar dades de ressonància magnètica funcional en repòs (rs-fMRI) i estudiar la FC en etapes preclíniques de l'AD.

Un conjunt de dades de més de 1.000 subjectes ha estat processat mitjançant un pipeline flexible, que consisteix en una sèrie de passos com l'extracció cerebral, la correcció de moviment i la reducció de soroll, dissenyats per preparar les dades per a posteriors anàlisis. A més, s'ha desenvolupat una aplicació de control de qualitat (QC), que combina la inspecció visual i mètriques quantitatives per garantir resultats fiables, amb 76% dels subjectes superant el QC. Tots els procediments desenvolupats en aquest projecte estan disponibles en un repositori de GitHub.

Els resultats demostren una relació significativa entre la FC i l'escala Centiloid, una mesura de la progressió de l'AD. També s'han identificat canvis en la connectivitat cerebral relacionats amb l'edat, mostrant potenciació i deteriorament de connexions, indicant una redistribució de les xarxes cerebrals. Finalment, s'ha introduït un marc exploratori per analitzar altres factors que poden influir en la funció cerebral, obrint la porta perquè es pugui utilitzar en futurs estudis de neuroimatge. Investigacions futures haurien d'avaluar el potencial d'aquests biomarcadors com a indicadors primerencs de malalties neurodegeneratives.

Paraules clau: Connectivitat Funcional, Malaltia d'Alzheimer, Ressonància Magnètica Funcional en Repòs, Xarxes en Repòs, Pipeline de Preprocessament.

Acknowledgments

I would like to thank Roser Sala for her guidance and support throughout this project. Her advice and feedback were essential to completing my work successfully, and I am very grateful for her mentorship. I would also like to thank Raúl Tudela for his generous help, collaboration and insights. His support made a big difference in carrying out this project.

Finally, I also want to thank my family and friends for their constant encouragement and patience, which have been a great source of strength during this project and my studies.

Glossary

A4 Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease study.

A β Amyloid Beta.

AD Alzheimer's Disease.

AI Artificial Intelligence.

ANTs Advanced Normalization Tools.

BET Brain Extraction Tool.

BIDS Brain Imaging Data Structure.

BIG Biomedical Imaging Group at the University of Barcelona.

BOLD Blood Oxygen Level Dependent.

CAGR Compound Annual Growth Rate.

CNN Convolutional Neural Network.

CSF Cerebrospinal Fluid.

CT Computed Tomography.

DMN Default Mode Network.

EPI Echo-Planar Imaging.

FC Functional Connectivity.

FD Framewise Displacement.

GLM General Linear Model.

GM Gray Matter.

HD-BET High Definition Brain Extraction Tool.

HbO Oxygenated Hemoglobin.

HbR Deoxygenated Hemoglobin.

IC Independent Component.

ICA Independent Component Analysis.

MRI Magnetic Resonance Imaging.

NFTs Neurofibrillary Tangles.

PET Positron Emission Tomography.

QC Quality Control.

qQC Quantitative Quality Control.

rs-fMRI Resting-State Functional Magnetic Resonance Imaging.

RSNs Resting-State Networks.

SUVR Standardized Uptake Value Ratio.

T1w T1-Weighted Imaging.

TE Echo Time.

TR Repetition Time.

vQC Visual Quality Control.

WBS Work Breakdown Structure.

WM White Matter.

List of Figures

Figure 1.	PET imaging showing non-uniform oxygen consumption across brain regions. . .	4
Figure 2.	Molecular representation of $A\beta$ structures and NFTs of AD compared to the normal state.	5
Figure 3.	T1w MRI shows gray matter loss, and Florbetapir PET highlights $A\beta$ accumulation in AD patients.	6
Figure 4.	Illustration of a BIDS structured dataset.	13
Figure 5.	Diagram illustrating sources of fMRI signal noise, including head motion, cardiac, and respiratory activity.	14
Figure 6.	General Linear Model for nuisance regression, separating BOLD timeseries into confounds to remove non-neuronal noise.	18
Figure 7.	ICA decomposition of preprocessed fMRI data into components with time courses and spatial maps.	20
Figure 8.	Ten pairs of networks from task-based BrainMap activation data (right) and rs-fMRI data (left), identified via ICA.	21
Figure 9.	Demonstration of the component selection problem in ICA, highlighting underfitting, optimal fitting, and overfitting.	21
Figure 10.	Dual regression schematic. Stage 1 extracts time courses, and Stage 2 generates spatial maps, with outputs highlighted in blue.	22
Figure 11.	Stage 2 of dual regression, showing the ICA-derived DMN and subject-specific spatial contribution patterns.	22
Figure 12.	Illustration of full vs. partial correlation, showing all versus direct connections. . .	23
Figure 13.	Network connectivity matrix from a single subject, transformed into a single line and stacked across subjects for group-level analysis.	24
Figure 14.	Overview of the AmyloidE dataset into BIDS format, structured by subject, session, and modality.	26
Figure 15.	Distribution of Centiloid values, PACC Cognitive scores, and Age across the three identified cluster groups.	26
Figure 16.	Diagram of preprocessing pipeline with the vQC and qQC checkpoints.	27

Figure 17. QC application screenshots showing anatomical and functional checkpoints, with review buttons.	29
Figure 18. Brain mask used to assess the accuracy of the brain extraction process, along with segmentation of brain regions for error detection.	29
Figure 19. Raw (a, c) and slice-timing corrected (b, d) fMRI images, showing unresolved acquisition artifacts in the right panels.	30
Figure 20. Proper (left) and improper (right) registration to the template, with errors caused by incomplete brain extraction.	30
Figure 21. Resting-State Networks extracted from the ICA analysis.	34
Figure 22. Examples of rejected ICA components classified as noise.	34
Figure 23. Example of dual regression stage 1 output timeseries and stage 2 spatial maps.	35
Figure 24. Partial correlation between Visual 2 and DMN vs. Centiloid scale.	37
Figure 25. Connectivity matrix of RSNs FC vs. age, showing significant full and partial correlations (red: positive, blue: negative).	38
Figure 26. Age-related changes in intra- and inter-network connectivity for DMN and Executive networks.	39
Figure 27. Heatmap showing Pearson correlations between partial FC features, Tau SUVR activities, and structural MRI measures (red: positive, blue: negative).	40
Figure 28. Significant correlations and average Pearson coefficients between FC features, Tau SUVR, and structural MRI measures.	41
Figure 29. WBS diagram illustrating main work packages and task subdivisions.	43
Figure 30. GANTT diagram of the global project execution.	44
Figure 31. PERT diagram illustrating the chronology of the project.	45
Figure 32. Motion parameters from MCFLIRT, showing rotational, translational displacements, and FD for QC.	59
Figure 33. Example of an HTML QC report, showing anatomical and functional outputs, motion correction, and segmentation metrics.	60
Figure 34. Power spectrum of network nodes from ICA decomposition, showing the 0.01–0.1 Hz range after Butterworth filtering.	62
Figure 35. Hierarchical clustering of RSNs using FSL Nets, grouping networks like DMN and visual networks together.	62

List of Tables

Table 1.	Descriptive statistics of the Amyloid-E group downloaded (N = 1,178).	25
Table 2.	Reviewer QC Results.	31
Table 3.	Overall QC Results.	32
Table 4.	Table of activities with duration and precedents.	44
Table 5.	SWOT analysis of the project.	46
Table 6.	Project Costs and Resources.	47
Table 7.	Quantitative QC Criteria Summary.	61
Table 8.	Reviewer Agreement Across Groups.	61

Contents

1	Introduction	1
1.1	Project Definition	1
1.2	Objectives	1
1.3	Motivation	2
1.4	Scope and Limitations	2
2	Background	3
2.1	General Concepts	3
2.1.1	Big Data and AI	3
2.1.2	Advancements in Medical Imaging leading to Functional MRI	3
2.1.3	Alzheimer's Disease	5
2.1.4	Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study	7
2.2	State of the Art	7
2.2.1	Current Approaches to fMRI Data Preprocessing	7
2.2.2	Current Trends in Biomarker Identification and Imaging Analysis	8
2.3	State of the Situation	9
3	Market Analysis	11
3.1	Medical Imaging Preprocessing Market	11
3.2	Medical Imaging Biomarkers Market	12
4	Concept Engineering	13
4.1	Data Structuring and Storage	13
4.2	Preprocessing steps	14
4.2.1	Brain Extraction	15
4.2.2	Anatomical Segmentation	15
4.2.3	Slice Timing Correction	16
4.2.4	Motion Correction	16
4.2.5	Spatial Smoothing	16

4.2.6	Spatial Registration	17
4.2.7	Temporal Filtering	17
4.2.8	Nuisance Regression	18
4.3	Quality Control	19
4.3.1	Report based Approach	19
4.4	Extracting brain Functional Connectivity patterns	20
4.4.1	Dual regression	22
4.4.2	Network Analysis	23
5	Detailed Engineering	25
5.1	Data Acquisition and Structuring	25
5.2	Preprocessing Pipeline	27
5.3	Quality Control	28
5.3.1	Development of the QC Application	28
5.3.2	QC Stages and Errors	28
5.3.3	QC Results	31
5.4	Group Independent Component Analysis	33
5.5	Dual Regression	35
5.6	Functional Connectivity Mapping and Analysis	35
5.6.1	Interrogating the Models: Exploring correlations between FC and external variables	36
5.6.2	Correlations between Centiloid and FC	36
5.6.3	Correlations between Age and FC	37
5.6.4	Exploratory Framework for Multivariate Connectivity Patterns	39
6	Execution Schedule	43
6.1	Work Breakdown Structure (WBS)	43
6.2	Execution Chronogram (GANTT)	43
6.3	Critical Path Method Diagram (PERT)	45
7	Technical Feasibility	46

8	Economical Feasibility	47
9	Legislation and Regulations	48
10	Conclusions and Future Lines	49
11	References	51
12	Annex	58
12.1	Annex A: GitHub Repository FlexfMRI	58
12.2	Annex B: Cognitive Assessments	58
12.3	Annex C: Overview and Analysis of the QC Framework	59
12.4	Annex D: Analyses of Temporal Filtering and RSN Clustering	62
12.5	Annex E: MultiBrain Congress Poster	63

1 Introduction

1.1 Project Definition

This project aims to develop a comprehensive set of procedures for processing large datasets of resting-state functional Magnetic Resonance Imaging (rs-fMRI) data and to explore their associations in various clinically significant contexts. Specifically, the project will utilize the developed processing pipeline to examine early alterations in functional connectivity (FC) associated with Alzheimer's disease (AD), using data from the large-scale A4 study sponsored by Eli Lilly and Company [1, 2].

The project involves organizing and preprocessing fMRI data through an automated and flexible pipeline to ensure consistency and reliability. Additionally, a Quality Control (QC) application will be developed to evaluate the preprocessing outcomes and detect any issues that could affect the analysis. Finally, the processed images will be analyzed to explore alterations in brain connectivity that could help identify early markers of AD, enhancing the understanding of its progression and supporting research into early detection and intervention strategies.

1.2 Objectives

The objectives of this project are centered around using fMRI data to better understand early-stage changes in brain connectivity associated with preclinical (pre-symptomatic) AD and establishing a robust framework for future analyses. The specific objectives are as follows:

1. **Data Structuring:** Download and organize the extensive dataset from the A4 study to facilitate efficient processing and analysis. This involves structuring the data in a systematic way that ensures accessibility and compatibility with downstream analyses.
2. **Preprocessing Pipeline Development:** Define and implement a flexible and automated preprocessing pipeline for fMRI data. The pipeline will incorporate essential steps such as motion correction, spatial normalization, and others, ensuring the data is clean and consistent across all subjects.
3. **Quality Control Application Development:** Develop a Quality Control (QC) application to evaluate the output of the preprocessing pipeline. The application will identify and address potential issues, such as artifacts or misalignments, to ensure the preprocessed images meet the highest standards for subsequent analyses.
4. **Functional Connectivity Analysis:** Investigate patterns of functional connectivity by analyzing the processed images and identifying relevant resting-state networks (RSNs) defined in the literature. By focusing on early-stage changes within these networks, the analysis aims to uncover subtle neural alterations that may occur prior to the clinical onset of Alzheimer's disease.

These objectives collectively aim to advance fMRI neuroimaging research and contribute to a better understanding of Alzheimer's disease.

1.3 Motivation

The brain, the organ that defines our thoughts, emotions, and identity, remains one of the greatest mysteries in science. Despite decades of research, we are only beginning to scratch the surface of its complexities. How does this complex network of neurons store memories, make decisions, and shape every aspect of our lives?

The emergence of rs-fMRI has provided an unprecedented window into the brain's intrinsic activity, even when not engaged in specific tasks. Functional connectivity (FC) refers to patterns of coordinated activity between different brain regions over time. This imaging technique can extract FC patterns, revealing how brain regions communicate, forming the foundation for cognition, behavior, and memory. These spontaneous neural fluctuations offer information into healthy brain function and disruptions associated with diseases like AD.

Alzheimer's Disease, a neurodegenerative disorder affecting millions, disrupts memory and identity and has been associated with altered brain connectivity. Using rs-fMRI data from the A4 study, which includes cognitively normal individuals with early signs of amyloid buildup, this project investigates subtle connectivity changes in preclinical cases. These findings may offer insights into the earliest stages of AD and strategies for early detection.

1.4 Scope and Limitations

This project will be conducted within the Biomedical Imaging Group (BIG) at the Faculty of Medicine and Health Sciences at the University of Barcelona (UB). The work will span six months, requiring careful planning and efficient use of resources to ensure all tasks are completed within the timeframe.

A key limitation is the restricted project duration, which requires prioritizing tasks and setting realistic goals. Additionally, while the project benefits from access to high-performance computing resources provided by the BIG, the economic scope is limited, as there is no external budget for additional tools. However, the available infrastructure is enough to support the computational demands of this research.

Another limitation lies in the A4 study dataset. All participants are cognitively normal but exhibit amyloid deposits, representing a preclinical stage of AD. The absence of a healthy control group (subjects without amyloid buildup) limits the scope of comparative analyses. Furthermore, as the A4 study is a clinical trial, the subjects are highly homogeneous, with little variability between them. This lack of differentiation restricts the ability to capture a wide range of FC changes, potentially reducing the sensitivity of the analyses. Additionally, as the dataset does not fully reflect the expected distribution in the normal population, caution should be taken when interpreting and extrapolating the results.

This project is not intended for commercial purposes, nor does it aim to develop a medical product for validation or patenting. The primary objective is to contribute to the scientific community by developing tools that enhance research capabilities, improve the application of fMRI techniques, and advance our understanding of AD.

2 Background

2.1 General Concepts

2.1.1 Big Data and AI

In today's world of technology, the integration of big data and artificial intelligence (AI) is driving innovation in many areas. In neuroscience, as in other fields, it is opening up new ways to conduct research and find solutions to neurological disorders.

Advances in technology have enabled the collection of vast and diverse datasets from sources like blood analyses, genetic information, and neuroimaging. While managing this data presents challenges, tools like cloud storage and scalable frameworks have significantly eased the process [3]. However, the real value of big data is achieved when AI and deep learning models are applied to analyze it, uncovering patterns, predicting trends, and generating insights that surpass traditional methods.

Deep learning, particularly convolutional neural networks (CNNs) [4], are increasingly used in medical imaging, including Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) scans, to enhance the efficiency of research projects. Originally developed for tasks like image classification [5], object detection [6], and segmentation [7], these methods have become more common due to the availability of large datasets and big data resources, making them valuable tools in neuroimaging research.

In neuroscience, platforms like OpenNeuro [8] provide access to over 406 terabytes of datasets, enabling AI applications for tasks such as brain extraction, tissue segmentation, and identifying FC patterns [9]. The growing availability of large-scale data has also driven the development of fully automated AI solutions for preprocessing and quality control [3], ensuring consistent, scalable workflows that enhance the efficiency and reproducibility of neuroimaging research.

2.1.2 Advancements in Medical Imaging leading to Functional MRI

Medical imaging has transformed healthcare, offering non-invasive methods to diagnose, treat, and monitor diseases. It began in 1895 with Wilhelm Röntgen's discovery of X-rays, producing the first radiograph of his wife's hand. This breakthrough earned him the first Nobel Prize in Physics in 1901, establishing the field of diagnostic radiology [10]. X-ray technology was quickly adopted, with Marie Curie developing mobile X-ray units during World War I to aid battlefield surgeons. By the 1920s, X-rays were routinely used for diagnosing skeletal injuries, chest conditions, and gastrointestinal issues [10].

The mid-20th century saw the advent of Computed tomography (CT), MRI, ultrasound, and PET, each providing groundbreaking capabilities. CT scans offered cross-sectional imaging. MRI captured soft tissue detail without radiation. Ultrasound enabled real-time imaging, particularly in obstetrics. PET introduced functional imaging, aiding in diagnosing diseases like cancer and AD. These advances revolutionized both anatomical and functional diagnostics.

Among the key advancements in medical imaging, MRI stands out for its ability to produce detailed images of soft tissues without using ionizing radiation. Using powerful magnetic fields and radio waves,

MRI offers exceptional resolution and contrast, making it essential for diagnosing conditions involving the brain, joints, and internal organs. While its high cost and limited availability remain challenges, its safety and precision ensure its importance in modern medicine.

MRI works by aligning hydrogen protons in the body using a strong magnetic field. A Radio Frequency (RF) pulse disrupts this alignment, and as the protons return to equilibrium, they emit signals that vary based on tissue properties. Two key parameters define these signals: T1 (longitudinal relaxation time) and T2 (transverse relaxation time) reflect how quickly protons realign with the magnetic field in the longitudinal and transversal axis, respectively. Adjusting parameters like Repetition Time (TR) and Time to Echo (TE) enables MRI to create specific contrasts and highlight tissue differences [11]. For example, T1-weighted imaging (T1w) provides detailed anatomical views by emphasizing fat-rich structures, while sequences like FLAIR suppress fluid signals to detect abnormalities such as vascular lesions. These specialized sequences enhance MRI's diagnostic capabilities across various clinical applications.

MRI provides a powerful tool for anatomical imaging, but functional MRI (fMRI) extends this capability by measuring indirect brain activity. The human brain, despite making up only about 2% of the body's weight, consumes approximately 20% of its energy and holds 15% of its blood supply, even during rest [12, 13]. This high energy consumption is unevenly distributed, with some regions exhibiting greater oxygen demand [14], as seen in Figure 1. Interestingly, during specific tasks or cognitive demands, the relative increase in brain activity is relatively small compared to its baseline activity, emphasizing the brain's constant energy-intensive operation. This was first described in 1995 by Biswal et al., who identified intrinsic brain activity patterns at rest, forming the basis for rs-fMRI [15].

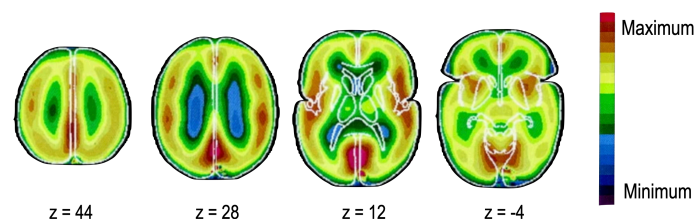


Figure 1: PET imaging technique of axial brain slices showing that oxygen consumption is not uniformly distributed across the brain. The color scale highlights regional differences, with some areas exhibiting significantly higher oxygen consumption than others. Adapted from [14].

Functional MRI emerged as a powerful tool for studying brain activity by tracking changes in blood flow and oxygenation. This MRI modality is a four-dimensional (4D) technique that captures spatial information in three dimensions over time, enabling the visualization of the brain's dynamic activity. By highlighting regions that consume more energy, it reveals brain activity during tasks or even at rest. It is based on the Blood Oxygen Level Dependent (BOLD) signal, which reflects changes in the levels of oxygenated (HbO) and deoxygenated (HbR) hemoglobin in the blood. When a specific brain region becomes active, it requires more oxygen, leading to an increase in local blood flow that delivers more oxygen than is immediately needed. This results in a decrease in deoxygenated hemoglobin, which is paramagnetic and disrupts the local magnetic field, and an increase in oxygenated hemoglobin, which is diamagnetic and enhances the MRI signal [16]. The BOLD signal captures these changes, providing an indirect yet effective way to map brain activity and identify regions involved in different tasks.

When the subject is at rest (not performing any specific task), fMRI allows analyses of spontaneous fluctuations in the signal. In this context, subjects lie on the MRI table with their eyes open or closed, remaining awake and letting their minds wander naturally, free of structured thought or external stimuli. This approach is known as resting-state fMRI. For each voxel in the image, the BOLD signal is measured, and when these signals are combined, they help us understand the brain's intrinsic activity and show patterns of FC.

2.1.3 Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder primarily affecting older adults, characterized by cognitive decline, memory loss, and behavioral changes. The theoretical basis of AD involves a multifactorial interaction of genetic, molecular, and cellular mechanisms that collectively contribute to its pathophysiology. These mechanisms are detailed below.

At the molecular level, a hallmark of AD is the accumulation of amyloid-beta ($A\beta$) plaques and neurofibrillary tangles (NFTs) composed of hyperphosphorylated tau protein. The amyloid cascade hypothesis suggests that overproduction or impaired clearance of $A\beta$ peptides, derived from the amyloid precursor protein (APP), leads to the formation of insoluble plaques in the extracellular space of the brain. This accumulation disrupts neuronal communication and triggers inflammatory responses, contributing to synaptic dysfunction and neuronal death. Complementing the amyloid hypothesis is the tau hypothesis, which focuses on tau protein, a stabilizer of microtubules. In AD, tau undergoes abnormal hyperphosphorylation, causing it to detach from microtubules and aggregate into NFTs within neuronal axons. This disrupts microtubule structure, compromising the structural and functional integrity of neurons, interfering with axonal transport, and leading to cell death. It is believed that the interaction between $A\beta$ and tau pathologies synergistically drives neurodegeneration, although the precise mechanisms remain under investigation [17, 18]. An overview of the molecular level of the disease can be seen in Figure 2.

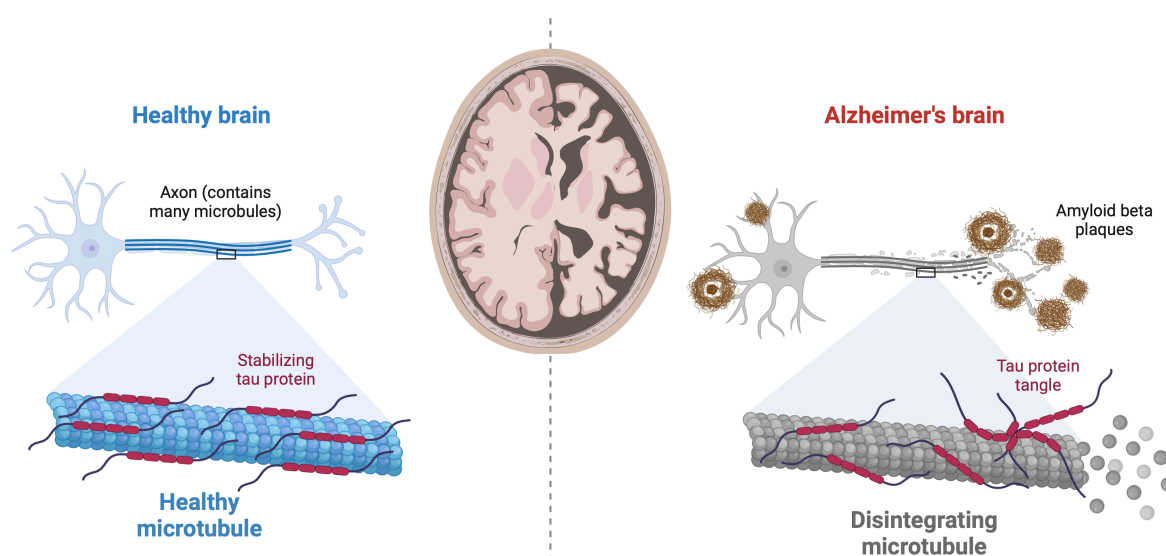


Figure 2: Molecular representation of $A\beta$ structures and NFTs of AD compared to the natural state. Adapted from [19].

In addition to molecular pathologies, several factors contribute to the complexity of AD. In familial AD, mutations in genes such as APP, PSEN1, and PSEN2 increase $A\beta$ production, leading to early onset. In sporadic AD, the apolipoprotein E (APOE) E4 allele is the most significant genetic risk factor, influencing $A\beta$ aggregation and clearance, and increasing disease risk [20].

Other mechanisms include neuroinflammation, where microglial activation transitions from protective to chronic, releasing pro-inflammatory cytokines that exacerbate neuronal damage [21]. Moreover, the cholinergic hypothesis emphasizes that a deficit in acetylcholine and the loss of cholinergic neurons are associated with cognitive decline [22]. Vascular dysfunction, driven by hypertension and diabetes, reduces cerebral blood flow and promotes $A\beta$ accumulation, connecting vascular health to AD pathology [18]. These factors highlight the multifactorial nature of AD, complicating its diagnosis and treatment.

Advancing the understanding of these mechanisms relies heavily on imaging technologies, which provide details into the structural and molecular changes associated with AD. Structural MRI and PET scans are widely used to assess disease progression and pathology. MRI can reveal cortical atrophy, particularly in the hippocampus and medial temporal lobe structures, which are early indicators of AD, as seen in Figure 3. PET imaging is used for both metabolic and molecular assessments. Radiotracers like FDG can assess glucose metabolism, while amyloid imaging tracers such as Pittsburgh Compound B (PiB) or Florbetapir (AV-45) visualize $A\beta$ plaques, which are detectable even before clinical symptoms appear (see Figure 3) [18]. More recently, Flortaucipir has been introduced as a PET radiotracer to visualize tau protein accumulation, another hallmark of AD [23]. Amyloid PET is valuable due to the Centiloid scale (0–100), based on Standardized Uptake Value Ratios (SUVR), which compare radiotracer uptake in amyloid-affected regions to a reference region, providing a normalized measure of amyloid burden. Values near 100 indicate high amyloid plaque burden, strongly linked to AD pathology, while lower values suggest minimal deposition. This scale ensures consistent quantification across studies and clinical settings [24].

