UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona

FIB

Optimization of tau-PET neuroimaging signal preprocessing: Application on Alzheimer's Disease

Alejandro Alarcón Torres

Bachelor Thesis

Specialization in Computing

Director: Juan Fortea Ormaechea

Ponent UPC: Luís Antonio Belanche Muñoz

April 19th, 2022

Contents

1 Context and Scope

1.1 Context

This is a Bachelor Thesis of the Computer Engineering Degree, specialization in Computing, taught by the Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya, and directed by Dr Juan Fortea and Dr Luís Antonio Belanche.

This thesis has been developed throughout an internship at Sant Pau Memory Unit, under the supervision of Dr Juan Fortea, Dr Jordi Pegueroles and Dr Victor Montal.

1.1.1 Introduction

For many years, research in the neuroscience field has urged to align different modalities (from different sources of data) of volumes and images to improve the information available from the clinical instruments used in neuroimaging. This process is called registration, and its complexity and variety have justified the development of a wide collection of methods.

A clear example that can benefit from such approaches is the study of a brain neurodegenerative disease: Alzheimer's Disease (AD). In recent years, many drugs and biomarkers have been developed to target specific proteins and molecules in order to cure this disease, or at least, stop its progression.

One of the key proteins that affect patients with AD, is the accumulation of tau protein, which will be the focus in this study. The in-vivo tracking of this protein is performed through tau-PET (Positron Emission Tomography) scans. The recent development of tau-PET imaging has allowed us to track the pathological accumulation of tau, and its topographical stereotypical pattern. Nevertheless, due to some technical limitations, there is a need for fusing its information with other imaging modalities in order to comprehensively study AD [10]. This aggregation of information is accomplished by using the registration process that will be explored in this thesis.

The main objective of this thesis is to determine the best possible approach to perform the registration process for tau-PET to structural MRI (Magnetic Resonance Imaging) scans, which is key in the neuroimaging field, allowing for a more precise extraction of the information encoded in these modalities of brain imaging. Moreover, we will compute and evaluate the diagnosis accuracy of using the Standardized Uptake Value ratio (SUVr) to differentiate between healthy controls and patients with AD.

We hypothesize that the outcomes from this project, which will optimize the registration of tau-PET to structural MRI scans, will help to comprehend the pathophysiological alterations that occur along the disease. Moreover, the resulting optimal pipeline would be of interest to be used in future clinical trials, as an ideal toolset to analyze tau-PET imaging data as a surrogate marker of drug efficiency in the AD field.

1.1.2 Terms and concepts

Alzheimer Disease

AD is the main cause of dementia, and is rapidly becoming one of the most expensive, lethal and burdensome diseases of this century [4]. Major developments in the last decade have helped to better understand the underlying pathology, the identification of multiple causative and protective genotypes, the identification of new blood-based and imaging biomarkers, as well as the initial signs of the benefits from disease-modifying treatments and lifestyle interventions. AD's strongest effects are on memory and executive functions, which impairs the daily activities from those suffering it. Depending on the stage of the disease, we can classify patients in three main groups: cognitively unimpaired subjects (HC), mild cognitive impairment (MCI) - which have moderate clinical alterations that slightly impact the daily activities - and Alzheimer's Dementia subjects (AD) [21].

Amyloid beta Protein

Amyloid beta (Aβ), is one of the pathological hallmarks of AD. Although scientists don't know the etiology of AD, several reports have pointed to the idea that Aβ deposits in the brain may play a role [23]. Aβ is generated in small amounts in the general population's brains. In most cases, the body is able to get rid of it, but for patients suffering from AD it can't be cleared out, and it builds up into extracellular Aβ plaques.

Tau Protein

Tau proteins are tubular structures that appear in the Central Nervous System. Their main role is strengthening the neurones' axon structure. When these proteins become abnormally hyperphosphorylated, they blend forming neurofibrillary tangles, leading to neuronal death, neuroinflammation and the subsequent neurodegeneration. These pathological alterations are believed to be related to several nervous system diseases, such as AD or Parkinson's [18], acting as triggers for neurodegeneration.

Concretely, recent research in AD has shown that the pathological accumulation of tau is strongly related to atrophy and cognitive decline [19]. Importantly, these associations are not found for the accumulation of Aβ. Thus, it is important to develop novel methodological frameworks to further study the relevance of tau in the disease process.

Neuroimaging Techniques

The study of the brain presents difficulties that do not appear in other bodily organs. As a critical and extremely delicate organ, its accurate study in-vivo is impossible given the invasiveness that would be required.

Neuroimaging techniques allow us to obtain a snapshot of the brain status through a non-invasive approach. There are many different imaging modalities that allow us to capture a variety of information, from anatomic and morphological to protein accumulation rate.

Consequently, it is crucial to propose a systematic framework that allows to integrate the diverse information of various imaging modalities, allowing the aggregation of morphological characteristics to the accumulation of specific proteins such as Aβ and tau.

Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is based on magnetic field gradients quantification, and is used in the medical field to capture information about the anatomical structure of the body, as shown in Figure 1.

This anatomical information is useful to track structural brain alterations, related to different neurodegenerative processes and pathological alterations, along the disease.

The most commonly used (and informative) brain MRI acquisition is fMRI-T1-weighted, which focuses on maximizing the differentiation between grey matter and white matter.

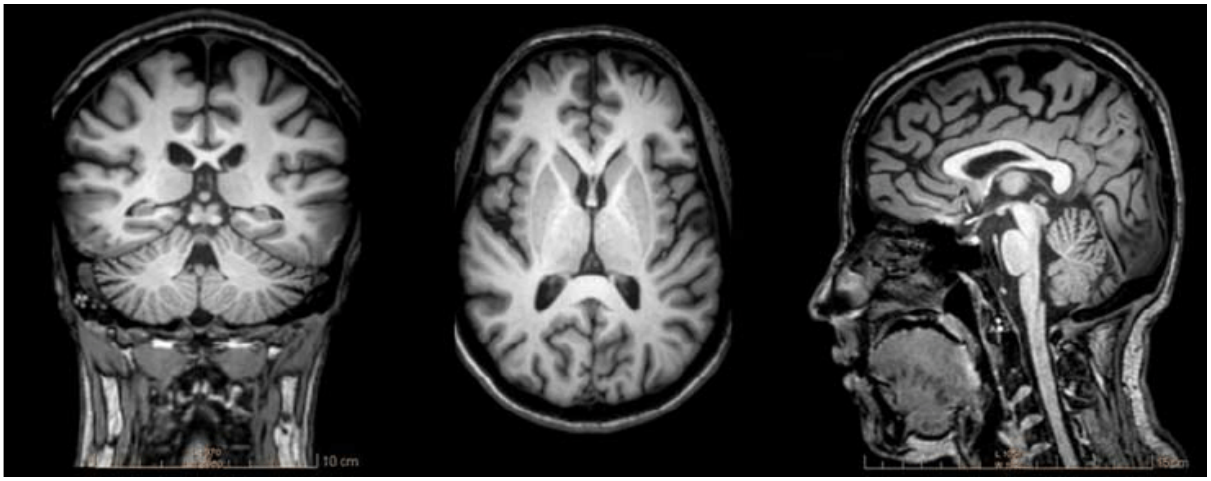


Figure 1: MRI-T1 scan[1].

Positron Emission Tomography

Positron Emission Tomography (PET) is a form of medical imaging that uses small quantities of radioactive tracers to measure and quantify the amount of specific molecules in the brain.

PET is a functional technique, as it provides a quantification as well as the topography of specific processes that occur within our organism. One of the most interesting utilities of PET in the neurodegenerative research field, is that it can be used to measure the amount of certain proteins in the brain. Nevertheless, PET data has lower resolution compared to other imaging techniques such as MRI (Figure 2). It is also more invasive than MRI, as a prior injection with radioactive isotopes that bind to specific proteins is needed.

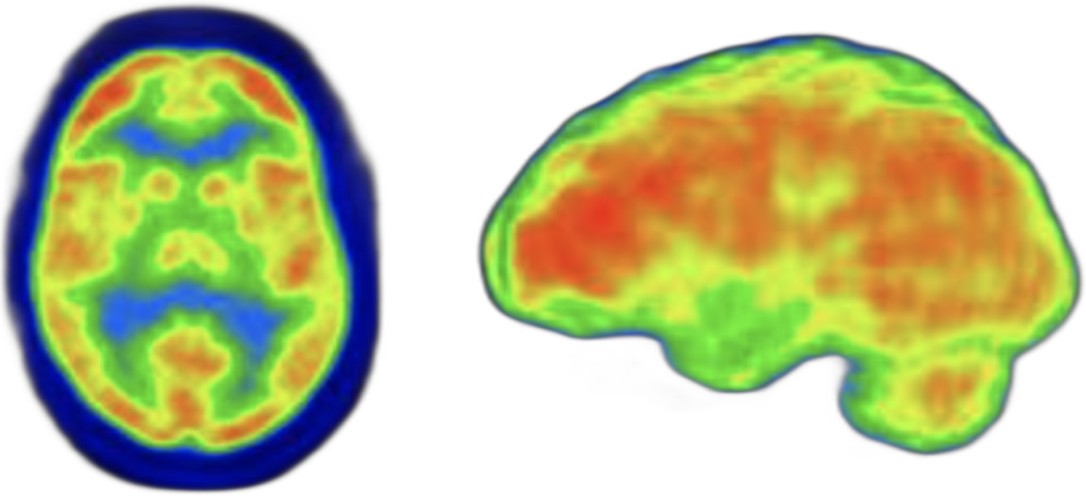


Figure 2: Slice and volumetric visualization of a PET scan [17]

Registration

Registration between two data sources consists on a geometrical alignment that allows the fusion of the datasets aggregating their information. In the field of neuroimaging, the registration process is performed between volumes (i.e. a 3D matrix composed of voxels, 3D representations of pixels, that encodes the imaging information using signal intensity) and is splitted into two categories: co-registration and standard normalization. In Figure 3 it can be observed how the registration process manages to align the original images (Figure 4).

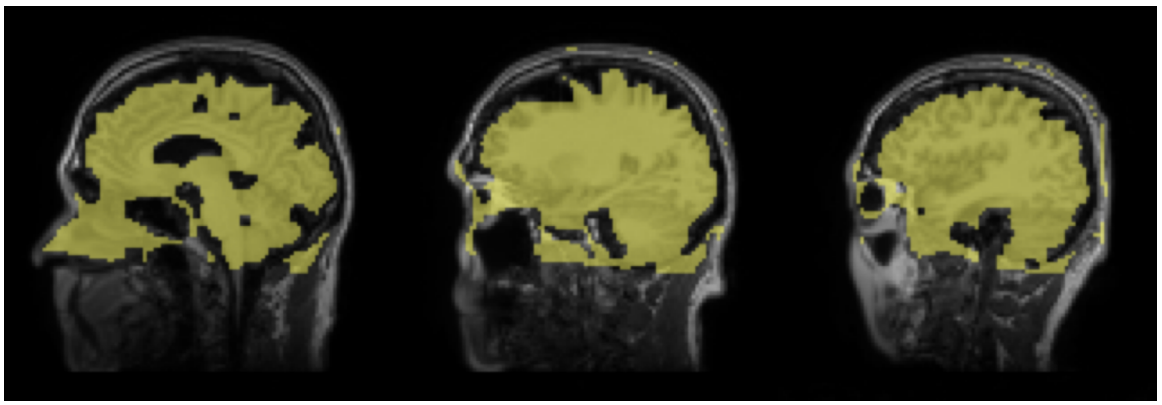


Figure 3: Superposition of tau-PET supra-threshold mask to MRI after registration.

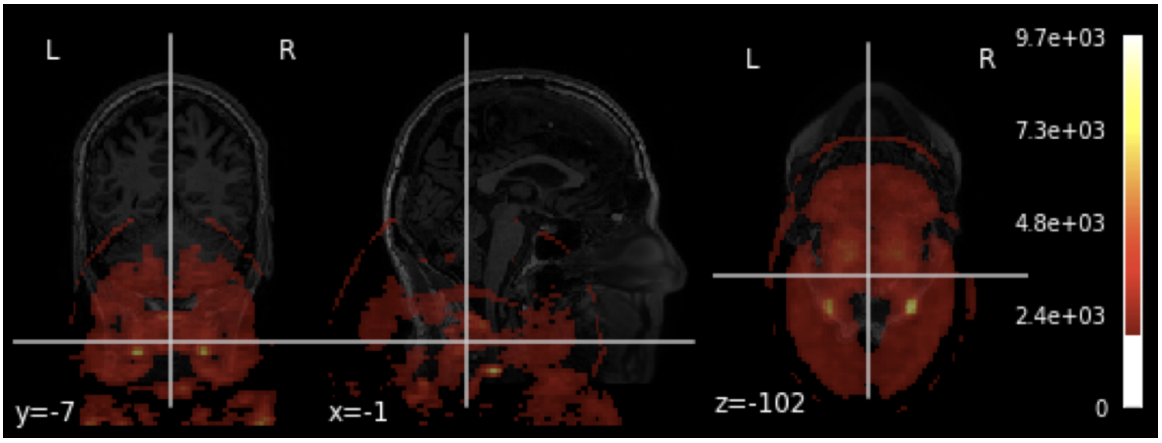


Figure 4: Superposition of tau-PET supra-threshold intensities to MRI without registration.

Co-register and Standard Normalization

On the one hand, co-registration [13] is the alignment of two images belonging to the same subject but obtained through different modalities. In this project, the focus will be on the co-registration of tau-PET images to MRI. Tau-PET generates very low anatomical information images, this is why in the fields of neuroimaging, it is of vital interest to achieve a good alignment between the PET and MRI images, as shown in Figure 5. Such process allows, by combination of both imaging modalities, to determine with high precision in which regions tau is prone to aggregate.

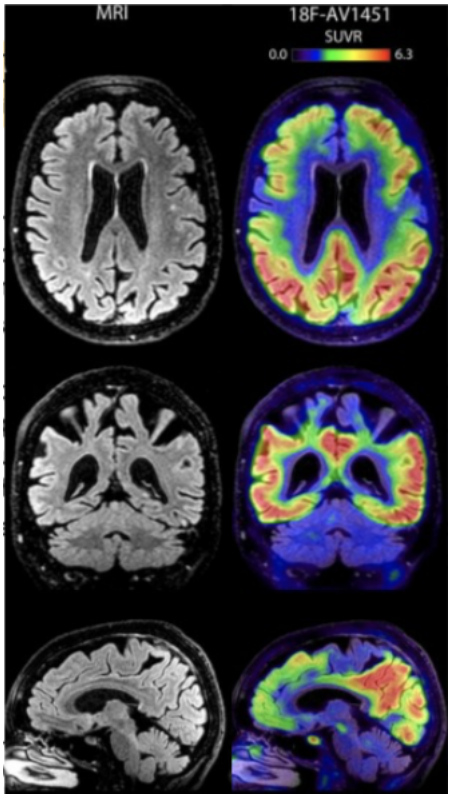


Figure 5: MRI T2-weighted-Fluid-Attenuated Inversion Recovery (FLAIR) in axial, sagittal, and coronal planes (Left) and 18F-AV1451 tau-PET superposition (Right) in patient with probable AD [18].

On the other hand, Standard Normalization, addresses the registration of individuals images to a standard space that facilitates their comparison. These two images, however, must belong to the same modality.

Standard spaces exist as a convention within the field of neuroimaging and have been generated from averaging MRI information from a set of subjects. There is a wide variety of templates available which have been generated from different sets of subjects, including specific subsets from all possible stage classifications throughout the AD continuum, or even from newborn population. The template we will be using in this project is the MNI152[1], which is one of the most commonly used in the field.

In Figure 6 we can observe the inputs of the standard normalization process as well as its output. We can see how the normalized subject's image adapts to the template while maintaining its anatomical information.

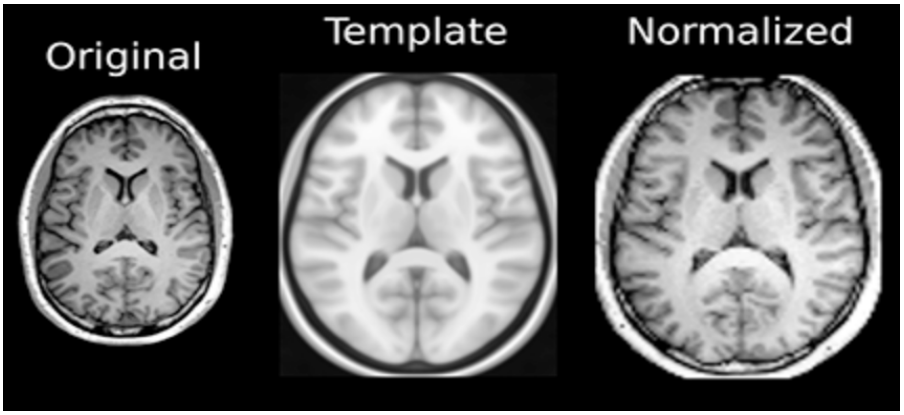


Figure 6: Standard Normalization inputs (left and middle) and output (right).

Standard Normalization and Co-registration are not exclusive. Their concatenation is indeed of great interest to the scientific community as it allows the possibility of moving/normalizing tau-PET images, and subsequently, their information, from the subject space into a standard space (Figure 7). This process facilitates the comparison between subjects.

[1]MNI152 standard-space T1-weighted average structural template image was created in the Montreal National Institute averaging 152 high-dimensional non-linear registrations.

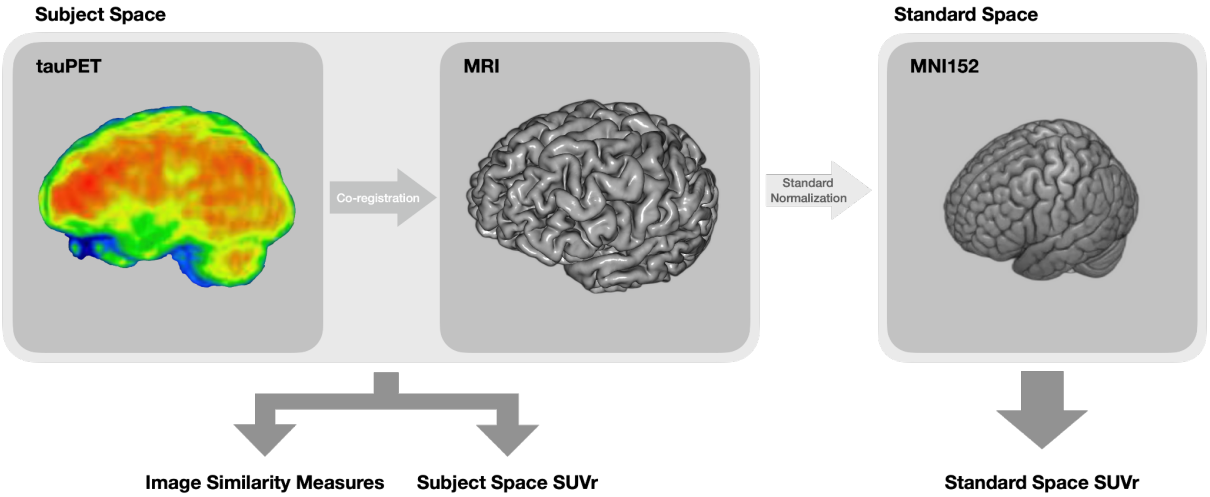


Figure 7: Illustration of the spaces and different images involved in the tau-PET to MRI registration process and the different measurements captured throughout the process.

Inputs and Outputs

In order to perform a registration, two images are required: a moving image and a fixed image. The modalities of the input images differ depending on the type of registration to be performed.

On the one hand, if a co-registration is intended, the two images should belong to the same subject. For the present project, the fixed image should be an anatomical image of high resolution, such as an MRI, and the moving image could be a PET image, which hardly encodes any anatomical information.

On the other hand, if a Standard Normalization is desired, the fixed image will be a template and the moving image should be the subject's capture. Both these images should belong to the same modality, and preferably, have high-resolution structural information.

Furthermore, if the registration parameters specify non-linear transformations (local map of small distortion/displacement), a gradient map (or Jacobian matrix) will also be generated, which would be applied to the volume resulting from the application of the transformation matrix, and which would apply the non-linear transformations.

Finally, and as a result of the registration process, a linear transformation matrix is obtained, which encodes the linear transformations to be applied in order to transform a point from one's image coordinate system to the other's.

1.2 Problem to solve

The main aim of this study is to determine which are the optimal parameters, between a sort of different registration and initialization approaches, to obtain the best quality registration between tau-PET and MRI scans in the context of AD.

1.3 Stakeholders

The project has many involved parties. First off, the scientific community, specially the areas related with neuroscience and neuroimaging, would be benefited from it when performing the analysis and diagnosis of new subjects with potential AD.

Secondly, the pharmaceutical industry could use the results shown in this thesis to better study the specific proteins and molecules that are thought to be the origin and trigger of the neurodegenerative processes that appear throughout the disease. Such improved pipeline could be also used in clinical trials, to track the effect of diverse drugs over the accumulation of tau.

Finally, all the potential AD patients would benefit from it, as a result of a better diagnosis, and potentially a better treatment of their disease.

1.4 Motivation

In the time being, the main imaging modalities used for the study and monitoring of AD are the Aβ-PET and tau-PET. However, in recent studies, it has been shown that tau protein deposition is more closely related to neurodegenerative factors, compared to Aβ [19].

