



UNIVERSITAT POLITÈCNICA
DE CATALUNYA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
TEORIA DEL SENYAL I COMUNICACIONS

This thesis is submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy (PhD)

STUDY OF EARLY STAGES OF ALZHEIMER'S
DISEASE USING MAGNETIC RESONANCE
IMAGING

by ADRIÀ CASAMITJANA DAZ

Advisor: Verónica Vilaplana
Barcelona, November 2019

Abstract

Alzheimer's disease (AD) is a neurodegenerative disorder that constitutes the most common dementia pathology. It represents a global epidemic that expands exponentially as the life expectancy increases with no yet useful treatment. Currently, it represents a huge social and economic burden for our societies and it is expected to tension public health infrastructures and finances in the near future.

AD is characterized by amyloid plaque deposition and neurofibrillary tangles measured by ex-vivo examination of the brain. Recent developments in fluid biomarkers and brain imaging allow in-vivo quantification of pathophysiological processes of amyloid deposition or tau tangles formation in the brain, providing the community with highly sensitive and specific in-vivo biomarkers for Alzheimer's disease diagnosis. Abnormal levels of these biomarkers are thought as the initiating event to a cascade of subsequent events that continue with synapse loss, cell death, memory impairment, functional dysfunction and cognitive decline. All these events constitute the Alzheimer's continuum which can be broadly split into two main parts: an initial long and silent preclinical stage characterized by abnormal AD biomarkers and cognition within the normal range that could last from 15 to 30 years and a posterior clinical stage where subjects develop dementia symptoms.

The etiology of AD is still poorly understood even though several risk factors are identified. Large observational studies can help the study of AD and its related biomarkers and risk factors. In this thesis we provide methodological tools for the analysis of Alzheimer's disease using magnetic resonance imaging (MRI). We focus on the study of subjects within the preclinical stage of AD by using statistical learning and pattern recognition frameworks to perform inferential statistics and develop predictive models.

The main outcomes of this thesis are three-fold: firstly, we develop an open-source toolbox for nonlinear neuroimage analysis in population studies. While nonlinear association between medical images and several factors is already known, standard neuroimaging softwares only provide linear statistical frameworks that limit the analyses. Secondly, we study the relationship between brain structure using MRI and the underlying Alzheimer's pathology along the disease continuum and at different stages. The close relationship between MRI and clinical symptoms has been widely studied but describing AD using biomarkers instead of clinical phenotypes allows us to study preclinical stages of AD. Finally, we present a framework to predict cognitively unimpaired and amyloid positive subjects using MR imaging and machine learning. We report the results in a cross-sectional study and in a longitudinal study that compares the volumetric rate-of-change between subjects with different amyloid status. We further test the proposed methodology as a part of the triaging process in clinical trials showing great potential benefits.

Acknowledgments

This thesis is a result from the ideas and contributions of many people and thanks to the support of many others. It can only be understood as a cooperational work that helped me to grow intellectually, scientificaly and humanly. My honest gratitude to all of them for accompanying me from the begining to the end and throughout the PhD journey with special appreciation to all my co-authors.

First, I must thank my supervisor, Verónica Vilaplana, for the unconditional support showed during all these years. Her passion to explore new fields and to continuously learn new concepts is very inspiring. She helped a lot at tough early stages, by providing organization, conceptual frameworks and good research practices. Throughout these years, she has been an invaluable help on increasing communicating skills, both writing and orally. She also gave me the first opportunity in teaching, one of my other passions.

Any of that would have been possible, though, without the initial contribution of Ferran Marqués, a professor from Universitat Politècnica de Catalunya (UPC) that helped a lot with the scholarship application and set the initial link with our collaborators at BarcelonaBeta Brain Research Center (BBRC).

I would also like to express my gratitude to all Neuroimaging Research Group at BBRC and specially to Dr. Juan Domingo Gispert (PI) for what I think has been a succesful collaboration. More than that, they have been an indispensable guidance and a source of ideas through really helpful discussions.

I acknowledge that this thesis would not have been possible without the financial assistance of the Image Processing Group (GPI) at UPC and the Spanish Government through the FPU 14/05988 Research Fellowship and the projects TEC2013-43935-R and TEC2016-75976-R. I must also acknowledge Albert Gil and Josep Pujal, for setting up and maintaining the computing server at GPI, and also for the high-quality technical support they steadily provide us with.

These years of my PhD would have been much harder without the support of my family and friends. Very special thanks to all colleagues at D5-120 that created a better work environment. I would also to extend thanks to all colleagues that I met at conferences, seminars, summer schools and during my visit at CMIC from whom I learned a lot.

To all of them and with love, thank you!

Contents

List of Figures	vii
List of Tables	ix
Glossary	xi
1 Introduction	1
1.1 Context and motivation	1
1.2 Contributions	4
2 Literature review	7
2.1 Alzheimer’s disease	7
2.1.1 Diagnostic criteria	7
2.1.2 Preclinical AD	12
2.2 Magnetic Resonance Imaging	14
2.2.1 Structural MRI	15
2.2.2 Diffusion MRI	16
2.2.3 Functional MRI	17
2.3 Pattern recognition in neuroimaging	17
2.3.1 Pattern recognition in Alzheimer’s research	21
3 NeAT: a nonlinear Neuroimaging Analysis Toolbox.	25
3.1 Introduction	25
3.2 The toolbox	26
3.2.1 NeAT overview	26
3.2.2 Model estimation	26
3.2.3 Statistical inference	30
3.2.4 Post-hoc analysis	31
3.2.5 Visualization	32
3.2.6 NeAT specifications	32
3.3 Case study 1: Atrophy patterns across the Alzheimer’s disease continuum	33
3.4 Case study 2: Effects of <i>APOE</i> - ϵ 4 in brain aging	34
3.4.1 <i>APOE</i> genotype effects on brain morphology in normal aging pop- ulation.	35
3.4.2 Interaction between <i>APOE</i> genotype and age in normal aging pop- ulation.	37
3.5 Conclusions	37
4 Brain structure in AD pathology	41
4.1 Introduction	41

4.1.1	Related work	42
4.2	Methodology	43
4.2.1	Partial least squares	43
4.2.2	Partial least squares orthogonalization	46
4.2.3	Statistical inference	47
4.3	Univariate brain structural effects	49
4.3.1	Material and methods	49
4.3.2	Results	50
4.3.3	Discussion	57
4.4	Multivariate brain structural effects	60
4.4.1	Material and methods	60
4.4.2	Results	61
4.4.3	Discussion	68
4.5	Conclusions	70
5	MRI and Machine Learning in Clinical Trials	71
5.1	Introduction	71
5.1.1	Motivation	72
5.2	Clinical trials	73
5.2.1	Past clinical trials	73
5.2.2	On-going clinical trials	74
5.2.3	Hypothesis	75
5.2.4	Savings	77
5.3	Cross-sectional analysis	82
5.3.1	Material and methods	82
5.3.2	Results	85
5.3.3	Discussion	88
5.4	Longitudinal analysis	92
5.4.1	Material and methods	93
5.4.2	Results	96
5.4.3	Discussion	101
5.5	Conclusions	105
	Conclusions	107
	Publications	109
	Bibliography	113
	Appendices	139
	Appendix A Chapter 3	139
	A.1 List of relevant regions along the AD continuum	139
	A.2 List of relevant regions in different clinical AD stages.	141
	A.3 List of relevant regions in presymptomatic AD	143
	A.4 Linear regression for prediction	145
	A.5 Statistical inference tables	145
	Appendix B Chapter 5	150
	B.1 Cross-sectional analysis	150
	B.2 Longitudinal analysis	153

List of Figures

1.1	Estimated number of people living with dementia worldwide (2015-2050)	2
1.2	An illustrative figure of amyloid plaque deposition and neurofibrillary tangles in a brain neuron compared to a healthy neuron.	3
1.3	Comparison of a normal aged brain and the brain of a person with Alzheimer's.	3
2.1	Different MRI contrasts: T1, T2 and FLAIR.	16
2.2	Typical machine learning pipeline.	19
3.1	Toolbox pipeline	27
3.2	Comparison between different curve fitting models: GLM, GAM, polynomial SVR and gaussian SVR.	34
3.3	Curve clustering algorithm run on relevant atrophy patterns along the AD-CSF index using GAM fitting.	35
3.4	Statistical inference using volumetric data and different curve fitting modules: GLM, GAM and polynomial SVR.	36
3.5	Statistical inference using cortical thickness data and GLM.	36
3.6	Differences between statistical maps of HE model using GLM and GAM at different brain ROI	37
3.7	Interaction between age and the <i>APOE-ε4</i> genotype using second order polynomial expansion of GLM and B-splines GAM.	39
4.1	Brain morphometric effect type related to confounders (age and sex)	51
4.2	Pairwise CSF biomarker relationship along the Alzheimer's continuum:	51
4.3	Effect type of AD pathology on brain morphology for subjects along the Alzheimer's continuum	52
4.4	Effect type of AD pathology on brain morphology for subjects in the CU stage	54
4.5	Effect type of AD pathology on brain morphology for subjects in the MCI stage	54
4.6	Effect type of AD pathology on brain morphology for subjects in the AD-dementia stage	55
4.7	Pairwise CSF biomarker relationship in asymptomatic subjects:	55
4.8	Effect type of AD pathology on brain morphology for subjects in the T0 tertile.	57
4.9	Effect type of AD pathology on brain morphology for subjects in the T1 tertile.	57
4.10	Effect type of AD pathology on brain morphology for subjects in the T2 tertile.	58
4.11	Absolute value of the effect size (single model) on several indicators evaluated at different cognitive stages.	62
4.12	Examples of relevant latent brain morphological patterns (effect-type) using volumetric features.	63
4.13	Examples of relevant latent brain morphology patterns (effect-type) using cortical thickness features.	64

4.14	Absolute value of the effect size (multiple models) on several indicators evaluated at the respective disease stages	66
4.15	AD-related effect-type of the brain latent model using volumetric features. . .	67
4.16	AD-related effect-type of the brain latent model using cortical thickness features.	68
5.1	AD prevention hypothesis: applying disease modifying therapies before neuronal loss.	72
5.2	Outline of clinical recruitment protocol of subjects with preclinical (PC) Alzheimer’s pathology.	78
5.3	Core model training and validation for automated MRI-based classification of $A\beta$ positive ($A\beta+$) subjects using logistic regression.	84
5.4	Examples of distributions of T1 features	86
5.5	Classification results using T1 features	87
5.6	Classification results using multimodal (T1+DTI) features	89
5.7	Workflow of the optimization and evaluation of the classification method. . .	95
5.8	Distribution of the interval Δt between reference and follow-up visits across the whole dataset.	97
5.9	AUC and savings (blue, green) reported using Jacobian determinant maps with different time intervals (Δt)	98
5.10	ROC and PR curves for Jacobian determinant maps with time spans in the range $2.5 < \Delta t < 3.5$ years using 0.5% of the features	99
5.11	Normalized feature maps of the 0.5% of features selected during the 100 different splits of the development/test sets	101
5.12	Statistical maps for group comparison between HC and PC (preclinical AD signature) and HC and MCI/AD-dementia (AD signature) subjects.	102
B.1	Top 10 most relevant features for (a) HC vs MCI and (b) HC vs AD-dementia models.	150

List of Tables

2.1	Historical overview of diagnostic criterias developed by two main working groups (NIA-AA and IWG) used in Alzheimer’s clinical and research practice.	12
4.1	Unsupervised cluster analysis of CSF effect types on brain morphology along the Alzheimer’s continuum	53
4.2	Unsupervised cluster analysis of CSF effect types on brain morphology along the Alzheimer’s continuum stratified by cognitive stage.	56
4.3	Unsupervised cluster analysis of CSF effect types on brain morphology in the asymptomatic stage of AD.	58
4.4	Predictive error of response variables for each diagnosis label and feature type using a single model for all Alzheimer’s continuum.	65
4.5	Predictive error of response variables for each diagnostic label and feature type using a separate models	68
5.1	Past phase III clinical trials that failed to show significant results in their primary outcomes.	74
5.2	Phase II trials conducted in Europe that are currently recruiting.	75
5.3	Phase III trials conducted in Europe that are currently recruiting.	76
5.4	Late-stage development trials (phase II/III or III) that attempt in cognitively unimpaired subjects.	76
5.5	Demographics of the ADNI subset and HCB cohort used for this study	83
5.6	Performance of the PC classification model according to the number of T1 features selected	86
5.7	Performance of the PC classification model according to the number of multimodal (T1+DTI) features selected	88
5.8	Savings summary using the simple and the complex model for population screening in clinical trials	91
5.9	Distribution of the number of 3D-T1 MRI acquisitions per subject.	93
5.10	Dataset demographics at baseline.	96
5.11	Demographics of the subset of the study cohort for which $\Delta t > 2.5$ used for machine learning classification	97
5.12	Performance of the system using a different number of features evaluated on the interval $2.5 < \Delta t < 3.5$ years	100
A.1	List of statistically relevant brain ROIs along the AD continuum.	141
A.2	List of statistically relevant brain ROIs at different cognitive categories along the AD continuum.	143
A.3	List of statistically relevant brain ROIs at preclinical stages.	145
A.4	Predictive error of response variables for each diagnosis label and feature type using a single linear regression model Alzheimer’s pathology.	145