In conclusion, the theoretical basis of Alzheimer's disease is based on a complex interaction of amyloid and tau pathologies, genetic predispositions, neuroinflammation, cholinergic deficits, and vascular contributions. Understanding these multifactorial mechanisms is essential for developing effective therapeutic strategies to prevent, slow, or halt the progression of this devastating disease.

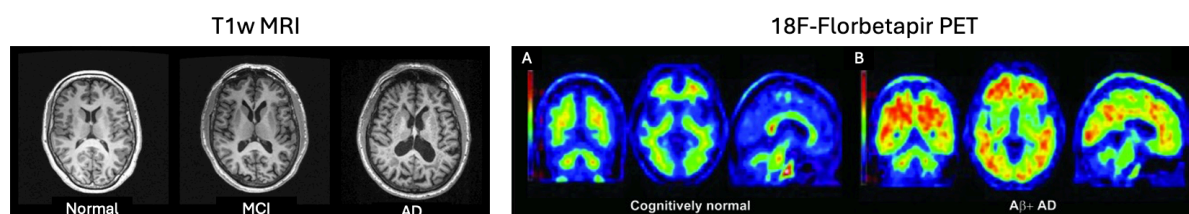


Figure 3: (Left) T1w MRI axial slices show reduced gray matter volume in an AD patient compared to a healthy control, with intermediate loss in a mild cognitive impairment (MCI) patient. Obtained from [25]. (Right) 18F-Florbetapir PET shows minimal $A\beta$ deposition in a normal subject (A) and significant $A\beta$ accumulation in an AD patient (B). Obtained from [26].

2.1.4 Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study

In the context of AD, numerous clinical trials have been conducted to evaluate potential treatments aimed at reducing or delaying the progression of symptoms. One of the most notable recent trials was the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) study. It was a Phase 3, double-blind, placebo-controlled trial targeting individuals in the preclinical stages of AD, conducted by Eli Lilly and Company [1]. This multicentric, worldwide study lasted four and a half years and involved participants selected based on evidence of amyloid buildup in the brain, identified through amyloid PET imaging, while remaining clinically normal but at high risk of future cognitive decline.

The study aimed to test whether solanezumab, a monoclonal antibody designed to bind to $A\beta$, could slow cognitive decline. The measure used to evaluate results was a composite cognitive score sensitive to early changes in preclinical Alzheimer's [27]. In addition to efficacy, the trial collected extensive data on solanezumab's long-term safety, cognitive performance, and imaging metrics, including structural MRI, functional MRI and PET scans. In 2023, the trial was concluded and the results showed that solanezumab did not slow the progression of cognitive decline [1]. Despite the negative outcome, the data collected over the trial remains highly valuable for further analysis, with the potential to reveal better understanding of AD pathology and contribute to future research.

The screening data from this clinical trial will be used in this project to test the preprocessing pipeline and conduct further analysis. Notably, the dataset does not include a healthy control group (participants without amyloid buildup) or individuals with developed AD. However, the Centiloid scale, which measures amyloid buildup, serves as a continuous indicator to assess the progression of AD pathology along a gradient rather than through categorical group comparisons.

2.2 State of the Art

Given the scope of this project, which encompasses both the preprocessing of fMRI data and the subsequent analysis of imaging biomarkers and fMRI metrics, this section will be divided into two distinct parts. Firstly, the state of the art in image preprocessing is examined, focusing on the techniques and methodologies used to prepare fMRI data for analysis. Secondly, the state of the art in imaging biomarkers is explored, highlighting current advances in identifying meaningful patterns and metrics derived from imaging data.

2.2.1 Current Approaches to fMRI Data Preprocessing

When conducting fMRI preprocessing for a study, several common steps are typically performed, such as slice timing correction, temporal filtering, and spatial normalization. However, there is no standardization regarding the order of these steps or the parameters that should be used. Many of these decisions depend on the specifics of the fMRI acquisition and the objectives of the study.

In this context, an open-source software called fMRIPrep was introduced in 2019 by researchers from Stanford University. This software is described as an analysis-agnostic tool designed to address the

challenges of robust and reproducible preprocessing for fMRI data. According to its creators: “fMRIPrep equips neuroscientists with an easy-to-use and transparent preprocessing workflow, which can help ensure the validity of inference and the interpretability of results.” [28].

This tool has significantly automated the preprocessing workflow and has become widely used, establishing itself as the current gold standard for fMRI preprocessing. Its popularity is largely due to its adoption within the scientific community, which enhances reproducibility and consistency in studies.

Additionally, it integrates state-of-the-art automatic methods from widely used neuroimaging software packages like FSL [29], ANTs [30], and FreeSurfer [31], ensuring high-quality preprocessing while minimizing the need for manual intervention. Moreover, fMRIPrep can be combined with other available software, such as SIMEXP [32], to perform visual and quantitative QC.

While fMRIPrep provides a standardized and robust preprocessing pipeline, it may lack the flexibility and customizability of manually performing specific correction steps with tailored parameters. Although fMRIPrep allows some adjustments, such as modifying the degree of spatial or temporal filtering, it does not offer full customization of the algorithms behind individual preprocessing steps. This limitation has motivated the development of alternative preprocessing approaches to some of the correction steps, many of which are driven by AI.

One example is HD-BET (High Definition Brain Extraction Tool), which uses CNNs to address challenges that traditional methods, like those used in fMRIPrep, may encounter [33]. Traditional automatic methods often rely on assumptions about healthy brain structures, which can result in suboptimal performance when processing atypical brains, such as those with lesions or other abnormalities. AI-driven tools like HD-BET aim to overcome these limitations by adapting more effectively to diverse data sets.

Despite these limitations, fMRIPrep remains a dominant tool in the neuroimaging community. Its widespread adoption is largely due to its ease of use, standardization of workflows, and open-source availability, which promote reproducibility and accessibility for researchers worldwide.

2.2.2 Current Trends in Biomarker Identification and Imaging Analysis

Imaging biomarkers are quantifiable indicators derived from medical imaging techniques such as MRI, CT, and PET. These biomarkers play a key role in modern medicine, aiding in the diagnosis, prognosis, and therapeutic monitoring of diseases like cancer, cardiovascular disorders, neurodegenerative conditions, and many others. They are objective and quantifiable features that enable accurate disease staging and monitoring of progression, highlighting their critical role in advancing precision medicine [34].

There has been a growing interest in more accessible biomarkers, such as those detected through blood tests, ocular scans or cerebrospinal fluid (CSF) analysis. In the context of AD, these methods are gaining popularity as they are less invasive and cost-effective alternatives to imaging [35].

However, imaging biomarkers remain necessary despite their lower accessibility and higher cost. While blood and CSF biomarkers, such as phosphorylated tau (e.g., p-tau217) or $A\beta$ levels, are valuable for early detection, they often provide a global measure of pathology without localization. For instance, CSF $A\beta$ tests can identify amyloid-positive individuals in the early stages of AD but may plateau during

pre-dementia stages, missing continued amyloid accumulation [36]. Additionally, imaging techniques, such as 18F-Florbetapir PET scans, offer some advantages by directly visualizing amyloid deposition and tracking its progression. Unlike blood or CSF tests, imaging allows for staging of the disease and provides detailed, localized details about the pathology [37]. This depth of information can guide more precise diagnoses and targeted interventions, underscoring the unique value of imaging biomarkers in neurodegenerative research and clinical practice [38].

Current trends in imaging biomarkers emphasize multimodal approaches, combining structural imaging techniques like CT and MRI with metabolic imaging such as PET and SPECT. This integration offers a comprehensive view of both anatomical and functional changes, enhancing diagnostic accuracy [39].

Significant progress is also being made in the development of new PET tracers targeting specific molecules. For instance, in 2019, Flortaucipir became the first FDA-approved tau-specific PET tracer for imaging tau pathology in AD [23], marking a milestone in precision diagnostics.

In parallel, computational advances, particularly AI-based methods, are being employed to analyze large datasets and extract patterns of disease progression. While these tools hold great promise, their adoption in clinical practice is challenged by concerns about interpretability, as many AI models function as "black boxes", making their reasoning difficult to explain and validate [40]. Despite this, the integration of imaging biomarkers and computational tools continues to drive innovation in this field.

In the context of this project, modalities like fMRI-derived biomarkers, such as FC measures, are expanding the horizon of imaging biomarkers. Although their clinical applications remain underexplored, they offer significant potential to advance our understanding of brain network dynamics. All of these advancements showcase how rapidly evolving imaging techniques and computational tools are transforming medicine, paving the way for more precise and personalized medical care.

2.3 State of the Situation

Alzheimer's disease affects over 55 million people worldwide, with this number projected to increase significantly in the coming decades, representing a critical global health challenge [41]. Recent advancements in treatment, such as the approval of lecanemab (commercially known as Leqembi) in November 2024 by the European Union, mark a major milestone in targeting $A\beta$ pathology [42]. While this drug offers promise, it has been widely discussed due to its notable secondary effects, emphasizing the need for robust and comprehensive approaches to drug evaluation.

Recent clinical trials for AD therapies, including monoclonal antibodies like solanezumab, lecanemab and donanemab, primarily rely on traditional cognitive assessments and PET imaging biomarkers to test drug efficacy and monitoring [1, 43, 44]. While cognitive assessments are valuable, they are subjective and prone to variability, limiting their reliability. Similarly, PET imaging, though effective for detecting amyloid deposition, has limitations when evaluating treatment response. For instance, monoclonal antibodies eliminate $A\beta$ proteins, leading to reduced PET signal without providing further insight into the broader effects on brain function.

To address these gaps, there is a growing opportunity to introduce new assessment tools, such as

fMRI FC biomarkers. These non-invasive measures can offer complementary data by capturing network-level brain changes, providing a more comprehensive view of disease progression and therapeutic impact [45]. However, for fMRI biomarkers to be effectively integrated, it is essential to establish standardized workflows to ensure reproducibility and reliability [46]. Additionally, rigorous validation of fMRI-derived biomarkers is required to confirm their clinical utility.

Nonetheless, incorporating fMRI FC biomarkers into AD research offers a promising path to address existing gaps in assessment methods, providing a better understanding of disease progression and enabling additional evaluation of treatment effects. These biomarkers could even predict the presence or absence of secondary effects after drug treatment, further enhancing the ability to assess therapeutic efficacy in clinical trials.

3 Market Analysis

To provide context for this project, a market analysis has been conducted to explore its key characteristics, major players, and future perspectives. As in the previous sections, the market analysis has been divided into two parts. The first part will focus on the medical imaging preprocessing market, examining the technologies, tools, and companies involved in preparing imaging data for analysis. The second part will focus on the imaging biomarkers market, highlighting its role in diagnostics, current trends, and potential innovations.

3.1 Medical Imaging Preprocessing Market

The medical imaging preprocessing market is growing significantly, driven by technological advancements and rising demand in both clinical and research applications. This market can be divided into clinical sectors such as neurology, cardiology, and oncology, as well as research-focused solutions, with software ranging from open-source platforms to proprietary tools. Analyses for 2024 are limited, so data from 2023 will be referenced. In 2023, the global medical image analysis software market was valued at USD 3.27 billion and was projected to grow at a compound annual growth rate (CAGR) of 7.8% from 2024 to 2030, reflecting the increasing reliance on advanced imaging technologies across diverse applications [47].

Open-source preprocessing software initiatives like AFNI [48], FSL [29], and fMRIPrep [49] have particularly contributed to the growth in neuroimaging research, providing accessible and reliable tools for the scientific community. Also, the wide availability of neuroimaging data projects, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) [50], Human Connectome Project (HCP) [51], and UK Biobank [52], has helped researchers by providing large, standardized datasets. These resources, both datasets and preprocessing tools, have significantly advanced neuroimaging research, enabling more varied studies.

On the proprietary side, industry leaders like Siemens integrate preprocessing software directly into their MRI or CT imaging machines, providing seamless workflows and standardized solutions. Meanwhile, smaller companies such as QMENTA or Advantis Medical Imaging focus on more specialized applications. These start-ups deliver customized preprocessing tools for areas like neuroimaging and radiomics, using advanced technologies such as AI, high-performance computing, and secure cloud platforms to enhance clinical and research capabilities.

QMENTA, a medical imaging and cloud software start-up based in Boston and Barcelona, offers solutions for preprocessing structural, functional, and diffusion MRI images. Their platform, designed for both doctors and researchers, features AI tools, streamlined workflows, and secure data sharing to improve collaboration in neurology [53].

Additionally, Advantis Medical Imaging, based in Greece, offers an FDA-cleared and CE-marked software for brain MRI preprocessing, diffusion tensor imaging fiber tracking and even fMRI preprocessing. Their platform is designed for clinicians, providing automated workflows for image preprocessing [54].

Proprietary solutions are more prevalent in clinical settings, where ease of use, regulatory compliance,

and precise analytics are essential, while open-source tools are adopted in research due to their flexibility, accessibility, and community-driven development. Additionally, pharmaceutical companies also use these tools for drug testing in clinical trials, highlighting their versatility beyond healthcare and research.

This market is evolving in two key areas: fully automated pipelines and faster computing solutions. Automation reduces the need for human involvement, especially in QC stages, making processes more efficient and less dependent on manual oversight. Deep learning technologies play a key role in this shift, replacing traditional methods with more adaptable, fully automated systems that can handle a wider range of applications [55]. Secondly, advancements in computing speed and cloud-based solutions are accelerating data processing and integration with diagnostic tools. Together, these trends are transforming how imaging data is analyzed, improving both research workflows and clinical care.

3.2 Medical Imaging Biomarkers Market

The imaging biomarker market is divided into several segments, including imaging modalities (MRI, CT, PET), application areas (such as oncology, neurology, and pulmonology), and end users, which include hospitals, research institutes, and pharmaceutical companies. Historically, this market has shown strong growth, driven by advancements in imaging technology and rising disease prevalence. Estimated at USD 7.8 billion in 2024, it is projected to reach USD 17.7 billion by 2030, with a CAGR of 14.6% [56].

Imaging biomarkers have significantly advanced medical research and diagnostics. Amyloid imaging, crucial in AD research, enables the extraction of SUVR, while FDG-PET remains a key tool in oncology and epilepsy diagnostics. Diffusion-weighted imaging revolutionized stroke detection by providing essential information about acute ischemic events. Looking ahead, fMRI FC biomarkers present promising opportunities for exploring brain network dynamics and could potentially be introduced into clinical practice in the future [46, 57, 58].

Many companies, like Qubitech, operate at the intersection of the preprocessing and biomarker detection markets, reflecting the industry's integration of data processing with biomarker identification. Qubitech's software for brain PET, SPECT, and MRI provides automated workflows and report generation for neurological diagnostics [59]. Similarly, collaborations like Bayer and Freenome in oncology biomarkers highlight the increasing convergence of imaging and drug development [60].

Future advancements in this market are expected to focus on combining imaging and blood-based biomarkers for more comprehensive diagnostics. Additionally, AI, particularly deep learning, will drive improvements in predictive analytics for early disease detection. Standardizing imaging biomarker protocols across modalities and institutions will be crucial for ensuring consistency and reliability [46]. These developments have potential to transform diagnostics and patient care, reinforcing the key role of imaging biomarkers in precision medicine.

4 Concept Engineering

This section describes the methodology followed to organize and preprocess neuroimaging data. It covers data download and standardization, the customization of the preprocessing pipeline for fMRI analysis, and the implementation of QC measures to ensure accuracy. Additionally, advanced analytical techniques, such as Independent Component Analysis (ICA) and network-based methods, are introduced to extract meaningful brain connectivity patterns from the previously cleaned fMRI data.

4.1 Data Structuring and Storage

In neuroimaging, researchers often work with large datasets that include various imaging scans and modalities, such as MRI or PET, along with complex metadata. The absence of a standardized framework for organizing heterogeneous data creates significant challenges in sharing, collaboration, and interpretation. This lack of consistency leads to inefficiencies and complicates the integration of datasets from various sources, ultimately restricting reproducibility and restricting scientific progress.

To address these issues, the Brain Imaging Data Structure (BIDS) was introduced as a standardized framework for organizing and sharing neuroimaging data. BIDS ensures datasets are structured consistently, enhancing data sharing, minimizing errors, and improving compatibility with automated processing tools like fMRIPrep [61]. Validator programs also ensure datasets conform to BIDS standards [62], promoting consistency and reliability across shared resources.

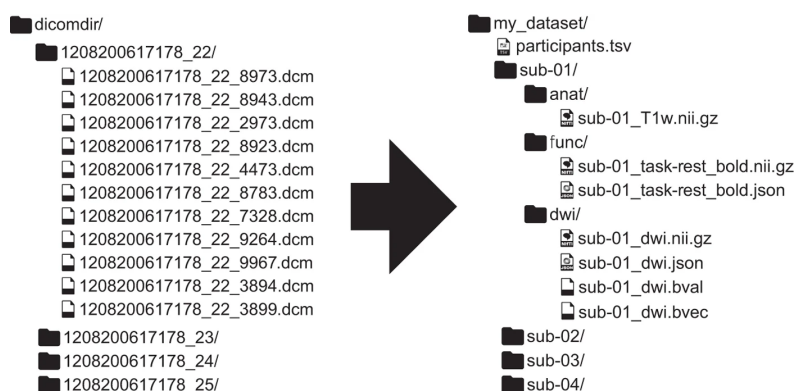


Figure 4: Illustration of a BIDS structured dataset. Obtained from [61].

BIDS specifies conventions for data formatting and naming, using NIfTI (.nii or .nii.gz) for imaging data and JSON files for metadata, as seen in Figure 4. It includes naming patterns that allow consistent identification of data attributes like subject IDs, session numbers, and modality types, facilitating easy navigation and retrieval. BIDS also distinguishes between raw data (unaltered acquisitions) and derived data (processed results), which is essential for maintaining data integrity.

While other structuring frameworks exist, BIDS is the most widely used in the neuroscience community. Therefore, the first step after downloading raw data is organizing it into the BIDS format, ensuring standardization and scalability. Although this can be time-consuming due to varying raw data formats, Python scripts can streamline the process. This upfront effort can simplify downstream workflows.

4.2 Preprocessing steps

Preprocessing rs-fMRI data is a critical step aimed at enhancing the quality of the data for subsequent analyses. The primary objective of this phase is to minimize the impact of various artifacts and structured noise, therefore ensuring that the neural signals of interest are accurately captured and analyzed [12].

Structured noise in rs-fMRI arises from two main sources: physiological processes and subject movement. Physiological noise includes periodic fluctuations from breathing and the heartbeat, which introduce signals that can confound neural activity measurements [63]. These biological rhythms, though relatively stable (heart rate of about one beat per second and breathing cycle every three seconds), are difficult to remove completely due to the limited temporal resolution of fMRI, often leading to aliasing [64].

Moreover, subject movement, particularly head motion, is another major source of structured noise in fMRI. Head motion can cause spatial misalignment between consecutive fMRI volumes, known as first-order motion effects, which are relatively easy to identify by examining successive images. However, higher-order motion effects, such as partial volume effects, spin history effects, and field inhomogeneities, are more complex and difficult to correct [65]. Since both subject movement and physiological processes like cardiac and respiratory control are influenced by brain activity—the same source of the BOLD signal—it becomes challenging to distinguish signal from noise, as shown in Figure 5. These complexities underscore the need for robust preprocessing techniques to improve data quality.

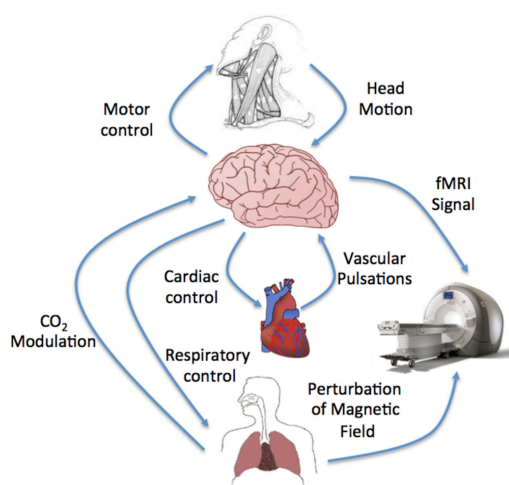


Figure 5: Diagram illustrating sources of fMRI signal noise, including head motion, cardiac, and respiratory activity, all influenced by brain activity. This overlap with the origins of the BOLD signal makes it more difficult to differentiate between signal and noise. Obtained from [66].

Beyond physiological and movement-related noise, artifacts can also arise from the fMRI scanner hardware itself. These artifacts may include electromagnetic interference, gradient instabilities, and other system-related noises that degrade the quality of the fMRI signal [12].

The field of noise reduction in fMRI is actively evolving, with research focused on developing new techniques to enhance data quality. Current advancements include improved motion correction algorithms, better physiological noise modeling, and machine learning approaches that can more effectively distinguish between neural signals and noise [67]. These innovations improve the reliability and validity of FC analyses, contributing to a better understanding of brain network dynamics.

There are different approaches to fMRI preprocessing. fMRIPrep has been extensively tested in previous BIG projects and is widely recognized in the scientific community. However, it does not fully meet the modularity and flexibility requirements needed for some projects. This limitation prompted the need for an approach that supports parameter adjustments and the integration of alternative algorithms, offering a high degree of adaptability to suit diverse research needs.

Additionally, fMRIPrep relies on FreeSurfer, a state-of-the-art software designed for advanced structural processing, which ensures precise alignment and robust anatomical references. However, this focus on structural and spatial processing demands significant time and computational resources. Since structural analyses were not the focus of this functional study, the use of FreeSurfer was unnecessary.

For these previous reasons, fMRIPrep was not used in this project. Instead, a modular preprocessing pipeline already available at the BIG was adapted to meet the specific needs of the study. Although the pipeline was previously available, it was disorganized and required some improvements, which were implemented to enhance its functionality and structure. This flexible pipeline allows for multiple algorithm choices for various steps. For instance, brain extraction can be performed using tools like HD-BET, BET implemented via Nipype, or BET from FSLpy. Similarly, other steps like registration methods also offer several alternatives to suit different needs. However, for this project, commonly used algorithms within the BIG group were selected. The evaluation of the different behaviors and comparative performance of these algorithms is beyond the scope of this project and will not be addressed here. The following sections outline the key steps involved in this customized preprocessing pipeline.

4.2.1 Brain Extraction

Brain extraction, or skull stripping, is a step that isolates brain tissue from surrounding structures like the skull, scalp, and dura mater in MRI scans. This step is essential to prevent non-brain elements from interfering with image analysis. Traditional methods often struggle when processing brains with abnormalities or handling diverse MRI datasets. Recently, the HD-BET algorithm has become widely used, using artificial neural networks and trained on a variety of data sources to address these challenges and provide more robust brain extractions [33].

4.2.2 Anatomical Segmentation

The next step in processing anatomical images is segmentation, where the brain is divided into distinct tissue types like gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). This process is essential for understanding the structural foundation of functional activity in fMRI analyses. The tool used is FSL's FAST (FMRIB's Automated Segmentation Tool) which combines tissue classification and bias field correction using a hidden Markov random field model [68]. Accurate segmentation is necessary to enhance fMRI analysis, as it helps remove non-neuronal signals, therefore, cleaning the BOLD signal.

4.2.3 Slice Timing Correction

A T1w 3T MRI image typically takes around 5 to 10 minutes to acquire and represents a single volume. In fMRI, however, we work with 4D data, incorporating the time dimension, which requires acquiring multiple volumes—often 100 to 200. To reduce acquisition time to the same 10 minutes, fMRI usually uses lower spatial resolution, and a specialized technique called Echo-Planar Imaging (EPI) is employed. EPI enables rapid acquisition of entire brain volumes by collecting data for each slice in quick succession. However, because slices are acquired at slightly different times within each volume, there is a temporal mismatch that can affect the alignment of neural signals across slices.

This is why EPI images often undergo slice timing correction, an optional preprocessing step in rs-fMRI. This process adjusts for differences in acquisition timing between slices within each repetition time (TR), ensuring proper temporal alignment. This correction is particularly useful when the TR is longer, such as 3 seconds, where the timing differences across slices can be significant. By using interpolation, the BOLD timecourses are temporally shifted to align the acquisition times, improving the accuracy of analyses that depend on precise timing. However, interpolation introduces slight temporal smoothing and a loss of high-frequency information [69]. Still, if we work with faster TRs, often below 1 second in modern multiband EPI sequences, these timing differences become negligible. In our study, the TR was approximately 3 seconds, and slice timing correction was applied using the 3dTshift tool from the AFNI software package [70].

4.2.4 Motion Correction

Motion correction is designed to address the effects of unavoidable head movement during scans. This step ensures that all volumes are spatially aligned by registering each volume to a reference volume, typically the first or middle volume. The process applies transformations to correct for translational movements (left-right, up-down, front-back) and rotational movements (pitch, yaw, roll), generating six motion parameters that track head motion throughout the scan. The algorithm used to apply this correction was MCFLIRT (Motion Correction FMRIB's Linear Image Registration Tool) [71].

4.2.5 Spatial Smoothing

Another commonly used preprocessing step is spatial smoothing of functional images. This involves applying a Gaussian kernel, which calculates a weighted average of neighboring pixels, resulting in a blurring effect on the images. It can be performed using various tools, such as SUSAN from FSL implemented through Nipype [72]. The primary advantage of spatial smoothing is that it helps reduce the influence of noise by averaging out random fluctuations. However, the downside is that it reduces spatial localization accuracy, as fine details in the data can be lost [73].

In our case, we decided not to apply this step because the functional data we are using has an isotropic spatial resolution of 3.3 mm³, which is already suboptimal. Furthermore, the registration process, which involves resampling the functional data to standard space, introduces a slight smoothing effect due to

interpolation. Applying additional smoothing would further degrade the resolution, resulting in the loss of critical spatial information necessary for our analysis.