In consequence, there has been a great increase in the amount of tau-PET scans performed in the recent years, generating the need for a robust post processing that allows for a better extraction of the information captured in these scans to enhance the study of the disease, as well as the development of new drugs.

1.5 Previous studies

There are very few studies that focus on the optimization of the parameters in the registration process in order to optimize the alignment of tau-PET and MRI images. However, there are similar studies that focus on the alignment of MRI with other PET modalities, such as FDG-PET [3] and Aβ-PET.

Previous reports on those different PET tracers cannot be applied directly to tau-PET data. Tau-PET has its particularities, one of which is the inconsistent topology of the tau protein depositions throughout the brain structure and its pattern along the disease [14], in contrast to the distinguishable pattern that has been described from Aβ [12]. Moreover, in recent studies, it has been characterized how tau-PET imaging suffers from off-target binding and quantification [14], which might impact the registration process.

1.6 Scope

In order to achieve its objectives, this project has been divided into several sub-objectives that will serve as mid-term goals throughout is development.

1.6.1 Objectives and sub-objectives

Completing initial formation

Due to the specificity and complexity of tasks that are carried out during the registration process in the field of neuroimaging, one of the first steps that serves as an introduction to this field is the habituation to the use of specific software.

This software includes both python libraries related to neuroimaging, as well as programs that complete the whole process of registration. This software also includes some visualization tools in addition to generic python libraries focused on data handling (i.e. Numpy and Pandas).

Understanding the necessary concepts related to the biologic field

The specificity of the field of neuroimaging makes it necessary to understand the biological processes taking place in the brain, for a meaningful interpretation of the results.

This project aims to focus on the accumulation of TAU protein. However, it is important to understand the causes and consequences of this accumulation, as well as contextualizing it within the continuum of AD for relating it to other factors [16].

Familiarization with the most used registration methods

In recent years, much emphasis has been put on the automation of the registration process. There are numerous studies that suggest the suitability of different methods to different types of registration.

The study of these already existing methods is of vital importance for a better understanding of both the context of the registration process as well as the particularities of the different types of methodologies, as it can help immensely in the creation of a new framework for the registration of tau-PET to MRI images.

Generation of scripts to perform co-registration and standard normalization

Once familiarized with the software and libraries used in the process, as well as the necessary concepts within the field of neuroimaging, it is possible to proceed to the automation of the registration process, splitted into the phases of co-registration and standard normalization.

Various scripts will be programmed using python ANTsPy and FSL libraries to automate the registration process and the parameterization of the necessary fields for subsequent statistical analysis.

Bulk execution of registrations in cluster

Due to the time required to process each registration, and given the size of the dataset, as well as all the parameter configurations that will be involved in the processing, a cluster set up at the Sant Pau Memory Unit will be used. For that matter, a previous setup is required, based on the creation of a hierarchy of files accessible as well as the parallelization of the process and adaptation to the environment.

Quality Control of outputs

After the registration processes have been completed, a quality control is needed to assure the results are good enough. Some images will be generated for every registration (figure 8) in order to facilitate the quality control process, based on the superposition of the transformed image on top of the moving image to visually determine the viability of the results.

It is also necessary to start preparing the datasets containing all the quality measurements that have been captured in order to perform a statistical analysis afterwards.

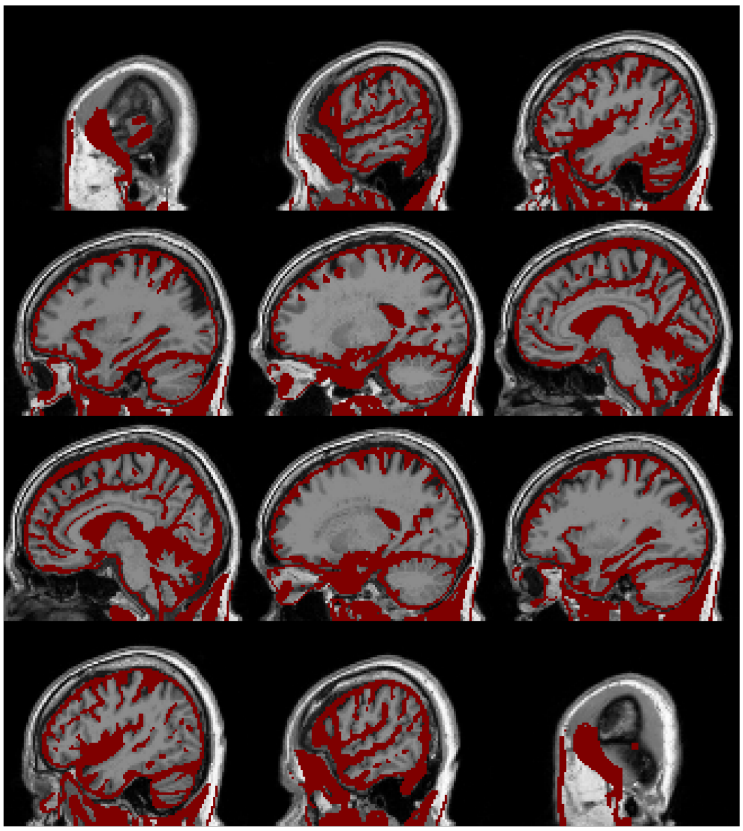


Figure 8: Quality Control image generated throughout the execution of the registration process. The red mask corresponds to the cerebrospinal fluid (CSF), which surrounds the brain and should align perfectly with the cranium.

Statistics for determining the best registration approach

In order to select the best approach, we will base our decision on statistical results, and not only on visual and quantitative measures. To that end, we will statistically compare similarity measures of the different configurations in order to optimize the registration algorithms.

Furthermore, once the best configuration is selected, we will also statistically compare the SUVr measures computed through two different schemes (Figure 7).

This statistical analysis will be structured in 2 blocks, each one addressing one of the main objectives of this project. The first block will be sub-structured in 3 studies, whereas the second will consist of 2 studies:

Block 1 - Determining the configuration of parameters that optimizes the registration process.

- **Study 1** - Characterization of the 10 best registration configurations
- **Study 2** - Statistically significant differences between registration configurations
- **Study 3** - Statistically significant differences throughout the AD Continuum

Block 2 - Evaluating the differences between computing the SUVr in different spaces: subject space vs. standard space.

- **Study 4** - Determining the best space to compute the SUVr
- **Study 5** - Classification and prediction capabilities of derived data-driven methods

These studies will frame the structure of the statistical analysis and exhibit the robustness of the procedures that have been implemented in order to assure the validity of the results.

1.6.2 Requirements

Regarding computer science, the requirements for this project are the knowledge of several programming languages (Python, R, Bash).

In the biomedical research field, the requirements are knowing and understanding the neurodegenerative processes that take place throughout AD, as well as the causes that have led to its development.

Finally, it is of great importance to understand the different information and specificity encoded in the different imaging techniques, more precisely, the meaning of the values stored for each voxel as well as the actual topography.

1.7 Methodology and rigor

1.7.1 Methodology

For the purpose of this study, the methodology used will be grounded on decades of work in the optimization of the registration between MRI to other PET modalities such as Aβ-PET or FDG-PET. We will take advantage of several registration algorithms implemented on the packages FreeSurfer, FSL and ANTsPy libraries. These are extensively used in the field of neuroimaging.

Regions of interest

Brain Regions Of Interest (ROIs) surge as an intent of segregating different brain sub-structures based on their functions, purpose or anatomy. This set of ROI identifications are usually referred to as atlas or parcellations. In this study, we will work with the Desikan atlas [5] (Figure 9), one of the most commonly used in the AD field.

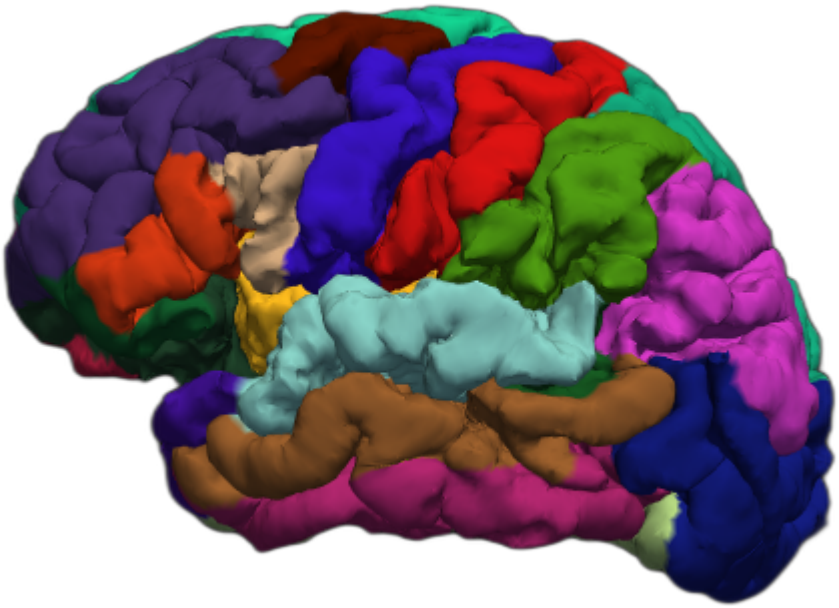


Figure 9: Pial representation of the regions of interest in one hemisphere generated with the Desikan atlas. [5]

A large number of studies have tried to define, contrast and characterize the topography of the accumulation of tau protein in-vivo inside the brain structure throughout the AD continuum [15]. As a result, some ROIs have been characterized for providing information on the tau protein accumulation spreading (Figure 10), whereas others are known for not accumulating this protein and serve as contrast.

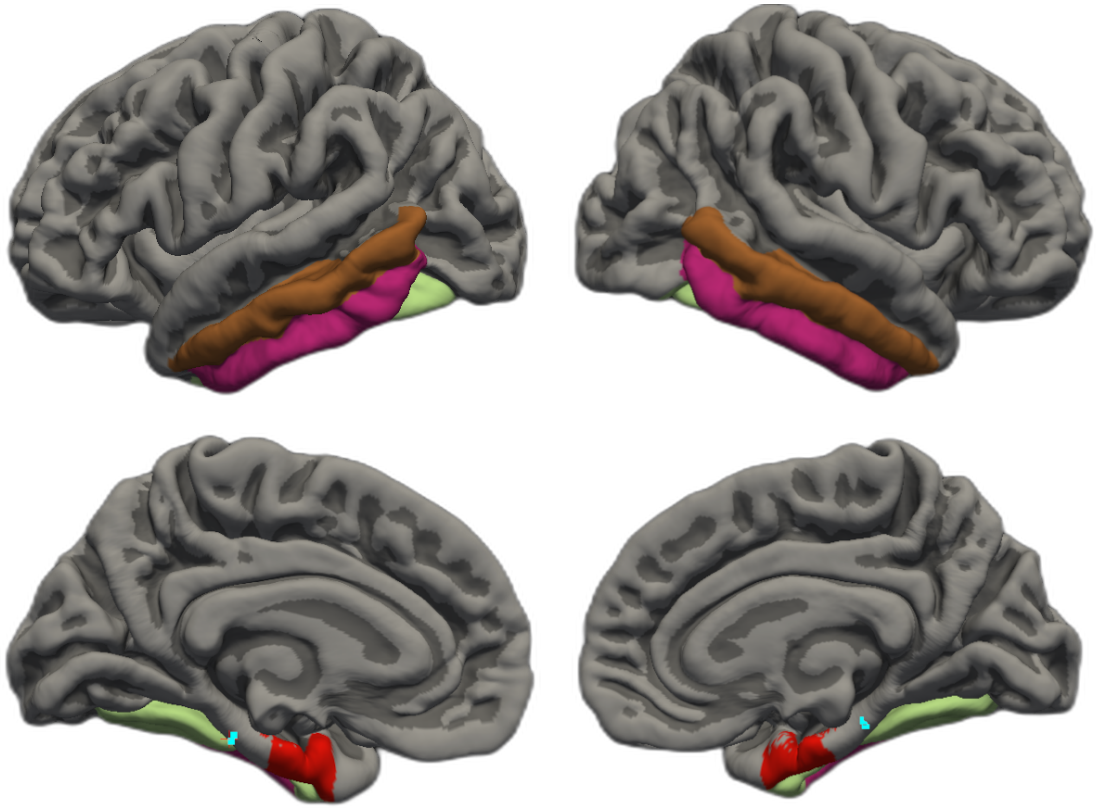


Figure 10: ROIs used to compute the mean tau SUVr dividend which stores information regarding the levels of tau deposition of a subject. [Víctor Montal]

Need for intensity normalization

The values derived from the capture of a tau-PET scans have a certain bias due to factors such as unequal basal depositions of the protein among the subjects [24], properties of the injected tracer or even the machinery involved in the capture. Intra-individual differences should be taken into account [14]. As an example, the basal level of tau accumulation of two participants of the same age and gender in analogous stages of the disease, might differ. Henceforth, and taking into account that the final intention of the process is to perform a comparison between subjects, a standardization of these values is necessary.

SUV (Standardized Uptake Value) is a term used for semiquantitative analysis in nuclear medicine. It is used by professionals to distinguish between "normal" and "abnormal" levels of uptake [2] (Figure 11).

Subsequently, dividing the target ROIs' SUV by the average of the SUVs of the contrast ROIs, a ratio known as SUVr (Standardized Uptake Value ratio) is generated (Figure 12), which diffuses the bias derived from the basal SUV of the subject [22] and facilitates their comparison. Importantly, the reference region that is used to normalize, is a region that has been shown to not accumulate tau along the AD progression. Thus, we are always correcting for basal inter-individual differences, without accounting for disease-related increments of tau-PET signal hyperintensities.

$$SUV = \frac{\sum_{i \in ROI} voxel_i}{|ROI|}$$

Figure 11: Expression used to compute the SUV values of the different brain ROIs.

$$SUV_R = \frac{\frac{\sum_{i \in ROI_{tau}} SUV_i}{|ROI_{tau}|}}{\frac{\sum_{i \in ROI_{Basal}} SUV_i}{|ROI_{Basal}|}}$$

Figure 12: Expression used to compute the SUVr value of a subject.

Register configurations

In this project, up to 27 different configurations for the tau-PET to MRI registration will be evaluated (Figure 16). These configurations will be generated by combining the different possible values for the relevant parameters needed to conduct the registration process.

Algorithms

There is a wide variety of algorithms that have been developed in recent years for the registration of images in the field of neuroimaging. However, as a result of the evolution of technology, more exhaustive algorithms that take advantage of local cost functions that generate more information in exchange for greater complexity have been flourishing over time.

Moreover, these algorithms present different paradigms for the approach of the registration process. Some treat the image directly, whilst others rely on a first computation of the cerebral white matter and grey matter surfaces in order to align them afterwards [9].

Bear in mind that the methods that have demonstrated better results for the registration of other PET modalities, such as Aβ-PET, do not necessarily generate a good result for tau-PET, as there are significant differences in the topographies and accumulations presented for different proteins, that eventually become decisive when facing the registration computationally.

In this study we will be using 3 different algorithms:

- MRI-coreg[2] - This algorithm performs a registration based on Information Theory. We will use the Freesurfer implementation of this algorithm.

[2]More information available at https://surfer.nmr.mgh.harvard.edu/fswiki/mri_coreg

- BB-register[3] - This algorithm is extensively used for within-subject cross-modal registration (corregistration) by using boundary-based cost function. We will use the Freesurfer implementation of this algorithm.
- W-register - This algorithm is indeed the same as the previous but with a specific set of parameters that try to align the white-matter surface instead of the grey-matter surface. This should have some benefits due to the topology of Tau protein deposition. It's based on the idea behind BB-register, but we have developed some tuning of the parameters in-house.

Degrees Of Freedom

Another one of the most relevant parameters when executing a registration, is the degrees of freedom (DOF), which state the different deformations that can be carried out for the linear transformations to be applied to the images, in the form of rotations and scalings for the different axes. The typical values are 6, 9 and 12. Furthermore, the registration can also be carried out elastically, which, in addition to the expected linear transformations, applies non-linear conversions, which are encoded through a Jacobian matrix. Importantly, such elastic transformations are usually performed in the standard normalization process, which involves images with high-resolution structural information. It is uncommon to use elastic algorithms for multimodal images, such as tau-PET and MRI, where the information encoded is significantly different.

Initialization

The initialization method is another crucial parameter when trying to enhance the registration process. It determines how an initial solution from where to start an optimization should be generated. Some of the simpler initialization methods are based on centroid or dimensional alignment, whilst the more complex undertake some linear deformations or shearings in order to generate a better fit.

In this study, 4 different initialization methods will be considered:

- FSL - This method performs an affine transformation in order to generate an initial solution.
- SPM - Similar to FSL, but with slight optimization criteria, as it is implemented within the SPM package.
- COREG - This method performs few iterations of the MRI-coreg algorithm in order to generate an initial solution.
- RR - This method performs a robust transformation to generate an initial solution.

Cost Functions

Lastly, we also need to introduce the cost functions. These are used for both optimization throughout the process as well as solution evaluation. Note that there is a large number of cost functions available, some work locally, considering only surrounding voxels, while others work globally, taking into consideration the entire volume at once.

[3]More information available at https://surfer.nmr.mgh.harvard.edu/fswiki/bbregister

In this study we will be using 6 different cost functions to measure the image similarity that should be enhanced during the registration process and, which should be indicative of the quality of the registration. The fact of contemplating different cost functions prevents our study from being biased by evaluating the registration quality with the same cost function that has been optimized throughout the registration process.

This 6 different cost functions are:

- Mattes Mutual Information
- Correlation
- Demons
- ANTs Neighborhood Correlation
- Joint Histogram Mutual Information
- Mean Squares

1.7.2 Validation

Throughout the fulfillment of the project, different methods of validation were involved.

First of all, during the registration executions for the entire database of subjects, quality control mechanisms were implemented in order to facilitate the process. These images were reviewed by specialized researchers in order to discard gross registration failures.

As far as statistics concern, the general structure as well as the specific tests and hypotheses to be contrasted, were evaluated by specialists in order to verify the structure and validity of the results.

Importantly, the implementation, and the scrutiny of the results will be reviewed by experts in the field to ensure its veracity as well as robustness.

2 Project Planning

The duration of this project is estimated at 5 months. Since it is part of a TFG + Internship agreement, it is expected to be a 40h weekly commitment, executable both in presentially and remotely.

2.1 Task definition

2.1.1 Initial formation

The first step is to get familiar with the work environment at the level of software. For this, some very generic libraries for data processing will be installed and I will be familiarized with their main functions, as well as libraries more concrete related to the field of neuroimaging.

The generic libraries for data processing are NumPy and Pandas. These libraries are characterized by working very efficiently with datasets and numeric operations.

On the other hand, the Nilearn and ANTsPy libraries are widely extended in the field of neuroimaging. Both provide implementations for the most commonly used methods for registration of cerebral images, providing great variability for possible parameters.

2.1.2 Field Concepts

It is important for the project to thoroughly understand some basic concepts related to AD. Furthermore, it is also of great relevance to study in depth the registration process functioning, as well as the different characteristics of each method, that could end up facilitating the optimization of tau-PET to MRI registration.

2.1.3 Registration Methods

The most frequently used algorithms for co-registration and standard normalization will be studied in detail. For standard normalization the ANTsPy library will be used, whereas for co-registration we will use some of the algorithms incorporated in the Freesurfer framework, namely BB-register and MRI-coreg.

2.1.4 Standard Normalization and co-registration

In this part of the process, the scripts to carry out the registration process will be created, separating on the one hand the standard normalization, and on the other hand, the co-registration.

For that matter, the parameter space should be determined, that is, the variety of parameters which will be taken into account and their domains, and for which all the possible permutations should be generated for its subsequent execution.

The final program should perform the full transformation of the tau-PET image from subject space into an MRI standard space for its analysis. Throughout the process, it is extremely important to generate the necessary quality control mechanisms that allow for a quick quality control of the results, in order to verify that the processes are fulfilled successfully.

2.1.5 Bulk execution

Once the programs are generated, they will be adapted to be executed within the cluster environment. Finally, these programs will be executed for the full sample of subjects that form the dataset. Importantly, we will include participants from different cognitive profiles to assure variability and robustness on the later analyses.

2.1.6 Quality Control of the results

Once all necessary executions have been performed, a brief quality control process should be carried out to ensure that the executions have been successful and that the results obtained are valid solutions, that is to say, to ensure that the optimization process has not strayed in local minima throughout the process generating an invalid registration.

2.1.7 Statistical Analysis

The final part of the project will focus on the statistical analyses of the data obtained after the registration processes have been completed. There are two objectives proposed.

The first is to determine the optimal configuration of parameters to perform the registration of tau-PET to MRI images by assessing the quality of their outputs as well as their robustness throughout the AD continuum.

The second objective is to appraise the classification and prediction capabilities of data-driven methods derived from the SUVr values computed in different spaces as well as some demographic co-variables.

2.1.8 Conferences

Throughout the internship, specific knowledge related to the project will be acquired by attending diverse talks and journal clubs where state-of-the-art reports in the field will be discussed. These conferences entail a better understanding of the field itself, as well as an extremely accurate contextualization of the methodologies and studies that are being carried out by the highest level institutions in the field. Such presentations will be performed by experts on the neuroimaging field from the Sant Pau Memory Unit.

Furthermore, I personally had the chance to perform two presentations for the Neuroimaging Research Group myself, one about the registration process, with a special consideration to Boundary-Based registration, and another laying the structure and results of this thesis.

Preparing both these presentations really helped structuring the concepts and mechanisms, and I believe performing them in front of the Neuroimaging Research Group really gave me an insight of how my thesis defense would go and how to prepare for it.