A.5	Predictive error of response variables for each diagnosis label and feature type using linear regression fitting a specific model for each Alzheimer’s cognitive stage.	145
A.6	Effect size (p-value) of each associated marker of interest on on average brain ROI volume using a single PLS model for Alzheimer’s pathology.	146
A.7	Effect size (p-value) of each associated marker of interest on on average brain ROI cortical thickness using a single PLS model for Alzheimer’s pathology.	147
A.8	Effect size (p-value) of each associated marker of interest on on average brain ROI volume using a multiple PLS model for Alzheimer’s pathology.	148
A.9	Effect size (p-value) of each associated marker of interest on on average brain ROI cortical thickness using a multiple PLS models for Alzheimer’s pathology.	149
B.10	Classification of (a) HC vs MCI and (b) HC vs AD-dementia using T1 features.	150
B.11	Classification of (a) HC vs MCI and (b) HC vs AD-dementia using T1+DTI features.	151
B.12	Features derived from T1 images as normalized GMV from ROIs.	151
B.13	Multimodal top ranking features as weighted by the LR decision function on the training set	153
B.14	Acquisition characteristics of the selected subjects on the interval $3.5 > \Delta t > 2.5$ years.	153
B.15	Percentage of discriminant voxels that correspond to each of the brain regions of interest (ROIs)	158
B.16	Performance of the system using a different feature selection method (l1-norm selection) and evaluated on the interval $3.5 > \Delta t > 2.5$ years	159
B.17	Performance of the system trained on the interval $3.5 > \Delta t > 2.5$ years and evaluated in all other cases.	159

Glossary

- AAL** Automated Anatomical Labeling.
- AD** Alzheimer's disease.
- ADNI** Alzheimer's Disease Neuroimaging Initiative.
- AIBL** Australian Imaging, Biomarker & Lifestyle Study of Ageing.
- APOE** Apolipoprotein E.
- AUC** Area under receiver operating characteristic (ROC) curve.
- CDR** Clinical Dementia Rating.
- CSF** Cerebrospinal fluid.
- CU** Cognitively unimpaired.
- CV** Cross-validation.
- DMN** Default-mode network.
- dMRI** Diffusion Magnetic Resonance Imaging.
- DTI** Diffusion-tensor Imaging.
- fMRI** Functional Magnetic Resonance Imaging.
- GAM** General Additive Model.
- GLM** General Linear Model.
- GM** Grey-matter.
- HC** Healthy controls.
- IWG** International Working Group.
- MCI** Mild Cognitive Impairment.
- ML** Machine Learning.
- MMSE** Mini-Mental State Examination.
- MRI** Magnetic Resonance Imaging.

NeAT Neuroimaging Analysis Toolbox.

NIA-AA National Institute of Aging & Alzheimer's Associations.

NMR Nuclear Magnetic Resonance.

PACC Preclinical Alzheimer's Cognitive Composite..

PET Positron Emission Tomography.

PLS Partial Least Squares.

PR Precision-Recall.

ROC Receiver Operating Characteristic.

ROI Region of Interest.

rs-fMRI Resting-state fMRI.

SBM Surface-based morphometry.

sMRI Structural Magnetic Resonance Imaging.

SVR Support Vector Regression.

TBM Tensor-based morphometry.

VBM Voxel-based morphometry.

WM White-matter.

Introduction

The aim of this thesis is to study and model Alzheimer’s disease (AD) signature in Magnetic Resonance Imaging (MRI), with special attention and focus in early, preclinical stages where subject’s brains present AD-related pathology without expressing dementia related symptoms [1, 2].

1.1 Context and motivation

Over the last decades, research in medical imaging has made significant progress in addressing challenging tasks using model-based approaches (e.g: segmentation or disease prediction). More recently, the hand-crafted and explicitly designed models are being replaced by learning-based approaches that use data-driven solutions and machine learning techniques. The success of deep learning and the improvements in computing power may accelerate this trend. At the same time, the number of observational and interventional studies is increasing, providing large-scale datasets, some of them publicly available (ADNI¹, UK Biobank²), and including both imaging and non-imaging data. All these constitute a great opportunity for methodological development, reproducibility studies or benchmarking. However, the principal challenge of the medical imaging community is the slow uptake of cutting edge methodology by both clinicians and industry. The main reason for this is the lack of robustness of many algorithms due to large variability across subjects and acquisition protocols and/or the low interpretability of the results. Hence, researchers should be focusing more on finding interpretable solutions that solve clinical problems [3].

One of the most widespread applications of medical imaging is the study of neurodegenerative diseases, an irreversible process that result in the progressive death of nerve cells in the brain. Neurodegeneration is the principle cause of dementia, a syndrome associated with a progressive loss of memory, cognition, change in behaviour and the ability to perform everyday activities. A recent estimate from the World Health Organization (WHO) indicates that there are nearly 50 million people worldwide living with dementia, currently being the 5th cause of death [4]. Nonetheless, there are still huge differences between low- to mid-income countries where the prevalence of dementia is much lower and high- income countries where it represents the 3th cause of death. A possible explanation is that neurodegeneration is a condition of the aging brain and hence it is linked with life expectancy that presents a huge gap between rich and poor countries (15-20 years as stated in [5]). Moreover, death causes in low- and mid- income countries are more related to communicable diseases, maternal causes or nutritional deficiencies, while in high-income countries non-communicable diseases are the main cause of death (71%).

¹<http://adni.loni.usc.edu/>

²<https://www.ukbiobank.ac.uk/>

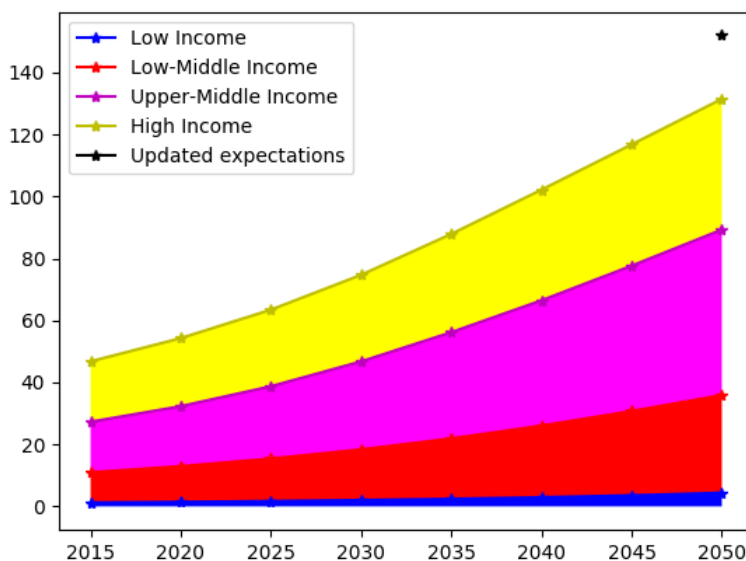


Figure 1.1: Estimated number of people living with dementia worldwide (2015-2050) stratified by countries of different income level. Source: World Alzheimer’s report in 2015 [6] (Table 2.8). In black the last projection for 2050 from [4].

However, as low- and mid- income countries have poor infrastructures and diagnostic systems, figures might be estimated from incomplete data.

In the World Alzheimer’s report in 2015 [6], the future projection estimate almost doubles the number of people living with dementia worldwide every 20 years (Figure 1.1), reaching 131 million people in 2050. A more recent estimate [4] projects 152 million people with dementia in the 2050, the triple of the current estimate (50m, 2018), and so will increase the associated economic and social burden. Last estimates and projections of the total worldwide cost of dementia [6] were US\$818 billion, reaching US\$1 trillion in 2018 and doubling to US\$2 trillion by 2030. If dementia were a country, it would be the 18th biggest economy in 2015, accounting for 1.1% of the global gross domestic product (GDP). Altogether, dementia poses significant global challenges for our societies, national health services, carers and families, making research into its prevention and treatment a major public health priority [7].

Among many other diseases, Alzheimer’s disease (AD) is the most common underlying pathology of dementia patients, accounting from 60% to 70% of subjects that develop dementia symptoms [4]. AD is a neurodegenerative disease characterized by brain extracellular amyloid plaques deposition and intracellular tau neurofibrillary tangles (Figure 1.2) and may eventually lead to dementia in later stages of the disease. The presence of these two pathologies may induce synapse loss and dead neurons (Figure 1.3). The etiology of AD is poorly understood, even though many risk factors are known such as genetic, cerebrovascular diseases or brain injury, among others [8], and has been associated to many factors, e.g: sleep quality [9] or menopause [10].

In current clinical practice, AD dementia is diagnosed as “probable AD” or “possible AD”

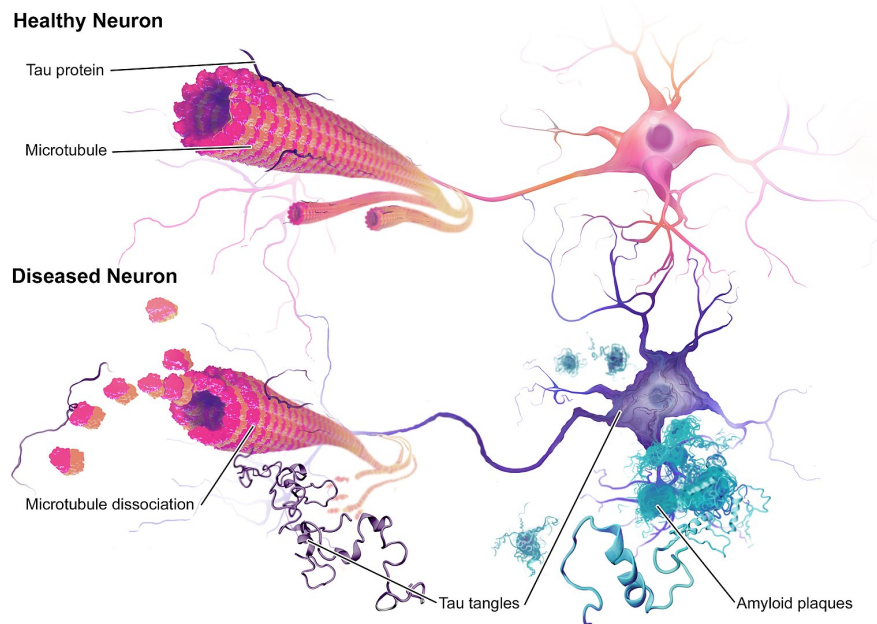


Figure 1.2: An illustrative figure of amyloid plaque deposition and neurofibrillary tangles in a brain neuron (down) compared to a healthy neuron (up). Source: Wikimedia Commons.

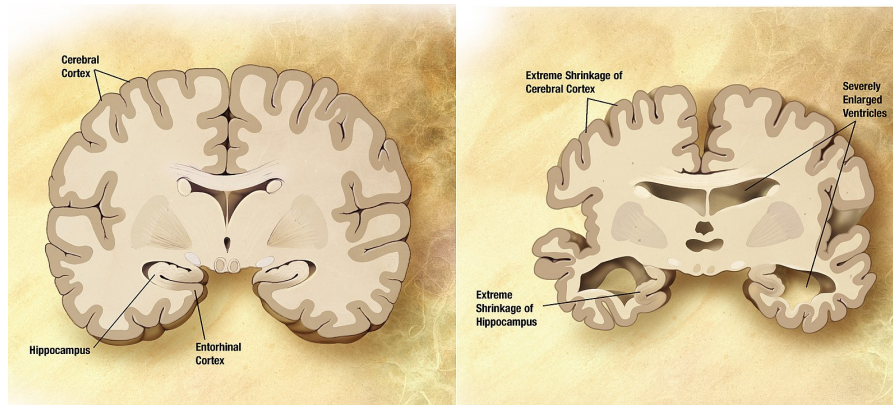


Figure 1.3: Comparison of a normal aged brain (left) and the brain of a person with Alzheimer's (right). Some of the typical characteristics are pointed out in both figures. Source: Wikimedia Commons.