4.2.6 Spatial Registration

Registration aligns brain images acquired in different spatial representations, such as functional, structural (T1w), and standard spaces (e.g., Talairach or MNI space). These spaces differ in image sizes and voxel dimensions, requiring transformations to match their spatial structures. The primary goal is to enable group-level comparisons by transforming subject-specific data into a common standard space.

In a FC analysis, we have both anatomical and functional images, and our aim is to align the functional images to a standardized template for group-level analysis. While it might seem logical to directly warp the functional images to the template, this approach is often ineffective due to their lower spatial resolution, which makes it challenging to align them accurately with the detailed features of the template. Instead, the higher-resolution anatomical image serves as a more suitable intermediary.

To normalize the functional scans to a standardized space, the process begins by aligning the anatomical image to the template. This step creates a transformation that can then be applied to the functional data. Functional scans are first registered to the anatomical image in their native space, ensuring proper alignment between the two. Since both scans are typically acquired during the same session, discrepancies in head position are minimized. By applying the anatomical transformation to the functional images, they are properly normalized to the template without requiring direct alignment of the lower-resolution functional scans to the standardized space.

Registration can be performed in two primary ways: volumetric space, which utilizes 3D voxel representations, or surface-based space, where the cortical GM is represented as a flattened or inflated surface for better alignment across folds. For our analysis, we used the ANTs (Advanced Normalization Tools) registration method, a widely used and reliable volumetric-based approach [74].

4.2.7 Temporal Filtering

Temporal filtering involves removing specific frequency components from the BOLD signal's time series at each voxel. The goal is to eliminate unwanted signal components, such as noise or artifacts, while preserving the neural signal of interest.

High-pass filtering is particularly common, as it removes low-frequency components below the range typically associated with BOLD signal fluctuations. One main artifact targeted by high-pass filtering is scanner drift—slow, which introduce gradual changes in the baseline of the BOLD signal caused by scanner hardware. In this analysis, we applied a high-pass filter with a cut-off frequency of 0.01 Hz to remove these low-frequency baseline artifacts while retaining the meaningful neuronal signal.

Low-pass filtering, though less common, can be applied in conjunction with high-pass filtering to create a band-pass filter. This approach removes both low and high frequency components, effectively isolating the frequency range associated with the BOLD signal. A commonly used cut-off frequency for

low-pass filtering is 0.1 Hz, which helps retain the neural signal while discarding rapid fluctuations typically attributed to noise. Using the Nilearn library in Python [75], a second-order Butterworth filter was applied to implement a band-pass filter on the BOLD signal. Additionally, the time series were demeaned and standardized to have a variance of one, ensuring consistency and comparability across voxels. See Annex 12.4 and Figure 34 for an example of the BOLD time series power spectrum after temporal filtering.

4.2.8 Nuisance Regression

Nuisance regression is the final step in the preprocessing pipeline, aimed at removing other unwanted sources of variability that could mask the true neural signals. During this process, timecourses representing variance from non-neuronal sources, such as fluctuations in WM and CSF (previously segmented), are identified and removed. These signals, often due to physiological noise or scanner-related artifacts, should not contribute to the BOLD signal [76].

The removal of nuisance signals can be achieved through multiple linear regression (see Figure 6), where unwanted timecourses from WM, CSF, and other confounding sources are treated as regressors. This modeling process separates the true BOLD signal from the noise, improving the signal-to-noise ratio. Using tools like the *clean_img* function from the Nilearn package [75], this step ensures that the fMRI data accurately reflects neural processes, free from non-neuronal noise, enabling more reliable interpretations of brain function. However, this approach has limitations, as it might be possible to unintentionally remove actual BOLD signal along with the nuisance components. This potential error may not be immediately detectable, potentially affecting the interpretation of the data [77].

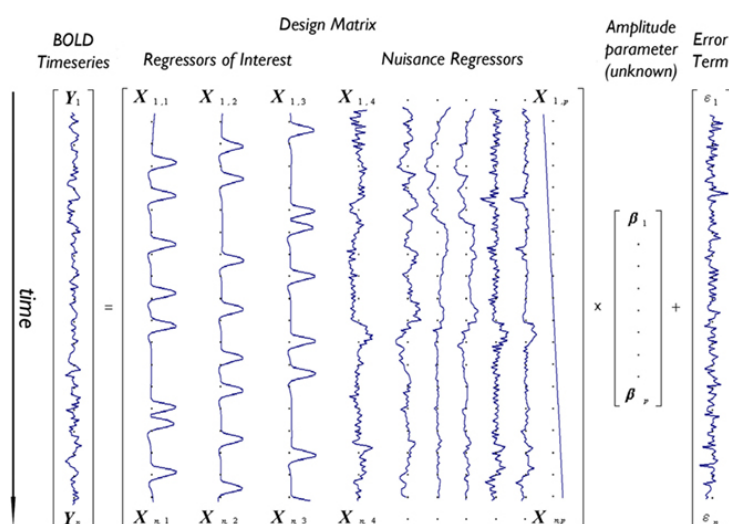


Figure 6: General Linear Model example for nuisance regression, separating BOLD timeseries into different confounds in a design matrix to remove unwanted noise from non-neuronal sources. Obtained from [76].

4.3 Quality Control

When using automated methods to process imaging data, it is essential to implement a quality control (QC) layer to ensure that preprocessing steps are applied correctly without compromising downstream analyses. This is particularly critical for large datasets, where manual inspection can be time-consuming but still remains necessary for maintaining data integrity. Analyses such as posterior Independent Component Analysis (ICA), which will be discussed later, are highly sensitive to data quality, making rigorous verification of preprocessing steps critical.

There are two primary approaches to QC in neuroimaging: quantitative (qQC) and visual (vQC). The quantitative approach involves extracting numerical features from the outputs and removing data points that deviate significantly from the mean. This method is efficient and objective, as it relies on predefined thresholds and metrics. However, it is limited in scope, focusing only on specific measurable features. In contrast, vQC involves manually inspecting the visual outputs of preprocessing steps. While this method is more thorough, as it can identify a wider range of artifacts beyond those captured by metrics, it is time-consuming and subjective, leading to variability in results.

To address the limitations of both methods, we developed a QC framework that combines visual inspection with quantitative metrics, striking a balance between efficiency and reliability. For each subject, reports were generated containing visual outputs from key preprocessing steps, allowing reviewers to assess reliability comprehensively. At the same time, quantitative parameters were extracted to identify and exclude subjects with significant deviations, reducing the need for extensive manual inspection. This combined approach ensures an efficient and robust evaluation of preprocessing steps, enhancing the accuracy and consistency of the data for subsequent analyses.

4.3.1 Report based Approach

Reports were generated for each preprocessed subject, summarizing key outputs from both anatomical and functional preprocessing steps. These reports were compiled into individual HTML files and stored within each subject's BIDS directory for easy access and review.

To streamline the manual QC process, a web application was developed to enable multiple reviewers to visually inspect these reports and classify them into three categories: 'Yes,' 'No,' and 'Maybe.' A 'Yes' indicates that no errors were identified in the preprocessing. A 'Maybe' suggests that the reviewer observed potential issues but is uncertain about whether to reject the image. Finally, a 'No' signifies the presence of clear errors, such as poor registration, faulty brain extraction, or evident artifacts (e.g., temporal or frequency artifacts). These errors will be commented in Section [5.3.2](#).

4.4 Extracting brain Functional Connectivity patterns

After preprocessing, rs-fMRI data can be analyzed to assess synchronization or de-synchronization between brain regions. Functional connectivity methods can be broadly categorized into two approaches: voxel-based and node-based [12]. Voxel-based methods estimate connectivity for each voxel across the brain, capturing fine-grained spatial patterns and generating detailed connectivity maps. In contrast, node-based methods focus on predefined brain regions or networks, summarizing connectivity at a higher level and emphasizing region-to-region or network-level interactions. For this project, a voxel-based analysis was chosen to explore similarities between regions at the voxel level and, therefore, to identify areas that exhibit coordinated activity, providing a more detailed understanding of brain connectivity.

Among voxel-based methods, Independent Component Analysis (ICA) is one of the most widely used techniques in the neuroimaging community and was selected for this project. ICA is a data-driven approach that decomposes multivariate signals into additive subcomponents. It identifies independent components (ICs) from the observed brain signal, producing a set of 3D spatial maps, or components, which are theoretically spatially independent of one another.

This technique works by analyzing the time series associated with each voxel across the brain, using their temporal dynamics to identify regions that are functionally coordinated. Regions grouped into the same component exhibit similar time series patterns, indicating that they are part of the same functional network. ICA assumes that the observed data is a mixture of multiple independent components, which, while not directly observable, can be mathematically separated into their spatial and temporal representations. For each component, ICA generates two outputs: a spatial map showing the brain regions linked to the signal and a time series that describes the signal's evolution over time, as shown in Figure 7.

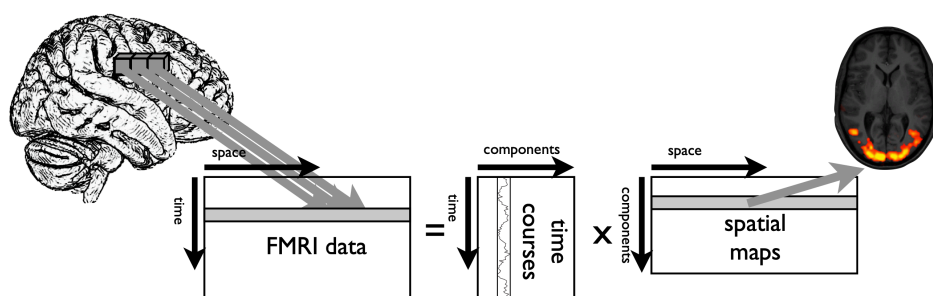


Figure 7: Illustration of ICA decomposition. Preprocessed fMRI data is separated into multiple components, each consisting of a time course and a corresponding spatial map. Obtained from [78].

This technique is widely used in group-level fMRI analysis to identify resting-state networks (RSNs), such as the default mode network (DMN) and dorsal attention networks, as well as cognitive, motor, and sensory networks, which are very useful in FC studies. RSNs were first defined by Smith et al. in 2009 [79], who discovered correspondences between task-based networks (networks active during specific tasks) and RSNs. This breakthrough has opened the door for numerous studies into FC. Even if performed at the individual level or with a large group, ICA is able to extract common networks that exhibit consistent patterns of activity across subjects. The main RSNs defined by Smith et al. [79] can be seen in Figure 8.

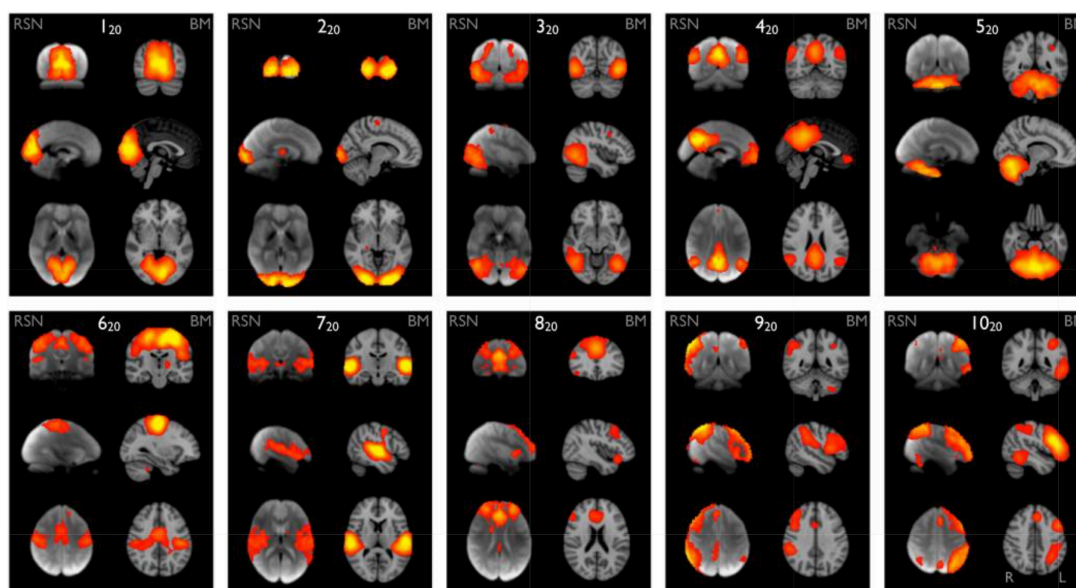


Figure 8: Ten pairs of networks from task-based BrainMap activation data (right) and rs-fMRI data (left), identified via ICA. Main RSNs like visual, DMN, frontoparietal can be seen. Obtained from [79].

ICA has some drawbacks when applied to fMRI data. One major issue is overfitting, as the algorithm cannot inherently distinguish between true signals and noise, nor determine the actual number of meaningful components in the data. When too many components are extracted, true signals may be split into smaller, fragmented components, complicating interpretation. Conversely, extracting too few components can overlook important signals, reducing the analysis's explanatory power, as shown in Figure 9. Determining the optimal number of components is therefore critical but challenging [78]. Another limitation is the interpretation of spatial maps, which assign values to each voxel based on its contribution to a component. Setting an appropriate threshold for visualization introduces subjectivity and variability into the analysis.

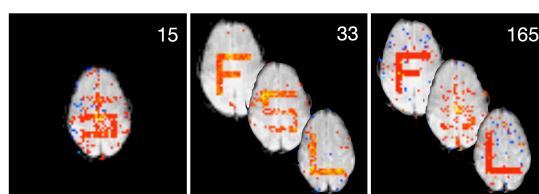


Figure 9: Illustration of the component selection problem in ICA. The left panel (15 components) shows underfitting, with signals mixed and poorly resolved. It only shows one network because there is a small number of components. The middle panel (33 components) represents optimal fitting, with distinct components. The right panel (165 components) demonstrates overfitting, where signals are fragmented, complicating interpretation. Obtained from [78].

4.4.1 Dual regression

The output of ICA analysis is a set of RSNs, or IC, representing patterns common to all subjects at the group level. However, ICA does not indicate individual contributions to these networks. This can be addressed using Dual Regression, which quantifies how strongly each group-level network is present in a subject or how much an individual contributed to the group-level network.

Dual Regression is a two-step process, as illustrated in Figure 10. In the first stage, the cleaned fMRI data from a subject and the group-level ICA spatial maps are combined to extract an individual time series for each component. This time series reflects how the group-level network is represented in the subject's data over time, enabling a detailed understanding of individual contributions. In the second stage, these subject-specific time series are used as regressors to generate spatial maps unique to the subject, illustrating how group-level patterns are distributed within the subject's brain. Thus, stage 1 focuses on the temporal domain, while stage 2 emphasizes the spatial domain [12, 78].

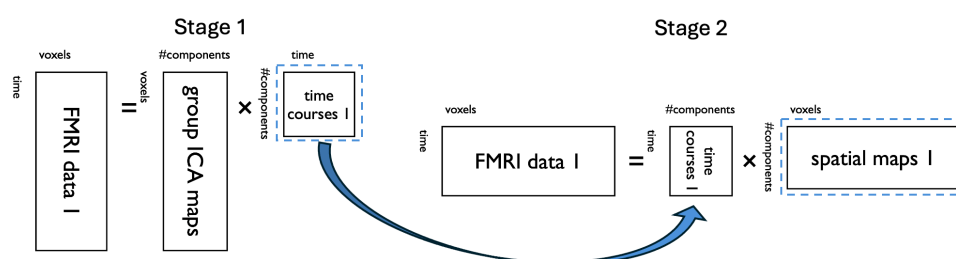


Figure 10: Dual regression schematic. Stage 1 extracts individual time courses, and Stage 2 generates individual spatial maps. The two outputs are highlighted in blue. Adapted from [78].

This two-stage process converts group-level ICA components into subject-specific versions, estimating each individual's contribution. The resulting spatial maps, while not directly used in this project, have potential for voxel-wise statistical analysis to explore spatial variability and the relationship between group-level networks and individual brain activity. In this study, only the time series output from stage 1 will be used, as the focus is on extracting connectivity measures between networks rather than analyzing structural changes within the networks. However, examining stage 2 outputs can serve as a form of quality assurance, as each subject's map should resemble to some degree the group-level network, as seen in Figure 11.

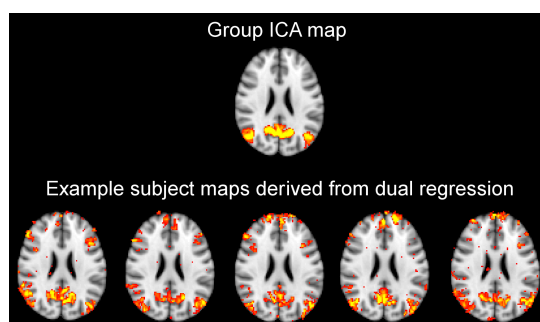


Figure 11: Stage 2 of the dual regression analysis displays the ICA-derived DMN above, with subject-specific contributions below, representing individual patterns within the group DMN. Obtained from [78]

4.4.2 Network Analysis

The human brain can be understood as a complex network of spatially distributed but functionally interconnected regions that continuously share information [80]. There are several types of brain connectivity, each capturing different aspects of the brain's organization and communication. Structural connectivity refers to the physical wiring of the brain, defined by WM tracts that can be mapped using diffusion MRI [81]. Other types, like effective connectivity, capture the directional influence one brain region exerts over another, enabling the modeling of causal interactions [82].

In the context of fMRI, functional connectivity is a key focus of study. It is defined as the temporal dependency of neuronal activation patterns between anatomically distinct brain regions [80]. FC provides information into how brain regions interact and coordinate activity, revealing intrinsic network dynamics that support cognition and behavior.

After identifying RSNs with ICA and subject-specific contributions using dual regression, network analysis can reveal individual network expression and interactions. Additionally, connectivity features derived from these analyses can be integrated with other data types, such as behavioral or clinical measures, to study differences across populations. These integrative analyses could be key to identify functional imaging biomarkers that improve our understanding of neurological conditions [46].

There are various methods to extract parameters that describe FC from fMRI data, all aimed at assessing relationships between predefined brain regions. Each brain region, or node, is represented by a timecourse signal, which is used to measure correlations and interactions with other regions. Full and partial correlation are widely used methods because they are reliable and computationally efficient. These techniques strike a balance between simplicity and effectiveness, making them suitable for large-scale datasets. While more advanced techniques can model connectivity in greater detail [82], their comparison is beyond the scope of this project.

Full correlation calculates Pearson's correlation coefficient to measure the similarity between Blood Oxygen Level-Dependent (BOLD) time series. This method is simple and computationally efficient but has a key limitation: it captures both direct and indirect connections. For example, if regions A and C are connected only through region B, full correlation may still suggest a direct link between A and C, making it difficult to distinguish direct from indirect relationships, as shown in Figure 12. Additionally, full correlation is sensitive to noise, which can reduce its reliability if preprocessing is not thorough.

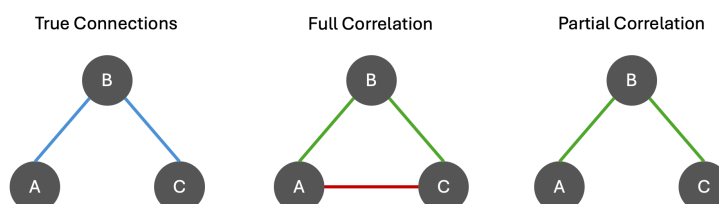


Figure 12: Comparison of full and partial correlation. Full correlation detects both direct and indirect connections (A and C through B), while partial correlation identifies only direct connections. Adapted from [12].

On the other hand, partial correlation focuses on direct connections by regressing out the influence of other regions before calculating correlations [82]. For example, it removes the effect of region B when

analyzing the connection between A and C, as shown in Figure 12, ensuring that only direct relationships are detected, with near-zero correlation if no direct connection exists. Tested on simulated fMRI data, partial correlation has shown high sensitivity for detecting network connections [82]. However, with short timeseries, it can yield unreliable results due to reduced degrees of freedom [12]. This issue is less problematic with longer timeseries, as is the case here.

There is no clear consensus in the scientific community regarding the preference for full or partial correlation. Therefore, both methods will be employed in this study, as their implementation is straightforward and allows for a more comprehensive analysis. For each pair of networks or nodes, a single feature is generated. This information is represented in connectivity matrices, which are symmetric because the connections are not causal and the diagonal values are zero since a region cannot have a connection with itself. This means that for N nodes, the total number of unique features is calculated using the formula $N \times (N - 1) / 2$, which represents all possible connections between pairs of nodes, as illustrated in Figure 13. These features serve as potential markers for studying brain networks by capturing patterns of FC. In the following analyses, these biomarkers will be used to identify patterns linked to the clinical and cognitive variables of the subjects.

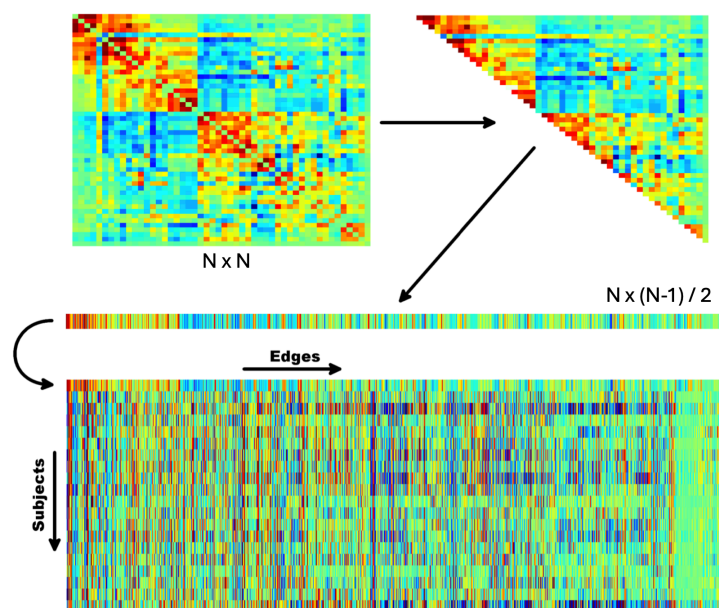


Figure 13: A network connectivity matrix from a single subject is transformed into a single line by removing the redundant entries from the symmetric matrix. These lines, representing edges, are then stacked across subjects to form a matrix prepared for group-level analysis. Adapted from [12].

5 Detailed Engineering

5.1 Data Acquisition and Structuring

Building on the previous methodology, this section applies the framework to a specific dataset from the A4 study, accessed through the Image and Data Archive [83]. This dataset includes three participant groups, all of whom underwent PET imaging to evaluate amyloid deposition: AmyloidE, AmyloidNE, and LEARN AmyloidNE:

1. **AmyloidE:** Participants with elevated amyloid levels (E) on Florbetapir PET imaging (N=1,323). Most had MRI (N=1,261) and fMRI (N=1,178), and a subset (N=392) underwent Flortaucipir PET imaging.
2. **LEARN AmyloidNE:** Participants with non-elevated amyloid levels (NE) screened for the LEARN (Longitudinal Evaluation of Amyloid Risk and Neurodegeneration) study (N=567). Most had MRI and fMRI data (N=496), and some (N=55) underwent Flortaucipir PET imaging.
3. **AmyloidNE:** Participants with non-elevated amyloid levels (N=2,596) who were not part of LEARN. They only underwent cognitive assessments and Florbetapir PET imaging.

This project focused on the AmyloidE group, consisting of participants with elevated amyloid levels. To perform a functional connectivity analysis, the required data includes structural T1w MRI and rs-fMRI. Additionally, Florbetapir PET imaging was necessary to link FC features to amyloid deposition. Of the initial 1,323 subjects in this group, only 1,178 met the criteria of having complete imaging datasets for the objectives of this project.

Characteristic	Value
Age - yr.	72.0 \pm 4.8
Female - no. (%)	691 (58.7%)
Education - yr.	16.6 \pm 2.8
Race White - no. (%)	1,104 (93.7%)
Family History Dementia: parent or sibling - no. (%)	882 (74.9%)
ApoE ϵ 4 carrier - no. (%)	681 (57.8%)
A β SUVR (18F-Florbetapir)	1.3 \pm 0.2
Centiloids (18F-Florbetapir)	65.8 \pm 33.0
PACC score	-0.4 \pm 2.7
Wechsler Memory Scale LMDR score	11.4 \pm 3.3
MMSE score	28.7 \pm 1.3
CFI combined score	3.8 \pm 2.9
CDR-SB score	0.1 \pm 0.2

Table 1: Descriptive statistics of the Amyloid-E group that included participants who underwent all required imaging modalities (N=1,178). Values are expressed as mean \pm standard deviation for continuous variables, and as count (percentage) for categorical variables. Cognitive test details and scoring methods are provided in the Annex section 12.2.

The participants in the dataset, aged 65 to 85 years, lived independently and demonstrated high cognitive abilities despite having elevated amyloid deposition. As explained in Section 2.1.3, the Centiloid scale, which quantifies amyloid buildup in the brain using Florbetapir PET imaging, showed a mean score of 65.8 with a range of approximately 35 to 200 among the participants. Cognitive assessments confirmed their eligibility, with no diagnosis of mild cognitive impairment or dementia. Key metrics included a mean Clinical Dementia Rating (CDR) score of 0, a Mini-Mental State Examination (MMSE) score between 25 and 30, along with other cognitive evaluations. Refer to Table 1 for detailed statistics of the study group. An explanation of the cognitive tests is provided in Annex Section 12.2.

After downloading the Florbetapir PET scans, structural MRI (T1w, T2w and others), rs-fMRI, and their respective metadata for 1,178 subjects, the data was organized into a BIDS-compliant structure, as previously outlined in Section 4.1. Managing such a large dataset with multiple imaging modalities required careful organization to ensure consistency and usability.