2.2 Resources

2.2.1 Database of subjects (S)

The data used in the preparation of this project was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

The dataset considered consists of a total of 180 subjects, of which images in both tau-PET and MRI have been acquired (Table 1).

Characteristic	CN, N = 70[1]	EMCI, N = 36[1]	LMCI, N = 48[1]	AD, N = 28[1]
Age	69 (66, 72)	73 (65, 77)	72 (65, 78)	74 (70, 80)
Unknown	1	0	0	0
Gender				
Female	43 (61%)	14 (39%)	20 (42%)	11 (39%)
Male	27 (39%)	22 (61%)	28 (58%)	17 (61%)
APOE4				
0	47 (67%)	21 (58%)	26 (55%)	7 (25%)
1	20 (29%)	11 (31%)	15 (32%)	13 (46%)
2	3 (4.3%)	4 (11%)	6 (13%)	8 (29%)
Unknown	0	0	1	0

[1] Median (IQR); n (%)

Table 1: Summary of the demographic data of the subjects.

2.2.2 Computational Cluster (C)

The cluster is composed of 6 intel i7 CPUs, adding up to 48 cores; 128GB of RAM and 48TB of storage. The operative system is CentOS 8 and it uses Slurm as workload manager for the parallelization of processes and management of resources.

2.2.3 Designated Computer (D)

In the assigned personal workspace inside the IIB facilities, a PC with an 8-core Intel i7 CPU, nVidia GTX 1050 GPU and 8GB of RAM has been provided. The operating system is PopOS 20.04.

2.2.4 Personal Laptop Computer (L)

Finally, a personal Apple Macbook Pro 2021 with a M1 Max processor with 32GB shared memory.

In Table 2 we can observe the resources' dependencies, as well as time estimations and dependencies, which will be chronologically exhibited in Figure 13 in a more visually appealing and informative manner through a Gantt diagram.

Code	Job	Time	Dependencies	Resources
T1	**Initial Formation**	**30h**		
T1.1	Numpy & Pandas	10h		L, D
T1.2	Nilearn	10h	T1.1	L, D
T1.3	ANTsPy	10h	T1.1	L, D
T2	**Field Concepts**	**110h**		
T2.1	Alzheimer's	20h		D
T2.2	Registration	30h	T2.1	D
T2.3	PET-TAU registration	30h	T2.1, T2.2	D
T2.4	Novel Techniques	30h	T2.1, T2.2	D
T3	**Registration Methods**	**40h**	**T2**	
T3.1	BB-register	10h		D
T3.2	MRI-Coreg	10h		D
T3.3	Flirt	10h		D
T3.4	ANTsPy	10h	T1.3	D
T4	**Standard Normalization & Co-registration**	**90h**	**T1,T2,T3**	
T4.1	Parameter Space	10h		L, D
T4.2	Standard Normalization	40h		L, D
T4.3	Co-registration	40h		L, D
T5	**Bulk Execution**	**30h**	**T4**	
T5.1	Adapting the scripts	10h		L, D
T5.2	Execution	20h		D, C, S
T6	**Pre-processing results**	**50h**	**T5**	
T6.1	Quality Control	30h	T1, T2	L, D
T6.2	Generating the dataset	20h		L, D, S
T7	**Statistical Analysis**	**100h**	**T6**	**L, D, S**
T8	**GEP & TFG**	**100h**		**L, D**
Total		**550h**		

Table 2: Quantification of the time costs and dependencies of the objectives and sub-objectives defined in the project.

2.3 Knowledge Integration

Throughout the realization of this project, a wide range of knowledge and skills developed during the degree have been put into practice:

- In the statistical analyses, concepts related to PE and ADEI have been extensively used as a structural basis. Hypothesis tests, experimental designs or the R programming language have been crucial in this project. Additionally, at the begging of the analysis part and in the final stages, I have had the pleasure of having some feedback from Josep Franquet Fabregas, associate professor in ADEI.

- As far as software architecture goes, the structure and modularity have been developed according to the strategies that were laid in subjects such as IES or PROP.

- For the adaptation of the developed programs to the cluster environment, parallelism concepts were implemented, developed in courses such as PAR or SO.
- The whole project code, aside from the statistical part, has been developed in the Python programming language, which is introduced in LP and extensively used in CAIM.
- For computational and time cost evaluation, concepts coming from A and EDA were used, as well as many kinds of data structures and classes.
- During the project, discussions on ethics, explainability, optimization, parameter space or knowledge representation have been assessed, as well as data governance. These topics were also discussed in the AI course.
- Despite not having had the chance to enroll for the APA course, I believe this projects' content is closely related to it, as it constantly considers cost functions, initializations and other metrics that are of crucial importance in the Machine Learning field.
- Last but not least, all the computers that have been used for the realization of this project, as well as the cluster's environment, were based in UNIX distributions, which are extensively studied in the SO course.

2.4 Estimations and Gantt

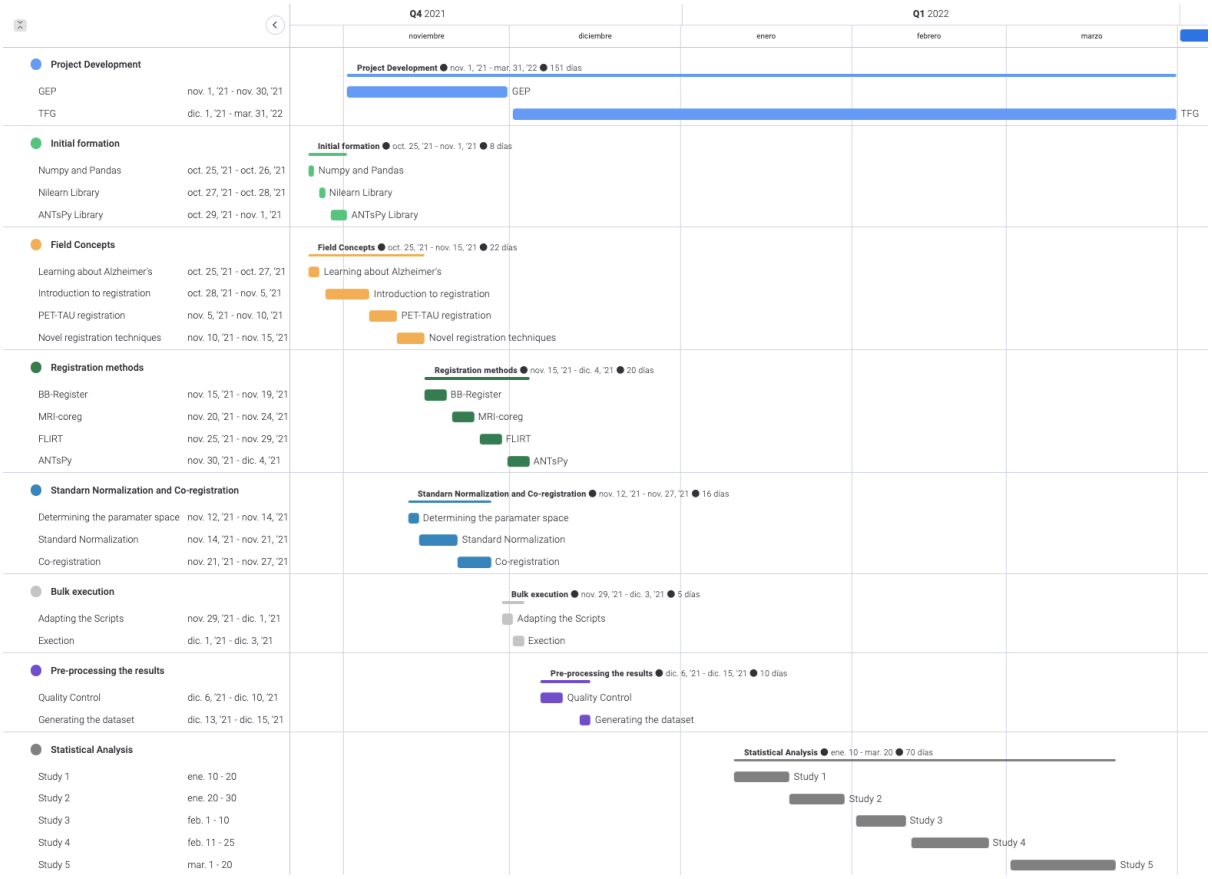


Figure 13: Gantt diagram showing the chronological layout of the objectives and sub-objectives defined in the project.

2.5 Risk management

Due to the strong relationship of this project with the biomedical field, it was of crucial importance to study superficially, and in some cases in depth, a wide variety of concepts and methodologies used in the areas of biomedical science, neuroscience and neuroimaging.

Given the specificity of the project and the extraordinary conditions in which it was carried out (including late enrollment), as well as the need of getting familiar with the biomedical field; the main risk is the non-completion of the project for the established deadlines and a subsequent adjournment.

If during the realization of the project, and considering the schedule presented in Figure 13, it is impossible to adapt to the proposed goals, the final calculations and statistics will be carried out from regions of interest (ROIs) instead of for each voxel of the volume in question.

The expected result for this project, in addition to the generation of a framework based on the most commonly used algorithms, is to develop and test novel techniques that incorporate out-of-the-box thinking to provide new methods for better information extraction throughout the process, such as the ones being developed by Islam K. et al., that include technologies such as Deep Learning [11].

Nevertheless, bearing in mind the time constraints, the project will be covering the usage of the already existing methods and algorithms for other kinds of imaging modalities registration in order to define a new framework for tau-PET imaging.

3 Implementation and Results

3.1 Registration

In this project, our objective is to analyze the different configurations that can be applied to the registration process of tau-PET to MRI images, and define how they affect the quality of the output. Each configuration consists of 3 parameters: Algorithm, Initialization and Degrees of Freedom.

Considering all the possible values for the set of parameters, and taking into account that the MRI-coreg algorithm does not require any initialization method, we obtain a total of 27 different registration configurations (Figure 16), which must be executed for each of the 180 subjects in our dataset, resulting in a total of 4860 executions.

If we analyze the execution costs for the whole process, given that the designated computer takes an average of 6 minutes to execute each one of these executions, we end up generating a total execution time of 20,160 minutes (Figure 14).

$$180\ subjects \times \frac{27\ executions}{1\ subject} \times \frac{6\ minutes}{1\ execution} = 20{,}160\ minutes \approx 20\ days$$

Figure 14: Time cost estimation for full batch execution on PC.

However, taking advantage of the Memory Unit's cluster and adapting the scripts to favor the parallelism between the different executions, we can generate a 1:8 parallelism ratio and complete the execution of all these processes in less than 2 days (Figure 15).

$$180\ subjects \times \frac{27\ executions}{1\ subject} \times \frac{6\ minutes}{1\ execution} \times \frac{1\ non-parallel}{8\ parallel} = 2{,}520\ minutes \approx 1.75\ days$$

Figure 15: Time cost estimation for full batch execution on cluster.

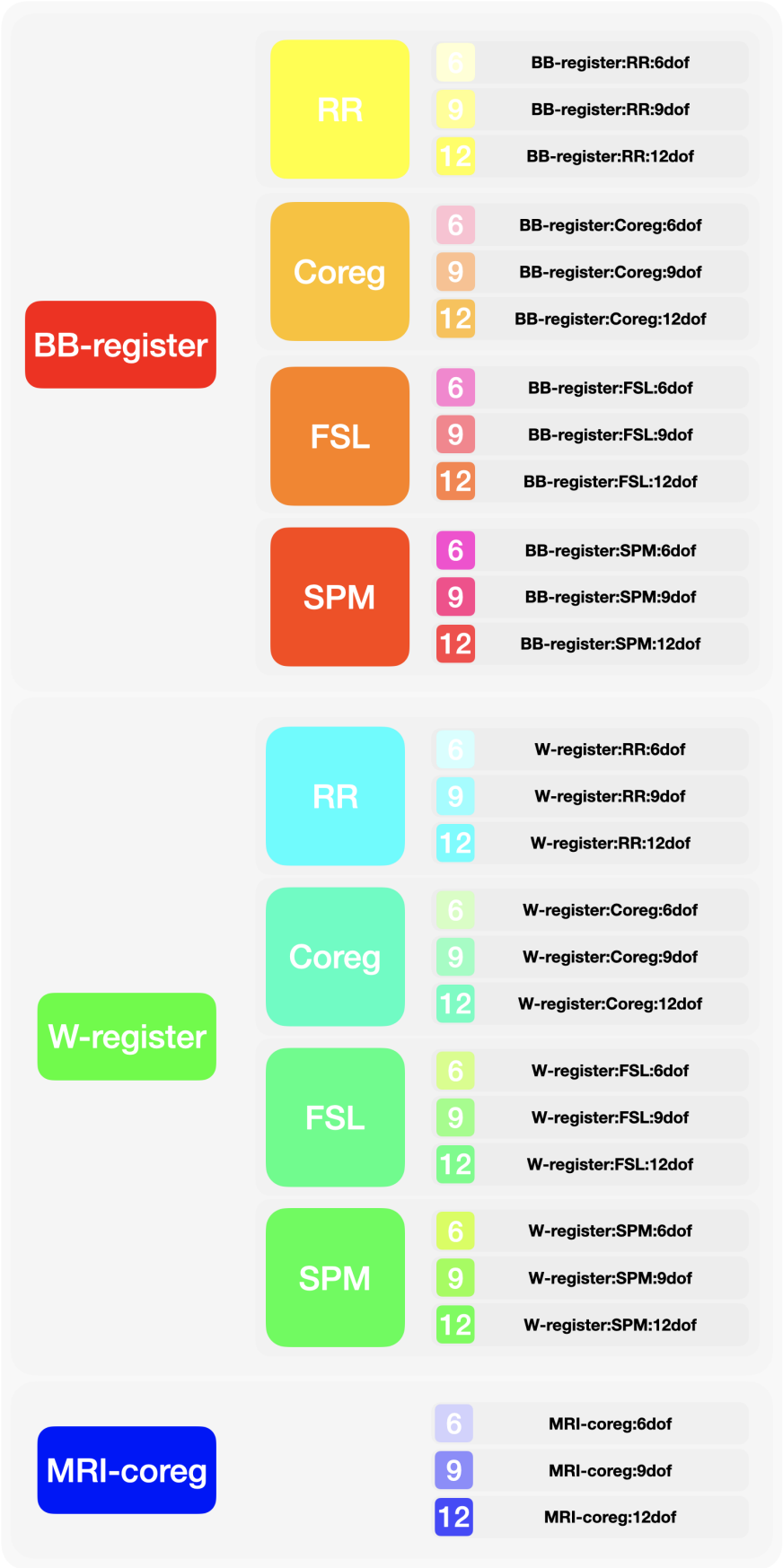


Figure 16: Scheme of the 27 different configurations that are generated for each subject.

To carry out the execution of all registration processes, a series of scripts were programmed to achieve modularity and simplicity in the code (Figure 17).

1. The main script is in charge of launching all the registration processes into the cluster. It has been programmed to scan the database of subjects available and call sub-processes that will execute all the different 27 configurations for each of the subjects.

2. The T1-Std script uses the ANTsPy framework to perform the standard normalization process, which computes the necessary transformations to convert information from the subject space into the standard space.

3. The PET-T1 executes a co-registration of the subject's tau-PET scan to the MRI.

4. The PET-Std script uses the previous scripts' outputs in order to perform the concatenation of transformations, transforming the tau-PET image into the MNI space.

5. Throughout the registration process, some image similarity metrics are computed in order to objectively measure the quality of the registration process.

6. After the registration process finishes, the SUVr values are computed in the subject and standard spaces.

7. All the quality control images that are generated through the previous steps are gathered in the form of soft-links in order to ease its review.

8. Finally, all the information related to registration quality metrics and SUVr values is condensed into a dataframe that will be statistically analyzed in the later stages of the project in order to assess the original objectives.

First, the main script is generated, which will be executed from the command line within the cluster unit. This script will generate the 27 configurations for each of the subjects and execute them.

For the execution of the registers, two different scripts will be programmed, one responsible for the co-registration and another in charge of the standard normalization and computation of the SUVr in the different spaces.

Additionally, a script that will run when the executions are finished will be generated, and will be responsible for gathering all the information that has been generated throughout the previous processes in the form of metrics and quality control images.

Quality control images will facilitate visual verification of the quality of the outputs, and will be analyzed by specialized personnel.

Once all the registration processes for the set of subjects have been executed, we can proceed to collect the data that has been generated during these executions. In particular, we will collect data on the quality of registration according to 6 different objective metrics, and the SUVr values computed in different spaces (standard space and subject space).

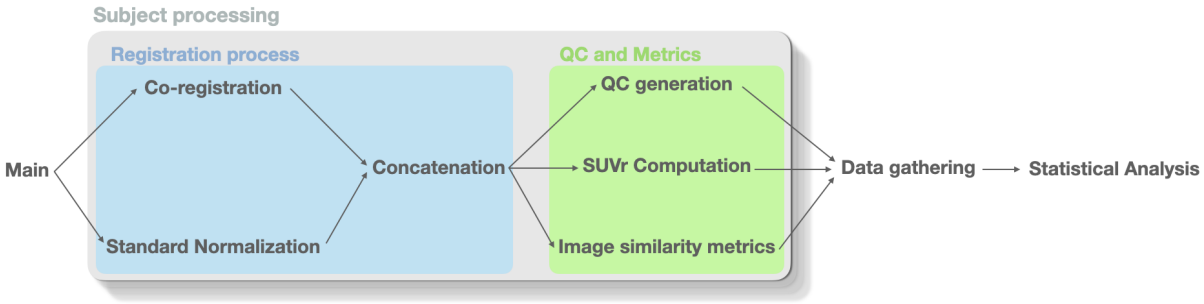


Figure 17: Visual representation of the structure and dependencies of the programming part of the project.

3.2 Statistical analyses

On the analytical side of the project, we can establish two main objectives:

1. Study the set of parameters (registration algorithm, initialization method and degrees of freedom) that optimizes the registration process outputs.

2. Evaluate the differences in computing the SUVr using an atlas computed in different spaces (native space vs standard space).

To evaluate these objectives, we will perform a series of statistical analyses, contemplating the set of similarity measures and aggregating some meaningful demographic data of the subjects, such as age, diagnostic group or APOE4, which have been proved to characterize AD [8].

This series of analyses will be grouped in 5 different studies, forming 2 different blocks, one per objective. These studies, and their main results and conclusions, are summarized in the subsequent paragraphs.

3.2.1 Block 1

The first block's purpose will be evaluating our first objective - determining the configuration of parameters that maximizes registration quality. It will consist of 3 studies:

Study 1 - Characterization of the 10 best registration configurations

First, by quality control and preliminary data visualization from the six quality metrics of registration, we were able to determine that there were two metrics of special interest due to the information they stored as well as their distributions. These two metrics are the focus of this and later studies related to registration quality and are Mattes Mutual Information (MMI) and Correlation.

We have selected these two metrics as they are both non-absolute and local. Firstly, they are non-absolute because they are indeed computing a coefficient of similarity between the two input images, which varies from 0 to 1. Secondly, the fact that they are computed locally allows for a better quantification of alignment between the two images as well as avoiding local minima.

The rest of the metrics were discarded for two main reasons. The first was the fact that some of the calculations generated errors due to their implementation in the ANTsPy library, resulting

in a series of missing values throughout the dataset. Secondly, and in relation to the previous paragraph, they generated absolute values, which are difficult to compare due to the inherent difference in the intensity values of the two modalities; or were computed non-locally, and thus generated a far from precise evaluation of the registration output.

Thenceforth, we proceed to statistically describe the registration quality for the different configurations and order the results from best to worst registration quality. In Table 3 it can be observed that for the two metrics that have been considered, the 10 configurations that maximize the registration quality are the same for both metrics. The only difference is the order of the best 3 configurations, which is alternated.

Configuration	Pos MMI	Pos Correlation	Mean MMI	Mean Correlation
MRI-coreg:9dof	1	3	0.718	0.675
MRI-coreg:6dof	2	1	0.717	0.677
MRI-coreg:12dof	3	2	0.716	0.675
W-register:coreg:6dof	4	4	0.701	0.668
BB-register:coreg:6dof	5	5	0.680	0.658
W-register:RR:6dof	6	6	0.649	0.636
BB-register:RR:6dof	7	7	0.626	0.619
W-register:RR:12dof	8	8	0.600	0.607
W-register:RR:9dof	9	9	0.598	0.605
BB-register:RR:12dof	10	10	0.542	0.562

Table 3: Best configurations for MMI and Correlation ordered by registration quality and their average values.

As it can be derived from Table 3, the best three performing configurations are the ones that involve the MRI-coreg algorithm, followed by BB-register and W-register with coreg initialization and 6 DOF.

Study 2 - Statistically significant differences between registration configurations

The preliminary step for initializing this study, is to determine whether or not the registration qualities are statistically different from one another. For that purpose, we will perform a Friedmann test, which is a non-parametric test that also accounts for repeated measures, contrarily to an ANOVA test.

This tests manifests a significant omnibus difference between these top 10 configurations. Henceforth, we perform a post-hoc pairwise comparison to determine which group-differences were statistically significant.