[11] based on medical history, clinical examination and neuropsychological testing. These clinical criteria have been widely used for clinical diagnosis providing modest yet useful to date sensitivity (71-81%) and specificity (70%) [12, 13] against post-mortem verification. Definite AD dementia can only be diagnosed through histopathological examination of brain tissue ex-vivo using autopsy (or, rarely, using in-vivo biopsy) for those patients that meet the criteria for “probable AD” [14]. A prodromal stage where subject present mild memory deficits and cognition impairment but do not reach dementia levels has also been defined as mild cognitive impairment (MCI). However, these diagnostic criteria are not enough for clinical research and do not provide a useful framework for clinical

trials. Increasing number of large and publicly available observational studies and the use of in-vivo biomarkers in research have helped a lot in the understanding of the disease and bring us closer to the development of beneficial therapies. Recent updates of diagnostic frameworks incorporate the use of pathophysiological and topographical biomarkers [15, 16] even though they are mostly intended for research and have been taking off in clinical trials for the past 5-10 years. The use of biomarkers in clinical practice will increase the sensitivity and specificity of the diagnosis.

AD symptoms have been the target of the development of several disease modifying therapies, even though all trials failed, so far. One hypothesis is that we are tackling the disease too late, when subjects have suffered irreversible major neuronal damage and synapse loss. By the time of diagnosis, patients already show a significant volume reduction of some brain anatomical regions compared with normal aging subjects. This suggests that the pathophysiological process of AD begins years prior to the diagnosis of clinical dementia. Following this rationale, therapeutic interventions should take place before the onset of symptoms. Recent developments in fluid and imaging biomarkers show evidence that subtle changes in the brain start to occur 15-30 years before the onset of clinical symptoms and cognitive decline. Cerebrospinal fluid (CSF) and positron emission tomography (PET) are used as in-vivo proxies for AD underlying pathology with increasing accuracy compared to clinical tests, being able to better capture the heterogeneity of AD among subjects. This early stage of the disease is called “preclinical stage of AD” where subjects present Alzheimer’s pathology in their brains without developing dementia symptoms. In this thesis we use the label “preclinical Alzheimer’s pathology” (PC) for cognitively unimpaired subjects that show in-vivo amyloid evidence in the brain either using CSF $A\beta_{1-42}$ or amyloid PET. Moreover, for clinical AD categories (MCI, dementia) we only consider subjects with abnormal levels of brain amyloid yielding to what we label as “MCI due to AD” (MCI) and “dementia due to AD” (AD-dementia). Besides PET, magnetic resonance imaging (MRI) is another relevant imaging technique that provides in-vivo macroscopic images of brain tissue configuration and it is also incorporated in diagnostic frameworks as a valid biomarker of neurodegeneration.

Preclinical AD represents a window of opportunity for disease modifying therapies and secondary prevention practices that hopefully will halt or slow down the progress of the disease. However, they face a common challenge: how to identify subjects at risk. Ongoing clinical trials use either increased genetic risk or positive amyloid biomarkers as valid strategies. Cerebral amyloid-beta deposition is considered a necessary yet not sufficient condition for sporadic Alzheimer’s disease [17] while the genetic is neither necessary nor sufficient condition. Overall, on one hand eligibility poses several issues both operational and ethical, such that overdiagnosis [18], and on the other hand, larger samples are required in order to cover the entire heterogeneity of the disease and get statistically significant results. Hence, we need to develop precise and cost-effective techniques that are able to detect subjects at risk of AD.

1.2 Contributions

Among many other factors, the heterogeneity of the disease and the different acquisition protocols among sites need analytically complex modeling techniques. The increasing amounts of data gathered in both observational studies and clinical trials allow us to develop advanced techniques based on pattern recognition and artificial intelligence. The

study of AD as biological continuum uses and will unequivocally involve brain imaging as a powerful biomarker on one hand and will benefit from the large literature about image analysis from the computer vision society on the other. Hopefully, brain imaging techniques will eventually translate to standard clinical practice despite some of their current drawbacks, such as universal availability or cost.

This thesis will use statistical and pattern recognition methods from the literature to develop new methodological frameworks for the study and prevention of AD using MRI, with special focus on the early stages of AD. This moves us from the definition of AD as a dementia syndrome using clinical outcomes to using in-vivo pathological biomarkers (e.g: cerebrospinal fluid) to account for the whole spectrum of AD biological continuum. We are interested in the analysis of brain structure using MRI and its relationship to Alzheimer’s pathology and Alzheimer’s related markers. Nonetheless, most of the techniques developed in this thesis are generalizable and can be also used for other studies.

In Chapter 2, we provide a common ground to set the basis of this thesis. we begin with an historical overview of Alzheimer’s disease diagnostic frameworks, comparing the terminology adopted by different research groups and stressing their main similarities and differences. From the historical perspective of Alzheimer’s disease as a dementia syndrome there is a recent trend towards defining Alzheimer’s as a purely biological continuum [16] that could help research and also improve the development of prevention therapies. Deriving a common terminology is key not only for the understanding of this thesis but also for the development of the field. After defining AD, we continue by introducing Magnetic Resonance Imaging (MRI), its different modalities and highlighting some of the main findings relating AD and MRI. Finally, we dedicate a section to introduce pattern recognition in the neuroimaging field and review the use of pattern recognition methods in Alzheimer’s disease research.

The following three chapters represent the main contributions of this thesis. We develop MRI based statistical and pattern recognition modeling techniques and use them to study brain morphology along the Alzheimer’s continuum. In Chapter 3, we present an open-source statistical analysis toolbox for neuroimaging studies. We extend the current neuroimaging software capabilities by introducing several non-linear modeling techniques followed by a statistical inference module with several available metrics. We complement the toolbox with model comparison strategies and an interactive graphical user interface (GUI) for visualization purposes. The overall design is compatible with standard neuroimage processing softwares used by the community and can be seen as complementary software for statistical analysis in population studies. We use the toolbox to examine the relationship between MRI and AD related markers reporting two different study cases: (i) atrophy patterns across the Alzheimer’s continuum and (ii) effects of APOE- ϵ 4 in brain aging.

After studying the possible nonlinear patterns of brain morphology related to several factors, in Chapter 4, we are precisely interested in studying the concrete relationship between AD pathology and brain structure and evaluate the capacity of MRI to predict pathophysiological markers. For that purpose, we use regional MRI-derived features (volume and cortical thickness) and cerebrospinal fluid (CSF) biomarkers as a proxy of AD pathology. As a multifactorial disease, we use a multivariate model for AD pathology involving amyloidosis, tauopathy and neurodegeneration and we investigate the use of an univariate and a multivariate models for brain structure. Our initial hypothesis is that

AD pathology may trigger few underlying processes driving brain structural changes. Hence, we model the joint variability of both measurements by means of latent variables that then can be further used for prediction.

The results in the two previous chapters highlight that MRI can constitute an earlier biomarker of AD that become abnormal a few years before the onset of clinical symptoms, suggesting its potential to detect subjects at risk for dementia. In Chapter 5, we further investigate this by studying the predictive capacity of AD pathology in asymptomatic subjects using MRI. Concretely, we use machine learning (ML) algorithms to detect amyloid positive subjects from a pool of cognitively unimpaired subjects using cross-sectional and longitudinal studies using MRI. This methodology could be used for triaging in clinical trials for secondary prevention of AD as gold standard measures (CSF/PET) are unsuited for screening of the general population given their invasiveness and cost. Hence, the main contribution of Chapter 5 is the proposal of a modified protocol for population screening and the development of a theoretical framework to assess its potential in reducing costs and participant burden.

We end up this thesis by summarizing the main results of this PhD thesis as well as highlighting the contributions and suggesting some directions for future work.

Literature review

In this chapter we review the characteristics of Alzheimer’s disease (AD) and different diagnostic criteria used in the last decades. We focus on the preclinical stage of AD (PC) where subjects experiment brain pathological changes before the onset of clinical symptoms of dementia. The literature on AD diagnosis is substantial and we do not attempt to cover all the wide variety of biomarkers ranging from genetics, blood or demographic factors to cerebrospinal fluid (CSF) or brain imaging. Instead, we focus on magnetic resonance imaging (MRI) as a biomarker of neurodegeneration that provides brain structural information regarding its morphology, connectivity and brain functional activity. We introduce the typical MRI modalities used in the study of neurodegenerative diseases and the most common type of analysis. MRI has extensively been used for statistical inference of the effect of several factors (e.g: age, dementia, risk factors) in brain structure. Lately, the appearance of larger observational studies have allowed the development of more complex models within the pattern recognition field to study disease progression and diagnosis. In the last section of this chapter, we provide an extensive yet ordered review of the methodology used to study the relationship between MRI and AD.

2.1 Alzheimer’s disease

Alzheimer’s disease (AD) is classically defined as a dual clinicopathological entity consisting of (i) neuropathological changes that usually involve the presence of intracellular β -amyloid neuritic plaques and extracellular neurofibrillary tangles (NFT) and (ii) clinical phenotypes of dementia such as progressive episodic memory loss and cognitive impairment. This dichotomy is a source of confusion whenever individuals match only one of the definitions, either because different pathological processes share common clinical phenotypes [19, 20] or because pathological changes exist without the concomitant clinical manifestations [21, 22]. This last case constitute an interesting group of individuals for research and specially in preventive studies, as we will see throughout this thesis.

From the initial belief of AD as a presenile dementia defined in Alzheimer’s original case [23], there have been numerous attempts to describe AD and standardize a unified diagnostic criteria. In the following, we provide a brief historical path of AD definition with focus on early stages of the disease.

2.1.1 Diagnostic criteria

Prior to 1970s, AD was seen as a presenile dementia where the symptoms appeared earlier in life. However, different research studies during the 70s and early 80s decade [24, 25, 26] did not show neurological nor pathological differences between presenile and most of the

senile dementia cases. Hence, despite temporal differences it was believed that both constituted a single form of dementia due to neuritic plaques and neurofibrillary tangles. This brought up the definition of AD as a dual clinicopathological entity by the Work Group on the Diagnosis of Alzheimers Disease created by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimers Disease and Related Disorders Association (ADRDA) [14] in 1984. The NINCDS-ADRDA Alzheimer's Criteria adopted the term "probable AD" to clinically diagnose in-vivo dementia between the ages of 40-90 years with deficits on at least two areas of cognition measured based on medical history, clinical examination and neuropsychological tests. The 1984 criteria of "probable AD" explicitly exclude subjects with evidence of other cerebrovascular and neurodegenerative disease that could cause dementia. Even though it was extensively accepted in the research community and clinical practice for the rest of the 20th century and the beginning of the 21th century, "definite AD" could only be accurately diagnosed by histopathological evidence from autopsy or biopsy of those subjects that met the clinical criteria for "probable AD". The overall modest sensitivity (71-81%) and specificity (70%) [12, 13] of the criteria has proven to be useful in clinical practice until nowadays yet not sufficient for drug and/or preventive therapies development.

The flaws of the 1984 criteria rapidly became apparent: "probable AD" diagnosis was not sufficiently sensitive to detect neuritic plaques and neurofibrillary tangles as it excluded (i) a wide and heterogeneous group of people that had very mild cognitive impairment [27] and (ii) nonamnesic syndromes that were related to the same pathologies such as the "visual variant AD" [28]. The term mild cognitive impairment (MCI) first appeared in [29] and accounted for those subjects discarded by the 1984 criteria filling the gap between "probable AD" and cognitively unimpaired subjects. Initially, MCI was defined as a predementia stage with memory impairment whilst preserving other cognitive domains [30] and it was rapidly presumed to be an early (prodromal) stage of probable AD [31]. However, it has been recognized that the concept of MCI is very heterogeneous comprising many subtypes affecting different cognitive domains and with different clinical phenotypes [32]. The most common subtype is amnesic MCI (aMCI) and involves subjects with prominent memory complaints while other subtypes involve either very mild impairment of multiple-domains with or without memory deficits or a single non-memory related domain. Concerned by the heterogeneity of the MCI concept and trying to increase the clinical relevance of MCI diagnosis, the authors in [33] propose to redefine MCI depending on the underlying disorder. In their article, they propose a clinical criteria to diagnose "MCI of Alzheimer's type" or "prodromal AD" based on amnesic syndrom diagnosis involving memory changes and recall deficits. However, subsequent longitudinal studies [34] showed that not all aMCI subjects had neuritic plaques and neurofibrillary tangles (71%). Moreover, people with aMCI might have multiple pathologic changes besides neuritic plaques and neurofibrillary tangles such that α -synucleinopathy or cerebrovascular diseases, among others [35]. To date, the diffuse boundary between MCI and dementia is still object of debate and out of the scope of this thesis.