To streamline this process, Python scripts were used to automate the data structuring. The final output included 1,178 folders, one per subject, with subfolders named *anat* (anatomical), *func* (functional), and *pet* (PET imaging) modalities. Each subfolder contained imaging data in NIfTI format along with JSON metadata files. The total dataset size exceeded 125 GB. The structured dataset in BIDS format is shown in Figure 14.

Filename	Size
CHANGES	796 B
dataset_description.json	273 B
participants.json	3.6 KB
participants.tsv	553.9 KB
README.md	2.5 KB
sub-B10081264	--
ses-01	--
anat	--
sub-B10081264_ses-01_FLAIR.json	1.4 KB
sub-B10081264_ses-01_FLAIR.nii.gz	5.2 MB
sub-B10081264_ses-01_T1w.json	1.4 KB
sub-B10081264_ses-01_T1w.nii.gz	37.4 MB
sub-B10081264_ses-01_T2star.json	1.4 KB
sub-B10081264_ses-01_T2star.nii.gz	2.7 MB
sub-B10081264_ses-01_T2w.json	1.4 KB
sub-B10081264_ses-01_T2w.nii.gz	4.6 MB
func	--
sub-B10081264_ses-01_task-rest_bold.json	1.4 KB
sub-B10081264_ses-01_task-rest_bold.nii.gz	39.9 MB
pet	--
sub-B10081264_ses-01_trc-fbp_pet.json	2.0 KB
sub-B10081264_ses-01_trc-fbp_pet.nii.gz	6.6 MB
sub-B10102783	--
sub-B10108368	--
sub-B10117310	--
sub-B10350512	--

Figure 14: Overview of the AmyloidE dataset into BIDS format, structured by subject, session, and modality.

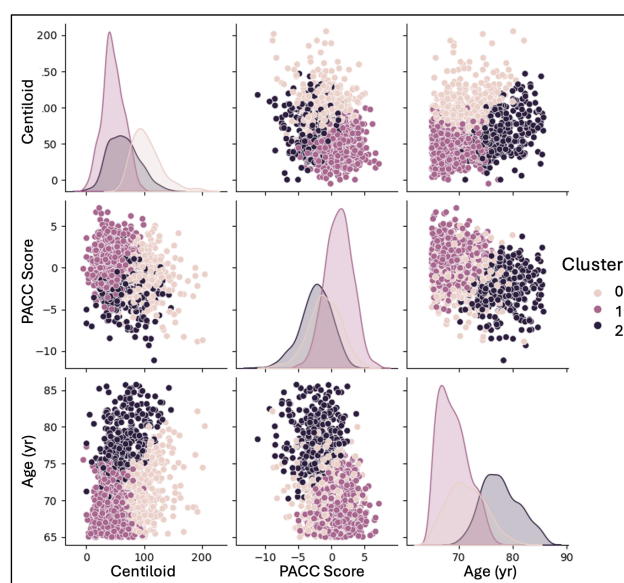


Figure 15: Distribution of Centiloid values, PACC Cognitive scores, and Age across the three identified cluster groups.

As discussed in Section 2.1.4, this dataset lacks distinct groups of subjects, as all participants are in the preclinical phase of AD, with no clear healthy control or fully diseased groups. To better characterize the dataset, a clustering technique was applied to classify subjects into three distinct groups (optimal number of clusters was determined using the elbow method). This was achieved using the K-means algorithm, with Centiloid values, age, and PACC scores (a composite cognitive score) as the clustering variables. This approach enabled the creation of three subject groups, which were subsequently used for further analysis. The resulting clusters are shown in Figure 15.

5.2 Preprocessing Pipeline

As previously mentioned, preprocessing of rs-fMRI data is a critical step to ensure accurate and reliable subsequent analysis. This process minimizes unwanted artifacts and noise while standardizing the data for downstream applications.

For this study, Python scripts originally developed by the BIG group were adapted. These scripts utilized various open-source tools, including FSL, AFNI, and ANTs, among others. All the code used is available in a GitHub repository, described in the Annex Section 12.1.

This preprocessing pipeline included all steps detailed in Section 4.2, with the exception of spatial smoothing. Given the relatively low resolution of 3.3 mm³ in our functional data, smoothing would have further degraded the resolution, potentially leading to the loss of critical information. The proposed preprocessing steps can be seen in Figure 16. Although QC uses several intermediate images, only the final cleaned functional image was used for the subsequent analyses.

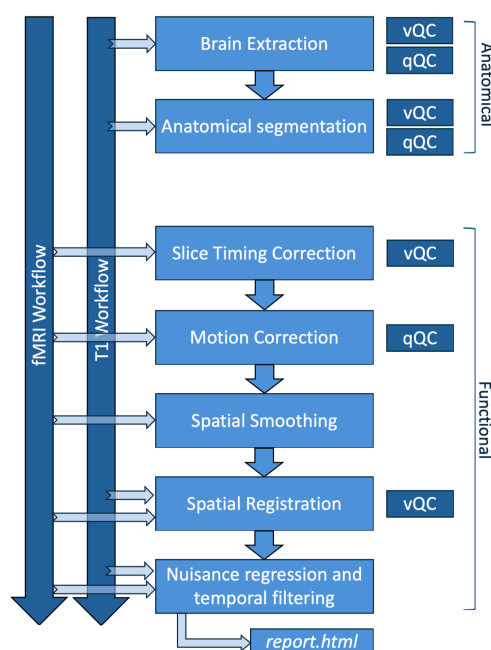


Figure 16: Diagram of preprocessing pipeline steps, including vQC and qQC checkpoints. Arrows on the left show the input images for each step. In this project, no spatial smoothing was performed.

While preprocessing might appear straightforward, it was a time-intensive process. With the available computational resources, each subject required approximately 12 minutes of processing time. With the large dataset, preprocessing 1,178 subjects required 10 days of continuous computation and the final size of the different preprocessing outputs was 480 GB. The output size exceeded that of the original data due to the incorporation of various intermediate images saved for QC purposes. In comparison, the standard fMRIPrep software averages 8 hours per subject, making our approach significantly faster. This highlights the need for an efficient pipeline that ensures reliability and scalability when working with large datasets while minimizing computational demands. Although QC uses several intermediate images, the final cleaned functional image was the only one used for the subsequent analyses.

5.3 Quality Control

As already mentioned, quality control is crucial for preventing errors in automated preprocessing pipelines. While tools like fMRIPrep and applications like SIMEXP facilitate preprocessing review [32], we developed a custom application adapted to our specific pipeline environment. This tool allowed for visual QC (vQC) and quantitative QC (qQC) and multiple reviewers to collaborate effectively, ensuring thorough evaluation.

5.3.1 Development of the QC Application

The first step in developing this application was designing a way to display each subject's preprocessing output. Python scripts were created to generate QC HTML reports for fMRI data, featuring visual representations of key preprocessing steps, such as anatomical segmentations, functional overlays, and registration to templates. Relevant images and animations were saved as PNG or GIF files, offering an intuitive and detailed summary of the data's processing status. An example HTML report can be seen in the Annex Section 12.3 in Figure 33.

In addition to visual outputs, the script calculated segmentation and motion correction metrics, which were assessed against predefined thresholds. These thresholds were set at more than 3 standard deviations from the mean to detect anomalies. Metrics outside this range triggered warning messages, indicating excessive movement during the scan, abnormal brain sizes, or poor segmentation of GM, WM, or CSF. Both visual and quantitative evaluations were combined into the final HTML report for comprehensive quality assessment, as illustrated in Figure 16.

Once the reports were created, the application was developed using Flask, a Python web library. This enabled the HTML reports to be viewed in a web browser, with three buttons (Yes, No, and Maybe) for making decisions, as seen in Figure 17. This approach significantly streamlined the QC process. Subjects previously rejected for quantitative reasons, such as poor segmentation or excessive head movement during scans, did not require further visual review. Additionally, the application allowed multiple members of the BIG group to connect remotely enabling them to contribute to QC decisions conveniently whenever they had available time. All the code developed for the application and reports is available in the GitHub repository, described in the Annex Section 12.1.

5.3.2 QC Stages and Errors

This next section outlines the stages of vQC, highlighting the common errors encountered and the specific features we evaluated to ensure high-quality processing.

Brain Extraction: This stage involved removing non-brain tissues, such as the skull, from brain MRI scans using the HD-BET algorithm. Common errors included residual neck tissue remaining after extraction or the improper removal of brain tissue, both of which could affect subsequent analyses. Visual evaluation focused on verifying the boundaries of the extracted brain mask and ensuring all non-brain tissues were correctly removed, as shown in Figure 18. Outliers in total brain volume voxel counts were also identified using qQC.

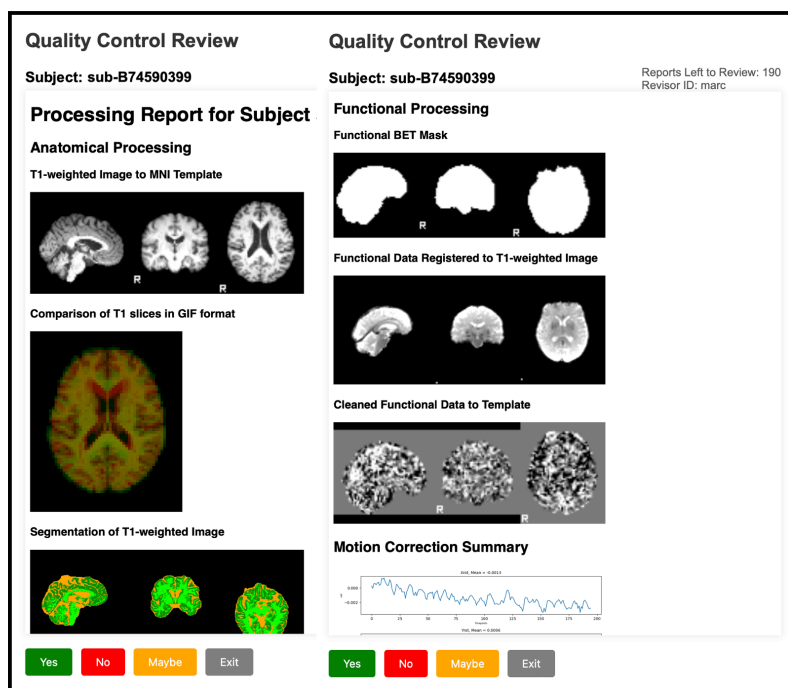


Figure 17: Screenshots of the QC application displayed in a web browser. Visual checkpoints for both anatomical and functional processing are shown, along with buttons for the reviewer's decision.

Anatomical Segmentation: The segmentation process identified WM, GM and CSF using FSL's FAST algorithm. Common errors included incorrect tissue classifications or incomplete separations. While vQC was used to verify that segmentation masks aligned with anatomical structures (see Figure 18), most errors were effectively identified through qQC. Subjects with an abnormal number of voxels—either too many or too few (beyond 3 standard deviations)—in specific tissues were automatically excluded.



Figure 18: (Left) Brain mask output used to identify errors in the brain extraction process, highlighting the areas of the brain. (Right) Brain segmented into three channels: WM (light green), GM (dark green), and CSF (orange), visually identifying any segmentation errors. The "R" indicates the right hemisphere.

Slice Timing Correction: This step corrected slice acquisition timing using AFNI's 3dTshift. Typical errors involved artifacts resembling scratches or distortions in the corrected images. vQC focused on confirming the absence of such artifacts and ensuring proper alignment across slices. Proper correction is shown in Figure 19, as seen in the transition from (a) to (b).

Some visually detected artifacts, as seen in the case of (c, d) in Figure 19, were not due to improper slice-timing correction but originated from the raw images themselves. These artifacts, often appearing as clear, periodic lines, were likely caused by issues during image acquisition. To investigate these anomalies, the 3D Fourier Transform of a fMRI volume was analyzed, revealing symmetrical peaks indicating periodic noise. Further analyses were conducted to develop methods for systematically detecting these peaks in Fourier space.

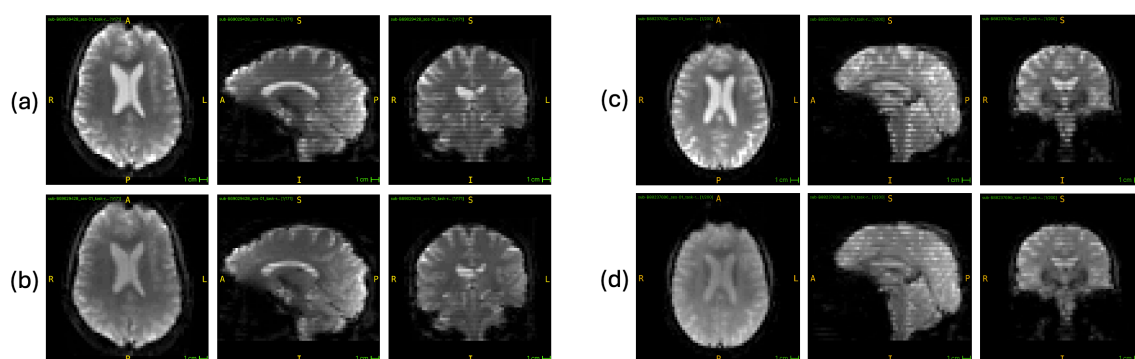


Figure 19: Comparison of raw (a, c) and slice-timing corrected (b, d) fMRI volumes. (a, b) are from one subject, while (c, d) are from another. Slice-timing correction improves alignment in (a, b) but leaves acquisition artifacts unresolved in (c, d).

The goal is to develop automated methods for detecting and correcting these artifacts while preserving image integrity and quality. This would reduce the need for subjective visual assessments and improve reproducibility. However, due to time constraints, this work is still under development by the BIG group.

Motion Correction: This step corrected small head movements using rigid-body transformations. Still, excessive motion could affect data reliability, and had to be excluded. Automatic QC flagged excessive movement based on calculations performed by the MCFLIRT algorithm, which quantified motion between consecutive time points in terms of translational changes (x, y, z) and rotational changes (radians). From this information, framewise displacement (FD) was computed as described in literature [84]. Mean FD and maximum FD were used as quantitative indicators for thresholding. In the Annex Section 12.3, Figure 32 illustrates an example of head motion parameters (translation, rotation, and FD) and their variation over time during the fMRI scan.

Spatial Registration: Spatial registration aligned fMRI data to a T1w image and subsequently to a standard template (MNI-152) using ANTs tools. To ensure accurate alignment and proper resampling, vQC was performed using a GIF visualization. This GIF uses red and green overlays to represent the template and registered images, respectively, allowing easy identification of misalignments or distortions. Misalignment is evident through color mismatches or inconsistencies in the overlay, offering a clear assessment of registration quality. An example is shown in Figure 20.

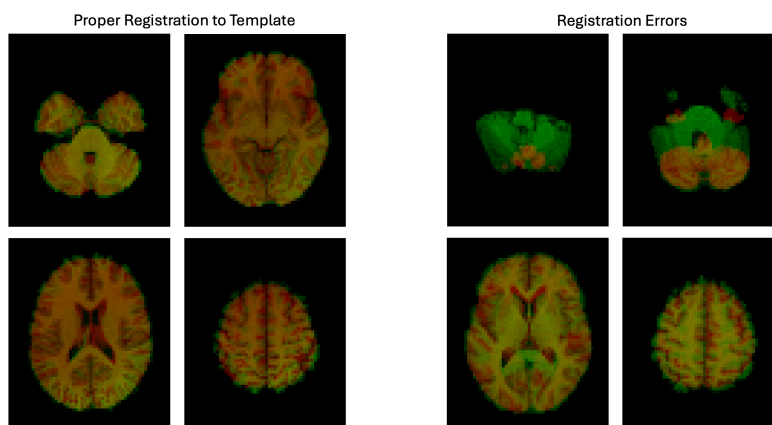


Figure 20: Comparison of registration quality. The images on the left show proper registration to the template, while the images on the right highlight registration errors caused by incomplete brain extraction, with residual neck structures affecting results.

Nuisance regression and temporal filtering: This step removed confounding signals from WM and CSF, while applying a frequency filter (0.01–0.1 Hz) to clean the data. A common error is the unintended removal of meaningful BOLD signal, which can compromise data integrity [77]. Since QC cannot directly evaluate these effects at this stage, it is essential to carefully set appropriate parameters (such as well-defined bandpass filter cutoffs in the literature [12]) to minimize the risk of signal loss.

5.3.3 QC Results

Although having a streamlined application significantly improved efficiency, reviewing a large database remained a time-consuming process. To recall from previous steps, we preprocessed 1,178 subjects. Of these, 65 were automatically rejected during quantitative QC for significantly deviating from the defined metric averages. This accounted for approximately 5%, which we consider an acceptable rate. Most automatic rejections were attributed to excessive motion during scans, while a smaller portion was due to segmentation issues. For detailed criteria regarding automatic rejections, refer to the Annex Section 12.3.

After automatic rejections, the remaining 1,113 subjects were reviewed visually by three researchers from the BIG group, including the author of this project. Since the process was highly time-consuming, not all reviewers evaluated the same subjects. Specifically, 471 subjects were reviewed by only one reviewer, 460 were reviewed by two reviewers, and 182 were reviewed by all three. The overlap in reviews allowed us to identify differences in evaluation styles and ensure consistency in final decisions. The results revealed distinct differences, as shown in Table 2 below.

Reviewer	Total Subjects Reviewed	Yes (%)	No (%)	Maybe (%)
Reviewer 1	881	684 (77.6%)	38 (4.3%)	159 (18.0%)
Reviewer 2	672	608 (90.5%)	12 (1.8%)	52 (7.7%)
Reviewer 3	384	212 (55.2%)	27 (7.0%)	145 (37.8%)

Table 2: Summary of subjects reviewed and their classification by each reviewer.

Reviewer 2 demonstrated a higher approval rate (90.5% Yes) compared to Reviewer 1 (77.6%) and Reviewer 3 (55.2%). Reviewer 3 also marked the largest proportion of subjects as "Maybe" (37.8%), indicating stricter criteria or more uncertainty in their assessments. Reviewer 1's results represented a middle-ground approach, with a balanced distribution across categories. These differences highlight the inherently subjective nature of vQC, as reviewers often prioritize different aspects, such as anatomical registration, segmentation quality, or the presence of functional artifacts.

To reach a final decision for each subject, a rule-based system was implemented. For cases with all three reviewers' inputs, a majority vote determined the outcome: two or more "Yes" resulted in a "Yes," and similarly for "No," while all other cases defaulted to "Maybe." When only two decisions were available, specific combinations resolved ties (e.g., "Yes" and "Maybe" defaulted to "Yes," while "No" and "Maybe" defaulted to "No"). If only one decision was available, it directly determined the outcome. This method ensured consistency while incorporating inputs from all reviewers. The final results, presented in Table 3,

show that 898 out of the initial 1,178 subjects (76.2%) passed the QC process and were therefore included in all subsequent analyses. For more detailed statistics and information into the vQC outcomes, please refer to the Annex Section [12.3](#).

Total Subjects	Yes (%)	No (%)	Maybe (%)
1,113	898 (80.7%)	35 (3.1%)	180 (16.2%)

Table 3: Overall summary of the visually reviewed subjects and their classification.

In conclusion, while qQC effectively reduces time and enhances consistency, vQC remains subjective and open to interpretation. Reviewers often differ in their thresholds and priorities, leading to variability in outcomes. By overlapping subject reviews among multiple evaluators, we were able to reduce individual biases and ensure a more reliable and robust QC process. To further address this issue, training reviewers with a subset of data before the full QC could ensure a shared understanding of the requirements and improve consistency.

This highly subjective nature of vQC suggests that we should gradually transition toward a more quantitative approach to enhance objectivity and reproducibility. However, it is important to recognize that some level of visual review will always be necessary, as certain aspects of data quality cannot be fully captured through quantitative metrics alone. Additionally, qQC can introduce bias by disproportionately excluding participants based on motion-related artifacts, potentially leading to skewed findings [\[85\]](#). Striking the right balance between vQC and qQC is essential for achieving both efficiency and reliability.

5.4 Group Independent Component Analysis

After completing the QC review of all subjects, the next step was to perform an ICA group decomposition. Only the subjects that passed the QC were included, excluding those marked as "No" or "Maybe". This resulted in a total of 898 subjects. The ICA analysis was conducted using FSL's MELODIC [29], a tool integrated into the FSL package that can be run efficiently via a simple command-line interface. The input for the MELODIC function consisted of the final preprocessed fMRI images.

Key parameters were carefully configured to optimize the analysis. First, the BET (brain extraction tool) function was disabled since this step had already been performed during preprocessing. Next, a background threshold of zero was applied to isolate the signal from the noise. Finally, the dimensionality, a critical parameter determining the number of components for decomposition, was set to 30. While projects within the BIG group have tested different numbers of components, this process was beyond the scope of this project, and a fixed, commonly agreed value of 30 was chosen to ensure reproducibility and avoid reliance on the software's automatic selection.

The main issue with running the ICA command on all 898 subjects was the algorithm's failure to decompose the data into meaningful networks. Despite attempts to troubleshoot and consult the community, the problem remained unresolved. However, the algorithm worked correctly and managed to extract meaningful networks on smaller subsets, such as 200 subjects. Due to time constraints, we opted to explore alternative approaches instead of processing all 898 subjects at once.

One alternative was to divide the 898 subjects into five smaller groups, run ICA separately for each group, and then merge the corresponding components. While this approach enabled the algorithm to generate identifiable networks, it introduced another challenge: the networks were ordered differently in each run, making alignment and merging challenging. For example, some networks appeared similar but were not identical across runs, complicating their pairing. Methods like cross-correlation or DICE similarity indices were considered for alignment, but these approaches risked introducing bias and arbitrary decisions, potentially compromising the study's objectivity.

A second, simpler solution (and ultimately chosen) was to randomly select a subset of 300 subjects from the 898. This approach aimed to create a manageable yet representative sample, allowing ICA to be run effectively with FSL's MELODIC. While fewer subjects were analyzed, the subset was randomly chosen to avoid biases. This method was preferred for its simplicity, reduced computational demands, and avoidance of the complexities involved in aligning networks across multiple ICA runs, ensuring the analysis remained both robust and unbiased.

The output of this analysis is a file named *melodic_IC.nii.gz*, which contains the 30 components selected, with each network represented as a separate volume. After review, 13 networks were identified as noise and discarded. The remaining 17 networks were visually inspected and categorized into functional groups, including visual, default mode network (DMN), executive, parietal, and other networks, as shown in Figure 21.

Using the FSL Nets Python library, we performed hierarchical cluster analysis to organize our 17 networks based on subject-specific time series. For details, refer to Figure 35 in the Annex Section 12.4.

Combining this cluster ordering with visual identification, we assigned network names to each independent component (IC) as follows: IC₁: Visual 2, IC₂: DMN, IC₃: Visual 1, IC₄: DMN Frontal, IC₅: Sensory Motor, IC₆: Executive 2, IC₇: Left Fronto-Parietal, IC₈: Right Fronto-Parietal, IC₉: Fronto-Parietal, IC₁₀: Sensory Temporal, IC₁₁: Visual 3, IC₁₂: Salience, IC₁₃: DMN Posterior 1, IC₁₄: DMN Posterior 2, IC₁₅: Executive 3, IC₁₆: Frontal, IC₁₇: Executive 1.

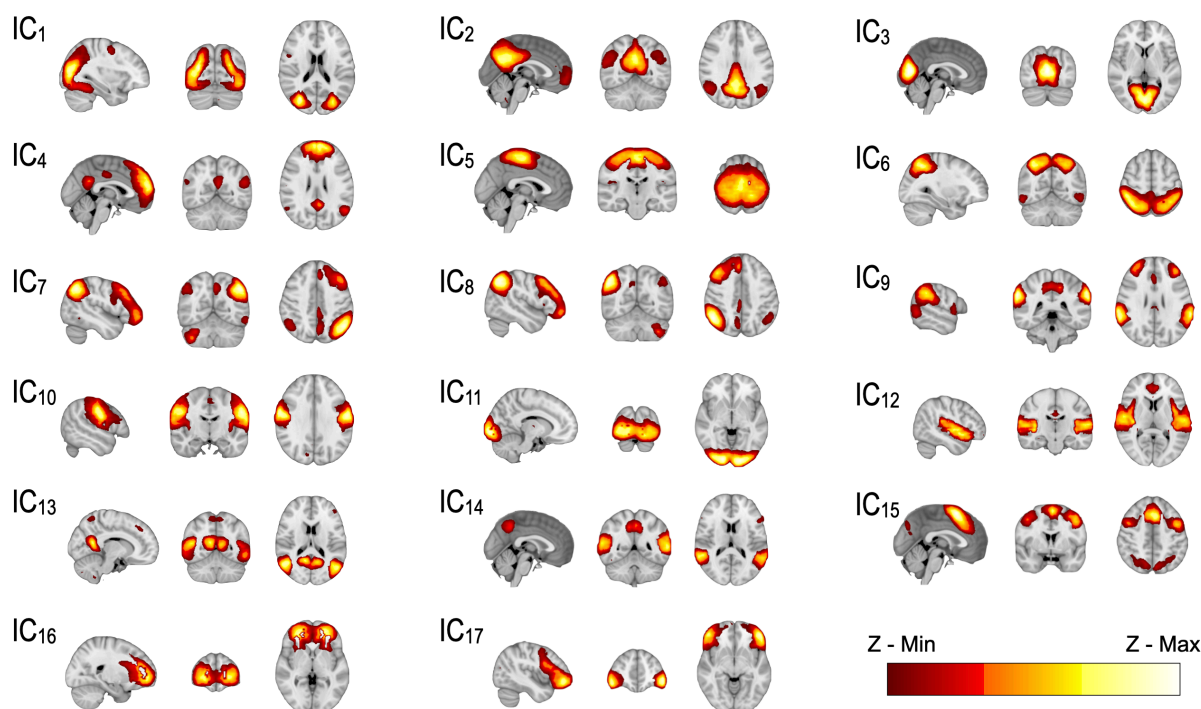


Figure 21: Resting-State Networks extracted from the ICA analysis. Each network was visually inspected to exclude components associated with noise. The color scale represents Z-values, scaled for each network based on its signal-to-noise ratio, as explained in [86].