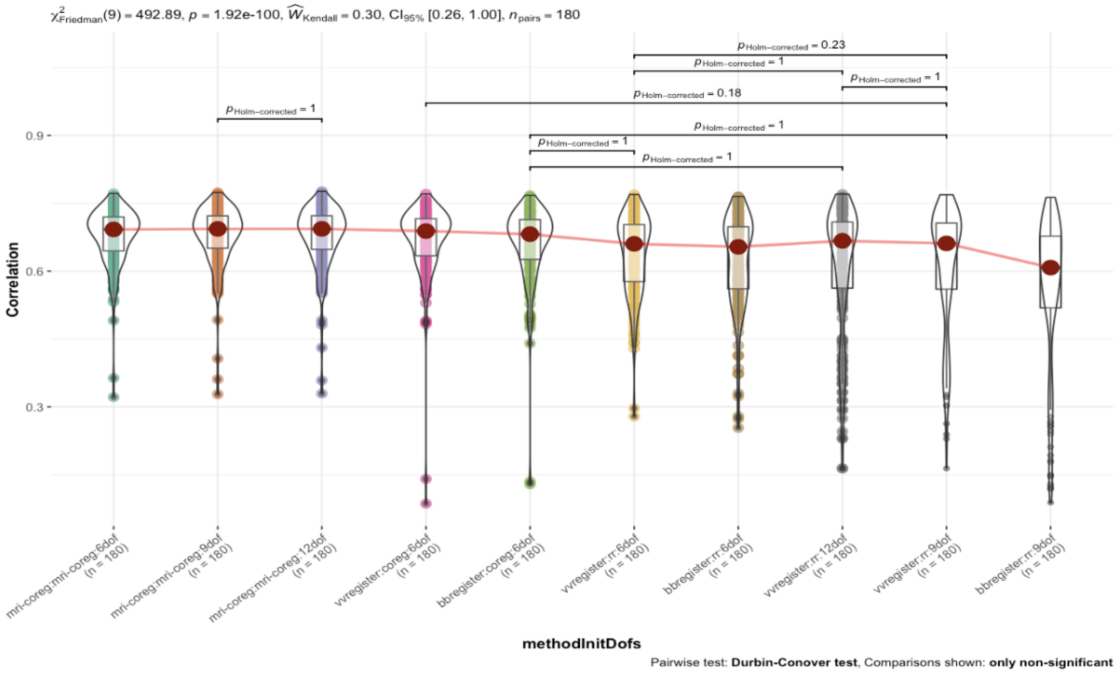


Figure 18: Statistically non-significant differences between configurations quality for Correlation metric.

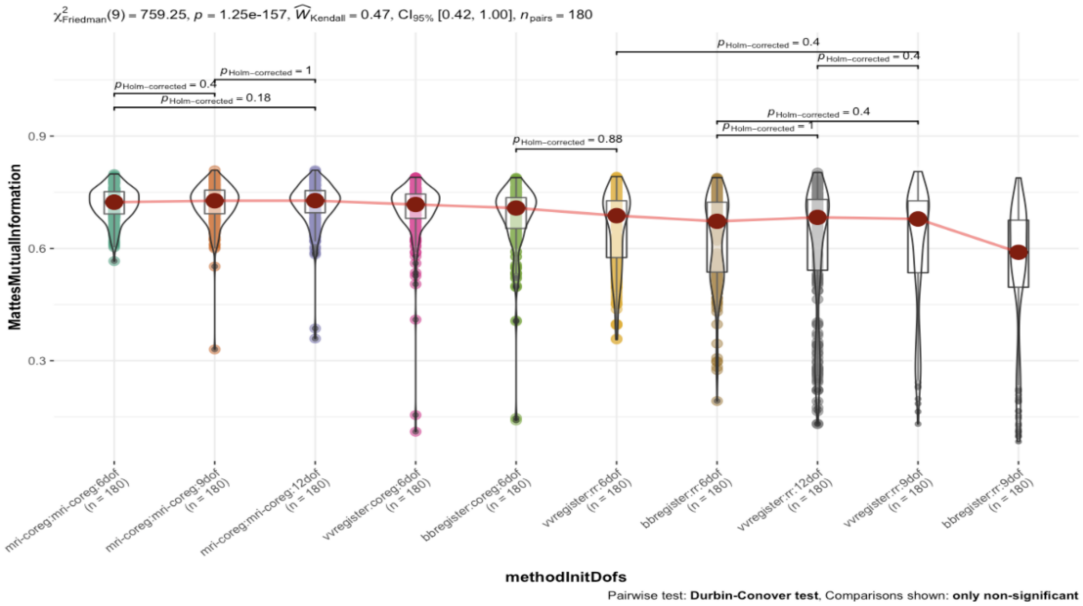


Figure 19: Statistically non-significant differences between configurations quality for Mattes Mutual Information metric.

From Figures 18 and 19, we can extract 2 observations. Firstly, since the plots show statistically non-significant differences for every pair of configurations, the fact that their relationship is not displayed means that their differences are statistically significant, and therefore generate different registration qualities.

Secondly, when observing the best 3 performing configurations, we can observe statistically non-significant differences between them, which suggests that they can be used indistinctly.

Study 3 - Statistically significant differences throughout the Alzheimer's Disease Continuum

To perform this study, we will first have to add a new dataset containing some demographic data of the subjects. Specifically, we will be considering the diagnostic groups throughout the AD continuum, which can be categorized in CN (Control Negative), EMCI (Early Mild Cognitive Impairment), LMCI (Late Mild Cognitive Impairment) or AD (Alzheimer's Dementia).

Throughout the continuum of AD, the brain structure tends to become atrophied, adding an extra difficulty to the registration process, and thus making readings imprecise, resulting in an accumulation of errors that, in later readings, could end up invalidating the results. Simultaneously, tau tends to accumulate, altering the distribution of signal intensity in the tau-PET images [6].

Provided there is no evidence of our series of registration quality metrics following a normal distribution, we will proceed to perform a Kruskal-Wallis test[4], which is a non-parametric variant of the ANOVA test, in order to assess the robustness of the ten best performing configurations.

Additionally, we computed a post-hoc Dunn test [7] [5] between every pair of diagnostic groups in order to evaluate which groups generate statistically significant differences.

Lastly, we quantified the effect size generated between the groups with the epsilon-squared[6] metric in order to assess which configuration generated the lesser separation between groups, and thus is the most robust for generating similar quality registrations for subjects all along the AD continuum.

[4]More information available at https://www.cienciadedatos.net/documentos/20_kruskal-wallis_test
[5]More information available at https://www.statisticshowto.com/dunns-test/
[6]More information available at https://www.statisticshowto.com/epsilon-squared/

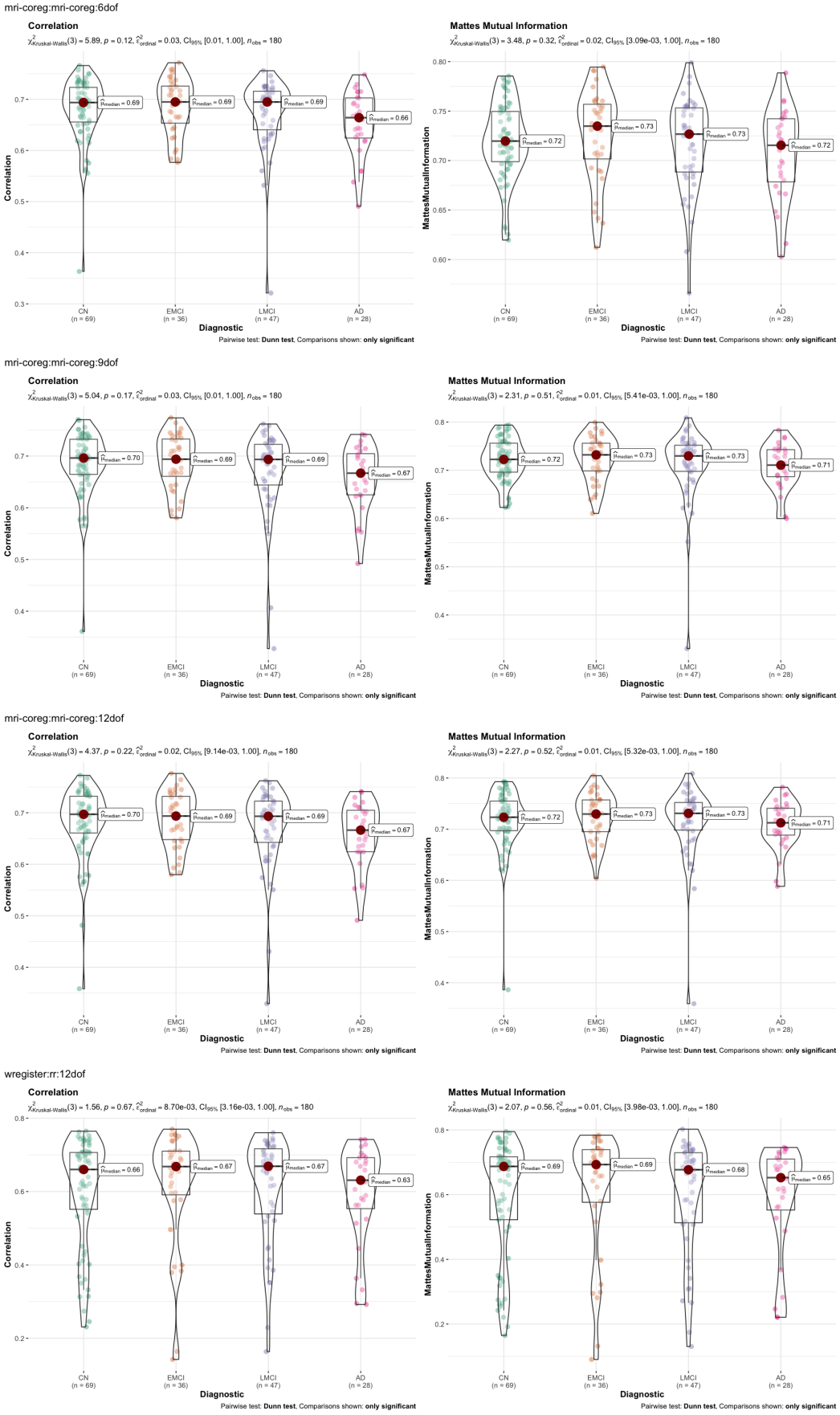


Figure 20: Comparison of the robustness of the 4 most relevant configurations between the different diagnostic groups for both Correlation and Mattes Mutual Information metrics.

In Figure 20 we can observe how the p-values associated to the null hypothesis of the Kruskal-Wallis test do not exhibit enough evidence to support the inequality of means among the different diagnostic groups for the most relevant configurations, which are the 3 best performing in terms of registration quality and the best performing in terms of robustness. Moreover, the post-hoc Dunn tests don't provide any evidence otherwise, as there are no statistically significant differences for any pair of diagnostic groups.

In the previous study, we stated that the top 3 configurations could be used indistinctly, and in the previous figures, we can appreciate that, of these three configurations, the one that maximizes the Kruskal-Wallis test p-values whilst minimizing the effect size is MRI-coreg with 12 degrees of freedom.

However, these test also suggest that the W-register with RR initialization and 6 degrees of freedom achieves better results in terms of rubustness.

Accordingly, and based in the aforementioned studies, we will proceed to evaluate these two different configurations, one providing the highest registration quality, and the other the highest robustness, in order to determine which one is preferable for tau-PET to MRI registration.

3.2.2 Block 2

The second block's purpose will be to evaluate our second objective - determining the best space (standard or subject's) for computing the SUVr as well as addressing its classification and prediction capabilities. It will consist of 2 studies:

Study 4 - Determining the best space to compute the SUVr

To proceed with this study, we will start by checking the equality of SUVr values computed in the different spaces for the entire set of subjects. Concretely, we will compare the quality as well as the group diagnostic separation generated by SUVr values computed in the subject space compared to the ones generated in the standard space. Those measures should differ as, on the one hand, the Desikan atlas is calculated individually for each subject, whereas on the other hand, it is previously computed and validated.

We must remember these values are two different observations of the same execution, so consequently, we will perform a paired Wilcoxon test[7], which is a non-parametric variant for the paired T-test.

This preliminary test suggests that when considering the entire set of subjects, the differences are statistically significant for both configurations (last row in Table 4). Therefore, we proceed to segregate our dataset by diagnostic groups and perform a post-hoc analysis for every group.

In this case, the results shown in Table 4 manifest that statistical significant differences in the SUVr values calculated for the different spaces for the HC group are non-existent. However, for the EMCI, LMCI and AD groups, there is no evidence of equality, thus we obtain statistically significant differences in the measurements when computing the SUVr in the subject's space vs. standard space for these two groups. The rationale is that, the topography of tau-PET changes along the AD continuum, and as a result, different diagnostic groups might be more impacted in how we compute the SUVr.

[7]More information available at https://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test

Diagnostic group	MRI-coreg:12dof		W-register:RR:6dof	
	paired Wilcoxon test p-value	SUVr are equal (95% CI)	paired Wilcoxon test p-value	SUVr are equal (95% CI)
CN	0.31	True	0.58	True
EMCI	0.04	False	0.03	False
LMCI	0.00	False	0.00	False
AD	0.00	False	0.00	False
Global	0.00	False	0.00	False

Table 4: Equality of SUVr segregated by diagnostic and global.

Henceforth, we must determine which space is better for performing this calculation, a task that we will determine by analyzing the group separation that is generated from the different SUVr values, since ideally, we would like this metric to be as differential as possible between the different groups in order to facilitate its classification and diagnosis.

With that purpose, we will perform a Kruskal-Wallis test, which is a non-parametric variant for the ANOVA test.

Moreover, we also compute the effect size generated between the different diagnostic groups by the two series of values with the epsilon-squared measurement.

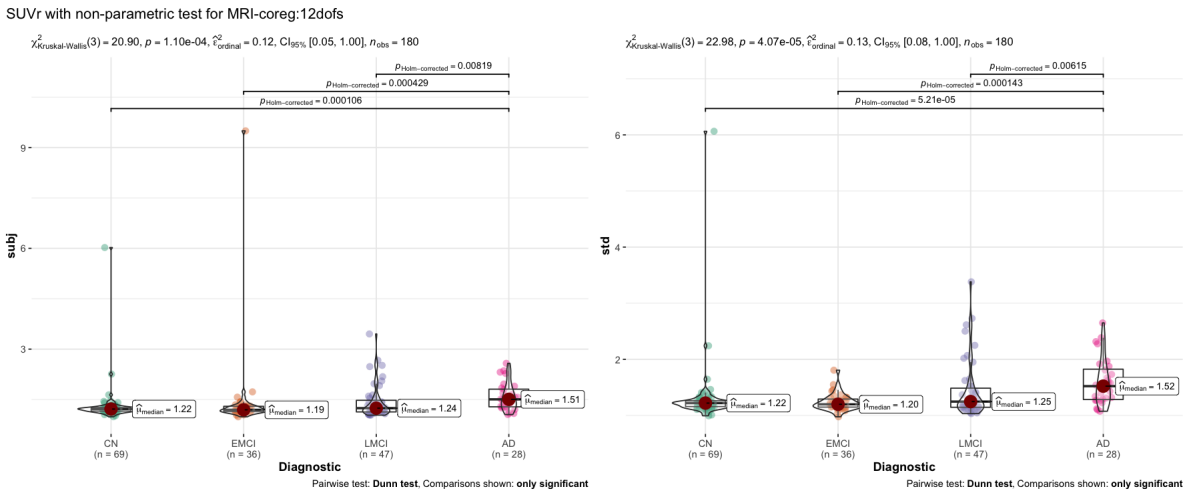


Figure 21: Kruskal-Wallis test, with post-hoc pairwise Dunn test and epsilon-squared effect size quantification to assess the differences between different diagnostic groups' SUVr values computed in standard and subject's space for MRI-coreg:12dof.

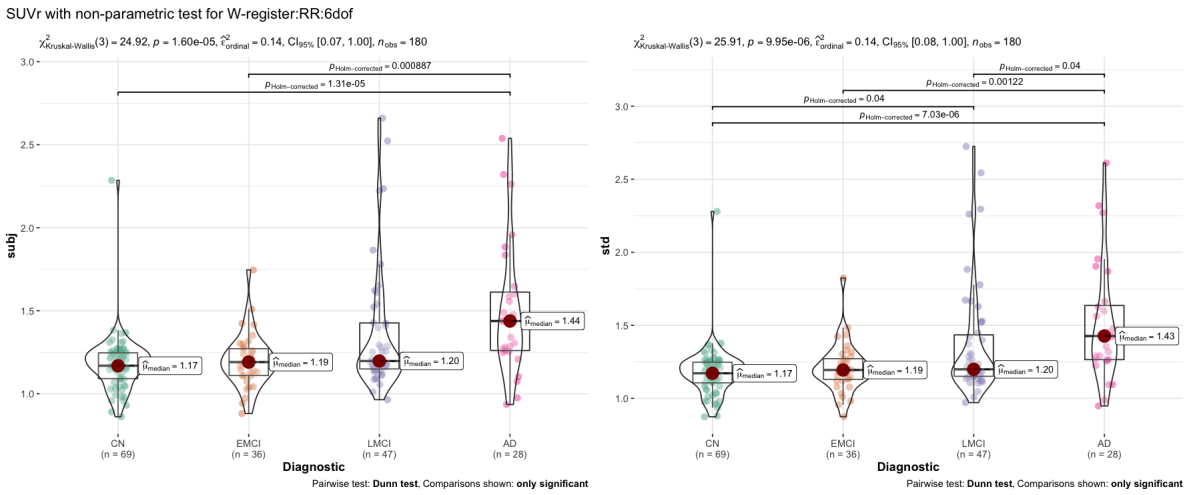


Figure 22: Kruskal-Wallis test, with post-hoc pairwise Dunn test and epsilon-squared effect size quantification to assess the differences between different diagnostic groups' SUVr values computed in standard and subject's space for W-register:RR:6dof.

In Figure 21 and 22 we can observe how, for both series of SUVr values, and both configurations, the null hypothesis of the test can be rejected, and therefore, there is evidence that the SUVr measurements are different among the diagnostic groups, as previously mentioned.

Furthermore, if we observe the epsilon-squared measure, we can establish that the SUVr values computed in the standard space generate a larger effect size as well as more significant differences between diagnostic groups than the ones computed in the subject's space, and therefore, should be preferable for generating a classification or prediction of a subject's diagnostic.

In conclusion, this study reveals that computing the SUVr in the standard space generates a better group separation that can boost a data-driven system's classification and diagnostic capabilities.

Study 5 - Classification and prediction capabilities of derived data-driven methods

To conclude this block as well as the statistical analysis, we are going to propose a logistic regression model that classifies subjects into CN or AD diagnostic groups.

We will use the data related to the SUVr computed in the standard space, as we proved to be more suitable, as well as other covariables such as age, and the APOE4 codification of the subjects. Importantly, those covariates will be introduced in our model, since it has been shown to strongly affect the amount of tau accumulation in the brain [8].

Furthermore, we will present two different models. One will be based on the SUVr values computed after the MRI-coreg:12dof registration, and the other will be based on the ones computed after the W-register:RR:6dof. We will analyze both models in order to determine which has a better fit on the data, and we will evaluate their capabilities by performing a ROC curve analysis.

To assess the goodness of fit of these models, we performed a Hosmer-Lemeshow test, which suggested that both models had a good fit to the data.

Subsequently, we proceeded to perform a ROC curve analysis for both models. We used boot-

strapping techniques, commonly used to reduce heuristic effects on the analysis by repetition, in order to achieve a higher precision for our results.

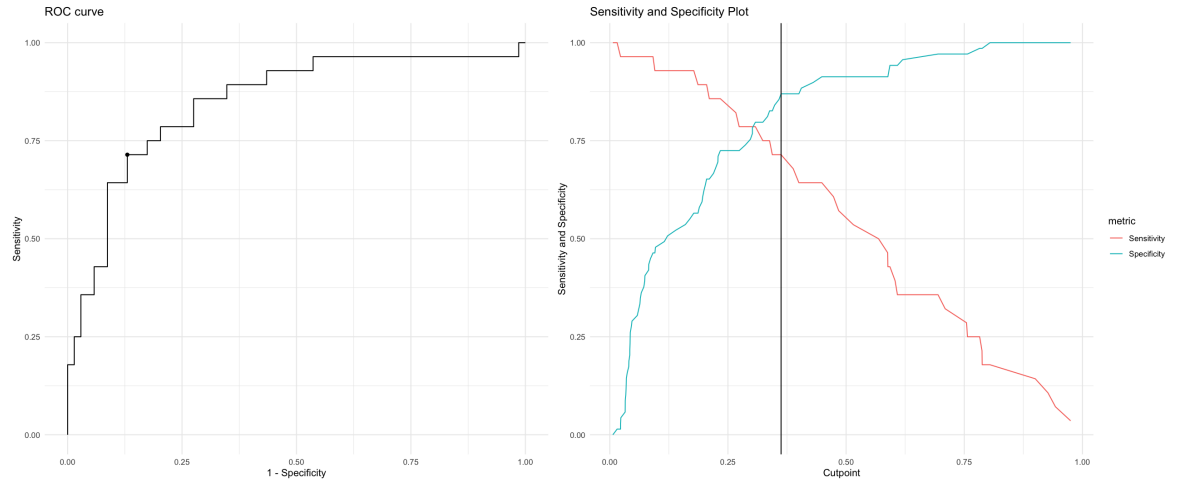


Figure 23: Density of the predicted values of the model for AD and CN groups and optimal cut-point representation (left). ROC curve representation (right) used to evaluate the performance of the MRI-coreg:12dof model.