More than two decades after the first AD diagnostic criteria, in 2005, the International Working Group (IGW) was created in order to explore a new diagnostic framework for AD that would extend the previous one by including the use of recently developed biomarkers and the definition of the "prodromal AD stage". They stated several factors that highlighted the need for an updated criteria: insufficient diagnostic specificity, improved AD phenotypes, improved non-AD phenotypes, histopathological definition of non-AD

dementias, need for early diagnosis, poor definition of MCI, unclear distinction between MCI and dementia and the appearance of new biomarkers (in-vivo evidence) that helped the study of the disease identifying molecular and structural changes in the brain [36]. Concretely, β -amyloid positron emission tomography (PET) imaging revolutionized the field by introducing a neuropathologically validated biomarker for neuritic plaques and hence, a proxy to the pathobiological path of AD. Other useful biomarkers are cerebrospinal fluid (CSF) and magnetic resonance imaging (MRI). The IGW proposed the first research criteria that used both clinical phenotypes and in-vivo biological evidence for AD diagnostic purposes [36]. They accounted for the typical representation of AD and defined “probable AD” as having early and significant memory impairment supported by at least one biomarker evidence among the following: (i) medial temporal lobe (MTL) atrophy using structural MRI, (ii) abnormal CSF biomarkers including amyloid β_{1-42} ($A\beta_{42}$), total tau (t-tau) and phospho-tau (p-tau) and (iii) abnormal PET biomarkers, including reduced glucose metabolism measured using fludeoxyglucose (FDG) or increased Pittsburgh compound B (PiB). We refer to positive/negative Alzheimer's biomarkers whenever their values lie within an abnormal/normal range compared to normal aging values.

With the inclusion of supportive features in the core diagnostic framework, the IWG moved AD from a clinicopathological to a clinicobiological entity and offered unified criteria applicable at all clinical states along the disease continuum, from “prodromal AD” to AD. However, it only focused on the typical amnesic representation of AD and it still missed situations where the clinicobiological duality is not present, such that when clinically asymptomatic subjects have positive Alzheimer's biomarkers or when clinically symptomatic subjects have partially positive or negative Alzheimer's biomarkers. The same IWG introduced a new lexicon in 2010 [37] aiming to provide a broader diagnostic coverage of the full AD clinical spectrum. In their definitions they used the most validated Alzheimer's biomarkers at that time: pathological markers (CSF $A\beta_{42}$, CSF p-tau, CSF t-tau and Amyloid PET) and topographical markers (FDG PET and MTL atrophy using MRI). They began by defining Alzheimer's disease as a clinical continuum that starts with the onset of clinical symptoms and encompassed both prodementia (labeled as “prodromal AD”) and dementia (labeled as “AD dementia”) stages. According to the clinical phenotype and supported by one or more biomarker evidence, AD can be labeled as (i) “typical AD”, referred to the typical pattern of early and progressive memory deficit that stays until later stages of the disease, (ii) “atypical AD”, referred to a less common pattern of deterioration of non-memory cognitive domains and (iii) “mixed-AD”, referred to patients that follow a typical AD pattern mixed with evidence of other clinical phenotypes or biological evidence of other comorbidities. Irrespective of the existence of clinical manifestation, they define “Alzheimer's pathology” as the underlying neurobiological changes responsible for AD that span the earliest pathogenic events comprising brain neuronal lesions (neuritic plaques and neurofibrillary tangles), synaptic loss and vascular amyloid depositions. Earlier in the Alzheimer's pathology continuum, they define the “preclinical stage of AD” as the period that spans from the earliest pathological events to the first appearance of cognitive changes. Two states were defined for preclinical AD: “asymptomatic at-risk state for AD” and “presymptomatic AD”. The former comprise subjects with in-vivo evidence of Alzheimer's pathology measured either using PET imaging with amyloid tracers or CSF concentrations, while the later comprise individuals who carry an autosomal dominant AD mutation that will inevitably develop clinical manifestations of AD. Finally, the “MCI” label was used to describe individuals with measurable MCI in the absence of significant effect on instrumental living activities

without a disease to which MCI can be attributed.

This new lexicon was then used for the IGW-2 criterion [15], an update to the previous IGW diagnostic criteria crystallising all refinements introduced in 2010. A broader clinical core diagnostic was defined by accounting for both amnesic and non-amnesic variants of AD and considered different weighting of Alzheimer’s pathology biomarkers along the disease continuum. Pathophysiological biomarkers (amyloid imaging or molecular CSF), as in-vivo indicators of underlying amyloidosis or tauopathy, were used as markers of Alzheimer’s pathology. They were included in the core diagnostic criteria providing high specificity for AD and allowing to better define the boundary between prodromal AD and undefined MCI. Moreover, the term “probable AD” was then discarded since they were able to provide in-vivo evidence of Alzheimer’s pathology. On the other hand, topographical biomarkers (FDG PET, structural MRI) are used for assessment of disease stage and progression rather than diagnostic markers. The IGW-2 criterion was initially designed for rigorous research purposes but intended to be applied in clinical practice at least in favourable cases (young-onset or atypical AD) in which biomarkers might help boost the diagnostic accuracy. However, the use of biomarkers poses other challenges that should be further addressed such that the disagreement between biomarkers or their use at older ages where the evidence is less salient and other comorbidities might be present.

In parallel to the IWG work, in 2009 the National Institute of Aging (NIA) and the Alzheimer’s Association (AA) sponsored a series of round tables that concluded with the creation of three separate working groups responsible for revising and updating the established 1984 criteria: one work group was assigned the task to formulate a diagnostic criteria for the dementia phase of AD, a second group focused on a diagnostic criteria for symptomatic pre-dementia stage of AD and a third group was asked to develop a research criteria for preclinical AD. Hence, a different diagnostic criteria was developed for each disease stage. The work from NIA-AA groups attempted to disambiguate the term AD by using AD-P when referred to Alzheimer’s pathology supported by in-vivo biomarker evidence and AD-C when referred to Alzheimer’s clinical dementia. The results were summarized in four different papers [38, 11, 39, 1]. An introduction of the NIA-AA criteria is found in [38], motivating the work done in the three separate groups and highlighting the need for an update of the criteria that incorporated biomarkers yet could still be used in clinical practice. The works in [11, 39] attempted to define a clinical criteria for AD dementia and MCI, respectively, with different levels of probabilistic likelihood. They used the “probable AD” definition from 1984 as the core clinical criteria for AD dementia with increasing level of certainty if cognitive decline in subsequent studies, genetic mutations and/or in-vivo pathophysiological evidence were present. For MCI subjects, they defined a general core clinical criteria as baseline and provided higher levels of certainty about the underlying pathology when amyloid and/or neuronal injury biomarkers were available. Finally, in [1], a conceptual framework was introduced using the pathophysiological model proposed in [40] to define the Alzheimer’s continuum. This work was intended only for research purposes providing the basis to define and characterize the preclinical stages of AD.

The works from NIA-AA and IGW share many features, such as the integration of biomarkers, the identification of a preclinical stage and the recognition of biomarker diversity along the disease based on [40]. However, both works conceptually differ on several points. The NIA-AA propose a stagewise diagnostic criteria while the IWG pro-

pose a simpler unified criteria for AD. The NIA-AA might be more adequate for clinical practice since it still works when biomarkers are not available, while the IWG might provide more accurate diagnosis when biomarkers are available. In the preclinical stage, subjects with positive amyloid biomarkers would be diagnosed with AD following the NIA-AA criteria while only considered as asymptomatic at-risk of AD using the IWG criteria.

Jack et al. in [41] went one step further in proposing a biomarker-based protocol independent of any clinical criteria that can be used in both cognitive-aging and dementia studies. They selected seven key Alzheimer's biomarkers used to define three binary (+/-) categories (A/T/N): "A" category refers to the β -amyloid biomarker (CSF $A\beta_{42}$ or amyloid PET), "T" category refers to the value of tau biomarker (CSF phospho tau or tau PET) and "N" category refers to biomarkers of neurodegeneration or neuronal injury (structural MRI, CSF total tau or [18F]-fluorodeoxyglucose PET). In total, they defined eight different biomaker profiles without assuming any causality among them and hence, accounting for different hypothesis about AD pathogenesis. Currently, the most used hypothesis states that proteinopathy "causes" the disease in the elderly [42], while other investigate other less $A\beta$ centric pathways [43].

The A/T/N system set the basis for the new research framework proposed in 2018 by the NIA-AA group as an attempt to update and unify the 2011 guidelines. This new research framework, [16], is intended for observational studies and interventional research but not for clinical routine care. It is a purely biological (biomarker-based) framework that uses the AT(N) system for profiling. While A and T categories use specific pathological biomarkers for AD, (N) uses biomarkers indicative of neurodegeneration that might not be specific for AD and therefore placed in parentheses. The 2018 framework defines the *Alzheimer's continuum* using all biomarker profiles with positive amyloid (A+) biomarkers, differentiating between *Alzheimer's disease* (AD) individuals that also present positive tau (T+) biomarkers, and individuals with *Alzheimer's pathologic changes* that present normal tau (T-) biomarkers and that are considered to be in early stages along the Alzheimer's continuum. Individuals with a (A-T-(N)-) profile are defined as to have normal Alzheimer's biomarkers while having amyloid negative (A-) and either tau positive (T+) or neurodegeneration ((N)+) is indicative of non-AD pathologic changes. Herein, genetics are excluded as gene variants do not measure neuropathological change but rather indicate an individual's risk to develop pathological changes (e.g: *APOE* ϵ 4 or the certainty provided by typical autosomal dominant mutations). The research framework is flexible to incorporate new biomarkers within the existing categories (e.g: CSF neurogranin [44, 45] within the (N) category) or create new categories, i.e. ATX(N), when new groups can clearly be defined. TDP43 [46], α -synuclein [47] and vascular biomarkers [48, 49] are some promising AD biomarkers currently under study. One of the limitations of this framework is the sensitivity of *in-vivo* biomarkers to detect the presence of neuritic plaques, pathologic tau tangles or the loss of neurons.

The NIA-AA research framework also use clinical (C) markers to define a cognitive continuum and its relation to the AT(N) system. They propose two different cognitive staging systems: a *syndrome cognitive staging* defined over the whole population (all biomarker profiles) that can be used in observational studies and a *numeric clinical staging* defined only over those subjects along the Alzheimer's continuum and mostly intended for interventional studies. The syndrome cognitive staging system also accounts for compatibility with NIA-AA 2011 guidelines where they defined three different cognitive entities

(dementia, MCI and cognitively unimpaired (CU)), and that is widely used in several ongoing studies such as ADNI [50] or AIBL [51]. The numerical clinical staging system was created for those subjects within the Alzheimer’s continuum as a sequential evolution from an initial stage with the onset of abnormal AD biomarkers in asymptomatic subjects. This numerical system splits the cognitive continuum into 6 stages, very similar to the recent FDA recommendations [52].