Figure 22 shows examples of components identified as noise and excluded from analysis. Types of noise include: (a) edge effects from motion or reconstruction artifacts at the brain boundaries, (b) signal from CSF, which is irrelevant to the analysis, (c) noise in the frontal regions due to field inhomogeneities, and (d) artifacts introduced during motion correction or from raw images. Excluding such noise is crucial for maintaining the integrity of subsequent analyses.

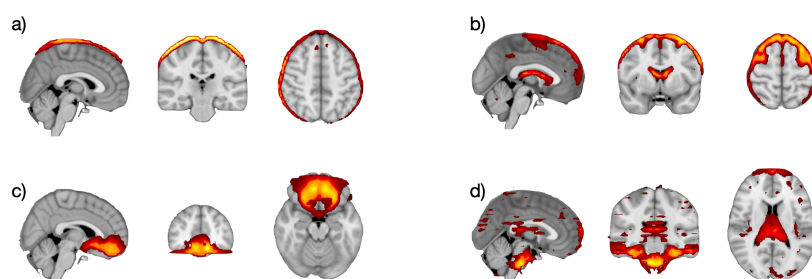


Figure 22: Examples of rejected ICA components classified as noise. (a) Edge effects caused by motion or reconstruction artifacts near the boundaries of the brain. (b) Signal originating from CSF, which is not relevant to the analysis. (c) Noise in the frontal regions, commonly attributed to field inhomogeneities observed in fMRI analyses. (d) Artifacts potentially introduced during motion correction or from raw images.

5.5 Dual Regression

The next step involved performing dual regression, where each group ICA network was mapped onto individual data to extract subject-specific spatial maps and associated time series. This process used FSL's dual regression command, taking as inputs the preprocessed functional images, ICA maps, and a brain mask to reduce the influence of noise on the results.

Dual regression involved two stages: Stage 1 extracted the time series, while Stage 2 generated individual spatial maps. For the analyses in this project, only the time series from Stage 1 were utilized, as Stage 2 outputs did not align with the project's objectives. An example of outputs from both stages is shown in Figure 23.

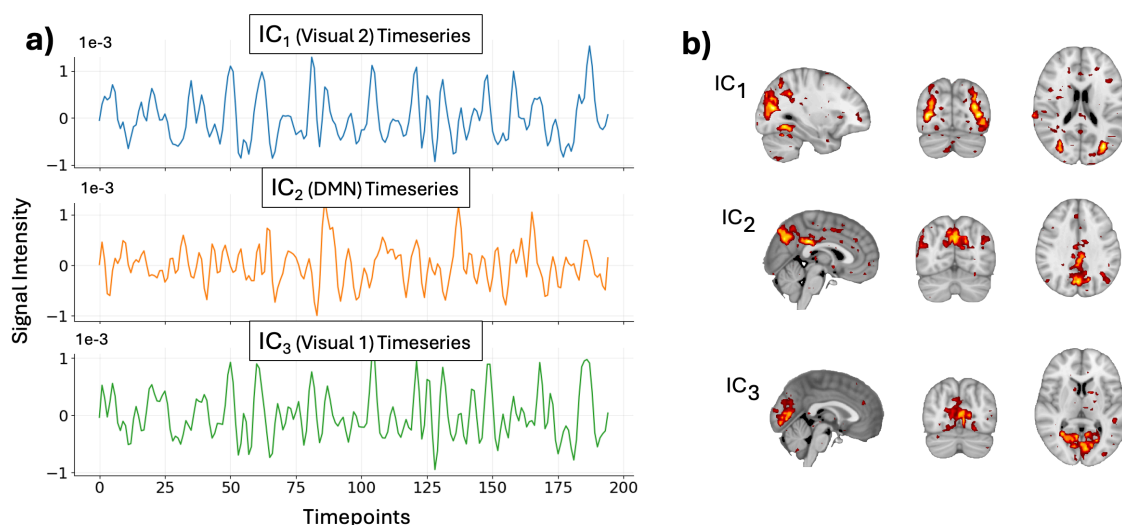


Figure 23: a) Dual regression stage 1 output timeseries for a single subject, showing the temporal fluctuations of the first three components across fMRI time points. IC₁, IC₂, and IC₃ correspond to Visual 2, DMN, and Visual 1 components, respectively. b) Dual regression stage 2 spatial maps for the same subject, displaying the subject-specific spatial patterns associated with IC₁, IC₂, and IC₃. These spatial maps represent subject-specific patterns from dual regression and are not group-level maps.

5.6 Functional Connectivity Mapping and Analysis

At this stage of the study, the main RSNs were identified, and dual regression was used to calculate each individual's contribution to these networks. Functional connectivity features between regions could then be extracted from this data. The Nets library from FSL, integrated into Python, was used to calculate full and partial correlations between networks (or nodes). As we selected 17 RSNs, for each subject, two 17x17 network matrices were generated: one for full correlations (computed using the Pearson correlation coefficient) and one for partial correlations (derived via L2-norm Ridge Regression). As explained in Section 4.4.2, since the matrices are symmetrical and the diagonal is irrelevant, this resulted in 136 unique partial correlation features and 136 full correlation features per subject.

Full and partial correlations were analyzed to explore their relationships with clinical and demographic variables. These measures, capturing different aspects of FC, served as potential biomarkers for understanding brain connectivity in relation to clinically relevant features.

5.6.1 Interrogating the Models: Exploring correlations between FC and external variables

A General Linear Model (GLM) was used for this analysis, employing various contrasts to assess associations between FC metrics and subject characteristics. The GLM was implemented using FSL's randomise tool (non-parametric permutation test), running 5,000 iterations to ensure robust and reliable testing. The results included corrected p-values that accounted for multiple comparisons (using family-wise error rate correction via the maximum statistic approach [87]), providing a clear measure of the significance of the observed correlations.

Two key relationships were investigated in this analysis. First, FC variables were correlated with Centiloid values, which were treated as a continuous measure to characterize the amount of amyloid deposition. Centiloid values served as indicators for the stage of Alzheimer's disease, providing a way to assess the relationship between connectivity and proximity to advanced AD pathology. Second, FC variables were analyzed in relation to age, aiming to capture potential age-related changes in brain connectivity. This dual approach allowed for a comprehensive exploration of both disease progression and age as factors influencing FC.

It is important to acknowledge that while these analyses seek to identify associations between FC metrics and subject features, correlation does not imply causation. The observed relationships may result from shared underlying factors or complex interactions rather than direct causal effects. Therefore, caution is required when interpreting these findings, and further studies are needed to validate these associations.

5.6.2 Correlations between Centiloid and FC

There is growing evidence that alterations in FC are associated with AD, particularly in the DMN, one of the first regions affected by amyloid deposition [45, 46]. This is why it is interesting to explore the relationship between Centiloid values and FC parameters in the preclinical state of AD, to identify early changes that may contribute to disease progression.

To address this, a multivariable approach was used that included Centiloid values as explanatory variables and age and gender as variables of no interest. Gender was coded as a binary variable (0 for male, 1 for female), and all variables were demeaned to reduce collinearity. To isolate the effect of Centiloid on FC parameters, specific contrasts were applied within the GLM. Contrasts such as [1, 0, 0] and [-1, 0, 0] were applied to test for positive and negative associations between Centiloid and FC, while controlling for age and gender. Assigning a contrast of 0 to age and gender ensured their influence was excluded from the analysis.

The GLM results revealed only a single significant correlation between network connections and FC variables. While some associations appeared significant before correcting for multiple comparisons, they disappeared after applying such correction (using the maximum statistic approach). This outcome was unexpected, as previous studies have consistently reported clear changes in network structures and connectivity in AD [45, 88]. The only significant association identified was a negative correlation between the Centiloid scale and the partial correlation between a visual network (IC_1) and the DMN (IC_2), as shown in Figure 24. However, this finding is not conclusive, as variability in visual stimuli across subjects during the

scan might have influenced the results. Nevertheless, the use of a highly restrictive randomise test with 5,000 iterations and an almost zero p-value suggest that There may be genuine interactions between the DMN and visual networks influenced by the Centiloid scale. This aligns with previous findings suggesting altered DMN activity during visual tasks in AD [89], highlighting potential disruptions in DMN-visual connectivity as part of disease progression.

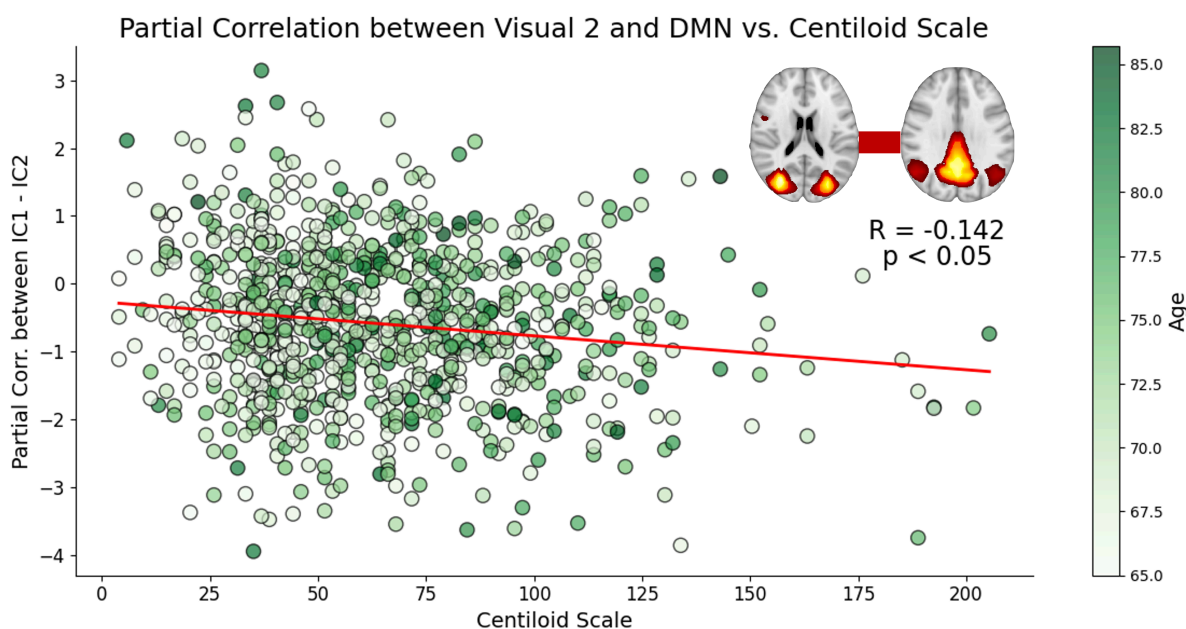


Figure 24: Scatter plot showing the partial correlation between Visual 2 and DMN versus the Centiloid scale, with age represented on the color scale. The analysis reveals a significant negative correlation ($R = -0.142$, $p < 0.05$), indicating that higher Centiloid values are associated with reduced FC between these regions.

Additionally, we explored an alternative analysis using the categorical cluster groups defined in Section 5.1 as explanatory variables, instead of relying on the continuous Centiloid scale. This cluster categorized subjects into three groups based on Centiloid, age, and a cognitive variable for differentiation. However, this approach did not yield significant results.

One possible explanation for these limited findings is the homogeneity of our preclinical subjects sample. At this early stage, connectivity changes with Centiloid may not yet be detectable. Including a broader spectrum of participants, such as healthy controls and individuals with more advanced stages of AD, could help reveal more pronounced differences. This hypothesis is supported by a recent study that used the A4 dataset [90], which identified significant connectivity correlations with amyloid deposition only after analyzing data collected at the conclusion of the clinical trial (not the screening data we are using). By that time, participants had become more differentiated in terms of disease progression, highlighting the importance of variability in capturing meaningful associations.

5.6.3 Correlations between Age and FC

Age-related changes in brain structure and function are well-documented [91]. To validate our preprocessing pipeline and QC processes, we analyzed patterns of brain connectivity associated with age and

compared our findings to established scientific literature. This approach ensured the reliability of our pipeline while demonstrating its ability to identify meaningful connectivity patterns in line with the existing research.

To explore the relationship between age and FC, a similar multivariable approach was used, including Centiloid and gender as variables of no interest. All variables were demeaned to reduce collinearity, and again, contrasts $[1, 0, 0]$ and $[-1, 0, 0]$ were used to assess positive and negative associations between age and FC while accounting for Centiloid and gender. Figure 25 shows the significant ($p < 0.05$) Pearson correlation coefficients between network connections and age, corrected for multiple comparisons (using family-wise error rate correction via the maximum statistic approach [87]). Both partial (upper triangle) and full (lower triangle) correlations are included, with positive correlations in red and negative correlations in blue, highlighting age-related changes in FC.

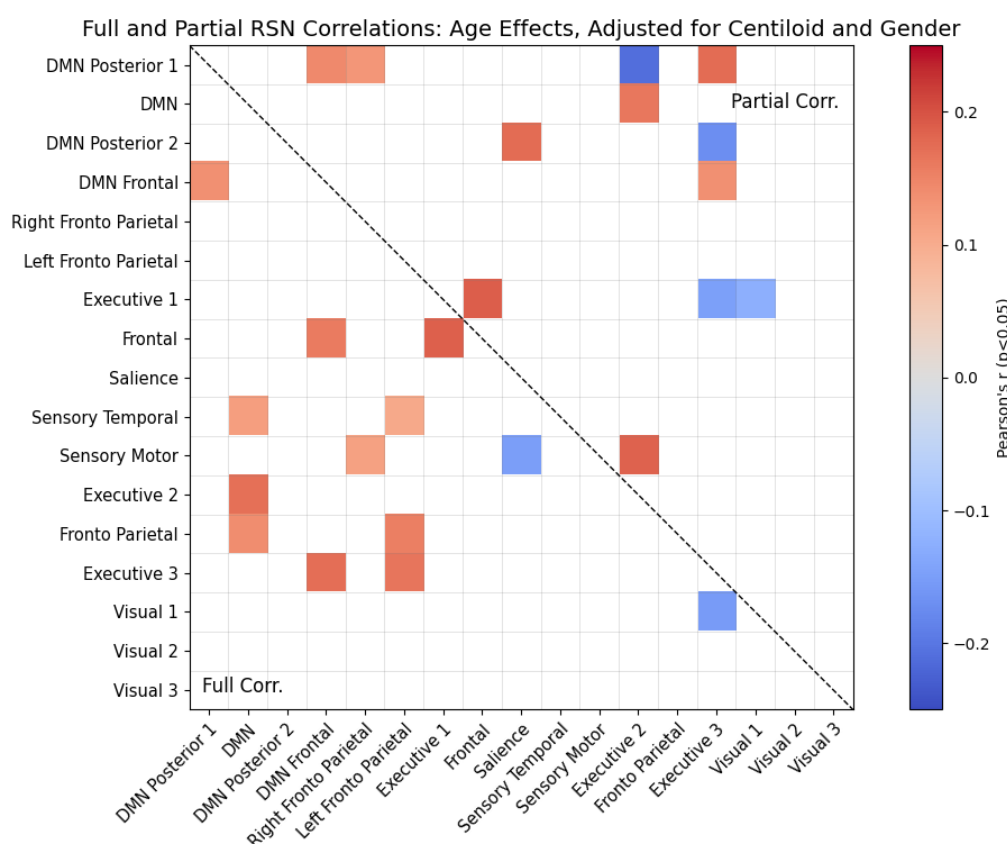


Figure 25: Connectivity matrix showing significant associations (corrected $p < 0.05$) between 17 RSNs and age, using full and partial correlations. Significant connections (corrected for multiple comparisons) are represented by the color scale (Pearson correlation coefficients), with red for positive and blue for negative correlations. Analysis was conducted using a GLM with randomise, accounting for Centiloid scale and gender effects. Dashed lines separate full and partial correlation results.

Figure 25 highlights a set of significant network pairs whose connectivity is either positively or negatively correlated with age. In other words, some networks show increased connectivity in older subjects, suggesting age-related strengthening of certain connections, while others display higher connectivity in younger individuals, indicating a decline with aging. For example, Figure 26 shows two instances where the connectivity between two networks increases with age.

The graph on the right of Figure 26 illustrates increased connectivity within the Frontal and Posterior

regions of the DMN. As both regions are sub-components of the DMN, this finding suggests intranetwork strengthening. In contrast, the graph on the left demonstrates increased connectivity between the DMN and the Executive Network, indicating internetwork strengthening.

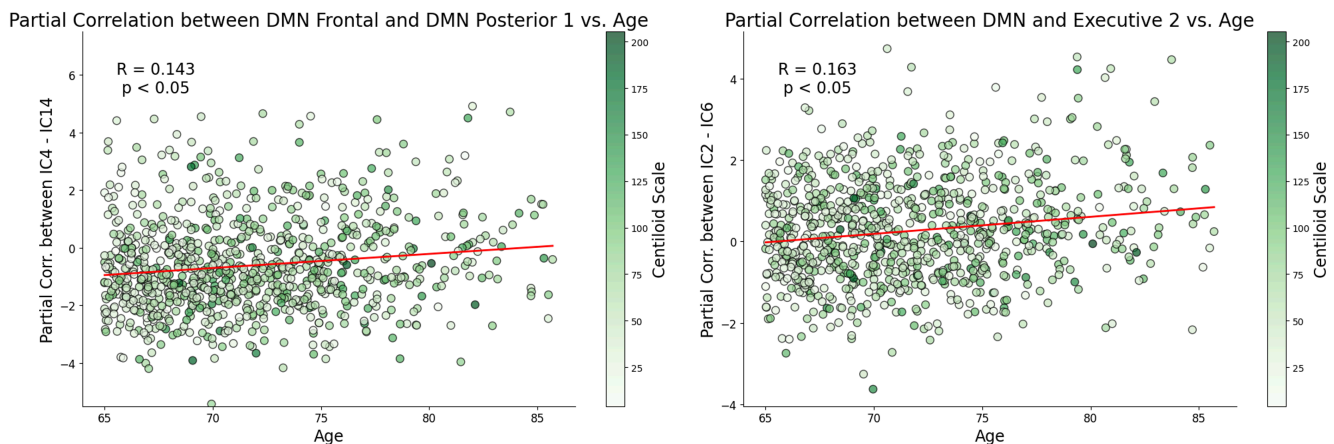


Figure 26: (Left) Scatter plot showing the partial correlation between DMN Frontal and Posterior 1 regions vs. age, with Centiloid represented by the color gradient. The plot indicates that intra-network connectivity within the DMN increases with age ($R = 0.143$, $p < 0.05$). (Right) Scatter plot showing the partial correlation between DMN and Executive 2 networks vs. age, with Centiloid scale represented by the color gradient. The plot demonstrates that inter-network connectivity between these networks increases with age ($R = 0.163$, $p < 0.05$).

These findings highlight both inter-network and intra-network connectivity changes, suggesting that brain connectivity adapts with aging. While some connections weaken, others strengthen through compensatory mechanisms. This pattern reflects the well-documented phenomenon of brain reorganization with age, demonstrating the brain's ability to adapt and maintain functionality [92]. Further analysis could be performed to explore these changes and reorganizations in greater detail.

Future studies could benefit from including additional covariates, such as cognitive scores or years of education, to better account for their potential influence on the results. This consideration is particularly relevant given the established theory of brain reserve, which suggests that higher educational or occupational background may reduce the risk of developing dementia by providing greater resilience against neuropathological changes [93].

5.6.4 Exploratory Framework for Multivariate Connectivity Patterns

The previous analyses focused on examining correlations between FC features, age, and Centiloid. While these analyses provide meaningful findings, they are largely hypothesis-driven and leave many other features and clinical factors unexplored, limiting our understanding of changes in FC. Manually testing relationships across a large set of features is both time-consuming and inefficient. To address this, we propose a framework designed to quickly identify multiple relationships between various features and FC parameters, streamlining connectivity analyses. This approach is especially useful in scenarios where all FC features have already been preprocessed and extracted, but new data (such as genetic or additional clinical information) becomes available. By enabling the exploration of these new features' relationships

with previously derived FC parameters in a non-hypothesis-driven manner, the framework ensures flexibility and adaptability for evolving datasets.

This framework uses a heatmap approach, with one axis representing features of interest (e.g., tau PET activity, anatomical measurements, or genetic data) and the other representing FC parameters. By calculating Pearson correlation coefficients, it identifies features with strong positive or negative associations with FC. For instance, genetic features such as gene expression levels or specific variants can be included to explore their connections with FC. This method simplifies the identification of key relationships across numerous potential features.

As a proof of concept, this framework was applied to examine two types of features in relation to FC. Firstly, tau PET activity features, which represent Flortaucipir tracer uptake across different brain regions, were analyzed. Secondly, anatomical MRI features, including regional brain volumes, were examined for correlations with FC. Figure 27 presents a heatmap summarizing the correlations between these two feature sets and FC parameters, providing a general overview of their associations.

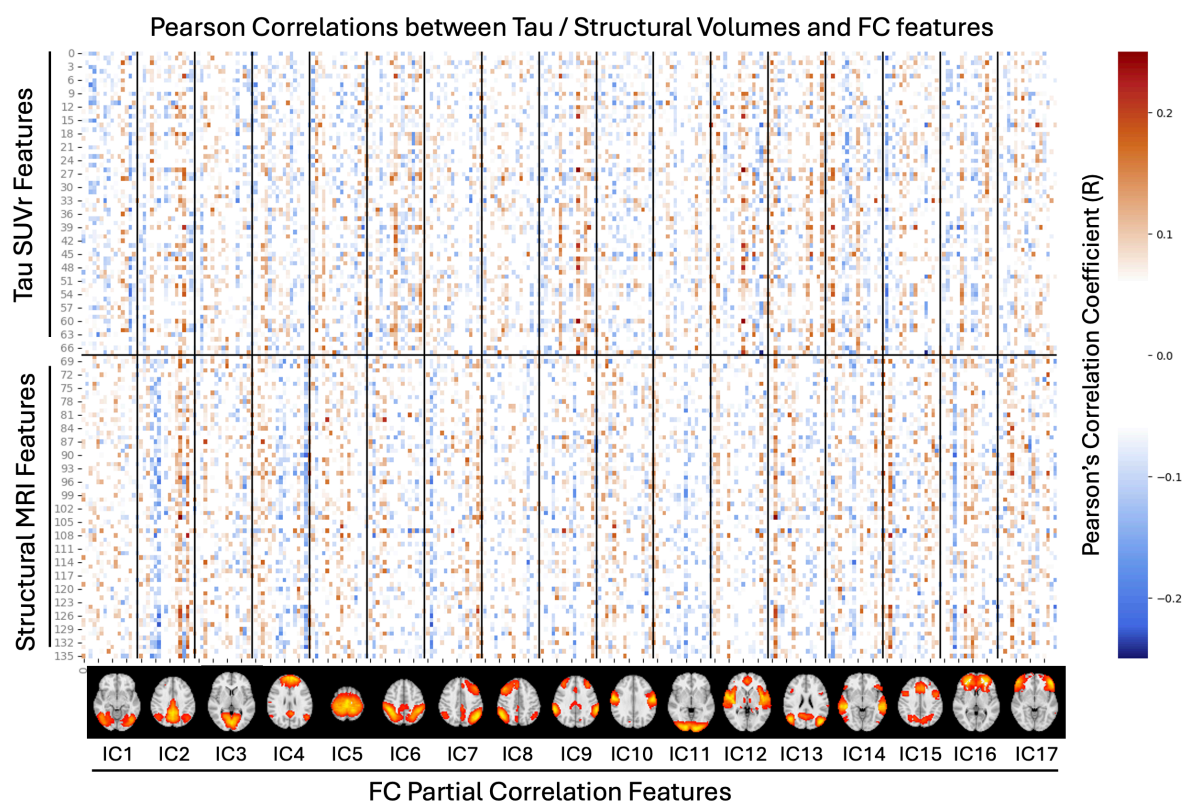


Figure 27: Exploratory heatmap illustrating Pearson correlations between partial FC features and two feature sets: Tau SUVR activities and structural MRI measures. Correlations are shown on a color scale, with red representing positive correlations and blue indicating negative correlations. FC features are grouped and ordered by network. Subdivisions delineate the two feature sets and different network connections for clarity.

As shown in Figure 27, each RSNs is associated with a set of FC features, totaling 16 per RSN, as each network connects with others but not with itself. The subdivisions within the heatmap represent distinct sections where the FC of a specific network is correlated with one of the two feature sets: tau PET or structural MRI features.

For example, the IC₂ Component (DMN) exhibits a notable pattern of both positive and negative correlations with structural MRI features. This suggests that some connections with this DMN component show increased FC associated with larger brain regions, while others display decreased FC. Conversely, other subdivisions of the heatmap appear predominantly white, indicating little to no meaningful correlation within those regions.

It is important to note that before interpreting these patterns, the data must be filtered to include only significant correlations ($p < 0.05$ and corrected for multiple comparisons using Bonferroni). The filtered version of this heatmap, displaying only significant correlations, is presented in Figure 28 (top).