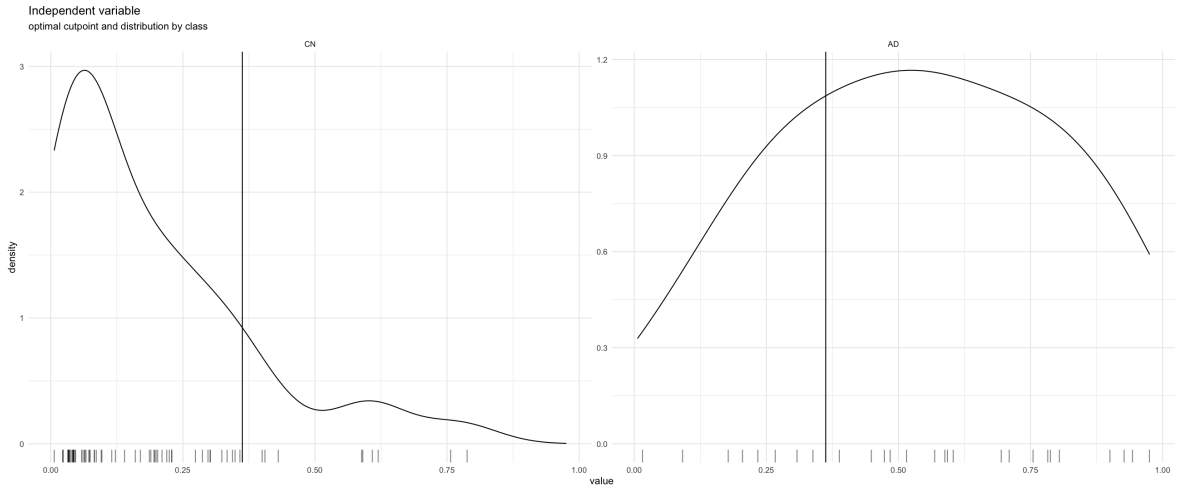


Figure 24: Representation of sensibility and specificity metrics that are used to evaluate the performance of the MRI-coreg:12dof model.

In Figures 23 and 24, we can observe how our optimal cut-point establishes the boundary to be set for the model in order to maximize the CN vs AD classification as well as the evolution of sensibility and specificity measures given that cut-point. We can also see a graphical representation of the ROC curve, which generates an AUC of 0.85, achieving an excellent performance in classification purposes. The accuracy of the model is 83%.

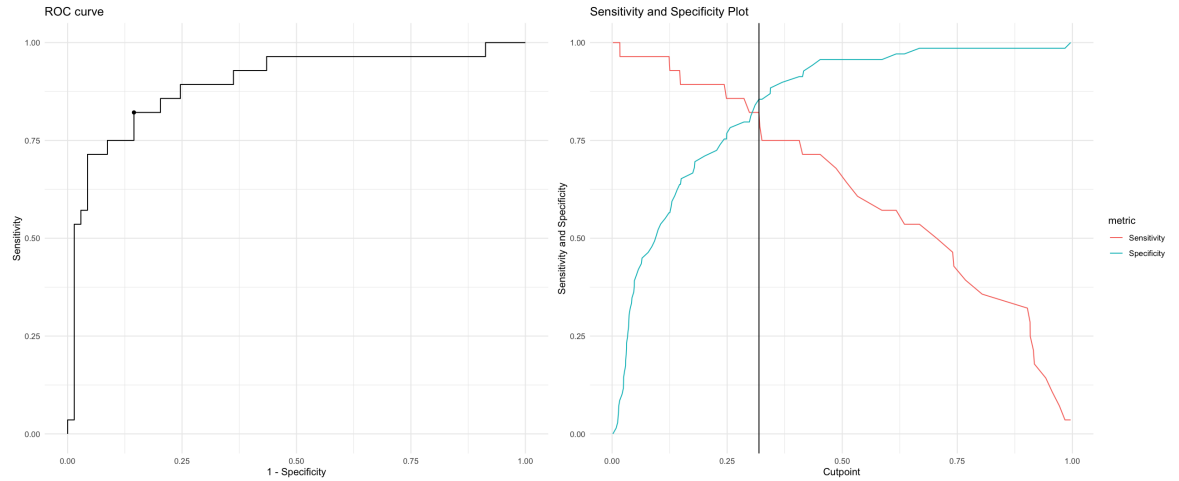


Figure 25: Density of the predicted values of the model for AD and CN groups and optimal cut-point representation (left). ROC curve representation (right) used to evaluate the performance of the W-register:RR:6dof model.

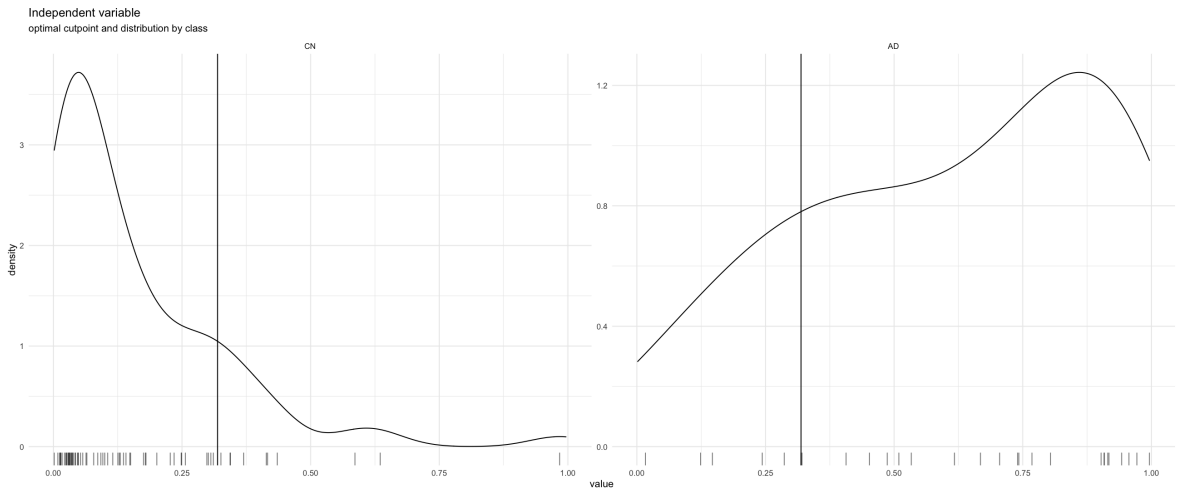


Figure 26: Representation of sensibility and specificity metrics that are used to evaluate the performance of the W-register:RR:6dof model.

Figures 25 and 26 show similar results for the ROC curve analysis generated with the W-register:RR:6dof model. Specifically, the AUC metric of 0.90 suggests a better performance in terms of classification, achieving a higher accuracy (85%).

Finally, we perform a statistical comparison of the two ROC curves that we have generated in order to assess the existence of statistically significant differences between these 2 models. We do that by performing another bootstrap analysis. The results show that the ROC curves generated for both models do not statistically differ, and subsequently, both models could be used indistinctly, despite one having a slightly higher AUC than the other.

3.3 Conclusions

In studies 1, 2 and 3, we establish that the best 2 possible configuration to perform the registration of tau-PET to MRI images are MRI-coreg algorithm with 12 degrees of freedom or W-register with RR initialization and 6 degrees of freedom. Although the first model achieves better registration qualities while preserving the robustness throughout the AD continuum, the second does a better job in terms of robustness, which is a key factor as the progression in the AD continuum tends to involve structural brain atrophy in the later stages, adding an extra difficulty that needs to be thoroughly considered [6].

Moreover, in study 4, we determine that the standard space is preferable for computing the SUVr values, as it generates a better separation of the diagnostic groups.

Finally, we propose two logistic regression models to assess the classification and prediction capabilities of a data-driven method derived from tau SUVr and demographic co-variables. We evaluate both models with a ROC curve analysis and we can affirm both models demonstrate excellent performances by achieving AUC values of 0.85 and 0.90. Therefore, we have stated that data-driven methodologies can have a huge impact in AD research, specially proving that an in-house tuning of an already existing methodology can achieve better performances in terms of robustness and subsequent evaluations than the state of the art praxis.

3.4 Future Work

The possible next steps for this project could be divided into research and engineering. On the one hand, in the neuroimaging research field, where explicability is key, my opinion is that this project could be very much improved by the way the SUVr values are computed.

Given the specificity of the topography of the tau depositions, the fact that we are averaging up to three times our data in order to compute the SUVr values, implies an enormous loss of information that ends up being evaluated by the model, specially on what refers to its topography.

Recent studies on Aβ-PET that apply a PCA on a voxel level [20], have demonstrated to be able to characterize the topography of these depositions in such a way that allows for a direct quantification of Aβ that does not require neither a registration nor a SUVr computation.

I believe that this approach is also possible for tau-PET even though the specificity of its topography, and that it would offer a much more precise approach on the tau protein characterization along the AD continuum.

On the other hand, as regards to engineering, I believe the precision and classification capabilities of possible models involving the information that we have been able to generate and gather, could be very much improved by implementing more complex modeling techniques, such as Neuronal Network architectures or even 3D convolutions. This approach, however, would come at the expense of explainability. In the research field, these more complex modeling techniques are viewed as "black boxes" that don't allow for much explanation of the results.

4 Budget and Sustainability

4.1 Budget

In the following sections, there is an in-depth description of the budget necessary for this project, divided into personnel costs and generic costs.

4.2 Personnel costs per activity

According to the IIB researchers, the annual salary for a pre-doctorate is 24.000€. Assuming 42 working weeks per year and 40 working hours per week, we can estimate an hourly wage of 15€/h. In Table 5 you can find a categorization of the time and budgets planned for every objective and sub-objective considered in the project.

Code	Job	Time	Cost
T1	**Initial Formation**	**30h**	**450 €**
T1.1	Numpy & Pandas	10h	150 €
T1.2	Nilearn	10h	150 €
T1.3	ANTsPy	10h	150 €
T2	**Field Concepts**	**110h**	**1650 €**
T2.1	Alzheimer's	20h	300 €
T2.2	Registration	30h	450 €
T2.3	PET-TAU registration	30h	450 €
T2.4	Novel Techniques	30h	450 €
T3	**Registration Methods**	**40h**	**600 €**
T3.1	BB-register	10h	150 €
T3.2	MRI-Coreg	10h	150 €
T3.3	Flirt	10h	150 €
T3.4	ANTsPy	10h	150 €
T4	**Standard Normalization & Co-registration**	**90h**	**1350 €**
T4.1	Parameter Space	10h	150 €
T4.2	Standard Normalization	40h	600 €
T4.3	Co-registration	40h	600 €
T5	**Bulk Execution**	**30h**	**450 €**
T5.1	Adapting the scripts	10h	150 €
T5.2	Execution	20h	300 €
T6	**Pre-processing results**	**50h**	**750 €**
T6.1	Quality Control	30h	450 €
T6.2	Generating the dataset	20h	300 €
T7	**Statistical Analysis**	**100h**	**1500 €**
T8	**GEP & TFG**	**100h**	**1500 €**
Total		**550h**	**8250 €**

Table 5: Costs per task estimation.

4.2.1 Generic costs

Amortizations

One aspect to take into account is the amortization of the resources used. Considering an average of 6 working hours per day, for a total of 100 days and assuming a life expectancy of 4 years for each of the resources, we can estimate an approximate cost of amortization for each resource with the expression beneath, as well as the expected amortizations displayed in Table 6.

$$Amortization(N) = Resourceprice \times \frac{1}{4years} \times \frac{1}{100days} \times \frac{1}{6hours} \times hoursused$$

Hardware	Price	Time used	Amortization
Designated computer	1500 €	550h	343 €
Laptop	4000 €	400h	666 €
Cluster	15000 €	2h	1'2 €
Total			**1100** €

Table 6: Amortization costs estimation.

Incidences and contingency plans

For this project, the contingency plans do not involve any additional costs, as they would consist in changing the final results methodologies rather than extending the working hours.

Travel costs

I bought a T-Jove to perform daily trips from home to work and from work to home. As the T-Jove lasts for 3 months, which is the expected duration of the project, it should be enough.

A T-Jove of only 1 zone costs 80€.

Software costs

All software used during the development of the project is open source. Therefore, the software cost for the project is 0€.

4.2.2 Total Costs

The total costs of the project are presented in Table 7, summarizing the aforementioned components.

Concept	**Price**
Personnel Costs	8250 €
Amortization	1100 €
Travel Cost	80 €
Software Cost	0 €
Total	**9430** €

Table 7: Total costs estimation.

4.3 Sustainability

4.3.1 Self-assessment

The grasp of sustainability is of key importance for this project, especially in the social dimension, as this projects' main objective is developing a methodology that ultimately helps to better characterize patients, and track the utility of drugs to treat AD.

Regarding the other dimensions of sustainability, this project also manages to help the environment by reutilising equipment and previously acquired open-source data.

Lastly, in the economical dimension, the projects' total cost is relatively low (and I say that as criticism) due to the surprisingly low wages assigned to the personnel in charge of performing this kind of research.

On that scope, since my first days in the research group, I have been able to perceive the lack of IT professionals in the field. Biomedical researchers, with very narrow insights on Computer Science and Software Engineering, are in charge of developing Machine Learning models as well as pipelines or project structures, to perform their studies, which is not their area of expertise.

I believe all this lack of involvement by Computer Science professionals is due to the low salaries that would be perceived in contrast to the private sector. Therefore, with a bigger budget in public research, a lot of Computer Science professionals would be considering new fields of research, improving the quality as well as the pace of research, and so giving faster and better results that would end up improving the populations' quality of life.

As Computer Scientists, we have to raise our own awareness of all the possible opportunities we have for helping society. It is hard to find any research that wouldn't be improved by our involvement. We are key figures for the optimization of many processes from which society would hugely benefit.

4.3.2 Economical dimension

Regarding PPP: Reflection on the cost you have estimated for the completion of the project

The total cost of the project has been estimated in Section 3.1. It has taken into account all the costs derived from personnel, hardware and software resources. The yearly salary has been obtained by personally asking staff.

From my point of view, the personnel budget is rather low, due to the low investment in research in the public sector.

Regarding Useful Life: How are currently solved economic issues (costs...) related to the problem that you want to address (state of the art)?

In the present, the problem this project is trying to solve is actually avoided and researchers tend to use sub-optimal methodologies that they think give valid results. Furthermore, researchers usually have to test different methods for every kind of imaging, which results in a time consuming process.

How will your solution improve economic issues (costs ...) with respect to other

existing solutions?

This project's solution pretends to avoid that trial-error process by determining a methodology that works well for all tau-PET registrations.

4.3.3 Environmental dimension

Regarding PPP: Have you estimated the environmental impact of the project?

This project does not waste any material resources. Furthermore, it helps determine the best approach for the tau-PET registration so that other researchers can directly utilize that approach instead of having to try different methods wasting their time and resources.

Regarding PPP: Did you plan to minimize its impact, for example, by reusing resources?

From a hardware perspective, the personal laptop has many other use cases other than the development of the project. The cluster is property of the Sant Pau Memory Unit and is used to perform other projects apart from this one. The designated desktop computer is also part of the Sant Pau Memory Unit and is designated for students who develop their thesis inside the Unit.

Regarding Useful Life: How is currently solved the problem that you want to address (state of the art)?

The best-registration problem is usually performed by assuming certain methods work well enough or by performing a variety of executions with different parameters until the registration is valid.

How will your solution improve the environment with respect to other existing solutions?

The final results of this project will determine the best approach for tau-PET registration, and therefore, prevent a lot of trial-errors when this process is performed by researchers.

4.3.4 Social dimension

Regarding PPP: What do you think you will achieve -in terms of personal growth- from doing this project?

I believe this project will help me develop my problem solving capabilities as well as translational skills that allow me to participate in different scientific fields, contributing my computer science knowledge and skills to research in different areas.

Before starting this project, I was genuinely interested in neuroscience, but my knowledge on the topic was almost irrelevant. This project has demonstrated me that I am capable of assuming new knowledge from new scientific fields and provide new perspectives to a problem that is usually approached computationally by professionals lacking that kind of skills. We, computer engineers and scientists, can hugely impact all these areas.

Regarding Useful Life: How is currently solved the problem that you want to address (state of the art)?

This problem is currently solved by comparing the performance of the different methods used in other kinds of registration in order to determine a good-enough approach.

How will your solution improve the quality of life (social dimension) with respect to other existing solutions?

This project's solution will help society in very different perspectives.

First of all, it will help researchers to extract better results from brain imaging techniques. That will be directly translated into a better diagnosis for patients throughout the AD continuum and therefore a better treatment and adaptation.

Lastly, the pharmaceutical industry will be able to perform a more in-depth study of the TAU protein in order to come up with new drugs and treatments that may improve the patients life's quality and expectancies.

Regarding Useful Life: Is there a real need for the project?

For the time being, AD doesn't have a cure nor a treatment that can slow the neurodegeneration pace. The studies and treatments that are being developed based on Aβ proteins seem to be approaching a dead-end. Studies based on tau protein are, contrarily, giving a better explanation for the neurodegeneration and thus giving a further insight into the disease. New treatments and, potentially, a cure for AD could be possible by tackling the tau protein hyperphosphorylation.

Appendix A

The objective of this first block it to determine the best configuration of parameters for optimizing the registration of tauPET to MRI images. Each of the configurations consists of one value for each of the three parameters that have been considered in this project (algorithm, initialization method and Degrees of Freedom, and are referred to as

<algorithm> : <initialization> : <DOFs>.

A total of 27 configurations are generated and have to be evaluated in order to determine which of them is the optimal provided two different factors:

- **Registration quality**: Registration quality can be quantified by 6 objective metrics. However, after a through evaluation on the way they are computed as well as come caracteristics of each one of them, we have reduced the dimensionality of the project by considering just two of the metrics: Correlation and Mattes Mutual Information. Our objective will be to maximize both of them.
- **Robustness**: This factor refers to the capability of each one of the configurations to generate similar-quality registrations for subjects in different states of the AD continuum, as this disease involves neurodegeneration and atrophy in the later stages, which may add an extra complication to the process.

In summary, our objective is to determine the configuration of parameters that not only maximizes the registration's quality, but is also able to generate similar-quality registrations for subjects all along the AD continuum. For that matter, we propose 3 different studies.

Study 1 - Characterization of the 10 best registration configurations

The objective of this first study is to characterize the 10 best performing configurations in order to reduce the dimensionality of the project.

With that objective in mind, we have computed the mean values for the Correlation and Mattes Mutual Information metrics for each of the configurations as well as a v.test, which is a criterion with a Normal Distribution.

We proceed to order the best 10 configurations by their quality according to the Correlation metric.

	Name	v.test Corr	mean Corr	v.test MMI	mean MMI
5	mri-coreg:mri-coreg:6dof	15.307837	0.7181242	13.200749	0.6753869
6	mri-coreg:mri-coreg:9dof	15.293987	0.7179006	13.342648	0.6773178
4	mri-coreg:mri-coreg:12dof	15.232074	0.7169012	13.217259	0.6756116
7	wregister:coreg:6dof	14.248613	0.7010262	12.667723	0.6681339
1	bbregister:coreg:6dof	13.000382	0.6808773	11.968817	0.6586238
9	wregister:rr:6dof	11.048984	0.6493779	10.316538	0.6361409
2	bbregister:rr:6dof	9.650358	0.6268014	9.127144	0.6199566
8	wregister:rr:12dof	8.014968	0.6004030	8.182235	0.6070991
10	wregister:rr:9dof	7.908013	0.5986765	8.080106	0.6057094
3	bbregister:rr:9dof	4.454464	0.5429295	4.922815	0.5627476

In particular, we must note that both metrics' best 10 performing configurations are the same, moreover, in an almost identical order. The only differences appear among the top 3 configurations, which come out in different order.