While IGW and NIA-AA groups are going towards the same direction when defining AD and AD biomarkers, still some differences exist between their last published guidelines. The IGW work in 2014 [15] is not purely a biological entity, as the diagnosis of AD requires both cognitive symptoms and AD biomarker signature. In contrast, NIA-AA in 2018 [16] proposes a biomarker-only based system for AD diagnosis. The former requires either positive amyloid PET or positive CSF $A\beta$ and t-tau to diagnose AD, while the latter considers CSF total tau as a non-AD specific marker of neuronal injury and, apart from amyloid markers, requires abnormal tau PET or CSF p-tau values to define AD. Nonetheless, tau PET is a very recent and promising AD biomarker and did not exist prior to the IGW definition in 2014. Another difference is found on the lexicon for early stages of the Alzheimer’s continuum: while IGW considers subjects with only amyloid positive biomarkers to be “asymptomatic-at-risk for AD”, NIA-AA considers them part of the “Alzheimer’s continuum” presenting “Alzheimer’s pathological changes”. Other nomenclature differences are present in the preclinical stages of AD between NIA-AA and a more recent article about preclinical AD from IGW group in 2016 [2]

An overview of the diagnostic criteria from established research groups can be found in Table 2.1

Diagnostic criteria of Alzheimer’s disease				
Year [Ref.]	Research Group	Diagnostic labels	Biomarkers	Intended use
1984 [14]	NINCDS-ADRDA	Probable and possible AD	No	Clinical & Research practice
2011 [11]	NIA-AA	Dementia due to AD	Not required	Clinical & Research practice
2011 [39]	NIA-AA	MCI due to AD	Not required	Clinical & Research practice
2011 [1]	NIA-AA	Preclinical AD	Yes	Research practice
2014 [15]	IWG	Preclinical AD and typical/atypical/mixed AD	Yes	Clinical & Research practice
2016 [2]	IWG	Preclinical AD	Yes	Research practice
2018 [16]	NIA-AA	A/T/(N)	Yes	Research practice

Table 2.1: Historical overview of diagnostic criterias developed by two main working groups (NIA-AA and IWG) used in Alzheimer’s clinical and research practice.

2.1.2 Preclinical AD

The preclinical stage of AD is thought to be the temporal span that starts with the first neuropathological changes and extends until the onset of clinical and cognitive symptoms.

However, in practice, setting both upper and lower boundaries is challenging. Moreover, the mismatch between different definitions of AD may induce confusion in the nomenclature. On the one hand, the NIA-AA group propose a purely biological framework that exclude clinical outcomes in the categorization losing the concept of preclinical/clinical AD. In an attempt to integrate past diagnostic categories [38], they propose a categorical cognitive staging that relates traditional syndromal categories with their proposed framework. Here, preclinical AD is defined over CU subjects that present both Alzheimer’s pathologies (amyloid plaques and neurofibrillary tangles), while CU subjects with only amyloid plaques are described as “preclinical Alzheimer’s pathologic change”. On the other hand, the IGW group in their last published work [15] suggest to differentiate two different preclinical states of AD: (i) an asymptomatic AD state, that groups people without clinical evidence of prodromal AD that carry an autosomal dominant AD mutation and (ii) an asymptomatic at-risk for AD state, that groups subjects without clinical evidence of prodromal AD and that have in-vivo evidence of Alzheimer’s pathology showing either decreased CSF $A\beta_{1-42}$ and increased CSF p-tau/t-tau or increased uptake of fibrillar amyloid PET. In a more recent update that can be attributed to the IGW group [2], they propose to restrict the term “preclinical AD” to subjects with high risk to progress to clinical AD (i.e. showing both pathological brain lesions related to AD) while preserving the “asymptomatic at risk for AD” to those subjects with only one biomarker evidence (either amyloidosis or tauopathy). This last claim differs from the NIA-AA research works based on the work in [40] that suggests amyloid positive marker triggers the start of Alzheimer’s pathology and hence exclude subjects showing only tauopathy evidence from the Alzheimer’s continuum. The NIA-AA framework is based on the hypothetical biomarker model of Jack et al. in [40]. The model suggests that an abnormal level of β -amyloid in the brain is the initiating event in the Alzheimer’s continuum, followed by tau concentrations, brain structural alterations, memory impairment and clinical symptoms.

Both works agree on the pathophysiological biomarkers for AD: CSF $A\beta_{1-42}$ or amyloid PET as a proxy for amyloidosis while CSF p-tau and tau PET are used for in-vivo evidence of tauopathy. The IGW group also includes CSF total tau as a biomarker of tauopathy even though they acknowledge that both p-tau and t-tau are less specific for AD than tau PET. Biomarkers used to asses neurodegeneration are MRI and FDG-PET biomarkers to asses neuronal injury (N) and the NIA-AA group include CSF t-tau in this category.

The preclinical stage of AD is also a window of opportunity for disease modifying therapies [53]. The continuous failure of clinical trials for drug development targeting mild-to-moderate AD dementia subjects combined with the increasing evidence of a long asymptomatic stage of AD supported by new biomarker developments have produced a shift towards AD prevention initiatives [54]. The challenge it poses is how to identify subjects at risk to develop AD. A current solution is to detect subjects with β -amyloid in the brain as it is thought to start accumulating 15-30 years before the onset of clinical symptoms. The prevalence of amyloid pathology in the brain is associated with age, APOE- ϵ 4 status and the presence of cognitive impairment [55]. In late middle-age adults, i.e. 65 years, the expected ratio of amyloid positivity among cognitively unimpaired subjects is 20%, increasing to around 65% for subjects carrying two copies of the APOE- ϵ 4 allele.

A recent study from the INSIGHT-preAD group [56] explores the relation between brain β -amyloid depositions and cognitive and neuroimaging features. They recruited 318

participants at baseline aged between 70-85 years and that reported subjective memory complaints but had normal memory and neuropsychological tests. The mean age at baseline was 76.0 (SD=3.5) years, amyloid PET scans revealed 88 (28%) amyloid positive subjects and 62 (20%) subjects were APOE- ϵ 4 carriers of whom 33 (38%) were amyloid positive. No difference was found in cognition, structural metabolism or connectivity at baseline between amyloid positive and negative participants and only differences in CSF biomarkers ($A\beta$, p-tau, t-tau) were found. A total of 274 subjects remained in the study after 30 months of follow up time with cognitive and clinical testing every 6 months and brain imaging (FDG PET, structural MRI, functional MRI and amyloid PET) every 24 months. After 30 months no differences in cognition were found between amyloid status groups and only four subjects (2%) progressed to prodromal AD, all of whom were amyloid positive at baseline. Hence, the authors conclude that participants in clinical trials of preclinical AD should be followed up for longer times. These findings are in line with the longitudinal study in [57] where they observed individuals for a median time of 3.1 years with maximum follow up times of 10.3 years. They show different trajectories for clinical and cognitive markers - Preclinical Alzheimer's Cognitive Composite (PACC), Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR) - between positive and negative amyloid groups. However, these differences began apparent at follow up times $t > 4$ years, being PACC the most sensitive marker showing differences almost 1 year prior to CDR. Both studies had an increased likelihood of amyloid positivity within APOE- ϵ 4 carriers and the expected prevalence of amyloid pathology in cognitively unimpaired subjects increases with the APOE- ϵ 4 status [55, 58]. The relation between β -amyloid depositions, clinical progression and APOE- ϵ 4 status suggest that the latter could be used as a first-level screening marker prior to brain $A\beta$ assessment. Other risk factors for sporadic AD are currently investigated and summarized in the report from the Lancet Commission [59], accounting for little education in early life, mid-life hypertension, obesity or hearing loss as well as social and physical activity, smoking habits, diabetes or depression in late life.

Even though brain amyloidosis is the strongest risk factor for sporadic AD, $A\beta$ pathology increases steeply with advancing age. Consequently, it results in comorbid $A\beta$ positivity in non-AD neurodegenerative disorders or increasing prevalence in elderly cognitively unimpaired population reducing the specificity of $A\beta$ biomarkers. The [18F]florotau PET tracer (tau PET) arise as a potentially more specific biomarker as its prevalence is less associated with age and is well suited for differential diagnosis [60]. Tau PET is more linked to anatomical and clinical variation due to Alzheimer's disease than amyloid PET [61]. Hence, the penetrance of tau PET in both observational and interventional studies is increasing rapidly. A couple of studies analyze the prevalence of both amyloid and tau biomarkers by using the AT(N) framework [58, 62]. The prevalence of each group changes substantially with age following non-monotonically increasing/decreasing pattern as it could have been expected.

2.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is an imaging technique used to capture brain anatomy in-vivo [63]. It is based on the Nuclear Magnetic Resonance (NMR) effect that non-zero spin particles produce in the brain when excited by a strong constant magnetic field (B_0) and an oscillating, weak and orthogonal magnetic field (B_1). Different tissue configurations will appear to have different signal intensities in the resulting image. More

diffuse mediums, such as intracranial cerebrospinal fluid (CSF) will produce different signal than white matter (WM) or grey matter (GM) tissues, which are less diffuse and contain higher proportion of other particles (e.g: fat). Brain lesions (e.g: tumors, white matter hyperintensities or traumatic injuries among others) would also appear differently than their surrounding. MRI can produce several image modalities adapting the acquisition parameters to produce different contrasts between brain tissues and thus providing complementary information. The ability to use multiple contrasts in the same session is a powerful technique interesting for many brain imaging applications. Moreover, one could use contrast agents that alter the configuration of brain tissue and produce more contrast between different mediums (e.g: the use of gadolinium [64] to better detect brain tumors). A clinical overview of MRI can be found in [65].

MRI is used as a non-invasive biomarker of neurodegeneration not specific to AD. However, there exists a vast literature relating MRI and AD looking for different temporal and spatial changes related to AD that could define the so-called *AD signature* and, more recently, the *preclinical AD signature*. Atrophy in Alzheimer’s disease patients is thought to appear years after the beginning of amyloid plaques deposition and neurofibrillary tangles creation and right before the appearance of the first symptoms [40]. Different spatial patterns of atrophy correlate with neuropathological subtypes of AD [66] that display different clinical phenotypes [67]. However, there exists significant neuroanatomical heterogeneity among subjects in the Alzheimer’s continuum [68] due to variability in underlying genetics, brain pathologies, environmental factors or cognitive reserve and may lead to large variations in clinical phenotypes. Cognitive reserve is a still poorly understood concept used to explain cases where patients with advanced brain damage do not develop clinical symptoms or cognitive decline [69].

Different MRI imaging techniques provide information about magnetic properties of brain tissue (structural MRI), diffusivity of molecules in brain tissues (diffusion MRI) or brain activity (functional MRI).

2.2.1 Structural MRI

Structural MRI provides static anatomical information about the brain. It describes qualitatively and quantitatively the shape, size and texture of gray and white matter structures in the brain. It can be used either in clinical practice for radiological reporting or for detailed analysis.

Different acquisition parameter configurations can be used to provide complementary information as seen in Figure 2.1

- **T1-weighted:** it emphasizes the contrast between GM and WM. CSF appears dark, WM appears bright and GM appears gray.
- **T2-weighted:** it emphasizes the contrast between brain tissue (GM and WM) and CSF. GM and WM appear light and dark gray while CSF appears bright.
- **T2*-weighted:** it is a modified T2 sequence to correct for magnetic field inhomogeneities. It is used to better detect hemorrhage or calcifications and it forms the basis of functional MRI.

- **FLAIR:** it emphasizes the contrast between brain tissue and abnormalities. It is very similar to T2-weighted contrast but CSF appears darker while abnormalities (e.g: white matter hyperintensities) are kept bright.

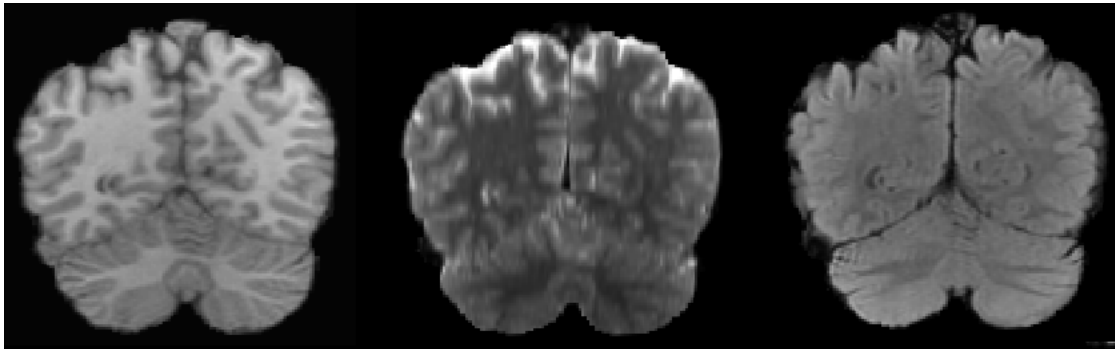


Figure 2.1: Different MRI contrasts: T1 (left), T2 (middle) and FLAIR (right).