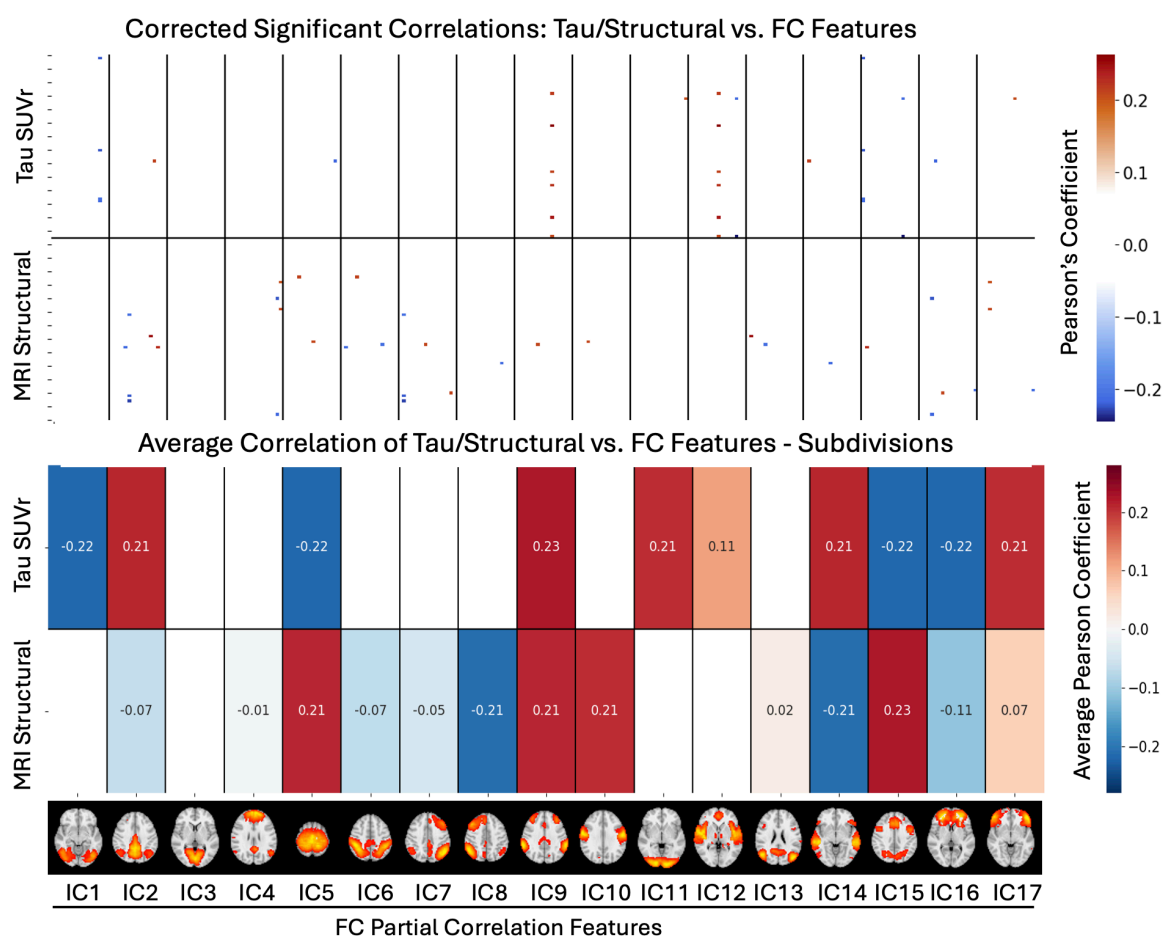


Figure 28: (Top) Heatmap displaying statistically significant correlations between partial FC features and two feature sets: Tau SUVR activities and structural MRI measures. (Bottom) Plot showing average Pearson correlation coefficients within each subdivision for Tau SUVR and structural MRI features, grouped by FC networks. Subdivisions correspond to the network-level organization of FC features, providing a summary of key trends in the data.

It can be seen in Figure 28 (top) that among the thousands of possible combinations, only a small subset of connections are statistically significant, highlighting areas worth further investigation. However, the limited number of significant results make interpretation challenging.

To address this, the bottom panel of Figure 28 presents the average Pearson correlation coefficients across subdivisions, providing an overall summary of the interactions between FC features and the other datasets. For example, FC features in the IC₉ component (Fronto-Parietal Network) show positive correlations with tau uptake (0.23), suggesting promising areas for further study.

While useful, this averaging approach has limitations. Collapsing values into a single mean can obscure important details, such as the presence of both positive and negative correlations or the total number of significant connections. For instance, IC₂ (DMN) exhibits significant positive and negative correlations with structural MRI features in the top graph, but these are obscured in the averaged results, yielding a small overall correlation (-0.07). Alternative methods, such as using absolute values or more advanced summarization techniques, could help retain critical information while simplifying data representation.

Despite these limitations, this framework is a valuable exploratory tool for understanding complex datasets. The figures are not intended to provide definitive conclusions but rather to guide further investigation by identifying patterns and relationships worth deeper analysis.

Once key features or global patterns are identified in this exploratory stage, a more rigorous analysis, such as a GLM, can be applied. This approach could incorporate additional covariates to thoroughly evaluate the relationships between the identified features and FC, as demonstrated in the analyses of Centiloid and age in Sections 5.6.2 and 5.6.3. The framework also shows potential for exploring diverse datasets (such as genetic information, cognitive measures, lifestyle factors, and blood biomarkers) in relation to FC. In addition, the framework's flexibility allows it to integrate data from different research focuses. For example, it can bridge studies where one researcher investigates FC while another examines structural connectivity using diffusion MRI, enabling a unified analysis that combines findings from both areas. However, further development and validation are needed to adapt and extend its application to other contexts effectively. All scripts and associated resources supporting this framework and the two previous analyses are available in the GitHub repository, detailed in the Annex Section 12.1.

6 Execution Schedule

This section describes the diagrams used to effectively plan the project's timeline. It covers defining the project's main stages, organizing work packages, breaking tasks into smaller components, and scheduling these tasks within a clear timeline.

6.1 Work Breakdown Structure (WBS)

The success of a project relies significantly on effective planning and detailed execution. One way to achieve this is by subdividing the work into smaller, more manageable components. This has been accomplished using the Work Breakdown Structure (WBS), a key project management tool that organizes all tasks within the project's scope. In this project, the WBS categorizes the project into six main work packages: project management, scientific research, data management, data preprocessing, analysis and interpretation, and finally, project closure.

This structured approach not only clarifies the sequence of tasks but also facilitates effective monitoring of the project's progress. Figure 29 presents the WBS developed for this study, offering an overview of the project's work packages and their associated tasks.

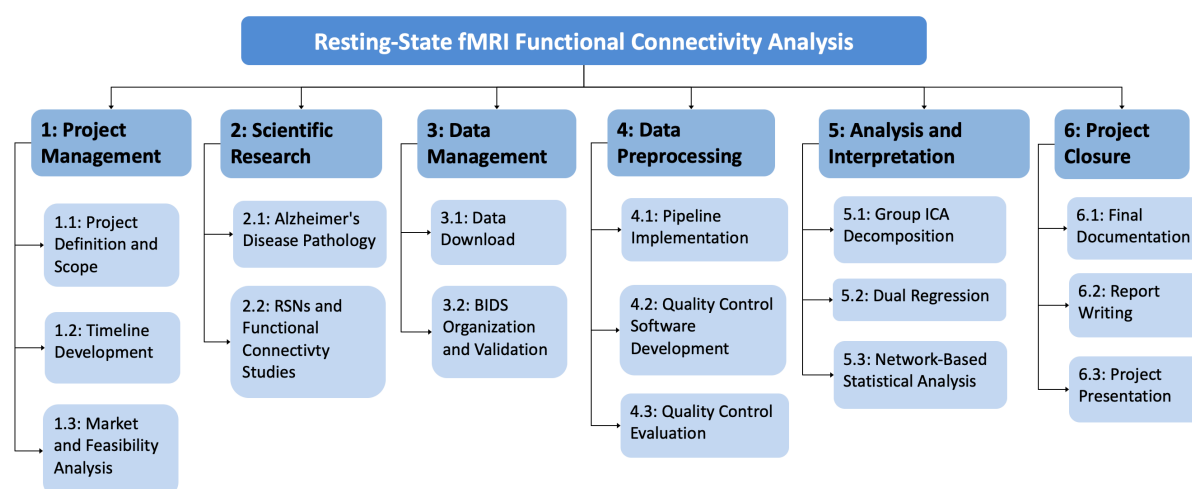


Figure 29: WBS diagram for the project, illustrating the main work packages and their associated tasks.

6.2 Execution Chronogram (GANTT)

In addition to the WBS, another tool used for effective project planning is the temporal chronogram, which assesses the timeline of tasks and their interdependencies. This tool is key to visually interpreting the project sequence, identifying activities that are critical and must adhere to strict timelines, and distinguishing those that allow for flexibility and have some margins. Table 4 provides an overview of the precedences and durations of each activity defined in the previous WBS diagram.

ID	Activity	Precedents	Duration (hours)
A	1.1 Project Definition and Scope	-	10
B	1.2 Timeline Development	A	4
C	1.3 Market and Feasibility Analysis	A	4
D	2.1 Alzheimer's Disease Pathology	A	10
E	2.2 RSNs and Functional Connectivity Studies	D	18
F	3.1 Data Download	B, C, E	12
G	3.2 BIDS Organization and Validation	F	20
H	4.1 Pipeline Implementation	G	50
I	4.2 Quality Control Software Development	G	60
J	4.3 Quality Control Evaluation	H, I	40
K	5.1 Group ICA Decomposition	J	30
L	5.2 Dual Regression	K	20
M	5.3 Network-Based Statistical Analysis	L	60
N	6.1 Final Documentation	M	20
O	6.2 Report Writing	M	30
P	6.3 Project Presentation	M	12

Table 4: Table of activities with their duration and precedents in hours. Total: 400 hours.

Using the data from Table 4, a GANTT chart was created to represent the project activities over time. Critical tasks are highlighted in red, and flexible ones in blue. Figure 30 provides a clear visualization of the overall plan, aiding in task management and monitoring throughout the project.

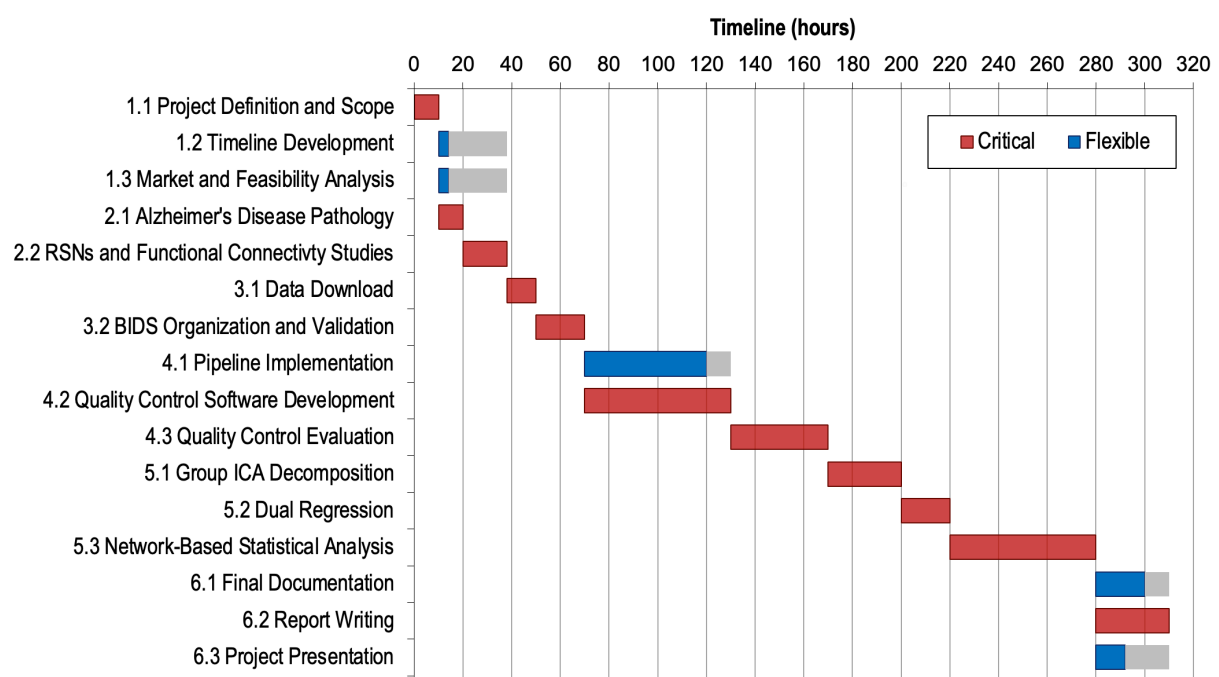


Figure 30: GANTT diagram of the global project execution.

6.3 Critical Path Method Diagram (PERT)

Complementary to the GANTT diagram, the PERT or Critical path method diagram also provides a visual representation of the chronology and interdependencies between activities, as shown in Figure 31. Using information from Table 4, the diagram represents activities as arrows. The critical path, which highlights activities with no flexibility that must be completed on time, is shown in red. Flexible activities, which can tolerate delays without affecting the overall timeline, are displayed in black. For improved clarity, the diagram is divided into two sections: the upper part illustrates activities from node 0 to node 6, while the lower part continues from node 6 to node 15.

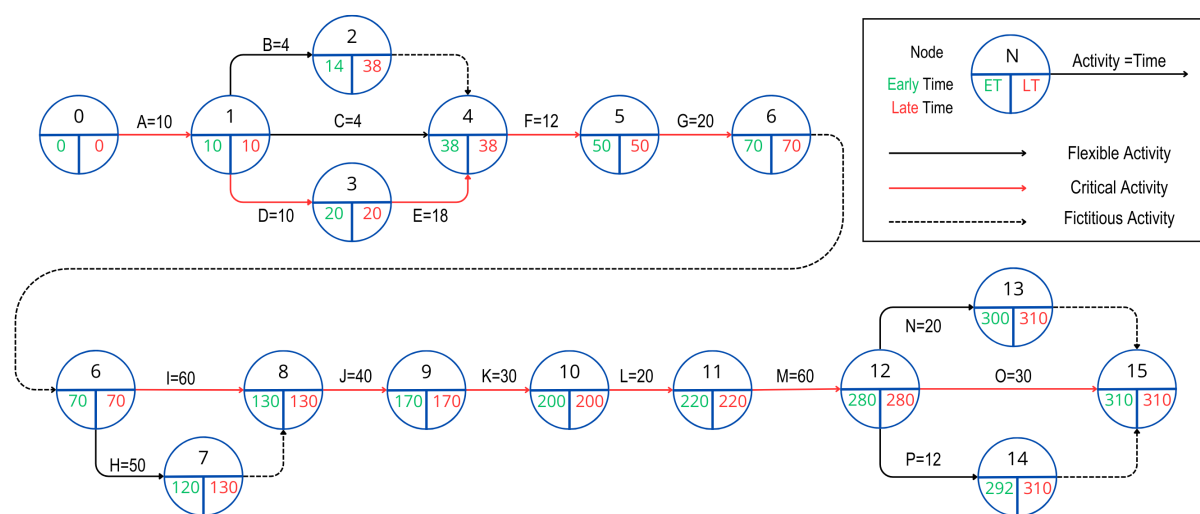


Figure 31: PERT diagram illustrating the chronology and interdependencies of activities within the project.

7 Technical Feasibility

This section evaluates the technical viability of the project, identifying factors that have influenced its development positively and negatively, as well as those that may affect its future evolution and application. This analysis is presented in a SWOT matrix (Strengths-Weaknesses-Opportunities-Threats), shown in Table 5, which outlines the key aspects of the project. Strengths include access to advanced tools, a flexible preprocessing pipeline that enhances scalability. However, weaknesses such as dataset homogeneity, and multisite variability and costly techniques present challenges. There are opportunities in the increasing demand for better diagnostics in preclinical AD and the potential to incorporate future high-quality datasets, while threats include competition from cost-effective biomarkers and potential resistance to new methodologies. This analysis offers a comprehensive perspective on the project's current and future potential, guiding its further development.

	Positive	Negative
Internal	Strengths <ul style="list-style-type: none"> • Access to Advanced Tools: High computational power and availability of cutting-edge software. • Strong Scientific Team: Collaboration with the BIG, a highly skilled research team. • Flexible Preprocessing Pipeline: Implementation of an adaptable pipeline using innovative and recent techniques. • Scalability of Results: Methods and findings could be extended to other datasets or conditions. • High-Quality Dataset: The A4 study dataset is recent, extensive, and includes a high number of subjects and imaging modalities, such as diffusion imaging. • Clinically Relevant Techniques: Use of established imaging techniques already implemented in clinical practice, avoiding the need for new modalities. 	Weaknesses <ul style="list-style-type: none"> • Dataset Homogeneity: <ul style="list-style-type: none"> ◦ Subjects have similar profiles due to trial recruitment, with little differentiation and no healthy controls included. ◦ The dataset is mostly well-educated, white, and cognitively healthy, reducing generalizability to other populations at risk of dementia. • Multisite Heterogeneity: Variability in acquisition parameters across sites affects data quality. • Costly Imaging Techniques: MRI, fMRI and, PET are expensive and less accessible. • Lack of Ground Truth: The dataset lacks distinct healthy vs. diseased groups, focusing instead on amyloid progression in cognitively normal subjects. • Resource Limitations: Restricted time and financial resources may difficult deeper exploration.
External	Opportunities <ul style="list-style-type: none"> • Growing Prevalence of Neurological Diseases: Increasing cases of AD and other neurodegenerative disorders highlight the need for diagnostic advancements. • Gap in Flexible Preprocessing Pipelines: While standardized pipelines exist, adaptable workflows could benefit a wider range of studies. • Growing Role of Functional Biomarkers: Personalized medicine and underutilized functional parameters offer opportunities for innovation and research. • Expanding Neurodegenerative Drug Market: Demand for AD treatments drives the need for better diagnostics and monitoring, including imaging biomarkers. • Preclinical Disease Focus: While most studies target AD, fewer explore preclinical stages, offering opportunities for biomarker development. • Future Datasets: Emerging large-scale datasets will enable in-depth analyses and integration into the proposed framework. 	Threats <ul style="list-style-type: none"> • Competition from Accessible Biomarkers: Blood and ocular based biomarkers are gaining traction in clinical trials due to their cost-effectiveness and accessibility. • Technological Obsolescence: Current imaging technologies may become outdated as new diagnostic methods emerge. • Regulatory Challenges: Strict regulations and research policies could slow down the adoption of novel biomarkers or techniques. • Scientific Skepticism: Resistance within the scientific community to adopting new methodologies or technologies could limit the project's impact.

Table 5: SWOT analysis of the project, with the Strengths, Weaknesses, Opportunities and Threats.

8 Economical Feasibility

This section aims to provide a detailed estimate of the budget required to cover all expenses associated with the complete execution of the project. These costs are specified in Table 6.

Item	Description	Unit Price	Total Price
Data Acquisition			
A4 Study Data	Free access under registration and project approval	Free	0 €
Preprocessing Tools			
Python	Programming language for data analysis	Free	0 €
FSL	Neuroimaging analysis software	Free	0 €
Visual Studio Code	Code editor for workflows	Free	0 €
Material Resources			
Workstation	Supermicro AMD Threadripper PRO 3000WX, 16 cores, 128 GB RAM, 2 SSDs	5,802 €	5,802 €
Graphics Card	PNY Nvidia Quadro RTX A5000, 24GB GDDR6	2,400 €	2,400 €
Storage Units	20 TB SATA hard drives (4 units, total 80 TB)	345 €	1,380 €
Personal Laptop	Secondary tool for research and report writing	1,500 €	1,500 €
Human Resources			
Biomedical Engineer	Total time: 400 hours	20 €/hour	8,000 €
Tutor	Total time: 40 hours	50 €/hour	2,000 €
Total Cost			21,082 €

Table 6: Breakdown of costs and resources for the project, including hardware upgrades and human resources.

As shown in Table 6, the project costs are divided into four categories: data acquisition, preprocessing tools, material resources, and human resources.

Firstly, the data used in the study was obtained at no cost, as the A4 study team made their dataset freely available to support scientific research. Although the clinical trial conducted by Eli Lilly & Company involved substantial costs, researchers can access the data by registering and submitting a justified project proposal. Similarly, the preprocessing tools had no associated costs, with Python, FSL, and Visual Studio Code being open-source and freely accessible.

For material resources, a personal laptop was used for tasks such as literature review and report writing, while computationally intensive analyses were performed on a high-performance workstation. These resources totaled 11,082 euros, with 1,500 euros for the laptop and 9,582 euros for the workstation.

Lastly, human resources accounted for 10,000 euros. This included 400 hours of work by a biomedical engineer at a rate of 20 euros per hour and 40 hours of tutoring at a rate of 50 euros per hour, based on salary references [94]. In total, the project cost 21,082 euros, reflecting the efficient use of open-source tools and free datasets alongside investments in computational and human resources.

9 Legislation and Regulations

The following section analyzes certain legal and regulatory aspects associated with the context of the project. While some of these elements may not be directly related to the project's objectives (such as ethics in medical research) they are closely tied to the clinical trial from which the data for this study is derived. Therefore, it is considered relevant to address them.

Medical research in neurodegenerative diseases like Alzheimer's poses significant ethical challenges. Informed consent is a critical component, ensuring participants understand the study's objectives, procedures, risks, benefits, and their right to withdraw at any time [95]. Participants must also be informed about potential findings and decide whether they wish to receive this information, particularly during presymptomatic or preclinical stages [96]. Obtaining informed consent in patients with AD is often challenging due to cognitive impairments. However, participants in the A4 study are cognitively normal, enabling them to fully understand the study's purpose and implications and make well-informed decisions about their participation. Additionally, all studies must undergo review and approval by an ethics committee, which evaluates data protection, fairness in participant selection, and the risk-benefit analysis [96, 97]. The validation of such studies depends on showing that the potential benefits, such as the use of the drug solanezumab, outweigh the risks associated. This balance is a cornerstone of ethical medical research.

Data protection is another key consideration for this project, especially when managing sensitive information from MRI imaging and subject-related data. Compliance with both Regulation (EU) 2016/679 General Data Protection Regulation (GDPR) [98], applicable to international standards, and the Health Insurance Portability and Accountability Act (HIPAA) [99], relevant due to the inclusion of data from U.S. sites is required to ensure personal data processing is lawful, fair, and transparent. Participants must be informed about how their data will be used, and explicit consent must be obtained for its inclusion in the study. For the A4 study, data anonymization is implemented by removing facial features from imaging data and assigning unique subject codes instead of names, eliminating identifying information and ensuring privacy. Participants are also fully informed about what data is collected, its use, who will access it, and how long it will be stored as explained in their study protocol [2]. Lastly, researchers and users working with this data must comply with the usage policies. These include restrictions on commercial use and information redistribution, a prohibition on attempts to identify participants, and requirements to implement security measures for data protection [100]. These policies ensure the data is used ethically, responsibly, and in compliance with legal standards.

Lastly, another key legal requirement of this project is respecting intellectual property rights by accurately citing scientific articles and databases, such as the A4 study data used, as mandated by the Spanish Law 14/2011 on Science, Technology, and Innovation (articles 14 and 15) [101]. Recognizing these sources not only ensures transparency in the research process but also respects the intellectual contributions of those who developed these valuable resources.

10 Conclusions and Future Lines

This project was driven by the need for adaptable and robust methodologies to analyze complex neuroimaging data. The outcomes not only address specific research goals in Alzheimer's disease but also demonstrate the potential of flexible frameworks to enhance the scalability and precision of future neuroscience studies.

We have successfully adapted and implemented a flexible, automated preprocessing pipeline designed for rs-fMRI, a technique that requires high-quality processing to reduce its susceptibility to noise and artifacts. This pipeline, although applied to a specific dataset, is highly adaptable and can be tailored to meet the varied requirements of different neuroimaging projects. Its modularity allows for the inclusion, exclusion, or modification of processing steps to suit individual study needs—for instance, opting out of spatial smoothing to maintain data resolution, or utilizing a new AI-based algorithm for brain extraction. These features ensure the pipeline's applicability across a range of methodologies and research priorities without sacrificing scientific rigor. Moreover, its efficiency in processing times, when compared to other softwares like fMRIPrep, makes it particularly suitable for handling large-scale datasets. As neuroimaging research evolves with the introduction of more advanced algorithms and new datasets, this pipeline stands ready to integrate and adapt these innovations, providing a robust tool that combines flexibility, precision, and speed to support the demands of contemporary neuroimaging analysis.

This project has successfully optimized the QC process in neuroimaging research through the development of an application that efficiently integrates both quantitative and visual methods. Traditionally, QC in neuroimaging relied heavily on manual inspections that were often subjective and inefficient. Our application not only streamlines this process but also enhances its objectivity by incorporating multiple reviewers. This improvement makes our QC practices more suited to handling the complexities of large-scale datasets, enhancing both consistency and scalability. The development of this QC application has opened the door for further research within the BIG, encouraging the exploration of innovative methods to move towards a more quantitative QC process and enhance automatic artifact detection. By reducing reliance on purely visual methods, this could improve the objectivity and scalability of QC. Furthermore, the QC application is highly adaptable, making it a versatile resource for future projects. All resources related to the preprocessing pipeline and QC application are available in the GitHub repository, as detailed in Annex Section 12.1.

We employed a data-driven ICA approach to identify RSNs from our cleaned data, successfully producing high-quality, well-established networks recognized in the literature [79]. Despite the A4 dataset's diverse acquisition parameters, which could have introduced variability and noise, our methods consistently delivered excellent results. This proves the robustness of our approach and the effectiveness of our preprocessing and QC methods. Furthermore, the consistency between visual identification and automated classification via hierarchical clustering emphasizes the reliability of our methods. Future studies could explore the integration of more advanced supervised and clustering techniques to improve the accuracy of RSN identification in neuroimaging research.

The hypothesis that FC would correlate with the Centiloid scale did not produce the expected results.

Although we did observe some noteworthy connectivity changes in the DMN, a topic often discussed in research [45, 89, 88], these findings were not widespread. This outcome may suggest that the Centiloid scale might simply indicate whether pathology is present or absent, rather than showing different stages of disease progression. Future research could explore alternative molecular markers, such as Tau PET, which could better capture disease progression. Additionally, the homogeneity of the dataset, comprising subjects in similar disease stages, likely influenced the results. Including a broader range of subjects, such as healthy controls and advanced AD cases, could provide more robust findings.

On the other hand, significant age-related changes in FC were observed, with many node pairs showing either connectivity losses or gains. These findings align with the literature, suggesting a pattern of brain reorganization in aging, where weakened connections are compensated for by the strengthening of others [92]. This further validates the pipeline and QC process as effective tools for detecting meaningful patterns. Future studies could build on these results by including more variables like education level and cognitive measures to explore their impact on brain FC changes related to aging.