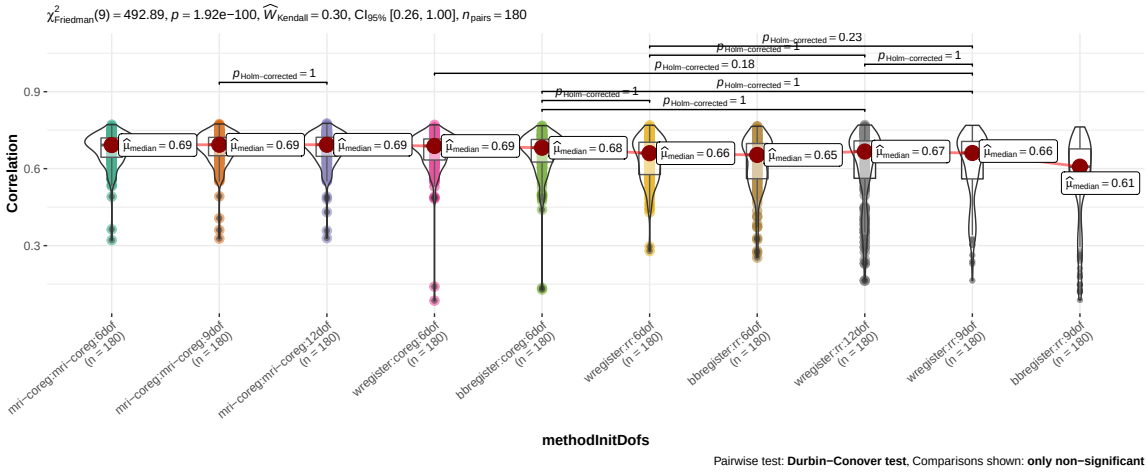
Configuration	Pos MMI	Pos Correlation
MRI-coreg:9dof	1	3
MRI-coreg:6dof	2	1
MRI-coreg:12dof	3	2
W-register:coreg:6dof	4	4
BB-register:coreg:6dof	5	5
W-register:RR:6dof	6	6
BB-register:RR:6dof	7	7
W-register:RR:12dof	8	8
W-register:RR:9dof	9	9
BB-register:RR:12dof	10	10

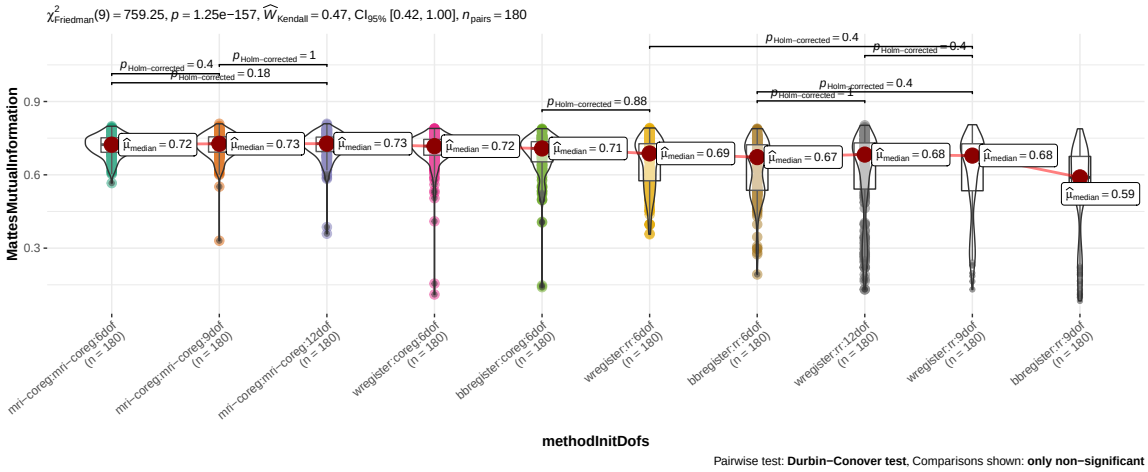
Study 2 - Statistically significant differences between registration configurations

Once we have selected the 10 best performing configurations in terms of registration quality, we will proceed to assess the differences between them, in order to determine if they can be used indistinctly.

Given the fact that we have several registrations for the same subject, a simple ANOVA test wouldn't be valid, as it wouldn't take into account those repetitions. Therefore, in order to assess if any of the configurations' registration quality is different from the rest, we will use a Friedman test, which does take in to account the various observations for each subject. This test is also non-parametric, as according to the Shapiro test previously performed, we cannot assume the normality of our series of data.

Furthermore, we will also perform post-hoc pairwise Durbin-Conover tests in order to evaluate the statistical differences between every pair of configurations. We display the statistically non-significant as we want to determine which of the configurations can be used indistinctly, shown in the plots below.





Study 3 - Statistically significant differences throughout the Alzheimer's Disease Continuum

To conclude this block, study 3 will assess the robustness of the different configurations for the different diagnostic groups defined along the AD continuum. These groups are:

- CN - Control Negative
- EMCI - Early Mild Cognitive Impairment
- LMCI - Late Mild Cognitive Impairment
- AD - Alzheimer's Dementia

In the following table we have a brief description of the subjects' demographic information.

Characteristic	CN, N = 70[1]	EMCI, N = 36[1]	LMCI, N = 48[1]	AD, N = 28[1]
Age	69 (66, 72)	73 (65, 77)	72 (65, 78)	74 (70, 80)
Unknown	1	0	0	0
Gender				
Female	43 (61%)	14 (39%)	20 (42%)	11 (39%)
Male	27 (39%)	22 (61%)	28 (58%)	17 (61%)
APOE4				
0	47 (67%)	21 (58%)	26 (55%)	7 (25%)
1	20 (29%)	11 (31%)	15 (32%)	13 (46%)
2	3 (4.3%)	4 (11%)	6 (13%)	8 (29%)
Unknown	0	0	1	0

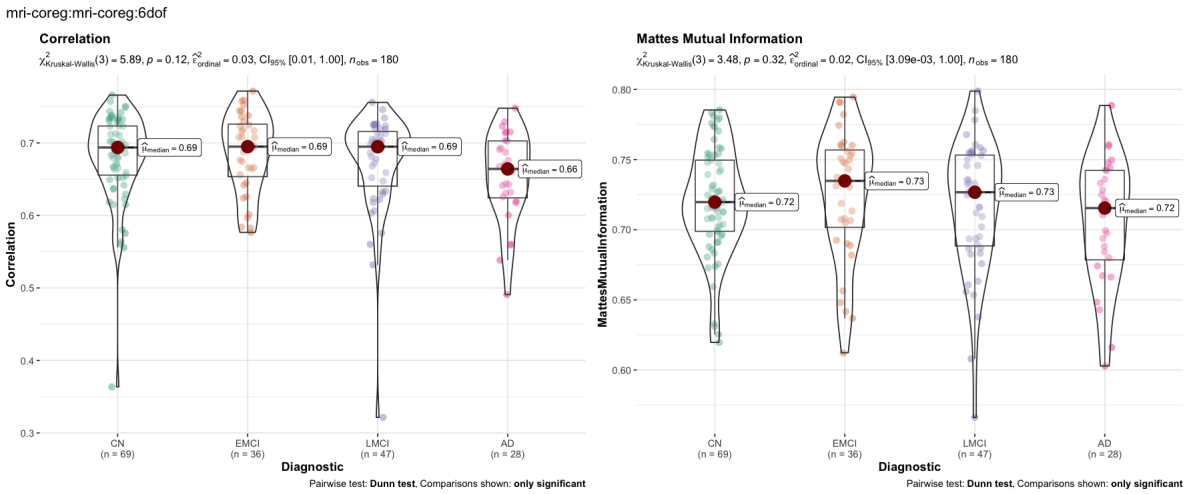
[1] Median (IQR); n (%)

Given the fact of not having enough evidences to affirm that our series of data follow a normal distribution, we will proceed to perform a Kruskal-Wallis test, which is a non-parametric variant of the ANOVA test.

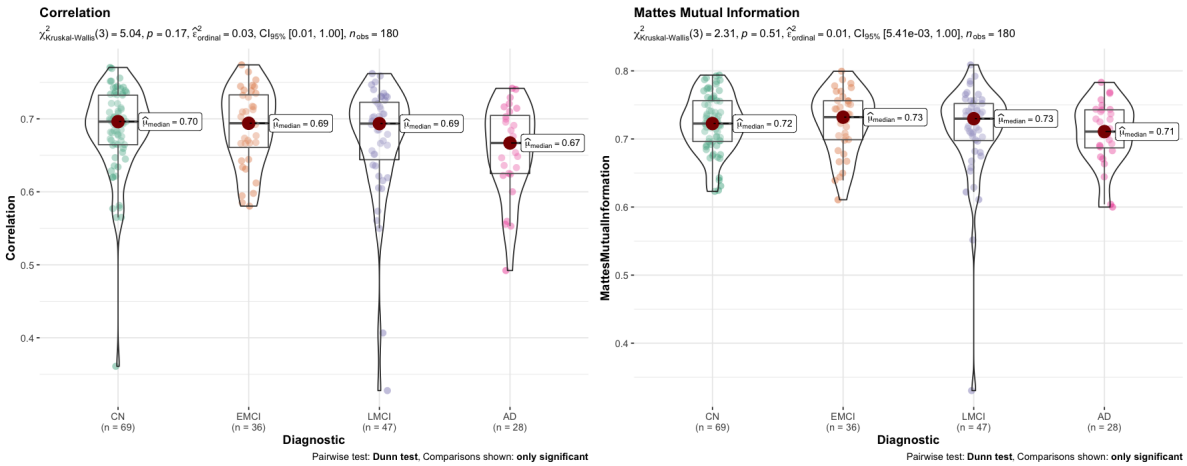
We will apply this test to the best 10 performing configurations, as on the previous study we have determined that they can be used indistinctly.

Furthermore, we will also perform a pairwise Dunn test for every pair of diagnostic groups in order to evaluate between which groups statistically significant differences tend to appear.

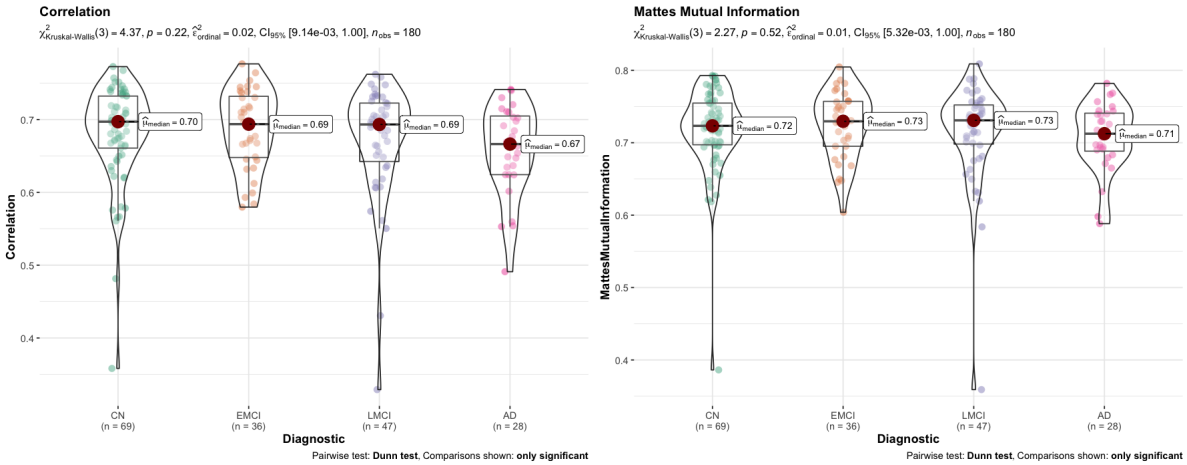
To conclude, we will compute the epsilon-squared metric in order to quantify the effect size existent among the different diagnostic groups for each of the configurations that are being considered. We have chosen epsilon-squared as it is the non-parametric variant of the multi-modal effect size metrics eta-squared and omega-squared.



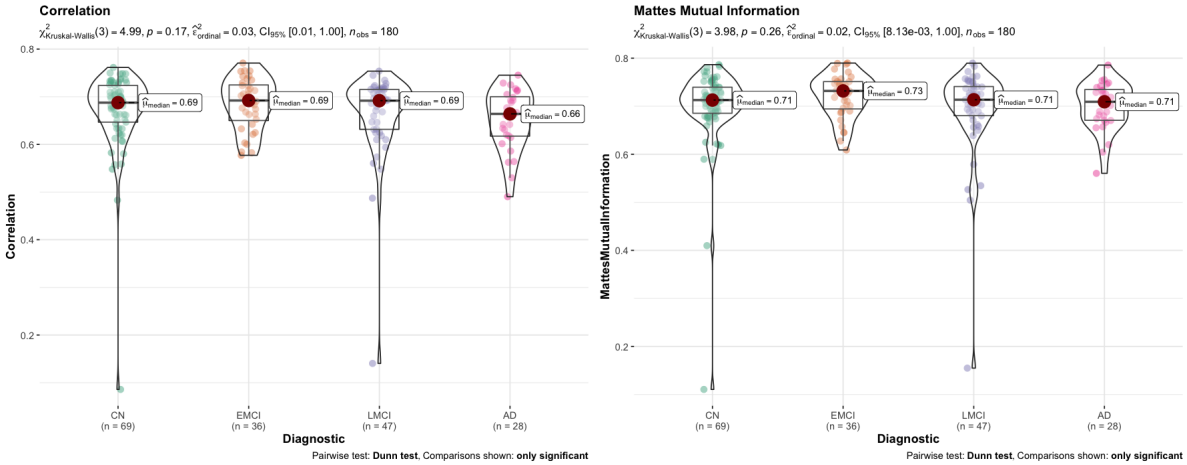
mri-coreg:mri-coreg:9dof



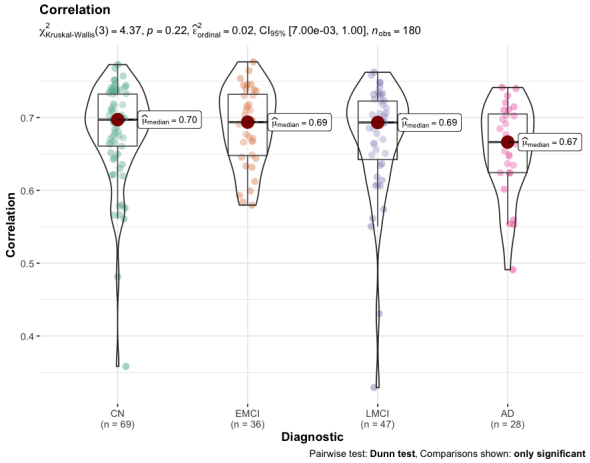
mri-coreg:mri-coreg:12dof

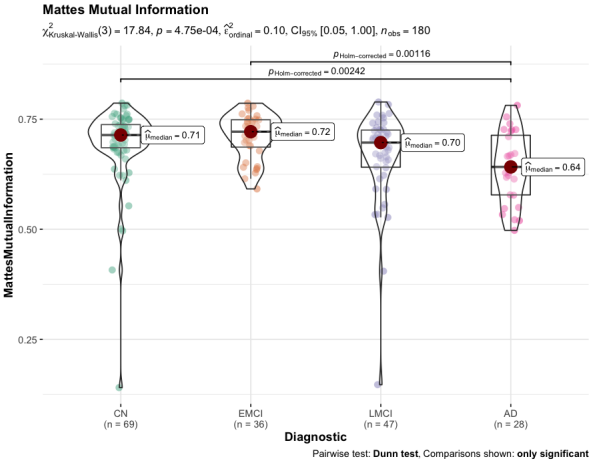


wregister:coreg:6dof

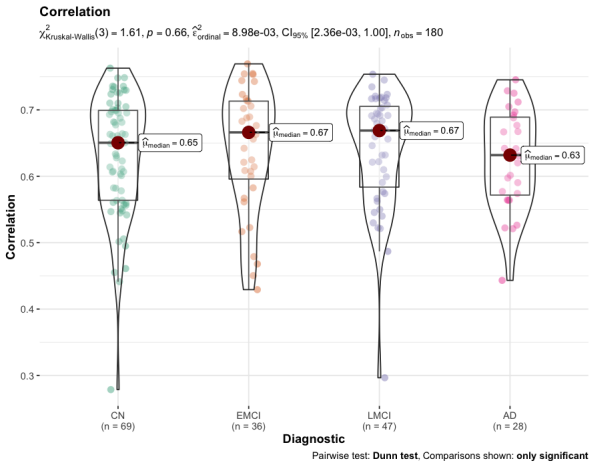


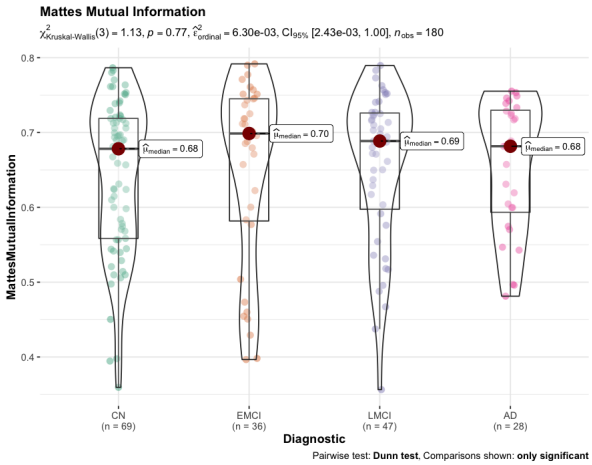
bbregister:coreg:6dof



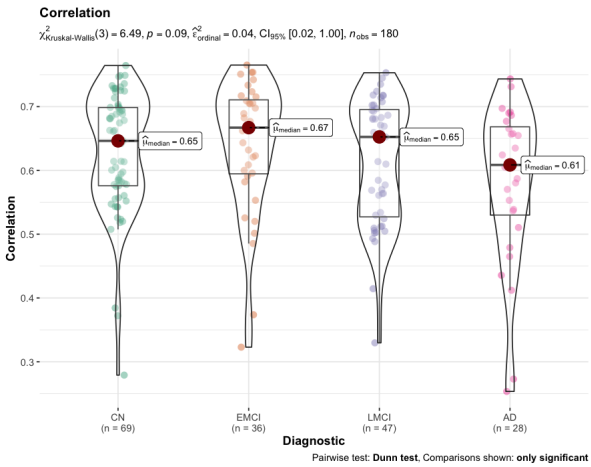


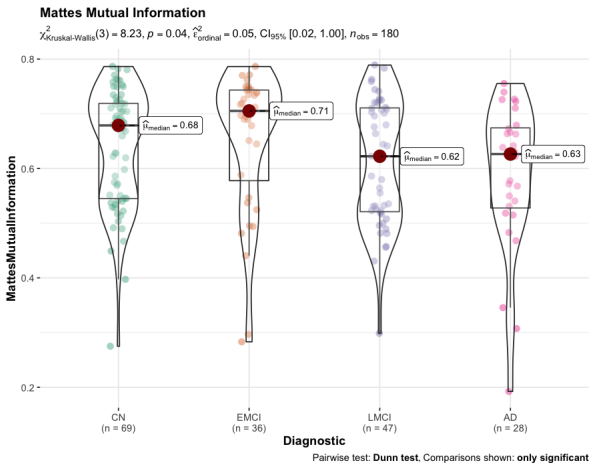
wregister:rr:6dof



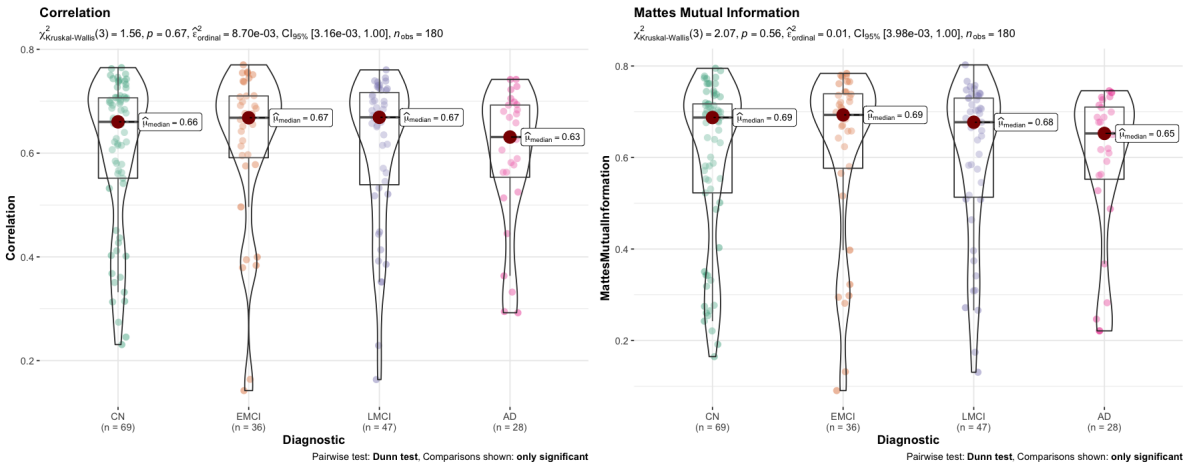


bbregister:rr:6dof

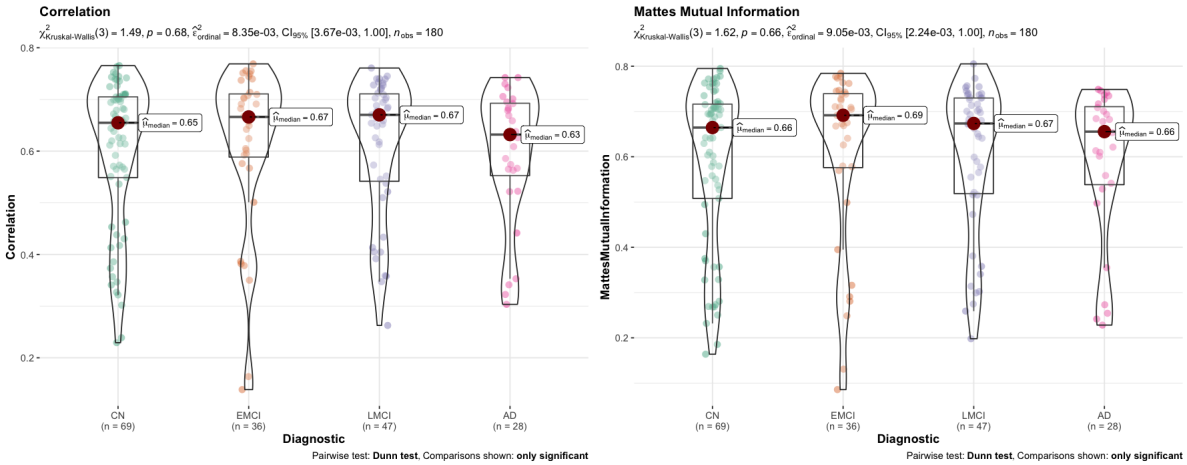




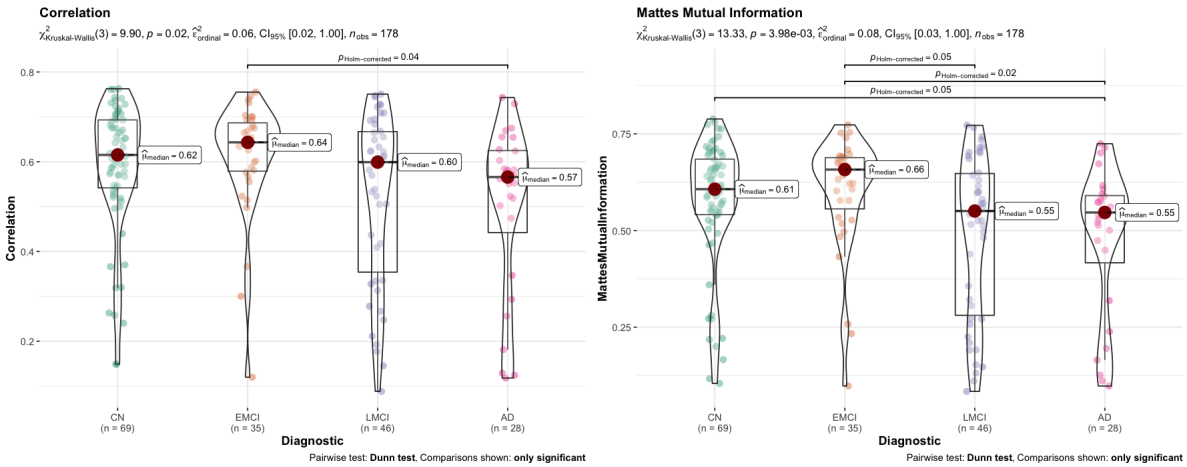
wregister:rr:12dof



wregister:rr:9dof



bbregister:rr:9dof



In the previuos figures, we can observe, for each of the ten best performing configurations, all the previously mentioned tests and their results.