The T1-weighted sequence is the most widely used due to the contrast between brain tissue (GM and WM) and it is used to assess brain anatomy, specially the cerebral cortex. The T2-weighted, T2*-weighted and FLAIR images are most widely used for abnormality detection, (e.g: brain tumors [70], multiple sclerosis [71] or white matter hyperintensities [72]).

Population studies relate structural MRI to any factors describing, for example, biological processes, behavioral assessments or clinical phenotypes. Quantitative analysis pipelines require extensive preprocessing of MRI images encompassing different processes such as denoising [73], bias field correction [74, 75] or tissue segmentation [76], among others. Morphometric studies measure the volume and shape of grey matter structures and the surface of the cerebral neocortex. Different types of morphometric techniques are used in medical image analysis: voxel-based (VBM), tensor-based (TBM) and surface-based (SBM). All techniques require a registration step to a template (spatial normalization step) in order to be able to compare different subjects. VBM measures voxel-by-voxel differences in brain tissue concentrations [77] and it is the most widely used technique due to its simplicity compared to the other methods. To increase statistical robustness and account for potential registration errors, a smoothing step is usually used to average neighbouring voxels intensities. TBM study brain shape directly by means of deformation fields from the registration step [78]. SBM focuses on the analysis of the brain neocortex unfolding the gyrus and sulcus as a 2D surface [79]. Other type of analyses are based on region of interest (ROIs) and are usually based on predefined anatomical atlases of the brain [80, 81, 82]. Robust segmentation algorithms that may involve a registration step are used to process MR images prior to the analysis. Segmentation and registration are two of the main processing steps for structural MRI studies. Several publicly available software [83, 84, 85, 86] help with the preprocessing and processing steps.

2.2.2 Diffusion MRI

Diffusion weighted imaging (DWI) provides information about the medium diffusivity. It uses the rate of diffusion of molecules, mainly water, to generate a contrast in MRI. It is specially useful in areas such as cerebral infarction or tumor detection. In cognitive aging and dementia-related neuroimaging studies, diffusion tensor imaging (DTI) appears

to be relevant because it builds a 3D representation of the likelihood of the direction of diffusion at every single voxel. Due to the underlying structure of white matter tissue where the architecture of axons in myelinated parallel bundles facilitate the diffusion to a certain direction, DTI can help to reconstruct in-vivo the main white matter pathways as well as to assess the integrity of white matter bundles.

Related to Alzheimer's disease, DTI studies show increased mean diffusivity (MD) in white matter and decreased fractional anisotropy (FA) in people with dementia, indicating loss of directionality in fiber bundles [87]. Moreover, the use of tractography algorithms indicate degeneration of cortical [88] and subcortical [89] networks. Moreover, it shows early disconnections in the temporal lobe at preclinical stages, followed by disconnections in the parietal and frontal lobes at the MCI stage [90].

2.2.3 Functional MRI

Functional magnetic resonance imaging (fMRI) is a technique that measures brain activity by targeting changes in the associated blood flow. This technique relies on the assumption that cerebral blood flow and neuronal activation are coupled. It is used to map task-related functional areas, such as to identify regions related to critical functions such as speaking, planning or reasoning. Surgery planning can use fMRI to avoid altering critical brain functions when removing brain tissue due to tumor or lesion. Moreover, it can be used to assess therapy development from patients with depression or that suffered a stroke.

Resting state fMRI (rs-fMRI) is a fMRI method that maps brain activity at rest, i.e. when an explicit task is not being performed. A number of conditions are identified in the brain, such as the default mode network (DMN) that comprise highly correlated regions that experience an increase of brain activity when at rest. The resting-state approach is useful to examine neurological conditions or psychological disorders.

In Alzheimer's disease, rs-fMRI is used to compute functional connectivity revealing network dynamic changes and the association between activity in distant regions regardless of the underlying structural connectivity [91]. Most of the studies reported decreased activity in some regions (e.g: medial temporal lobe) and increased activity in others (e.g: frontal lobe) within the DMN [92]. This idea of increased activity could be interpreted as compensatory mechanisms from the loss of activity within the DMN. In the same study, they show that patients in advanced dementia stages show decreased activity in the whole DMN bringing up the hypothesis of hyper-metabolism preceding hypo-metabolism. Moreover, other studies show interesting results beyond the DMN with functional changes in other brain networks such as early changes in medial parietal brain activity [93]. Despite interesting results, fMRI studies suffer from several limitations such as the acquisition noise or the low statistical power [94].

2.3 Pattern recognition in neuroimaging

Traditionally, in epidemiological studies, statistics have played a central role by looking at group-level (usually case-control) differences related to certain outcome measure (e.g: the exposure to certain condition or risk factor). In the neuroimaging field, statistical inference has significantly increased the understanding of brain structure, function or

metabolism and their link to brain disorders. Concretely, mass-univariate analysis using the general linear model (GLM) has been the dominant paradigm in neuroimaging studies for several decades [95] to make separate group-level inferences for each voxel and/or region in the brain. However, statistically significant group-differences have limited translation in clinical practice and other analysis techniques have been developed in the field. With the advances in neuroimaging techniques and computer hardware as well as the increasing amounts of data available, pattern recognition methods emerged as complimentary approaches to classical inference.

The problem of searching regularities in data has a long history and has been useful in order to find empirical models of the world. As Bishop wrote in [96], “pattern recognition is the field concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories”. The field of machine learning (ML) is closely related to pattern recognition and we will use both terms interchangeably. Bishop [96] continues stating that the main difference between them is that “pattern recognition has its origins in engineering, whereas machine learning grew out of computer science”. In contrast, despite its apparent similarity, machine learning and statistics are two different fields. Statistics is the branch of mathematics that deals with data and hence, machine learning is built upon a statistical framework. However, machine learning draws upon a large number of other items from mathematics (e.g: optimization, algebra and calculus) or computer science and engineering (e.g: computational frameworks and implementation efficiency). The main difference between classical statistical models and pattern recognition lies in their purpose (inference vs. generalization) even though they may use the same mathematical underlying model. The goal of pattern recognition and machine learning models is to make predictions about new (unseen) data as accurately as possible and trying to understand the behaviour of the learning algorithms [97]. On the other hand, statistical models aim at making inferences about data creating a mathematical model that characterizes the relationship between data and outcome variables. An overview of the pattern recognition field from a statistical perspective can be found in [98].

Machine learning has proven successful in a wide range of fields such as computer vision, speech processing, text analysis and also neuroimaging. Image-based diagnosis and disease prognosis are amongst the most successful applications of machine learning in neuroimaging. The authors in [99] found more than 500 papers between 1990-2015 tackling the problem of single subject prediction of brain disorders such that Alzheimer’s disease (AD), schizophrenia, depressive disorders or attention-deficit hyperactivity disorder. The TADPOLE challenge [100] aims at forecasting future evolution of AD-related biomarkers. More recently, deep learning frameworks have been applied in a wide range of neuroimaging problems, with rather notable success in registration and segmentation tasks that usually required computationally intensive optimization techniques. However, machine learning poses many challenges in the field such as the costly process of manual labeling or interpretation and evaluation of the methods [101]. From the medical imaging community perspective, the main challenge remains the slow uptake of methodological advances by both clinicians and industry. The lack of robustness of many algorithms due to large variability across subjects and acquisition protocols and/or the low interpretability of the results are amongst the reasons that hinder methodological translation between scientific development and the clinic. Accounting for that, researchers should

be focusing more on finding interpretable solutions that solve clinical problems [3].

Broadly speaking, a machine learning algorithm has, at least, two separate phases: the *training* and the *testing* phases. The former is used to tune model parameters and adapt them to the problem in hand while the latter uses the trained model to predict new samples and assess the generalization of the model to unseen examples during the training phase. A good and comprehensive coverage of pattern recognition and machine learning methods can be found in [96]. It covers both supervised learning where for each observation there is a tuple available consisting of input data and its corresponding target measures and unsupervised learning techniques where there are no target values for input data. Supervised learning techniques learn associations between input data (e.g: hippocampal volume) and target measures in the training phase and predict target values for the test examples. Depending on the nature of target values, we can distinguish between *classification* if the target values are discrete (e.g: disease status) and *regression* if the target is continuous (e.g: age). The goal of unsupervised learning techniques is to discover associations between input data points in order to group similar examples (e.g: clustering), estimate the distribution of the data (e.g: density estimation) or to discover underlying low-dimensional relationships (ex: dimensionality reduction).

Typical machine learning pipelines consist of several parts: feature extraction, feature-selection, ML algorithm and error estimation Figure 2.2. However, recent success of deep learning methods (specially, convolutional neural networks) embed all modules into one single module. A good implementation of machine learning methods for neuroimaging can be found in [102].

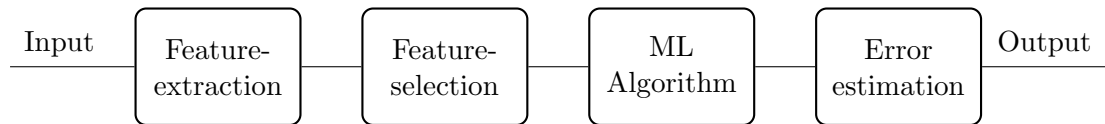


Figure 2.2: Typical machine learning pipeline.

Feature-extraction Feature-extraction uses the raw data from initial measurements, possibly unstructured, untractable, noisy and high-dimensional, and derive features intended to be informative, non-redundant and structured to facilitate the subsequent analysis. Feature extraction methods can also be seen as a way to reduce the dimensionality that could lead to better interpretation and/or visualization. The physical process and the high-dimensional nature of medical imaging data makes this step crucial.

The main aim of feature-extraction is to retrieve and quantify a set of accurate and proper information from neuroimaging data, such as size, shape or volume, among others. These features can vary from single-voxels, ROI-level or whole-brain features. Here, we include voxel-based morphometry (VBM, [103]), tensor-based morphometry (TBM, [78]) and surface-based morphometry (SBM, [104]), as we saw in section 2.2.1. Getting the most relevant imaging phenotypes is crucial for different types of problems [105]. However, voxel-wise and vertice-wise analysis is limited by noise, registration errors and large

inter-individual variability. To account for this and to increase the robustness of the extracted features against noise, some methods use a fixed anatomical brain structure built upon several regions of interest (ROI), and extract features at the ROI level after performing registration to a template (e.g. AAL¹ template [106]). In [107], 93 ROIs are defined and grey matter (GM) volume averages are computed for each ROI. In [108], 7 features are extracted for each of the 90 ROIs: tissue percentile (GM, WM, CSF and background) and the averaged voxel-wise Jacobian determinant in WM, GM and CSF, used for registration. A combination of multiple types of features can be found in [109], where they used cortical thickness and volumetry of a-priori selected ROIs as well as hippocampal texture and shape for automatic clinical diagnosis of the AD syndrome.

A common problem using predefined ROIs is that neurodegeneration patterns may not necessarily follow standard definitions of anatomical regions. Hence, predefined regions with regular shapes may reduce the power of the features. In [110], a data-driven approach that adaptively define regions of discriminative features with spatial consistency is proposed. Moreover, with the increasing quality and diversity of images, higher-resolution atlases can be used in order to track regional subfields [111].

Feature-selection After the extraction, it is possible to obtain many features from neuroimaging data. However, curse of dimensionality might be a problem when dealing with high-dimensional medical images with only few samples available (normally ranging from a few dozens to a few hundreds) resulting in poor generalization results [112]. Thus, feature selection becomes an important stage prior to the machine learning algorithm of choice. Feature selection methods aim to select the most relevant features for a given task omitting the improper ones in order to reduce the number of input variables to the classifier. The main approaches are filter-based, wrapper-based and embedded techniques [113]. Filter-based approaches rank features according to a given metric and select the k -most relevant ones. In [114] a paired t-test to select the most discriminative feature subset is used, while the work in [115] is based on the mRMR (minimum redundancy maximum relevance) feature selection algorithm that identifies the minimal set of features that jointly maximizes the discriminative power of the subset. Wrapper-based approaches use a predefined machine learning method, such as a classifier with a certain regularization (L_1 , L_2 , elastic net, ...) and rank and select features according to their relevance (i.e: weights) in the final classification model [116]. Finally, the embedded methods perform feature selection as a part of the classifier (e.g. random forests) [117].