The exploratory framework proved highly valuable, especially when no specific hypothesis was initially set. Its versatility is key, not just for adapting to evolving datasets and integrating various data modalities but also for exploring complex relationships between variables. For instance, when we gain access to the complete dataset from the A4 study's clinical trial, this framework could aid in analyzing relationships between FC at the screening stage and various outcomes measured later, such as cognitive performance, structural changes, or amyloid uptake. Such analysis could help identify early biomarkers for AD progression, providing crucial information for early detection and intervention. This adaptability supports potential discoveries and keeps our research in line with the dynamic nature of science, enabling broad applications and meaningful findings about disease mechanisms.

In a time when Alzheimer's research requires constantly evolving tools and methods, this project emphasizes the importance of adaptability and innovation. As neuroimaging and AI continue to transform the field, it's essential for researchers to embrace these technologies in their work. Equally important is the anticipated availability of large-scale, high-quality datasets, which offers a unique opportunity to advance our understanding of AD and its mechanisms. However, data alone is not sufficient; it must be paired with robust and adaptable frameworks to unlock its full potential. Therefore, the frameworks developed in this study tackle current challenges with their flexibility and scalability, ensuring they stay relevant as AI algorithms continue to evolve. Combined with emerging datasets, these frameworks lay a strong foundation for future research, paving the way for the discovery of new biomarkers for the early detection and progression of Alzheimer's Disease.

11 References

- [1] Reisa A. Sperling et al. “Trial of Solanezumab in Preclinical Alzheimer’s Disease”. In: *The New England journal of medicine* 389 (12 Sept. 2023), pp. 1096–1107. ISSN: 1533-4406. DOI: [10.1056/NEJMOA2305032](https://doi.org/10.1056/NEJMOA2305032).
- [2] *Study Details | Clinical Trial of Solanezumab for Older Individuals Who May be at Risk for Memory Loss | ClinicalTrials.gov*. URL: <https://www.clinicaltrials.gov/study/NCT02008357> (visited on 12/27/2024).
- [3] Elizabeth DuPre and Russell Alan Poldrack. “The future of data analysis is now: Integrating generative AI in neuroimaging methods development”. In: *Imaging Neuroscience* 2 (Oct. 2024), pp. 1–8. DOI: [10.1162/IMAG_A_00241](https://doi.org/10.1162/IMAG_A_00241).
- [4] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86 (11 1998), pp. 2278–2323. ISSN: 00189219. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems* 25 (2012). DOI: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [6] Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6 June 2015), pp. 1137–1149. ISSN: 01628828. DOI: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [7] Evan Shelhamer, Jonathan Long, and Trevor Darrell. “Fully Convolutional Networks for Semantic Segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (4 Nov. 2014), pp. 640–651. ISSN: 01628828. DOI: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683).
- [8] *OpenNeuro*. URL: <https://openneuro.org/> (visited on 10/22/2024).
- [9] Nguyen Thanh Duc et al. “3D-Deep Learning Based Automatic Diagnosis of Alzheimer’s Disease with Joint MMSE Prediction Using Resting-State fMRI”. In: *Neuroinformatics* 18 (1 Jan. 2020), pp. 71–86. ISSN: 15590089. DOI: [10.1007/S12021-019-09419-W/METRICS](https://doi.org/10.1007/S12021-019-09419-W/METRICS).
- [10] James H. Scatliff and Peter J. Morris. “From Roentgen to magnetic resonance imaging: the history of medical imaging”. In: *North Carolina medical journal* 75 (2 2014), pp. 111–113. ISSN: 0029-2559. DOI: [10.18043/NCM.75.2.111](https://doi.org/10.18043/NCM.75.2.111).
- [11] Vijay P.B. Grover et al. “Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians”. In: *Journal of Clinical and Experimental Hepatology* 5 (3 Sept. 2015), p. 246. ISSN: 22133453. DOI: [10.1016/J.JCEH.2015.08.001](https://doi.org/10.1016/J.JCEH.2015.08.001).
- [12] Janine. Bijsterbosch, Stephen M.. Smith, and Christian. Beckmann. *An introduction to resting state fMRI functional connectivity*. First. OUP Oxford, 2017, p. 157. ISBN: 9780192535740.
- [13] Marcus E. Raichle et al. “A default mode of brain function”. In: *Proceedings of the National Academy of Sciences of the United States of America* 98 (2 Jan. 2001), pp. 676–682. ISSN: 00278424. DOI: [10.1073/pnas.98.2.676](https://doi.org/10.1073/pnas.98.2.676).

- [14] Debra A. Gusnard and Marcus E. Raichle. "Searching for a baseline: Functional imaging and the resting human brain". In: *Nature Reviews Neuroscience* 2:10 2 (10 2001), pp. 685–694. ISSN: 1471-0048. DOI: [10.1038/35094500](https://doi.org/10.1038/35094500).
- [15] Bharat Biswal et al. "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI". In: *Magnetic resonance in medicine* 34 (4 1995), pp. 537–541. ISSN: 0740-3194. DOI: [10.1002/MRM.1910340409](https://doi.org/10.1002/MRM.1910340409).
- [16] Gary H. Glover. "Overview of Functional Magnetic Resonance Imaging". In: *Neurosurgery clinics of North America* 22 (2 Apr. 2011), p. 133. ISSN: 10423680. DOI: [10.1016/J.NEC.2010.11.001](https://doi.org/10.1016/J.NEC.2010.11.001).
- [17] Zdeněk Fišar. "Linking the Amyloid, Tau, and Mitochondrial Hypotheses of Alzheimer's Disease and Identifying Promising Drug Targets". In: *Biomolecules* 12 (11 Nov. 2022). ISSN: 2218-273X. DOI: [10.3390/BIOM12111676](https://doi.org/10.3390/BIOM12111676).
- [18] Ujala Sehar et al. "Amyloid Beta in Aging and Alzheimer's Disease". In: *International journal of molecular sciences* 23 (21 Nov. 2022). ISSN: 1422-0067. DOI: [10.3390/IJMS232112924](https://doi.org/10.3390/IJMS232112924).
- [19] *Pathology of Alzheimer's Disease*. URL: <https://app.biorender.com/biorender-templates/figures/all/t-5d8baeb4f7e1a5007dd46b18-pathology-of-alzheimers-disease> (visited on 05/19/2024).
- [20] *What Causes Alzheimer's Disease?* | National Institute on Aging. URL: <https://www.nia.nih.gov/health/alzheimers-causes-and-risk-factors/what-causes-alzheimers-disease> (visited on 12/27/2024).
- [21] Elina Zotova et al. "Inflammation in Alzheimer's disease: relevance to pathogenesis and therapy". In: *Alzheimer's research and therapy* 2 (1 2010). ISSN: 1758-9193. DOI: [10.1186/ALZRT24](https://doi.org/10.1186/ALZRT24).
- [22] Xiaoguang Du, Xinyi Wang, and Meiyu Geng. "Alzheimer's disease hypothesis and related therapies". In: *Translational neurodegeneration* 7 (1 Jan. 2018). ISSN: 2047-9158. DOI: [10.1186/S40035-018-0107-Y](https://doi.org/10.1186/S40035-018-0107-Y).
- [23] Saswata Roy et al. "The Role of 18F-Flortaucipir (AV-1451) in the Diagnosis of Neurodegenerative Disorders". In: *Cureus* 13 (7 July 2021). ISSN: 2168-8184. DOI: [10.7759/CUREUS.16644](https://doi.org/10.7759/CUREUS.16644).
- [24] William E. Klunk et al. "The Centiloid Project: Standardizing Quantitative Amyloid Plaque Estimation by PET". In: *Alzheimer's and dementia : the journal of the Alzheimer's Association* 11 (1 2015), p. 1. ISSN: 15525279. DOI: [10.1016/J.JALZ.2014.07.003](https://doi.org/10.1016/J.JALZ.2014.07.003).
- [25] Avinash Chandra, George Dervenoulas, and Marios Politis. "Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment". In: *Journal of Neurology* 266 (6 June 2018), p. 1293. ISSN: 14321459. DOI: [10.1007/S00415-018-9016-3](https://doi.org/10.1007/S00415-018-9016-3).
- [26] P. Murali Doraiswamy et al. "Amyloid- β assessed by florbetapir F 18 PET and 18-month cognitive decline: A multicenter study". In: *Neurology* 79 (16 Oct. 2012), p. 1636. ISSN: 1526632X. DOI: [10.1212/WNL.OB013E3182661F74](https://doi.org/10.1212/WNL.OB013E3182661F74).
- [27] Reisa A. Sperling et al. "The A4 study: stopping AD before symptoms begin?" In: *Science translational medicine* 6 (228 Mar. 2014). ISSN: 1946-6242. DOI: [10.1126/SCITRANSLMED.3007941](https://doi.org/10.1126/SCITRANSLMED.3007941).

- [28] Oscar Esteban et al. “fMRIPrep: a robust preprocessing pipeline for functional MRI”. In: *Nature methods* 16 (1 Jan. 2019), pp. 111–116. ISSN: 1548-7105. DOI: [10.1038/S41592-018-0235-4](https://doi.org/10.1038/S41592-018-0235-4).
- [29] FSL. URL: <https://fsl.fmrib.ox.ac.uk/fsl/docs/#/> (visited on 12/27/2024).
- [30] ANTs by stnava - Advanced Normalization Tools. URL: <http://stnava.github.io/ANTs/> (visited on 12/22/2024).
- [31] FreeSurfer - Free Surfer Wiki. URL: <https://surfer.nmr.mgh.harvard.edu/fswiki> (visited on 12/22/2024).
- [32] SIMEXP/fmriprep-qc: Lightweight QC for fMRIPrep outputs. URL: <https://github.com/SIMEXP/fmriprep-qc/tree/main> (visited on 12/28/2024).
- [33] Fabian Isensee et al. “Automated brain extraction of multisequence MRI using artificial neural networks”. In: *Human brain mapping* 40 (17 Dec. 2019), pp. 4952–4964. ISSN: 1097-0193. DOI: [10.1002/HBM.24750](https://doi.org/10.1002/HBM.24750).
- [34] L. Martí Bonmatí et al. “Imaging biomarkers, quantitative imaging, and bioengineering”. In: *Radiología (English Edition)* 54 (3 May 2012), pp. 269–278. ISSN: 2173-5107. DOI: [10.1016/J.RXENG.2012.05.001](https://doi.org/10.1016/J.RXENG.2012.05.001).
- [35] M. Altuna-Azkargorta and M. Mendioroz-Iriarte. “Blood biomarkers in Alzheimer’s disease”. In: *Neurología (English Edition)* 36 (9 Nov. 2021), pp. 704–710. ISSN: 2173-5808. DOI: [10.1016/J.NRLENG.2018.03.006](https://doi.org/10.1016/J.NRLENG.2018.03.006).
- [36] Sebastian Palmqvist, Niklas Mattsson, and Oskar Hansson. “Cerebrospinal fluid analysis detects cerebral amyloid- β accumulation earlier than positron emission tomography”. In: *Brain : a journal of neurology* 139 (Pt 4 Apr. 2016), pp. 1226–1236. ISSN: 1460-2156. DOI: [10.1093/BRAIN/AWW015](https://doi.org/10.1093/BRAIN/AWW015).
- [37] Andreas Hahn et al. “Association Between Earliest Amyloid Uptake and Functional Connectivity in Cognitively Unimpaired Elderly”. In: *Cerebral Cortex* 29 (5 May 2019), pp. 2173–2182. ISSN: 1047-3211. DOI: [10.1093/CERCOR/BHZ020](https://doi.org/10.1093/CERCOR/BHZ020).
- [38] Sebastian Palmqvist et al. “Accuracy of brain amyloid detection in clinical practice using cerebrospinal fluid β -amyloid 42: a cross-validation study against amyloid positron emission tomography”. In: *JAMA neurology* 71 (10 Oct. 2014), pp. 1282–1289. ISSN: 2168-6157. DOI: [10.1001/JAMANEUROL.2014.1358](https://doi.org/10.1001/JAMANEUROL.2014.1358).
- [39] Stefan J. Teipel et al. “Perspectives for multimodal neurochemical and imaging biomarkers in Alzheimer’s disease”. In: *Journal of Alzheimer’s disease : JAD* 33 Suppl 1 (SUPPL. 1 2013). ISSN: 1875-8908. DOI: [10.3233/JAD-2012-129030](https://doi.org/10.3233/JAD-2012-129030).
- [40] Liam G. McCoy et al. “Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based”. In: *Journal of clinical epidemiology* 142 (Feb. 2022), pp. 252–257. ISSN: 1878-5921. DOI: [10.1016/J.JCLINEPI.2021.11.001](https://doi.org/10.1016/J.JCLINEPI.2021.11.001).
- [41] *Alzheimer’s and Dementia* | Alzheimer’s Association. URL: https://www.alz.org/alzheimer_s_dementia (visited on 12/21/2024).
- [42] *Leqembi recommended for treatment of early Alzheimer’s disease* | European Medicines Agency (EMA). URL: <https://www.ema.europa.eu/en/news/leqembi-recommended-treatment-early-alzheimers-disease> (visited on 12/16/2024).

- [43] van Dyck CH et al. "Lecanemab in Early Alzheimer's Disease". In: *The New England journal of medicine* 388 (1 2023), pp. 142–143. ISSN: 1533-4406. DOI: [10.1056/NEJMOA2212948](https://doi.org/10.1056/NEJMOA2212948).
- [44] Mark A. Mintun et al. "Donanemab in Early Alzheimer's Disease". In: *The New England journal of medicine* 384 (18 May 2021), pp. 1691–1704. ISSN: 1533-4406. DOI: [10.1056/NEJMOA2100708](https://doi.org/10.1056/NEJMOA2100708).
- [45] Emily L. Dennis and Paul M. Thompson. "Functional brain connectivity using fMRI in aging and Alzheimer's disease". In: *Neuropsychology review* 24 (1 Mar. 2014), pp. 49–62. ISSN: 1573-6660. DOI: [10.1007/S11065-014-9249-6](https://doi.org/10.1007/S11065-014-9249-6).
- [46] Christian Hohenfeld, Cornelius J. Werner, and Kathrin Reetz. "Resting-state connectivity in neurodegenerative disorders: Is there potential for an imaging biomarker?" In: *NeuroImage. Clinical* 18 (Jan. 2018), pp. 849–870. ISSN: 2213-1582. DOI: [10.1016/J.NICL.2018.03.013](https://doi.org/10.1016/J.NICL.2018.03.013).
- [47] *Medical Image Analysis Software Market Size Report, 2030*. URL: <https://www.grandviewresearch.com/industry-analysis/medical-image-analysis-software-market> (visited on 12/15/2024).
- [48] *afni.nimh.nih.gov*. URL: <https://afni.nimh.nih.gov/> (visited on 12/27/2024).
- [49] *fMRIPrep: A Robust Preprocessing Pipeline for fMRI Data — fmriprep version documentation*. URL: <https://fmriprep.org/en/stable/> (visited on 12/26/2024).
- [50] *ADNI | Alzheimer's Disease Neuroimaging Initiative*. URL: <https://adni.loni.usc.edu/> (visited on 12/15/2024).
- [51] *Connectome - Homepage*. URL: <https://www.humanconnectome.org/> (visited on 12/15/2024).
- [52] Cathie Sudlow et al. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age". In: *PLoS Medicine* 12 (3 Mar. 2015). ISSN: 15491676. DOI: [10.1371/JOURNAL.PMED.1001779](https://doi.org/10.1371/JOURNAL.PMED.1001779).
- [53] *QMENTA, The all in one imaging platform for your clinical trial*. URL: <https://www.qmenta.com/> (visited on 12/14/2024).
- [54] *Advantis | Brain (Brainance MD)*. URL: <https://advantis.io/brain/> (visited on 12/27/2024).
- [55] Pawan Kumar Mall et al. "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities". In: *Healthcare Analytics* 4 (Dec. 2023), p. 100216. ISSN: 2772-4425. DOI: [10.1016/J.HEALTH.2023.100216](https://doi.org/10.1016/J.HEALTH.2023.100216).
- [56] *Imaging Biomarker Market Report: Trends, Forecast and Competitive Analysis to 2030*. URL: <https://www.lucintel.com/imaging-biomarker-market.aspx> (visited on 12/27/2024).
- [57] James P.B. O'Connor et al. "Imaging biomarker roadmap for cancer studies". In: *Nature Reviews Clinical Oncology* 14:3 14 (3 Oct. 2016), pp. 169–186. ISSN: 1759-4782. DOI: [10.1038/nrclinonc.2016.162](https://doi.org/10.1038/nrclinonc.2016.162).
- [58] Peter N.E. Young et al. "Imaging biomarkers in neurodegeneration: current and future practices". In: *Alzheimer's research and therapy* 12 (1 Apr. 2020). ISSN: 1758-9193. DOI: [10.1186/S13195-020-00612-7](https://doi.org/10.1186/S13195-020-00612-7).

- [59] Qubitech | Cuantificación automática de biomarcadores de neuroimagen. URL: <https://qubitech.com/> (visited on 12/14/2024).
- [60] Freenome Partners with Siemens Healthineers for Breast Cancer Research - Freenome. URL: https://www.freenome.com/newsroom/freenome-partners-with-siemens-healthineers-for-breast-cancer-research/?utm_source=chatgpt.com (visited on 12/27/2024).
- [61] Krzysztof J. Gorgolewski et al. "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments". In: *Scientific data* 3 (June 2016). ISSN: 2052-4463. DOI: [10.1038/SDATA.2016.44](https://doi.org/10.1038/SDATA.2016.44).
- [62] BIDS validation - The Brain Imaging Data Structure. URL: <https://bids-website.readthedocs.io/en/latest/tools/validator.html> (visited on 12/28/2024).
- [63] Norman Scheel et al. "Evaluation of noise regression techniques in resting-state fMRI studies using data of 434 older adults". In: *Frontiers in Neuroscience* 16 (Oct. 2022), p. 1006056. ISSN: 1662453X. DOI: [10.3389/FNINS.2022.1006056/BIBTEX](https://doi.org/10.3389/FNINS.2022.1006056/BIBTEX).
- [64] Jingyuan E. Chen et al. "On the analysis of rapidly sampled fMRI data". In: *NeuroImage* 188 (Mar. 2019), pp. 807–820. ISSN: 1095-9572. DOI: [10.1016/J.NEUROIMAGE.2019.02.008](https://doi.org/10.1016/J.NEUROIMAGE.2019.02.008).
- [65] Sanam Maknojia et al. "Resting state fMRI: Going through the motions". In: *Frontiers in Neuroscience* 13 (JUL Aug. 2019), p. 462471. ISSN: 1662453X. DOI: [10.3389/FNINS.2019.00825/BIBTEX](https://doi.org/10.3389/FNINS.2019.00825/BIBTEX).
- [66] Thomas T. Liu. "Noise contributions to the fMRI signal: An overview". In: *NeuroImage* 143 (Dec. 2016), pp. 141–151. ISSN: 1053-8119. DOI: [10.1016/J.NEUROIMAGE.2016.09.008](https://doi.org/10.1016/J.NEUROIMAGE.2016.09.008).
- [67] Ali M. Golestani and J. Jean Chen. "Comparing data-driven physiological denoising approaches for resting-state fMRI: implications for the study of aging". In: *Frontiers in Neuroscience* 18 (Feb. 2024), p. 1223230. ISSN: 1662453X. DOI: [10.3389/FNINS.2024.1223230/BIBTEX](https://doi.org/10.3389/FNINS.2024.1223230/BIBTEX).
- [68] Yongyue Zhang, Michael Brady, and Stephen Smith. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm". In: *IEEE transactions on medical imaging* 20 (1 Jan. 2001), pp. 45–57. ISSN: 0278-0062. DOI: [10.1109/42.906424](https://doi.org/10.1109/42.906424).
- [69] David B. Parker and Qolamreza R. Razlighi. "The Benefit of Slice Timing Correction in Common fMRI Pre-processing Pipelines". In: *Frontiers in neuroscience* 13 (JUL 2019). ISSN: 1662-4548. DOI: [10.3389/FNINS.2019.00821](https://doi.org/10.3389/FNINS.2019.00821).
- [70] AFNI program: 3dTshift. URL: https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dTshift.html (visited on 01/15/2025).
- [71] Mark Jenkinson et al. "Improved optimization for the robust and accurate linear registration and motion correction of brain images". In: *NeuroImage* 17 (2 2002), pp. 825–841. ISSN: 10538119. DOI: [10.1016/S1053-8119\(02\)91132-8](https://doi.org/10.1016/S1053-8119(02)91132-8).
- [72] *Neuroimaging in Python - Pipelines and Interfaces* — nipy pipeline and interfaces package. URL: <https://nipy.readthedocs.io/en/latest/api/generated/nipy.interfaces.fsl.preprocess.html#susan> (visited on 01/15/2025).

- [73] Cemre Candemir. "Spatial Smoothing Effect on Group-Level Functional Connectivity during Resting and Task-Based fMRI". In: *Sensors (Basel, Switzerland)* 23 (13 July 2023). ISSN: 1424-8220. DOI: [10.3390/S23135866](https://doi.org/10.3390/S23135866).
- [74] Brian B. Avants et al. "A reproducible evaluation of ANTs similarity metric performance in brain image registration". In: *NeuroImage* 54 (3 Feb. 2011), pp. 2033–2044. ISSN: 1095-9572. DOI: [10.1016/J.NEUROIMAGE.2010.09.025](https://doi.org/10.1016/J.NEUROIMAGE.2010.09.025).
- [75] *Nilearn Documentation - nilearn.signal.clean*. URL: <https://nilearn.github.io/dev/modules/generated/nilearn.signal.clean.html> (visited on 01/14/2025).
- [76] Martin M. Monti. "Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach". In: *Frontiers in human neuroscience* 5 (MARCH 2011). ISSN: 1662-5161. DOI: [10.3389/FNHUM.2011.00028](https://doi.org/10.3389/FNHUM.2011.00028).
- [77] Jingyuan E. Chen, Hesamoddin Jahanian, and Gary H. Glover. "Nuisance Regression of High-Frequency Functional Magnetic Resonance Imaging Data: Denoising Can Be Noisy". In: *Brain connectivity* 7 (1 Feb. 2017), pp. 13–24. ISSN: 2158-0022. DOI: [10.1089/BRAIN.2016.0441](https://doi.org/10.1089/BRAIN.2016.0441).
- [78] *FSL Course Material*. URL: https://open.win.ox.ac.uk/pages/fslcourse/website/online_materials.html (visited on 12/02/2024).
- [79] Stephen M. Smith et al. "Correspondence of the brain's functional architecture during activation and rest". In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (31 Aug. 2009), pp. 13040–13045. ISSN: 1091-6490. DOI: [10.1073/PNAS.0905267106](https://doi.org/10.1073/PNAS.0905267106).
- [80] Martijn P. van den Heuvel and Hilleke E. Hulshoff Pol. "Exploring the brain network: a review on resting-state fMRI functional connectivity". In: *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 20 (8 Aug. 2010), pp. 519–534. ISSN: 1873-7862. DOI: [10.1016/J.EURONEURO.2010.03.008](https://doi.org/10.1016/J.EURONEURO.2010.03.008).
- [81] Lucina Q. Uddin. "Complex relationships between structural and functional brain connectivity". In: *Trends in cognitive sciences* 17 (12 Dec. 2013), 10.1016/j.tics.2013.09.011. ISSN: 13646613. DOI: [10.1016/J.TICS.2013.09.011](https://doi.org/10.1016/J.TICS.2013.09.011).
- [82] Stephen M. Smith et al. "Network modelling methods for FMRI". In: *NeuroImage* 54 (2 Jan. 2011), pp. 875–891. ISSN: 1095-9572. DOI: [10.1016/J.NEUROIMAGE.2010.08.063](https://doi.org/10.1016/J.NEUROIMAGE.2010.08.063).
- [83] *IDA - Image and Data Archive*. URL: <https://ida.loni.usc.edu/login.jsp> (visited on 01/08/2025).
- [84] Jonathan D. Power et al. "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion". In: *NeuroImage* 59 (3 Feb. 2012), pp. 2142–2154. ISSN: 1095-9572. DOI: [10.1016/J.NEUROIMAGE.2011.10.018](https://doi.org/10.1016/J.NEUROIMAGE.2011.10.018).
- [85] Matthew Peverill et al. "Balancing Data Quality and Bias: Investigating Functional Connectivity Exclusions in the Adolescent Brain Cognitive Development (ABCD Study) Across Quality Control Pathways". In: *Human Brain Mapping* 46 (1 Jan. 2025), e70094. ISSN: 1097-0193. DOI: [10.1002/HBM.70094](https://doi.org/10.1002/HBM.70094).
- [86] *MELODIC FSL Documentation - Post-Stats*. URL: https://web.mit.edu/fsl_v5.0.10/fsl/doc/wiki/MELODIC.html (visited on 01/15/2025).