The Kruskal-Wallis tests' associated p-values, provide evidences of some configurations being more robust than others. Specifically, we can appreciate how the ones generated with the BB-register algorithm tend to be the least robust, whilst the ones generated with W-register are

usually the most robust.

Moreover, we can appreciate, by contrast, that most of the pairwise comparisons considered have had associated p-values smaller than 0.05, except in some cases using the BB-register algorithm.

Finally, by taking a look at the epsilon-squared effect size metrics, we can see how the configuration that minimizes them for both metrics, thus should be considered as the most robust, is W-register with RR initialization and 12 Degrees of Freedom.

However, given the Kruskal-Wallis test associated p-value, we have enough evidence to support the fact that our 3 best performing configurations in terms of registration quality are also very robust.

To conclude this block, and given the results obtained for the three studies that structure it, we can affirm that, given the registration quality and robustness factors that should be considered for evaluating the best 2 configuration of parameters in order to perform tau-PET to MRI registration are **MRI-coreg with 12 degrees of freedom** and **W-register with RR initialization and 6 degrees of freedom**.

Appendix B

In this second block, our objective will be to evaluate the SUVr values obtained in two different stages of the registration process. This values quantify the amount of Tau protein existent in the brain structure and are used to spot tauopathy (aggregation of tau protein into neurofibrillary tangles). Specifically, we will determine the best space where to compute these values (Subject Space vs. Standard Space) in order to maximize the diagnostic groups' separation. Lastly, we will evaluate the classification and prediction capabilities of a data-driven method based on these SUVr values.

This block will be splitted into 2 studies.

Prior to the initialization of this block, we well reduce the dimensionallity of our data frame by filtering it and selecting only the SUVr values computed for the registrations that have been performed by using the most suitable configurations, derived from Block 1, MRI-coreg with 12 degrees of freedom and W-register with RR initialization and 6 degrees of freedom.

Study 4 - Determining the best space to compute the SUVr

To start this study, we will perform a Shapiro test in order to determine whether our series of data follow a normal distribution or not.

MRI-coreg:12dof Standard Space SUVr

```
    Shapiro-Wilk normality test

data:  df.best$std
W = 0.51341, p-value < 2.2e-16
```

MRI-coreg:12dof Subject Space SUVr

```
    Shapiro-Wilk normality test

data:  df.best$subj
W = 0.35503, p-value < 2.2e-16
```

W-register:RR:6dof Standard Space SUVr

```
    Shapiro-Wilk normality test

data:  df.best$std
W = 0.51341, p-value < 2.2e-16
```

W-register:RR:6dof Subject Space SUVr

```
    Shapiro-Wilk normality test

data:  df.best$subj
W = 0.35503, p-value =4.293e-16
```

In the previous outputs we can see how the tests applied to the series of SUVr values that we are considering show enough evidences as to refute the null hypothesis, thus we can affirm that our series of data are not normally distributed.

Preliminary paired Wilcoxon test for all groups

Bearing in mind that our data consists in two different measurements for the each registration process, we will perform a paired Wilcoxon test in order to evaluate the similarity of these two series of data.

MRI-coreg:12dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 1.539e-06
alternative hypothesis: true location shift is not equal to 0
```

W-register:RR:6dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 0.0005507
alternative hypothesis: true location shift is not equal to 0
```

In the output above we can observe how the p-values associated to the paired Wilcoxon test exhibits enough evidence as to affirm that our series of data are statistically different. Therefore, we will proceed to segregate our data frame by diagnostic group in order to evaluate if that assumption is also true for each of the diagnostic groups.

Paired Wilcoxon test within Diagnostic groups

We have ordered the different groups according to the chronology in which they appear throughout the AD continuum.

CN

MRI-coreg:12dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj[inds.CN] and df.best$std[inds.CN]
V = 1039, p-value = 0.3152
alternative hypothesis: true location shift is not equal to 0
```

W-register:RR:6dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 0.5823
alternative hypothesis: true location shift is not equal to 0
```

For the CN diagnostic group, the paired Wilcoxon test doesn't exhibit enough evidences of the two series of data being statistically different.

EMCI

MRI-coreg:12dof

```
    Wilcoxon signed rank exact test

data:  df.best$subj[inds.EMCI] and df.best$std[inds.EMCI]
V = 206, p-value = 0.04597
alternative hypothesis: true location shift is not equal to 0
```

W-register:RR:6dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 0.02603
alternative hypothesis: true location shift is not equal to 0
```

For the EMCI diagnostic group, the paired Wilcoxon test exhibits enough evidences of the two series of data being statistically different.

LMCI

MRI-coreg:12dof

```
    Wilcoxon signed rank exact test

data:  df.best$subj[inds.LMCI] and df.best$std[inds.LMCI]
V = 268, p-value = 0.001359
alternative hypothesis: true location shift is not equal to 0
```

W-register:RR:6dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 0.003432
alternative hypothesis: true location shift is not equal to 0
```

For the LMCI diagnostic group, the paired Wilcoxon test shows enough evidences of the two series of data being statistically different.

AD

MRI-coreg:12dof

```
    Wilcoxon signed rank exact test

data:  df.best$subj[inds.AD] and df.best$std[inds.AD]
V = 51, p-value = 0.0002441
alternative hypothesis: true location shift is not equal to 0
```

W-register:RR:6dof

```
    Wilcoxon signed rank test with continuity correction

data:  df.best$subj and df.best$std
V = 4780, p-value = 0.01362
alternative hypothesis: true location shift is not equal to 0
```

For the AD diagnostic group, the paired Wicox test shows enough evidences of the two series of data being statistically different.

In summary, we can establish that, as we move along the AD continuum, the differences between the two series of data tend to increase. The rationale is that, the topography of tau-PET changes along the AD continuum, and as a result, different diagnostic groups might be more impacted in how the SUVr is computed.

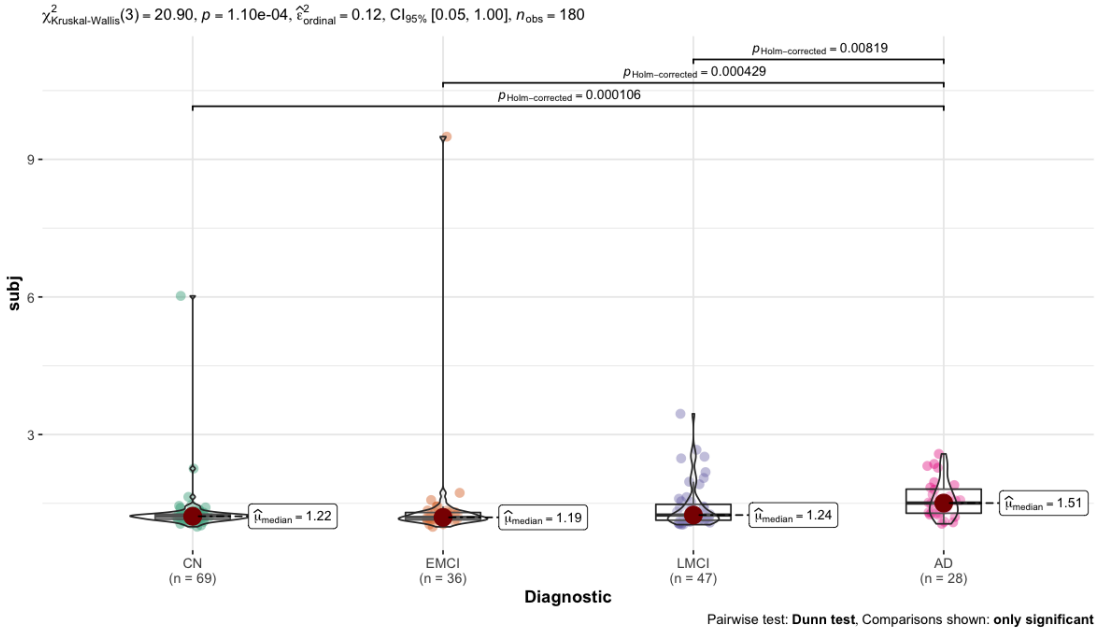
Post-hoc ggBetweenStats between SUVr data

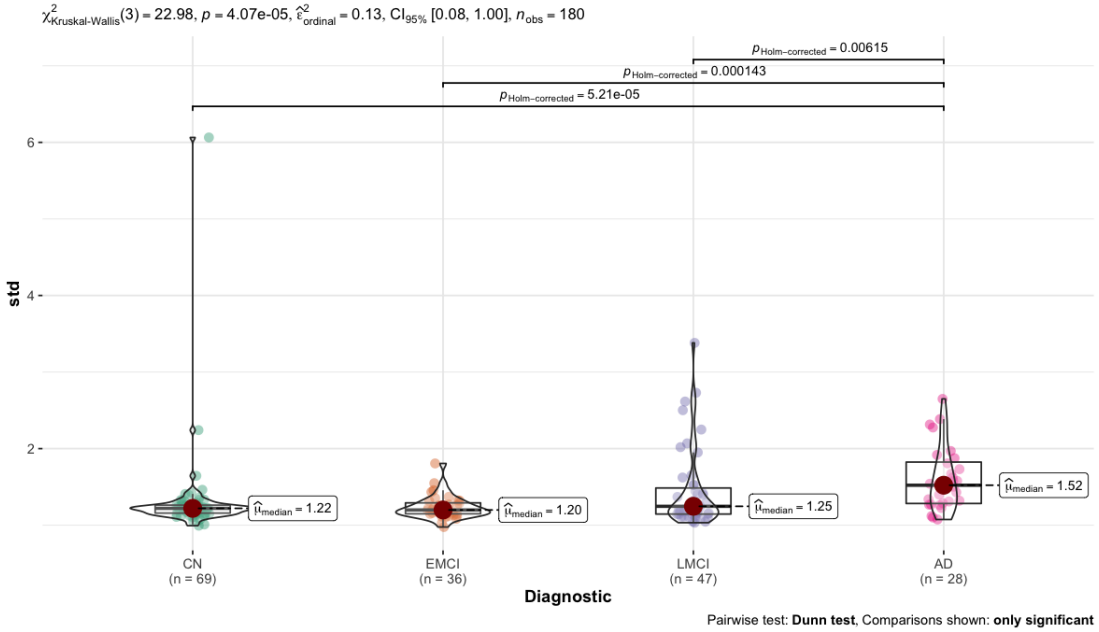
Once we have established that both spaces' SUVr values are different, we must determine which one is preferable, henceforth we will perform Kruskal-Wallis non-parametric tests in order to determine the differences between diagnostic groups for each of the SUVr series.

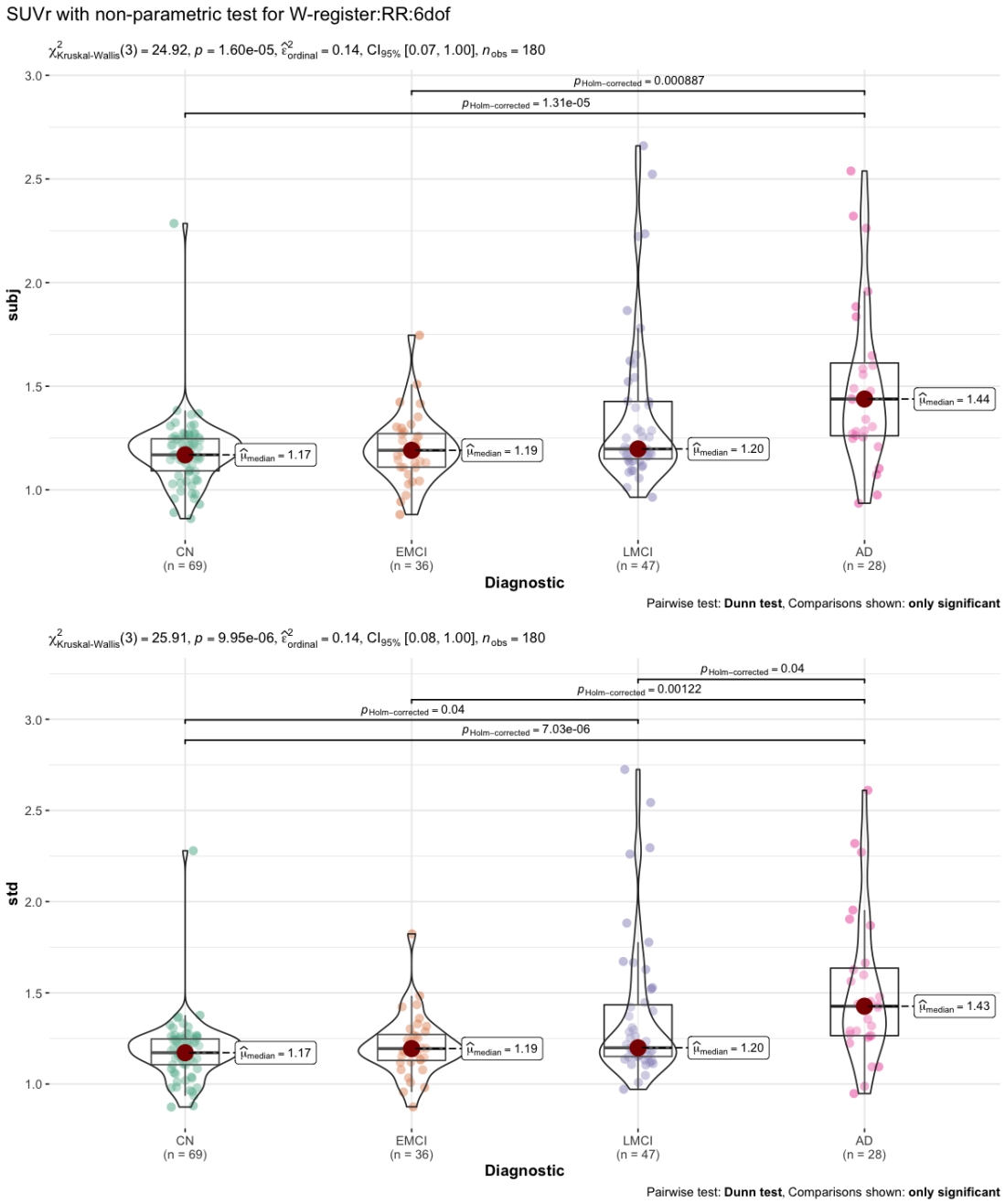
Moreover, we will also present pairwise Dunn tests for every pair of diagnostic groups in order to determine between which groups statistically significant differences do exist.

Finally, we will measure the effect size as a quantification of group separation with the epsilon-squared value.

SUVr with non-parametric test for MRI-coreg:12dofs







In the plots above, we can appreciate how statistically significant differences are displayed between AD and the rest of the groups. Furthermore, if we analyse the epsilon-squared statistic, we can see how it's value is bigger for the "std" series of data, hence we can establish that the Standard Space maximizes group separation and is preferable for classification and prediction purposes.

Study 5 - Classification and prediction capabilities of derived data-driven methods

In the previous study, we established that the most suitable space for computing the SUVr values was the Standard Space, henceforth, in this last study, we will only use the SUVr values

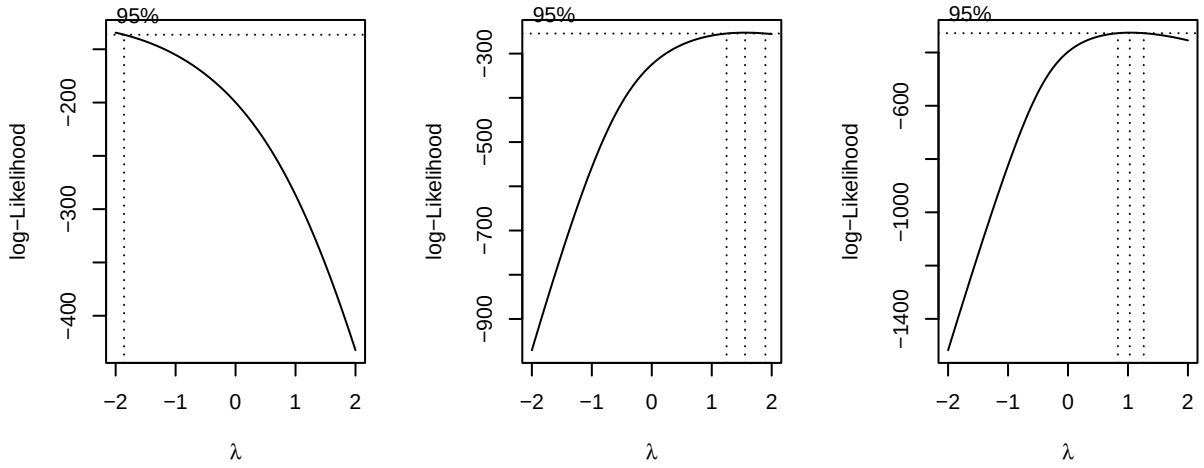
computed in that space, as they have been proved to serve better for classification and prediction purposes.

Transformations

Firsly, and a bearing in mind the the fact that our series of data isn't normally distributed, we well proceed to propose a series of transformations that should increase it's normality, with the objective of achieving a better fit when implementing the later logistic regression.

With that intention in mind, we use iterative BoxCox transformations by applying the the table below until we achieve a plot where $\lambda = 1$.

λ	Transformación
-2	$1/x^2$
-1	$1/x$
-0.5	$1/\sqrt{x}$
0	$\log(x)$
0.5	$\sqrt{x}$
1	x
2	x^2



$$f(SUVr) = \sqrt{\left(\frac{1}{SUVr_{std}{}^2}\right)^3}$$

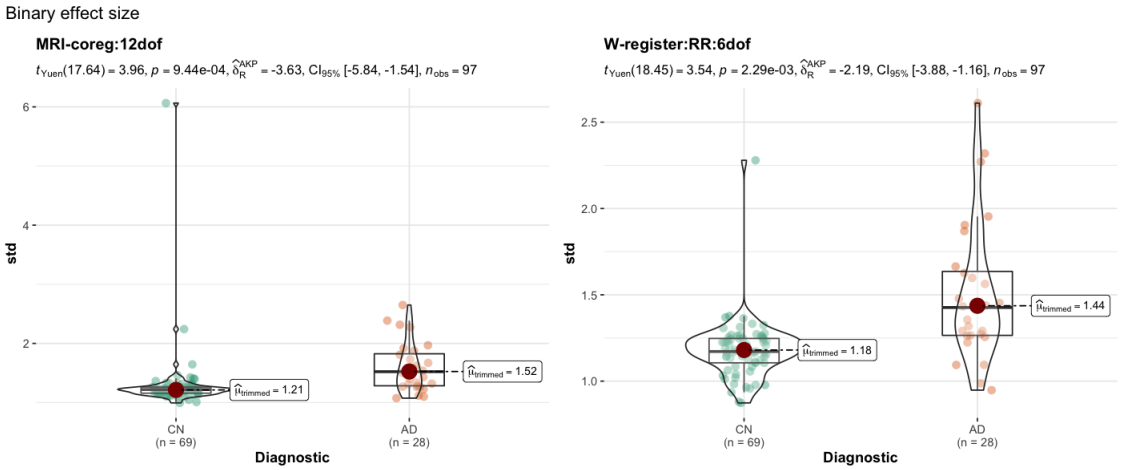
Once we have transformed our series of data according to the BoxCox transformations, we proceed to evaluate its normality with a Shapiro test.

```
    Shapiro-Wilk normality test
```

```
data:  df.best$std.transformed
W = 0.97473, p-value = 0.002342
```

The Shapiro tests shows evidence that our transformed series of data doesn't follow a normal distribution.

We will use the Yuen test, which is a robust independent sample t-test, in order to account for the difference in the distributions of our series of data. Furthermore, we will compute a robust variant of the Cohen's D effect size in order to quantify the diagnostic group separation.



We can appreciate how the absolute value of the effect size metric is bigger for the MRI-coreg:12dof series of data, which suggests it should have better classification and prediction capabilities.