Another way to tackle the curse of dimensionality problem is by using manifold learning, a set of machine learning techniques that aim at finding a low-dimensional representation of the high-dimensional data while preserving its intrinsic geometry. Manifold learning examples are principal component analysis (PCA) [118], partial least squares (PLS) [119], Laplacian eigenmaps [120] or isomap [121].

Machine learning algorithms A wide variety of machine learning algorithms could be used depending on the problem in hand. A complete survey can be found in [96]. To name a few, if we are dealing with a classification problem we could use logistic regression (LR, [122]), support vector machines (SVM, [96]), random forests (RF, [123]) or neural networks (NN, [98]), among others. Moreover, an ensemble of classifiers may boost the overall performance [124]. For regression problems we have the equivalent algorithms

¹Automatic Anatomical Labeling (AAL)

but suited for continuous target values. Regularization techniques are normally included in order to ensure smooth solutions and to prevent over-fitting, the scenario where your model fits accurately the training data but do not generalize to fit the testing set. It normally occurs when your input data is high-dimensional but only few observations are available for training. Other available models are graphical models, widely used when relationships between samples are known or can be inferred, and probabilistic models, preferred when a family of solutions is required. For unsupervised learning settings, we could use k-means for clustering or gaussian mixture models for density estimation.

Most of the aforementioned algorithms have hyper-parameters that need to be tuned during the learning phase. This is called model selection and can be done by model averaging or using a (nested) cross-validation strategy [125].

Error estimation The ability of the algorithm to predict unseen data is used to assess its performance. A common approach is to use the hold-out method, which randomly divides the dataset into two independent training and test subsets. The training set is used to train the classifier while the test set is used for evaluation purposes. However, the hold out estimation depends on the test subset size, requiring large sample sizes to generalize well. Due to the scarcity of annotated medical imaging data it can easily over-fit. Instead, k -fold cross-validation (CV) can be used [125]. In k -fold CV the dataset is split into k independent folds, using $k - 1$ folds for training and 1 fold for evaluating the algorithm. Iterating this scheme k times, the test fold is changed at each iteration and thus, each fold is used for model evaluation only once when the process is finished. Finally, gathering the model performance at each iteration, the overall system performance can be computed. Another approach is the Leave-one-out (LOO) method, which is equivalent to the k -fold when k is equal to the number of samples in the dataset. LOO is a much less biased statistical estimator of the algorithm performance compared to the k -fold approach, but produce larger variances if different training sets are sampled from the same distribution. The LOO scheme is also more computational intensive, so for rather large datasets is not recommended.

2.3.1 Pattern recognition in Alzheimer's research

Alzheimer's disease, as the most common form of dementia, has attracted a lot of attention from the clinical to the pattern recognition community. There exists a proliferation of observational studies gathering AD related markers with increasing efforts to deliver rich imaging biomarkers. Their increasing sample sizes allow to design more complex models. Some examples are ADNI [126], DIAN Study [127], AlfaStudy [128], INSIGHT-preAD [56], IDEAS [129], EMIF [130] or GAAIN [131]. Some of these initiatives are publicly available, such as the aforementioned ADNI, OASIS [132] or AIBL [51] datasets, which facilitates the comparison between models and performance. In this section we will focus on AD prediction using imaging biomarkers.

During the past 15 years, neuroimage-based prediction of Alzheimer's disease clinical diagnosis has been increasingly reported in the literature. The goal is two-fold: find image-based biomarkers related to AD and derive individual diagnosis using a single brain MRI and an already trained model in a larger pool of individuals [133]. An initial approach was to use univariate models, such as using the entorhinal or hippocampal volume as features [134]. However, neurological conditions are associated with large-scale

networks of distributed regions [135]. Hence, aggregating information across multiple regions using multivariate methods might increase the sensitivity and specificity of brain biomarkers. A wide range of modalities trying to predict AD clinical diagnosis have been published in the literature, as summarized in [133].

Structural MRI (sMRI) has been by far the most used imaging biomarker as a rather cheap, non-invasive yet powerful biomarker of neuronal loss [136] that is highly available in current clinical practice and research. Structural imaging biomarkers are able to detect atrophy which is closely linked to neuropsychological and cognitive deficits hence considered sensitive for diagnosis and prognosis of clinical AD [137]. The neuroimaging AD signature has been established as structural changes in AD-vulnerable regions (i.e. medial temporal lobe structures such as entorhinal cortex and hippocampus) that constitute diagnostic markers of cognitive impairment and AD progression [39, 138]. Several morphological features can be derived from structural images such as volume, cortical thickness or shape. The authors in [139] provide a good review of machine learning algorithms using structural MRI. Sabuncu et al. in [140] aimed at creating a benchmark to compare with when using different structural MRI features to predict AD clinical diagnostic categories, CSF $A\beta$ levels or Mini Mental State Examination (MMSE). They performed 12 different baseline experiments using 4 different types of morphological features (region of interest averages of grey matter volumes, cortical thickness, a combination of both and the cortical thickness from a mesh of 20k points) and 3 different multivariate algorithms (SVM, Relevant Voxel Machine and Neighborhood Approximation Forest). Any combination resulted significantly superior to any other even though they show consistently better results comparing multivariate to classical univariate models.

In the literature, many studies assessed the predictive capacity of sMRI to distinguish between cognitively unimpaired and dementia (due to AD) subjects. The study in [141] showed comparable performance between trained radiologists and pattern recognition algorithms (SVM) achieving both roughly 90% accuracy. Automatic methods can achieve high accuracies, such as 90.5% ([142]) or 92% ([143]) that are representative of the literature. Different morphological features are used achieving similar performances. Whole brain grey matter (GM) tissue maps were used in [144] with 89% accuracy. Downsampled grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) probability tissue maps were used to derive a Structural Abnormality Index Score (STAND-Score) that gives a 89.3% accuracy [145]. A combination of volumetric and cortical thickness within regions of interest averages and demographic features (age and APOE $\epsilon 4$) reported a total accuracy of 87.6% [146]. The heterogeneity in subjects within the MCI group hamper the use of automatic methods to differentiate from CU or dementia subjects. In longitudinal studies, MCI subjects are usually separated into those that converted to dementia (MCI-C) in the time span of the study and those that remained as MCI (MCI-NC). Nho et al. [142] and Costafreda et al. [147] trained a pattern recognition system to predict dementia from CU subjects and applied to predict MCI-C/MCI-NC with 72.3% and 80% of accuracy, respectively. The later used features from a 3D mesh representation of the hippocampus from both hemispheres. Eskildsen et al. [148] reported an accuracy of 76% using only cortical thickness measures and Davatzikos et al. [149] used a combination of morphological features and CSF achieving an accuracy of 73.4%. Nonetheless, most studies use the clinical MCI definition in [39] without in-vivo biomarker confirmation of underlying Alzheimer's pathology, yielding non-specific and probably less accurate studies.

In addition to neuronal loss, structural changes in AD are also characterized by the loss of axons and myelin sheaths ([150]), white matter tissue atrophy ([78],[151]) and loss of structural integrity in white matter pathways (Stebbins and Murphy, 2009). Diffusion MRI (dMRI) appear as a sensitive biomarker to these microstructural changes and has also been used in automatic clinical AD diagnosis studies. AD subjects are usually characterized by elevated mean diffusivity (MD) and decreased fractional anisotropy (FA) [152]. Tractography algorithms are also used upon dMRI to assess structural connectivity between brain ROIs. Nir et al. in [153] used a tractography algorithm to compute MD and FA along identified WM tracts and use those features for CU/dementia classification, achieving accuracies up to 80% in their experiments. Similarly, resting-state functional MRI (rs-fMRI) is emerging as an interesting biomarker for measuring connectivity in dementia studies. The default mode network (DMN) is the most studied large scale brain network and show decreased activity in typical AD regions that might be initially compensated by increasing activity in other regions at early AD stages [92]. Functional connectivity is often used as input features to pattern recognition systems and are derived by computing the mean activity in different brain regions of interest and the correlation between them. High accuracies can be found when discriminating CU individuals from those with dementia (100%, [154]), MCI converters from non-converters (91.4%, [155]) and MCI from CU or dementia subjects (75% and 97%, [156]). Interestingly, several studies exploit the complimentary information between different MRI modalities for clinical AD diagnosis. Several strategies are explored, such as multi-task learning [157], neural networks [158] or graph-based techniques [159].

The study of preclinical (PC) stages of AD has lately become a major research focus for the community as early intervention and preventive therapies may have better chances of treatment success. This preclinical stage constitutes a long and “silent” stage along the Alzheimer’s continuum characterized by abnormal pathological, metabolic and neuronal changes before the onset of clinical AD symptoms. Hence, detecting brain changes due to preclinical Alzheimer’s pathology is gaining interest in the pattern recognition community. According to several studies in the literature [160, 161, 162, 163, 164, 165] subjects in the preclinical stage of AD experiment subtle changes in brain morphology. Recently, some studies have investigated the use of MRI to predict amyloid positivity in the brain [166] and its potential impact in clinical trials [167].

Other imaging techniques (e.g: amyloid PET [168], tau PET [61], FDG PET [169]) or pathological markers (e.g: CSF [170]) provide different but complementary information to MRI. Risk factors (e.g: genetics) may influence the disease onset and progression. Hence, multimodal studies would provide a more complete picture of the disease helping its diagnosis and prognosis. Liu et al. provide a recent review of multimodal techniques in [171]. Some challenges these studies face are how to combine different sources of information or how to deal with missing data. According to the hypothetical model in [40] AD biomarkers become abnormal at different time points along the Alzheimer’s continuum. Disease progression models (DPM) are data-driven models that aim at explicitly model the temporal evolution of such biomarkers providing better interpretation of the evolution of the pathology. Gaussian processes (GP) have been used for this purpose [172]. Event Based modeling (EBM) are DPM computational algorithms that describe a disease as a series of events which may help to validate biomarker ordering helping understanding the progression pattern of disease [173, 174]. A recently published work from Young et al. [175] uses ideas from disease progression modeling and clustering to

develop an unsupervised machine learning framework able to capture both spatial and temporal disease heterogeneity for staging and subtyping different subjects.

NeAT: a nonlinear Neuroimaging Analysis Toolbox.

The etiology of Alzheimer’s disease is poorly understood, even though it is related to many risk factors. Brain imaging provides powerful biomarkers that measure brain metabolism, structure or activity, among others. In this chapter we introduce a toolbox for the analysis and study of nonlinear effects on medical images. We implement several linear and nonlinear curve-fitting methods and statistical inference metrics. We embed a graphical user interface (GUI) for 3D visualization of the results as well as some other functionalities for curve clustering or model comparison. Finally, we perform validation analysis with two cases of studies published in the literature.

This work was motivated by our BarcelonaBeta Brain Research Center (BBRC) colleagues, specially in the person of Dr. Juan Domingo Gispert, for the analysis of nonlinear effects on MRI data. Several people have been involved in the toolbox development, specially Mr. Asier Aduriz and Mr. Santi Puch during their BSc. thesis.

3.1 Introduction

The increase of computational power and advances in neuroimaging acquisition that enable faster scans and provide multiple image contrasts and modalities has motivated the development of complex modeling techniques. There are many neuroimaging tools available to the neuroscientific community, whose ultimate goal is to conduct statistical tests to identify significant effects in the images without any a priori hypothesis on the location or extent of these effects. In the literature, analysis at different levels of brain morphometry are found, involving voxel-based [84], surface-based [83] or boundary-based analysis [176].

Irrespective of these differences, the vast majority of them perform statistical inference upon different implementations of the General Linear Model (GLM). GLM has been shown to be flexible enough for conducting most of the typical statistical analysis [95]. However, it has a rather limited capability to model nonlinear effects. To this regard, it is worth noting that linear models are not sufficient to describe neuroimage variation with cognitive decline [177, 178] or pathological depositions in neurodegenerative disease [179, 180, 181, 182, 183, 127, 162] in neurodegenerative diseases progression. Moreover, many relevant confounders in neuroimaging are shown to be better described by nonlinear processes, such as the impact of aging on cognitive decline [184] or grey matter volume [185]. Under the GLM, the modeling of non-linear effects is limited to using polynomial expansion or transforming the variables of interest to linearize their effects. However, such approximations are suboptimal [186, 187, 188] and, with the increasing availability of neuroimaging data, a wide range of non-linear modeling methods can be reliably applied for statistical inference [189].