- [87] *FSL/fslnets* · GitLab. URL: <https://git.fmrib.ox.ac.uk/fsl/fslnets> (visited on 12/22/2024).
- [88] Rachel F. Buckley et al. "Functional network integrity presages cognitive decline in preclinical Alzheimer disease". In: *Neurology* 89 (1 July 2017), pp. 29–37. DOI: [10.1212/WNL.0000000000004059](https://doi.org/10.1212/WNL.0000000000004059).
- [89] Serge A.R.B. Rombouts et al. "Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: An fMRI study". In: *Human Brain Mapping* 26 (4 Dec. 2005), pp. 231–239. ISSN: 1097-0193. DOI: [10.1002/HBM.20160](https://doi.org/10.1002/HBM.20160).
- [90] R. Boyle et al. "Left Frontoparietal Control Network Connectivity Moderates the Effect of Amyloid on Cognitive Decline in Preclinical Alzheimer's Disease: The A4 Study". In: *The journal of prevention of Alzheimer's disease* 11 (4 Aug. 2024), pp. 881–888. ISSN: 2426-0266. DOI: [10.14283/JPAD.2024.140](https://doi.org/10.14283/JPAD.2024.140).
- [91] Naftali Raz et al. "Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter". In: *Cerebral cortex (New York, N.Y. : 1991)* 7 (3 1997), pp. 268–282. ISSN: 1047-3211. DOI: [10.1093/CERCOR/7.3.268](https://doi.org/10.1093/CERCOR/7.3.268).
- [92] Alexa M. Morcom and Wendy Johnson. "Neural Reorganization and Compensation in Aging". In: *Journal of Cognitive Neuroscience* 27 (7 July 2015), pp. 1275–1285. ISSN: 0898-929X. DOI: [10.1162/JOCN_A_00783](https://doi.org/10.1162/JOCN_A_00783).
- [93] Corinne Pettigrew and Anja Soldan. "Defining Cognitive Reserve and Implications for Cognitive Aging". In: *Current neurology and neuroscience reports* 19 (1 Jan. 2019). ISSN: 1534-6293. DOI: [10.1007/S11910-019-0917-Z](https://doi.org/10.1007/S11910-019-0917-Z).
- [94] *Biomedical Engineer Salary in Spain*. URL: <https://www.erieri.com/salary/job/biomedical-engineer/spain> (visited on 12/20/2024).
- [95] Dena S. Davis. "Ethical issues in Alzheimer's disease research involving human subjects". In: *Journal of medical ethics* 43 (12 Dec. 2017), pp. 852–856. ISSN: 1473-4257. DOI: [10.1136/MEDETHICS-2016-103392](https://doi.org/10.1136/MEDETHICS-2016-103392).
- [96] José L. Molinuevo et al. "Ethical challenges in preclinical Alzheimer's disease observational studies and trials: Results of the Barcelona summit". In: *Alzheimer's and dementia : the journal of the Alzheimer's Association* 12 (5 May 2016), pp. 614–622. ISSN: 1552-5279. DOI: [10.1016/J.JALZ.2016.01.009](https://doi.org/10.1016/J.JALZ.2016.01.009).
- [97] Tim G. Götzelmann, Daniel Strech, and Hannes Kahrass. "The full spectrum of ethical issues in dementia research: findings of a systematic qualitative review". In: *BMC medical ethics* 22 (1 Dec. 2021). ISSN: 1472-6939. DOI: [10.1186/S12910-020-00572-5](https://doi.org/10.1186/S12910-020-00572-5).
- [98] *Reglamento - 2016/679 - EN - GDPR - EUR-Lex*. URL: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679> (visited on 12/27/2024).
- [99] *HIPAA Home* | HHS.gov. URL: <https://www.hhs.gov/hipaa/index.html> (visited on 12/27/2024).
- [100] *Terms of Use* | The A4 Study. URL: <https://www.a4studydata.org/terms-of-use> (visited on 01/14/2025).
- [101] *BOE-A-2011-9617 Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación*. URL: <https://www.boe.es/buscar/act.php?id=BOE-A-2011-9617> (visited on 12/27/2024).

12 Annex

The annex provides supplementary materials for a better comprehension of the project, including Python scripts and pipelines (Annex A), cognitive assessments (Annex B), quality control processes (Annex C), data analysis results (Annex D), and a presented scientific poster (Annex E).

12.1 Annex A: GitHub Repository FlexfMRI

The Python scripts, notebooks, and data processing pipelines developed for this project are available in a GitHub repository. The repository includes:

- **Data Structuring Scripts:** Tools for converting raw neuroimaging data into BIDS format.
- **Preprocessing Pipelines:** Flexible preprocessing framework with configurable parameters.
- **Quality Control Framework:** Scripts and tools for visual and quantitative assessment of fMRI data.
- **Functional Connectivity Analysis:** Notebooks for exploring and analyzing functional connectivity data.

Repository URL: <https://github.com/mbioscma/FlexfMRI>

12.2 Annex B: Cognitive Assessments

These cognitive tests were undertaken by participants in the A4 study to assess their cognitive abilities and perform screening prior to the clinical trial. Detailed descriptive statistics for the study group used in this project are presented in Table 1.

1. **Preclinical Alzheimer Cognitive Composite (PACC) Score:** The PACC is the sum of four z-scores derived from other cognitive tests, designed to detect subtle cognitive changes. Higher scores indicate better cognitive performance. For the A4 study, the mean score was -0.4 ± 2.7 .
2. **Wechsler Memory Scale Logical Memory Delayed Recall (LMDR):** This test measures the ability to recall details from a story after a delay. Scores range from 0 to 25, with lower scores reflecting fewer details recalled and indicating poorer episodic memory performance. For the A4 study participants, the mean score was 11.4 ± 3.3 .
3. **Mini-Mental State Examination (MMSE):** A widely used screening tool for cognitive impairment. Scores range from 0 to 30, with lower scores signifying poorer cognitive function. It evaluates areas such as memory, attention, and orientation. For the A4 study, the mean score was 28.7 ± 1.3 .

4. **Cognitive Function Index (CFI):** A self-reported and partner-reported measure assessing subjective concerns about cognitive function. Participant and partner responses are combined (range: 0 to 30), with higher scores indicating greater cognitive concerns. For the A4 study, the mean score was 3.8 ± 2.9 .
5. **Clinical Dementia Rating – Sum of Boxes (CDR-SB):** This scale measures global functional impairment, assessing areas such as memory, orientation, and judgment. Scores range from 0 to 18, with higher scores reflecting greater cognitive and functional deficits. For the A4 study, the mean score was 0.1 ± 0.2 .

12.3 Annex C: Overview and Analysis of the QC Framework

This annex provides supplemental information about the QC framework implemented in this project. It includes detailed motion correction graphs used for quantitative assessment (Figure 32), an example of an HTML QC report for a specific subject (Figure 33), statistics on subject rejection due to quantitative criteria (Table 7), and some detailed statistics of visual QC results highlighting its subjectivity (Table 8).

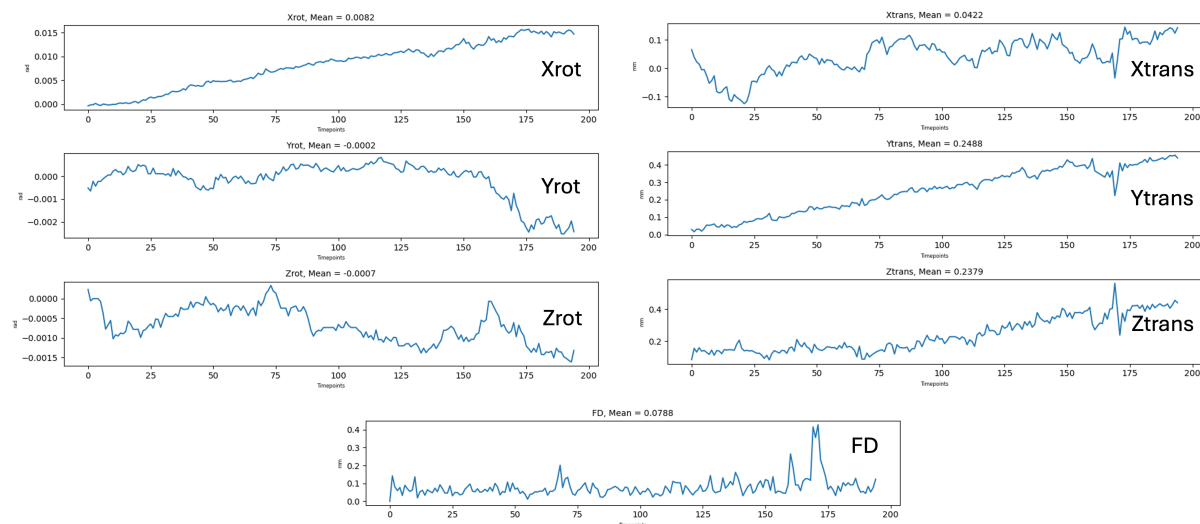
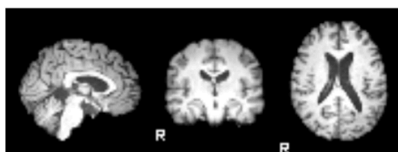


Figure 32: Motion parameters derived from MCFLIRT. The top six panels show rotational (Xrot, Yrot, Zrot) and translational (Xtrans, Ytrans, Ztrans) displacements across fMRI timepoints. The bottom panel displays framewise displacement (FD), integrating these six parameters to quantify motion per frame. Only FD was used for automated QC.

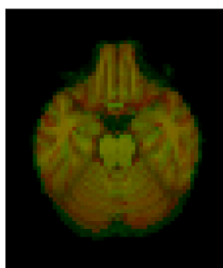
Processing Report for Subject sub-B82652828

Anatomical Processing

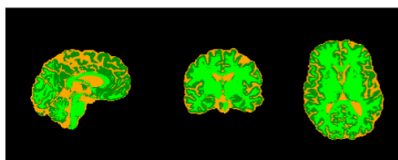
T1-weighted Image to MNI Template



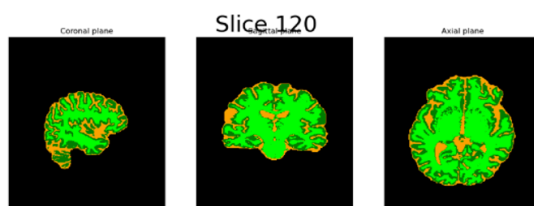
Comparison of T1 slices in GIF format



Segmentation of T1-weighted Image



Comparison of Segmented T1 slices in GIF format



Motion Correction Metrics

Parameter	Max	Mean	Std	Unit
Xrot	0.0028	0.0003	0.0005	rad
Yrot	0.0011	-0.0002	0.0003	rad
Zrot	0.0011	0.0001	0.0003	rad
Xtrans	0.0498	-0.0128	0.0216	mm
Ytrans	0.3772	0.0503	0.0763	mm
Ztrans	0.2737	0.0062	0.0483	mm
FD	0.9433	0.1441	0.1088	mm

T1 Segmentation Metrics

Region	Number of Voxels
GM	295708
WM	465127
CSF	441462
Total	1202297

Functional Processing

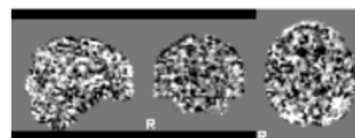
Functional BET Mask



Functional Data Registered to T1-weighted Image



Cleaned Functional Data to Template



Motion Correction Summary

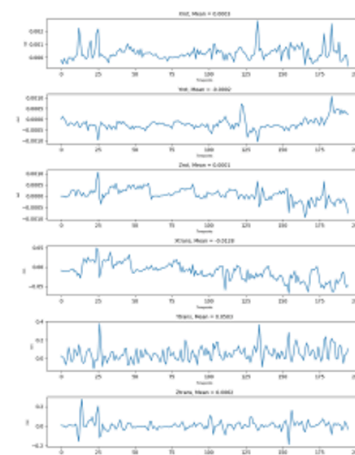


Figure 33: Example of an HTML QC report for subject *sub-B82652828*. The report displays outputs from anatomical and functional processing, along with motion correction metrics and T1 segmentation metrics, which are used as part of the qQC measures. These types of reports are integrated into the QC application.

Criterion	Pass (#, %)	Fail (#, %)	Threshold
FD max Pass (mm)	22 (43.1%)	29 (56.9%)	2.599
FD mean Pass (mm)	34 (66.7%)	17 (33.3%)	0.458
WM lower Pass (%)	47 (92.2%)	4 (7.8%)	0.341
CSF upper Pass (%)	47 (92.2%)	4 (7.8%)	0.402
GM upper Pass (%)	49 (96.1%)	2 (3.9%)	0.312
Total upper Pass (#)	49 (96.1%)	2 (3.9%)	1,320,292
WM upper Pass (%)	50 (98.0%)	1 (2.0%)	0.463
CSF lower Pass (%)	50 (98.0%)	1 (2.0%)	0.310
GM lower Pass (%)	51 (100.0%)	0 (0.0%)	0.171
Total lower Pass (#)	51 (100.0%)	0 (0.0%)	719,936

Table 7: Summary of qQC criteria results, outlining the reasons for rejection during this stage. "Upper" and "Lower" refer to thresholds set at deviations of 3 standard deviations from the mean. Most failures are attributed to motion-related measures, with maximum and mean FD being the primary contributors.

Metric	1 Reviewer (Group 1)	2 Reviewers (Group 2)	3 Reviewers (Group 3)
Subjects	471	460	182
"Yes" Ratings (Total)	341	784	379
"No" Ratings (Total)	22	24	31
"Maybe" Ratings (Total)	108	112	136
Perfect Agreement	N/A (only 1 reviewer)	363 (78.9%)	93 (51.1%)
All "Yes"	N/A	346	84
All "No"	N/A	3	1
All "Maybe"	N/A	14	8
Partial/Mixed Decisions	N/A	97 (21.1%)	89 (48.9%)
Total Unique Subjects			1,113

Table 8: Summary of reviewer agreement and ratings for subjects across groups with 1, 2, and 3 reviewers. The relatively low percentages of perfect agreement highlight the subjective nature of vQC. It is worth noting that achieving perfect agreement is inherently more challenging with three reviewers compared to two, but still these results highlight the subjectivity of vQC.

12.4 Annex D: Analyses of Temporal Filtering and RSN Clustering

This section presents the results of two analyses to better understand the data we worked with. Firstly, Figure 34 shows the power spectrum of network nodes timeseries which validates the filtering process applied. Secondly, Figure 35 shows how hierarchical clustering groups similar RSNs for further analysis.

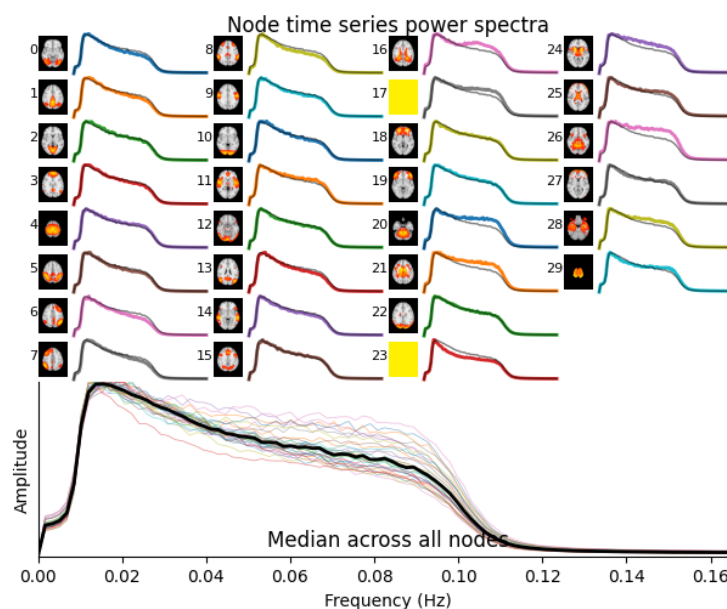


Figure 34: Power spectrum of time series for each network node derived from the ICA group decomposition. The lower panel shows the median power spectrum across all nodes, highlighting the frequency range between 0.01 Hz and 0.1 Hz. This range reflects the effect of the temporal Butterworth filter applied during preprocessing to remove noise, validating the filtering process.

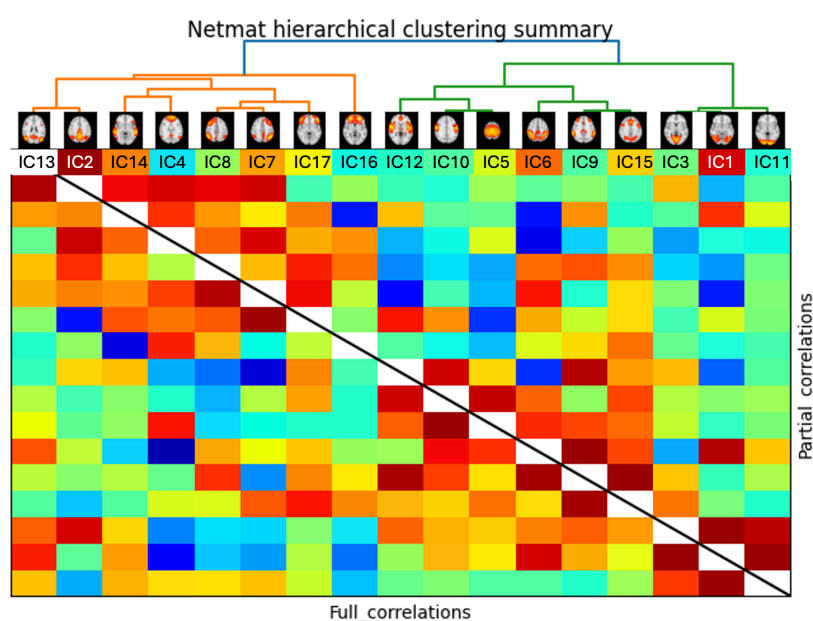




Figure 35: Hierarchical clustering of RSNs performed using the FSL Nets library, based on subject-specific time series. The resulting order, shown at the top of the matrix, aligns with visual identification, grouping similar networks together, such as the DMN and visual networks.

12.5 Annex E: MultiBrain Congress Poster


As part of this project, we had the privilege of presenting a poster at the International Multi-Brain Barcelona Congress II, held on November 13-14, 2024. Below are the poster and abstract that were presented.



Institut de Neurociències
UNIVERSITAT DE BARCELONA



EXCELENCIA
MARIA
DE MAEZTU



UNIVERSITAT DE
BARCELONA

From the cloud to the lab: Data organization and fMRI processing strategies for large-scale neuroimaging databases

Biosca, Marc ¹; Tudela, Raúl ^{1,2}; Wurm, Raphael ⁵; Ballestero, Marc ¹; Niñerola-Baizán, Aida ^{1,2,3}; Sala-Llloch, Roser ^{1,2,3}

¹ Institute of Neurosciences, University of Barcelona, Barcelona, Spain. ² CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain. ³ Fundació de Recerca Clínica Barcelona-Institut d'Investigacions Biomèdiques August Pi i Sunyer (FRCB-IDIBAPS), Barcelona, Spain. ⁴ Servei de Medicina Nuclear, Hospital Clínic de Barcelona, Barcelona, Spain. ⁵ Department of Neurology, Medical University of Vienna, Vienna, Austria.

Introduction

In recent years, **large multimodal neuroimaging datasets** have greatly advanced neurological research, however processing methods still vary widely.

- **Processed data:** Useful for direct analysis and hypothesis testing.
- **Raw Data:** Useful for methodological development and integrative studies.

Goal: Create a **standardized and flexible pipeline** to enable exploration of clinical and demographic factors, with **automated preprocessing** for consistent, reliable data analysis.

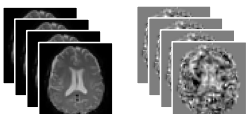


Figure 1. Example of fMRI raw data vs. processed fMRI data

Methods

1. Adaptation of the A4 study [1] raw dataset to the Brain Imaging Data Structure (**BIDS**).
2. Characterizing datasets using **k-means** clustering to group subjects by **demographic, cognitive, and clinical** traits.
3. Implementation of **automated and flexible** rs-fMRI preprocessing.
4. Development of a quality control (**QC**) tool utilizing **visual and quantitative** (FD and Segmentation Voxels) indicators.
5. Independent Component Analysis on preprocessed fMRI images

Characteristic	Mean ± SD	Characteristic	Mean ± SD
Age - yr	72.0±4.8	PACC score	-0.4±2.7
Female Sex – no. (%)	740 (58.7)	MMSE score	28.7±1.3
SUVr (¹⁸ F-Florbetapir)	1.3±0.2	CFI combined score	3.8±3.4
Centiloids (¹⁸ F-Florbetapir)	65.8±33	CDR-SB score	0.1±0.2

Table 1. Characteristics of subjects analyzed from A4 study dataset [1]

SUVr: Standardized Uptake Value Ratio. PACC: Preclinical Alzheimer Cognitive Composite. MMSE: Mini-Mental State Examination (0-30). CFI: Cognitive Function Index (0-30). CDR-SB: Clinical Dementia Rating – Sum of Boxes (0-18).

Results

1. We identified 3 cluster groups, characterized by distinct cognitive and amyloid profiles, using age, centiloid and PACC score as variables.
2. We implemented an automated fMRI preprocessing pipeline using Python scripts and open-source packages, summarized in Figure 3.
3. After evaluating the preprocessing results using the QC tool, we ran an ICA analysis. Results are shown in Figure 5.

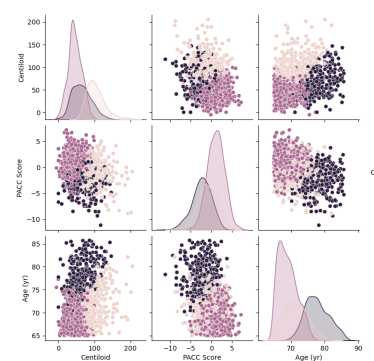


Figure 2. Distribution of Centiloid values, PACC Cognitive scores, and Age across identified clusters groups

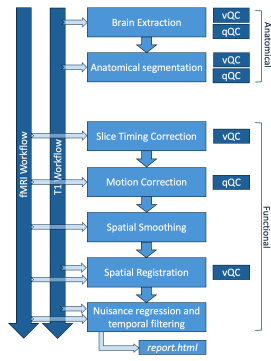


Figure 3. Summary of preprocessing pipeline with the visual (vQC) and quantitative (qQC) quality control checkpoints

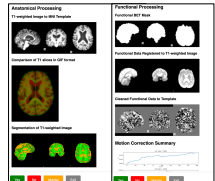
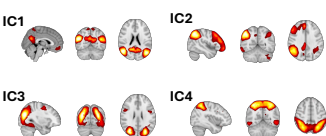


Figure 4. Examples of the quality control application



IC1
IC2
IC3
IC4

Figure 5. Spatial maps of some of the identified canonical resting-state networks

Conclusions

We utilized **raw images** from the A4 neuroimaging database in a laboratory setting and demonstrated that amyloid and cognitive variables can be used to delineate **cluster groups** among subjects. Additionally, we developed and tested a fully-automated and flexible **fMRI preprocessing pipeline**. Despite variability in acquisition parameters and clinical conditions, our approach successfully identified several **canonical resting-state networks** [2].


Acknowledgements

This work was supported by AGAUR, Generalitat de Catalunya (2021SGR00523) and the Spanish Ministry of Science and Innovation (PID2023-148152OB-I00/AEI/10.13039/501100011033 and María de Maeztu Unit of Excellence CEX2021-001159-M, within MICIU/AEI/10.13039/501100011033)


References

[1] Sperling, R. A., Donohue, M. C., Raman, R., Rafii, M. S., Johnson, K., Masters, C. L., van Dyck, C. H., Iwatsubo, T., Marshall, G. A., Yaari, R., Mancini, M., Holdridge, K. C., Case, M., Sims, J. R., Aisen, P. S., & A4 Study Team (2023). Trial of Solanezumab in Preclinical Alzheimer's Disease. *The New England journal of medicine*, 389(12), 1096–1107.

[2] Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), 13040–13045.



Agencia de Gestió d'Ajuts Universitaris i de Recerca



Grant CEX2021-001159-M funded by:

Page 63 of 64

Abstract



65 From the cloud to the lab: Data organization and fMRI processing strategies for large-scale neuroimaging databases

Biosca, M (1); Tudela, R (1,2); Wurm, R (1,5); Ballester, M (1); Niñerola-Baizán, A (1,2,3,4); Sala-Lluch, R (1,2,3)

- (1) Institut de Neurociències, Departament de Biomedicina, Universitat de Barcelona, Barcelona, Spain.
- (2) CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III.
- (3) Fundació de Recerca Clínic Barcelona-Institut d'Investigacions Biomèdiques August Pi I Sunyer (FRCB-IDIBAPS), Barcelona, Spain.
- (4) Servei de Medicina Nuclear, Hospital Clínic de Barcelona, Barcelona, Spain.
- (5) Departament of Neurology, Medical University of Vienna, Vienna, Austria.

INTRODUCTION: In recent years, large datasets of multimodal neuroimaging and clinical data have become available. While processed data aid hypothesis testing and statistics, raw images are valuable for methodological development and integrative studies. However, large variability exists in the processing of these datasets. We propose a pipeline for dataset standardization, performing clinical and demographics unsupervised exploration, and automated preprocessing for functional MRI (fMRI) data.

METHODS: We used the A4 study data, comprising older adults aged 65-85 with normal cognitive function or very mild cognitive impairment. Each participant had cognitive data, structural and functional MRI, and amyloid-PET available. Firstly, we adapted the raw dataset to the Brain Imaging Data Structure (BIDS). We characterized the database using k-means clustering on demographic, cognitive and clinical data. We implemented an automated fMRI preprocessing pipeline using Python scripts and open-source packages, allowing control over the different blocks, including motion correction, brain extraction and segmentation, spatial registration, spatial and temporal filtering, and nuisance regression. We also developed a quality control (QC) tool using visual and quantitative cues. Finally, we performed an independent component analysis (ICA) using FSL-MELODIC to evaluate the preprocessed images.

RESULTS: We included 1,323 individuals from the A4 database, with slightly elevated brain amyloid levels (as defined in the project protocols, using a Standard Uptake Value Ratio (SUVR) - threshold of 1.15). The clustering analysis identified 3 groups. A first subset of 250 passing the QC were included in a group ICA and showed common resting-state networks described in the literature.

CONCLUSIONS (1) We proved the use of raw images from the A4 neuroimaging database in the lab setting; (2) Identified 3 groups of subjects with different cognitive and amyloid profiles; (3) We developed a fully-automated, flexible fMRI preprocessing pipeline; (4) Further studies should explore the relationship between clusters and fMRI metrics.



Institut de Neurociències
UNIVERSITAT DE BARCELONA



EXCELEXIA
MARIA
DE MAEZTU