Model definition

We will now proceed to propose two different logistic regression models, one for the SUVr values computed in the MRI-coreg:12dof registration, and another one for the WW-register:RR:6dof. We will also include some demographic co-variates that have been proved to be explanatory when characterizing the subjects diagnosis. Specifically, we will be considering their age and their APOE4, which is a gene closely related to AD.

$$Diagnostic \sim SUVr_{std} + Age + APOE4$$

Goodness of Fit

We then proceed to evaluate the Goodness of Fit of each of the models predictions to the real series of diagnostics with a Hosmer-Lemeshow Goodness of Fit test, specifically designed for logistic regression models.

MRI-coreg:12dof

```
    Hosmer and Lemeshow goodness of fit (GOF) test

data:  as.numeric(df.ADHC$Diagnostic) - 1, pred_test.std
```

```
X-squared = 6.0766, df = 8, p-value = 0.6387
```

W-register:RR:6dof

```
    Hosmer and Lemeshow goodness of fit (GOF) test

data:  as.numeric(df.ADHC$Diagnostic) - 1, pred_test.std.t
X-squared = 8.8574, df = 8, p-value = 0.3724
```

In both cases we can see how the tests do not provide enough evidence as to affirm that our models' fit to the data is poor.

ROC Curve analysis

Subsequently, we will perform a ROC curve analysis for the two models we have generated. Moreover, we will use a bootstrapping method, which is based in repetition, in order to reduce the heuristic effects that could influence the process.

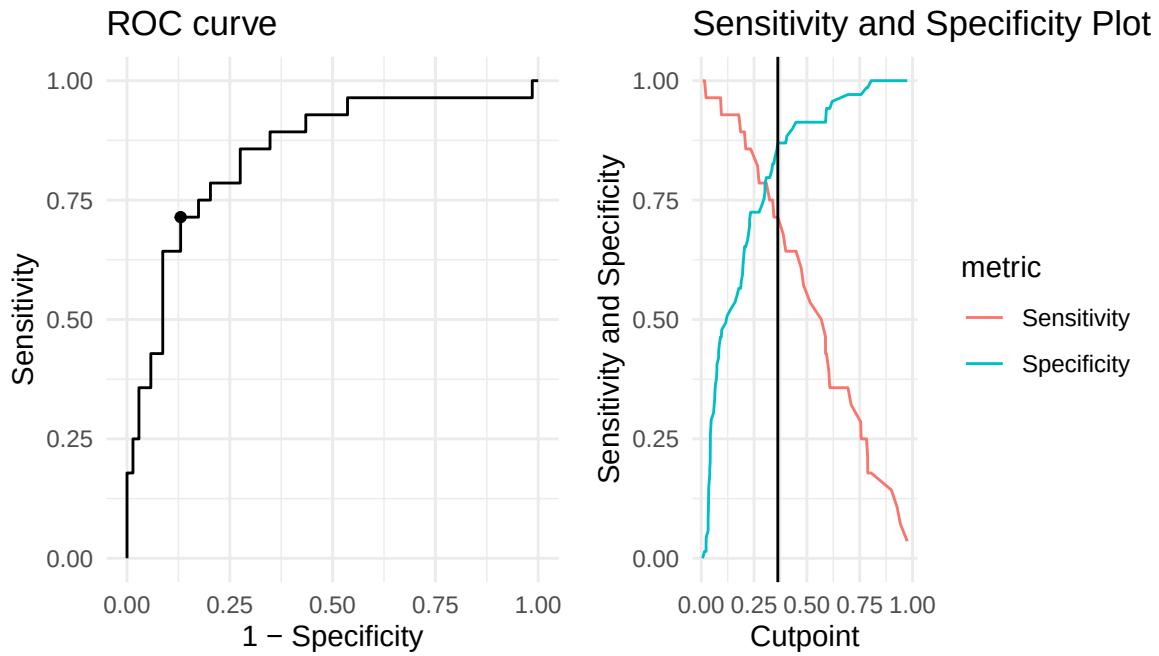
MRI-coreg:12dof model

In the output below we can numerically appreciate some of the metrics that are used to perform the ROC curve analysis for the model generated from the original series of SUVr values.

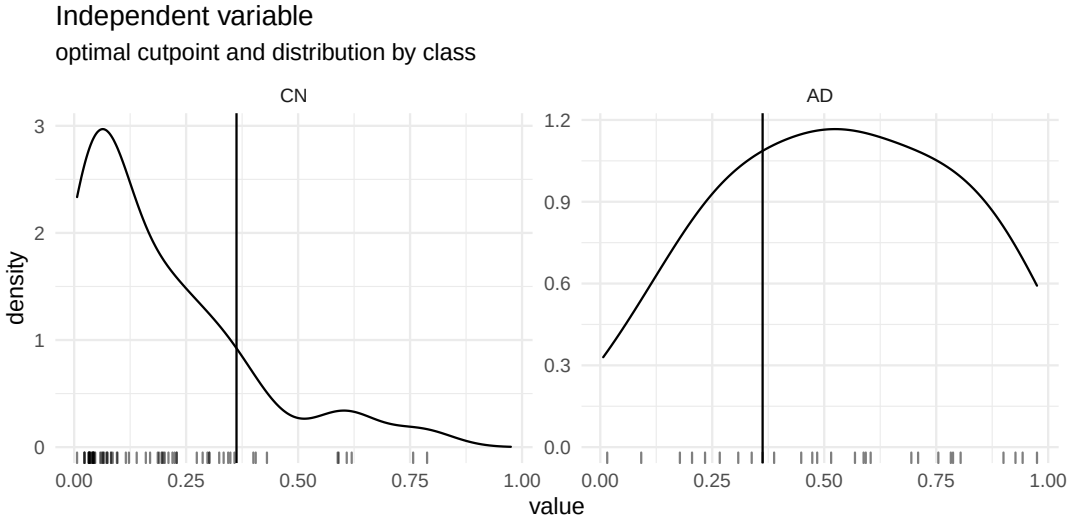
```
Method: maximize_metric
Predictor: pred_test.std
Outcome: Diagnostic
Direction: >=
Nr. of bootstraps: 1000

    AUC  n n_pos n_neg
 0.8483 97    28    69

 optimal_cutpoint sum_sens_spec    acc sensitivity specificity tp fn fp tn
           0.3625        1.5839 0.8247      0.7143      0.8696 20  8  9 60
```

Firstly, we can note that the analysis has been performed by executing a series of 1000 bootstraps. We can see that the AUC value for the optimized curve is 0.85, with 82% accuracy, 71% sensibility and 87% specificity.

In the figures below, we can visualize the ROC curve representation as well as the evolution of the sensibility and specificity of the model depending on the cut-point.



Finally, we can appreciate how the optimal threshold divides the CN and AD predicted values' distributions.



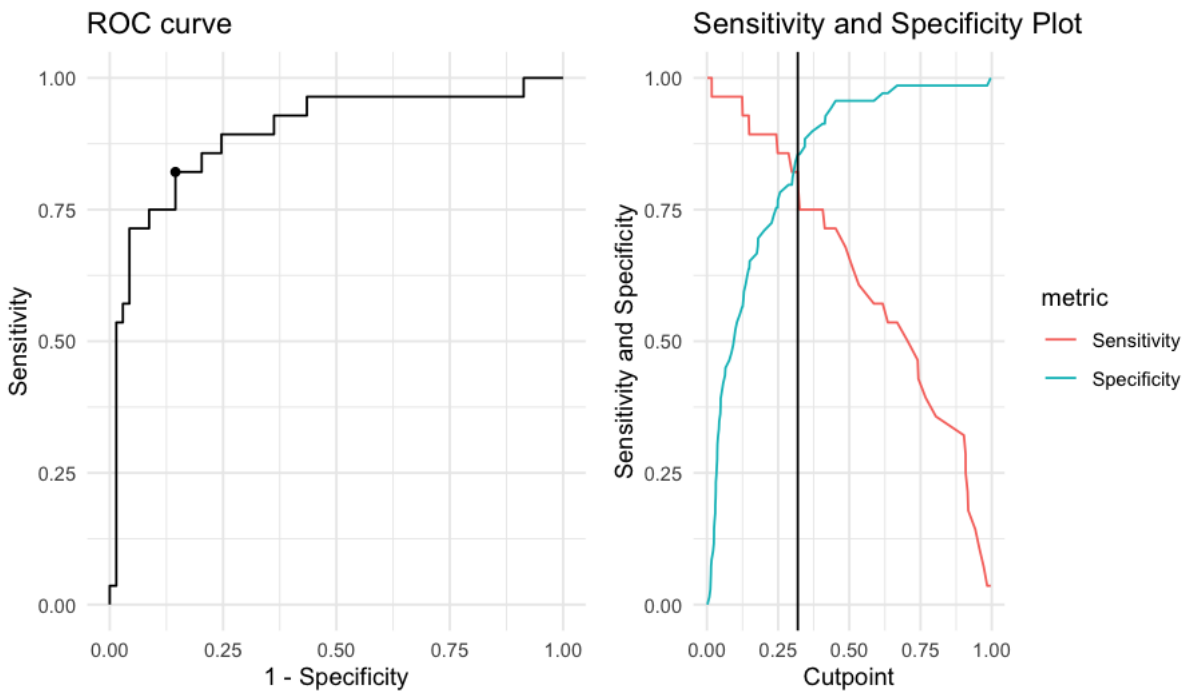
W-register:RR:6dof model

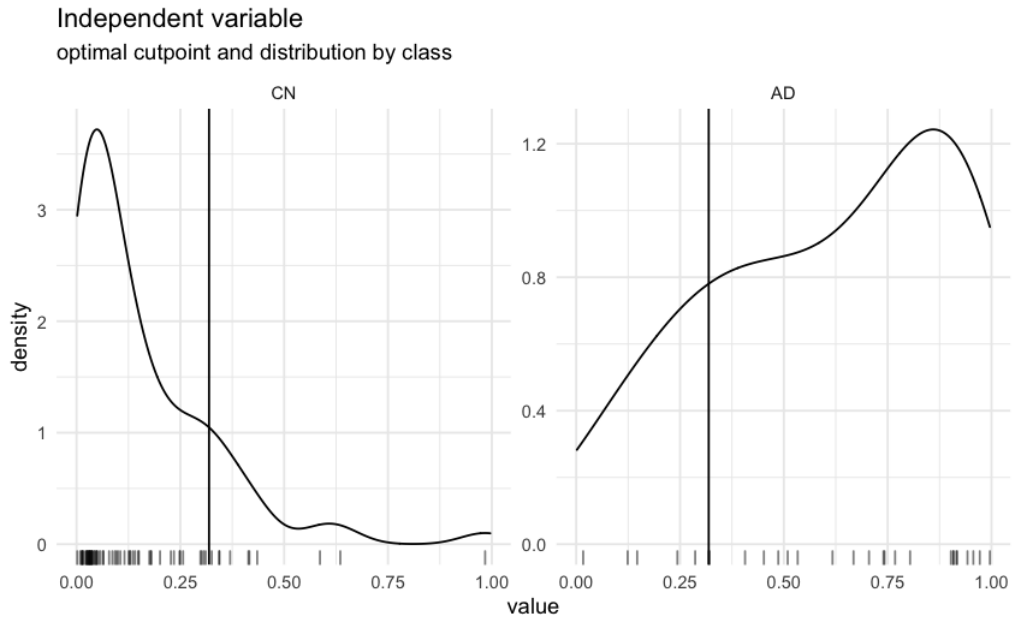
In the output below we can numerically appreciate some of the metrics that are used to perform the ROC curve analysis for the model generated from the transformed series of SUVr values.

```
Method: maximize_metric
Predictor: pred_test.std.t
Outcome: Diagnostic
Direction: >=
Nr. of bootstraps: 1000

    AUC  n n_pos n_neg
 0.8949 97    28    69

 optimal_cutpoint sum_sens_spec    acc sensitivity specificity tp fn fp tn
           0.3191         1.6765 0.8454       0.8214       0.8551 23  5 10 59
```

We can see that the AUC value for the optimized curve is 0.90, with 85% accuracy, 82% sensibility and 86% specificity.

In the figures below, we can visualize the ROC curve representation as well as the evolution of the sensibility and specificity of the model depending on the cut-point.





Comparison of ROC curves

To conclude, we will compare both ROC curves with a bootstrapping method in order to seek for statistically significant differences between them, and determine which of the models is the most suitable for the study of AD.

```
    Bootstrap test for two correlated ROC curves

data:  pred_test.std and pred_test.std.t by as.numeric(df.ADHC$Diagnostic)
- 1 (0, 1)
D = 0, boot.n = 2000, boot.stratified = 1, p-value = 1
alternative hypothesis: true difference in AUC is not equal to 0
sample estimates:
pAUC (0.880952380952381-0.848343685300207 specificity) of roc1
                                                  0.0326087
pAUC (0.880952380952381-0.848343685300207 specificity) of roc2
                                                  0.0326087
```

The comparison shows no evidence of statistically significant differences, thus both models could be used indistinctly for classification and prediction purposes.

References

[1] Niyazi Acer, Ahmet Tuncay, Yelda Ozsunar, and Mehmet Turgut. Quantification of volumetric changes of brain in neurodegenerative diseases using magnetic resonance imaging and stereology. *Neurodegenerative Diseases - Processes, Prevention, Protection and Monitoring*, 12 2011.

[2] Paige Bennett, Akiva Mintz, Brad Perry, Andrew Trout, and Paula Vergara-Wentland. *Specialty Imaging: PET*. Elsevier, 1 2017.

[3] Nicolaas I. Bohnen, David S.W. Djang, Karl Herholz, Yoshimi Anzai, and Satoshi Minoshima. Effectiveness and safety of 18f-fdg pet in the evaluation of dementia: A review of the recent literature. *Journal of Nuclear Medicine*, 53:59–71, 1 2012.

[4] Jeffrey L. Cummings and Greg Cole. Alzheimer disease. *JAMA*, 287:2335–2338, 5 2002.

[5] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, Marilyn S Albert, and Ronald J Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. 2006.

[6] Leonardino A. Digma, John R. Madsen, Emilie T. Reas, Anders M. Dale, James B. Brewer, and Sarah J. Banks. Tau and atrophy: Domain-specific relationships with cognition. *Alzheimer's Research and Therapy*, 11, 7 2019.

[7] Alexis Dinno. Nonparametric pairwise multiple comparisons in independent groups using dunn's test. *Stata Journal*, 15:292–300, 4 2015.

[8] E. Genin, D. Hannequin, D. Wallon, K. Sleegers, M. Hiltunen, O. Combarros, M. J. Bullido, S. Engelborghs, P. De Deyn, C. Berr, F. Pasquier, B. Dubois, G. Tognoni, N. Fiévet, N. Brouwers, K. Bettens, B. Arosio, E. Coto, M. Del Zompo, I. Mateo, J. Epelbaum, A. Frank-Garcia, S. Helisalmi, E. Porcellini, A. Pilotto, P. Forti, R. Ferri, E. Scarpini, G. Siciliano, V. Solfrizzi, S. Sorbi, G. Spalletta, F. Valdivieso, S. Vepsäläinen, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Boss, O. Hanon, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, F. Licastro, H. Soininen, J. F. Dartigues, M. I. Kamboh, C. Van Broeckhoven, J. C. Lambert, P. Amouyel, and D. Campion. Apoe and alzheimer disease: a major gene with semi-dominant inheritance. *Molecular Psychiatry 2011 16:9*, 16:903–907, 5 2011.

[9] Douglas N. Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48:63–72, 10 2009.

[10] Leonardo Iaccarino, Renaud La Joie, Lauren Edwards, Amelia Strom, Daniel R Schonhaut, Rik Ossenkoppele, Julie Pham, Taylor Mellinger, Mustafa Janabi, Suzanne L Baker, David Soleimani-Meigooni, Howard J Rosen, Bruce L Miller, William J Jagust, and Gil D Rabinovici. Spatial relationships between molecular pathology and neurodegeneration in the alzheimer's disease continuum. *Cerebral Cortex*, 31:1–14, 1 2021.

[11] Kh Tohidul Islam, Sudanthi Wijewickrema, and Stephen O'Leary. A deep learning based framework for the registration of three dimensional multi-modal medical images of the head. *Scientific Reports*, 11:1860, 2021.

[12] Seun Jeon, Jae Myeong Kang, Seongho Seo, Hye Jin Jeong, Thomas Funck, Sang Yoon Lee, Kee Hyung Park, Yeong Bae Lee, Byeong Kil Yeon, Tatsuo Ido, Nobuyuki Okamura,

Alan C. Evans, Duk L. Na, and Young Noh. Topographical heterogeneity of alzheimer's disease based on mr imaging, tau pet, and amyloid pet. *Frontiers in Aging Neuroscience*, 10:211, 2019.

[13] Stefan J Kiebel, John Ashburner, Jean-Baptiste Poline, and Karl J Friston. Mri and pet coregistration—a cross validation of statistical parametric mapping and automated image registration. *NeuroImage*, 5:271–279, 1997.

[14] Antoine Leuzy, Konstantinos Chiotis, Laetitia Lemoine, Per Göran Gillberg, Ove Almkvist, Elena Rodriguez-Vieitez, and Agneta Nordberg. Tau pet imaging in neurodegenerative tauopathies—still a challenge, 8 2019.

[15] Antoine Leuzy, Tharick A Pascoal, Olof Strandberg, Philip Insel, Ruben Smith, Niklas Mattsson-Carlgren, Andréa L Benedet, Hannah Cho, Chul H Lyoo, Renaud, La Joie, Gil D Rabinovici, Rik Ossenkoppele, Pedro Rosa-Neto, and Oskar Hansson. A multicenter comparison of [18 f]flortaucipir, [18 f]ro948, and [18 f] mk6240 tau pet tracers to detect a common target roi for differential diagnosis. *European Journal of Nuclear Medicine and Molecular Imaging*, 48:2295–2305, 2021.

[16] Mona-Lisa Malarte, Agneta Nordberg, and Laetitia Lemoine. Characterization of mk6240, a tau pet tracer, in autopsy brain tissue from alzheimer's disease cases. *European Journal of Nuclear Medicine and Molecular Imaging*, 48:1093–1102, 2021.

[17] Antonio Maldonado Suarez, Manuel Recio Rodriguez, Jimenez De La Peña Mar, David Ezpeleta Echavarri, Silvia Fuertes Cabero, and Vicente Martinez De Vega. Pet-rm multitrazador en el estudio de la enfermedad de alzheimer. *Seram*, 11 2018.

[18] Camille Noirot, Ismini Mainta, Aline Mendes, Paulina Andryszak, Hishayine Visvaratnam, Paul G. Unschuld, Giovanni B. Frisoni, and Valentina Garibotto. Tau pet imaging evidence in patients with cognitive impairment: preparing for clinical use. *Clinical and Translational Imaging 2018 6:6*, 6:471–482, 10 2018.

[19] Rik Ossenkoppele, Daniel R Schonhaut, Michael Schöll, Samuel N Lockhart, Nagehan Ayakta, Suzanne L Baker, James P O'Neil, Mustafa Janabi, Andreas Lazaris, Averill Cantwell, Jacob Vogel, Miguel Santos, Zachary A Miller, Brianne M Bettcher, Keith A Vossel, Joel H Kramer, Maria L Gorno-Tempini, Bruce L Miller, William J Jagust, and Gil D Rabinovici. Tau pet patterns mirror clinical and neuroanatomical variability in alzheimer's disease. *Brain*, 139:1551–1567, 5 2016.

[20] Jordi Pegueroles, Victor Montal, Alexandre Bejanin, Eduard Vilaplana, Mateus Aranha, Miguel Angel Santos-Santos, Daniel Alcolea, Ignasi Carrió, Valle Camacho, Rafael Blesa, Alberto Lleó, and Juan Fortea. Amyq: An index to standardize quantitative amyloid load across pet tracers. *Alzheimer's Dementia*, 17:1499–1508, 9 2021.

[21] Michael J Pontecorvo, Michael D Devous, Ian Kennedy, Michael Navitsky, Ming Lu, Nicholas Galante, Stephen Salloway, P Murali Doraiswamy, Sudeepti Southekal, Anupa K Arora, Anne McGeehan, Nathaniel C Lim, Hui Xiong, Stephen P Truocchio, Abhinay D Joshi, Sergey Shcherbinin, Brian Teske, Adam S Fleisher, Mark A Mintun, and for the 18F-AV-1451-A05 investigators. A multicentre longitudinal study of flortaucipir (18f) in normal ageing, mild cognitive impairment and alzheimer's disease dementia. *Brain*, 142:1723–1735, 6 2019.

[22] Christopher G Schwarz, Terry M Therneau, Stephen D Weigand, Jeffrey L Gunter, Val J Lowe, Scott A Przybelski, Matthew L Senjem, Hugo Botha, Prashanthi Vemuri, Kejal

Kantarci, Bradley F Boeve, Jennifer L Whitwell, Keith A Josephs, Ronald C Petersen, David S Knopman, and Clifford R Jack. Selecting software pipelines for change in flortaucipir suvr: Balancing repeatability and group separation. *NeuroImage*, 238:118259, 2021.

[23] Dennis J Selkoe and John Hardy. The amyloid hypothesis of alzheimer's disease at 25 years. *EMBO molecular medicine*, 8:595–608, 6 2016.

[24] Nicholas J Tustison, Brian B Avants, Philip A Cook, Junghoon Kim, John Whyte, James C Gee, and James R Stone. Logical circularity in voxel-based analysis: normalization strategy may induce statistical bias. *Human brain mapping*, 35:745–759, 3 2014.