In this work, we describe a new neuroimaging toolbox which is able to model nonlinear effects on data at two different levels: voxel-based and surface-based. We pool together several nonlinear parametric models, provide different model comparison strategies and implement a graphical user interface (GUI) for visualization purposes. In the following sections we briefly describe the main functionalities of the toolbox and illustrate its features with two studies: (i) nonlinear atrophy patterns across the Alzheimer’s disease continuum defined as a function of cerebrospinal fluid (CSF) biomarkers [162] and (ii) the effects of apolipoprotein $\epsilon 4$ genotype on brain aging, a risk factor to develop sporadic Alzheimer’s disease (AD) [190].

3.2 The toolbox

A general overview of the tool operatibility and its options and functionalities are introduced in this section. A more detailed mathematical description of the curve fitting methods and statistical inference metrics is provided, even though the reader is encouraged to read the original sources for a more deep understanding of such methods. More instructions on how to download and use the tool can be found in <https://imatge-upc.github.io/neat-tool/>.

3.2.1 NeAT overview

The NeAT toolbox is a modular and easy-to-use toolbox for the analysis of non-linear effects on medical images. Several curve fitting methods are used to model the relationship between certain factors (e.g: age, disease phenotype, genotype) and medical images (e.g: T1-weighted MRI). Those methods may include multiple covariates (factors) that can be split into confounder factors and variables of interest by using contrasts. A simple preprocessing step allow to orthogonalize, orthonormalize or simply normalize all covariates. A wide range of metrics can be used to assess the goodness of fit of each model. Statistical inference also allow the use of contrasts on modeling factors. The embedded 3D visualization GUI provide an unified and interactive environment to visualize both 3D statistical inference maps and the estimated curve at each voxel.

A high-level overview of the toolbox pipeline is provided in Figure 3.1. It consists of several interdependent modules connected through a *Processing* library that performs serialization between functionalities. Each other module (curve fitting, fit evaluation and visualization) is designed separately using abstract classes that facilitate both continuous adaptation and possible extensions of the toolbox. A description of each module/functionality is detailed in the following sections.

3.2.2 Model estimation

The model estimation step finds a parametric function of several explanatory variables that best fits the observations in terms of maximizing a quality metric or minimizing a loss function. Different specification of the latter two give rise to different models or fitters. To analyze the basics of each fitter, we consider the regression model

$$Y = f(X) + e \tag{3.1}$$

where $Y = [Y_1, Y_2, \dots, Y_N] \in \mathcal{R}^{L \times N}$ are the N dependent observations (e.g. number

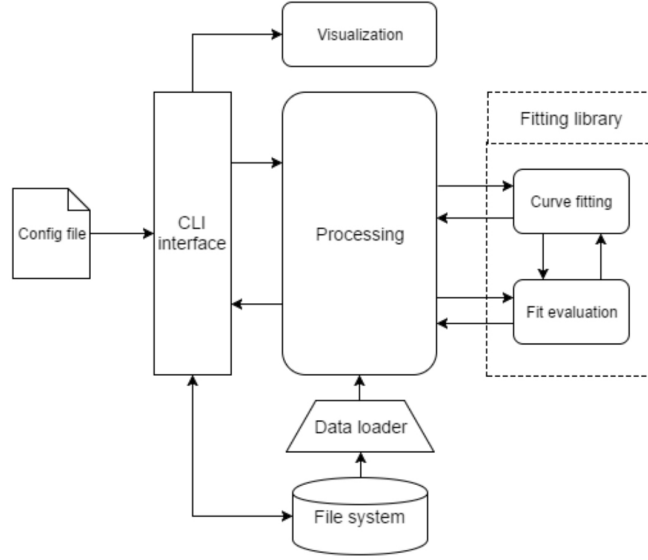


Figure 3.1: Toolbox pipeline. The *Processing* module govern the interaction between all other libraries that will be explained through the manuscript

of voxels), $X = [X_1, X_2, \dots, X_M] \in \mathcal{R}^{L \times M}$ are the M independent factors (covariates), $f(X) \in \mathcal{R}^{L \times N}$ is the fitted curve and $e \in \mathcal{R}^{L \times N}$ is the estimation noise. Each input variable (Y_i, X_j) is an L -dimensional vector corresponding to different measures (e.g. different subjects) of the same magnitude. Each covariate can be independently modeled using different models and the overall estimated model is found by adding up the contribution of each one:

$$\hat{Y} = \sum_{m=1}^M f_m(X_m) \quad (3.2)$$

being f_m the associated curve fitting method for each covariate. The available methods are detailed below. All observations are processed in chunks and fitted independently ($Y_i = f(X) + e_i$). Data processing (normalization and orthogonalization) techniques are optionally prepended to the overall analysis.

In this toolbox we consider the general framework that splits explanatory variables into variables of interest (predictor variables) and confounder factors (corrector variables) as explained in [191]. The goal of this scheme is to deduct confounder effects on the dependent variables to isolate the main effects of the variables of interest we want to analyze. This paradigm is widely used in neuroimaging: for example, using age (corrector) as confounder variable when analyzing the effect of Alzheimer's disease (predictor) on hippocampus volume (observation or dependent variable). Concretely, we split the initial space S , defined by all explanatory variables X , into two subspaces: predictor (S_p) and corrector (S_c) subspaces of dimensions P and C , respectively ($M = C + P$). The predictor subspace is defined using a contrast matrix C , described by $X_p = XC$, and its model is defined as $\hat{Y}_p = f_p(X_p)$. On the other hand, the corrector subspace is built using a null-contrast matrix (orthogonal to the contrast matrix), $C_0 = I - CC^\#$, where

$C^\#$ the pseudoinverse of C . Hence, the corrector subspace is described by $X_c = XC_0$, and its model is defined as $\hat{Y}_c = f_c(X_c)$. Even though C and C_0 are orthogonal, both subspaces are orthogonal only if the columns of X are orthogonal.

We model the contribution of each subspace on explaining the observations variance using an additive model $Y = \hat{Y}_c + \hat{Y}_p + e$, fitting first the corrector model on the observations ($Y = f_c(X_c) + e_c$) and then the predictor model on the residuals ($e_c = f_p(X_p) + e$). Since the fitting is done separately, both corrector and predictor functions, f_c and f_p can be any nonlinear model implemented in the toolbox. Note that each corrector and predictor variables can be modeled using different curve fitting methods:

$$\hat{Y} = \hat{Y}_c + \hat{Y}_p = f_c(X_c) + f_p(X_p) = \sum_{i=1}^C f_{c_i}(X_{c_i}) + \sum_{i=1}^P f_{p_i}(X_{p_i}) \quad (3.3)$$

Baseline curve fitting methods implemented in the toolbox are: (i) general linear model, (ii) generalized additive model and (iii) support vector regression. Each subspace (predictor and corrector) can be modeled by any of these techniques. While the first two methods model each dimension independently, the third allows for interactions between different dimensions.

3.2.2.1 General Linear Model: GLM

The General Linear Model [192] is the extension of multiple regression models to the case of multiple observations. The effect of each factor is independently analyzed without accounting for interactions between them. The model reads as follows:

$$Y = f(X) + e = X\beta + e \quad (3.4)$$

where β are the model parameters and e is the error of the model. GLM optimization involves minimizing the mean squared error $\|e\|^2$ between data points and the fitted curve. Nonlinear relationships can be modeled in the GLM framework by using a polynomial basis expansion of each regressor. The total number of degrees of freedom is the number of covariates in the analysis ($df = M$), including each basis expansion if used.

3.2.2.2 Generalized Additive Model: GAM

A Generalized Additive Model [193] is an extension of additive models (AM) to the case of multiple observations. In GAM, each observation depends on unknown smooth functions of each covariate:

$$Y = f(X) + e = f_1(X_1) + f_2(X_2) + \dots + f_M(X_M) + e \quad (3.5)$$

In the context of this toolbox, f_i refer to parametric smooth functions, called smoothers, that are iteratively estimated using the backfitting algorithm [194] to minimize the mean squared error $\|e\|^2$. If linear or polynomial smoothers are used, GAM is equivalent to GLM. Other smoothers available are B-splines or natural splines, implemented using the Patsy library¹. The total number of degrees of freedom is the sum of degrees of freedom

¹<https://patsy.readthedocs.io/en/latest/>

of each smoother $df = df_1 + df_2 + \dots + df_M$. For a linear smoother, the number of degrees of freedom is one $df_i = 1$, for polynomial smoother, the number of degrees of freedom is the polynomial order $df_i = d$ and for splines-based smoothers, the number of degrees of freedom is an input parameter set by the user.

3.2.2.3 Support Vector Regression: SVR

In opposition to GLM or GAM, that only account for the additive effect between covariates, Support Vector Regression [195] is a multivariate method that inherently accounts for interaction between covariates. In SVR, the goal is to find a function $f(x)$ that has at most ϵ -deviation from the observations and that is as smooth as possible. However, since the ϵ -deviation constraint might not be feasible, a hyperparameter C controls the balance between smoothness and errors greater than ϵ . SVR is a linear method in the parameters with a closed form solution. To introduce nonlinearities, SVR uses the *kernel trick* which implicitly transforms the inputs to a higher dimensional feature space by only specifying their inner product, i.e. the kernel function $k(X^i, X^j) = \langle \phi(X^i), \phi(X^j) \rangle$, where X^i and X^j are two different data points. Hence, the overall model is parameterized using parameters β as follows

$$Y = f(X) + e = \sum_{i=0}^{L-1} k(X^i, X)\beta_i + e \quad (3.6)$$

where X^i are all data points used to fit the model and X the feature vector to regress. Two kernel functions are implemented in this toolbox using the scikit-learn library [196]: polynomial and the radial basis function (RBF) defined as $k(X_i, X_j) = \exp(-\gamma\|X_i - X_j\|^2)$, where γ is a hyperparameter defining the width of the kernel. The total number of degrees of freedom depends on the kernel used and it is based on the solution proposed in [197].

Hyperparameter search: SVR relies on the election of several hyperparameters: ϵ and C for the general solution and kernel related hyperparameters, such as γ in RBF kernels. The hyperparameter values can be automatically determined by a grid search on the hyperparameter space [198]. This method consists of several steps: (i) sample M different value combinations from the hyperparameter space using one of the sampling strategies provided in this toolbox: random or deterministic sampling with linear or logarithmic scale, (ii) fit a subset T of the observations on all M hyperparameter combinations and (iii) select the hyperparameter combination that minimizes the metric of interest, q on the subset S ($Q = \sum_{i \in t} q_i$). The available metrics are: (i) minimum squared error, (ii) F-test goodness of fit and (iii) Mallows' C_p statistic [199, p. 211]. To avoid selection bias, this procedure is iterated varying the selected subset of observations (T). Larger subset sizes provide better hyperparameter estimations but increasing time and memory requirements, due to the intensive search performed. However, we allow parallelization of the second step and further iterations of the algorithm. To account for large between-subject variability, the voxelwise metric values are weighted by the inverse of the variance of each observation ($\hat{q}_i = q_i/\sigma_i$, $\hat{Q} = \sum_{i \in T} \hat{q}_i$). Moreover, due to the vast amount of background voxels, only those with minimum variance (σ_{min}) can be included in the subset of observations.

3.2.3 Statistical inference

Statistical maps evaluating the goodness of fit and penalizing by the complexity of the model can be computed for each of the fitting methods presented in Section 3.2.2. For statistical inference, two different thresholds can be used: numerical, over the metric used, and spatial, to ensure minimum cluster size of relevant voxels. For this purpose, several metrics are available in the tool:

- **Minimum squared error (MSE) and Coefficient of determination (R^2):** these two metrics evaluate the predictive power of the model without penalizing for its complexity.

$$\text{MSE} = \frac{1}{N} \|Y - f(X)\|^2 \quad (3.7)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_y}, \quad SS_{res} = \|Y - f(X)\|^2, \quad SS_y = \|Y - \bar{Y}\|^2 \quad (3.8)$$

where $\bar{Y} = 1/L \sum_{i=0}^{L-1} Y^i$ is the mean of the observations.

- **Akaike Information Criterion (AIC):** the AIC criteria [200] is founded on information theory. It is useful for model comparison as it provides a trade-off between the quality or goodness of fit and the complexity of the model, which is proportional to the number of parameters.

$$\text{AIC} = 2k - 2LLR, \quad LLR = -\frac{N}{2} (\log(2\pi \cdot \text{MSE}) + 1) \quad (3.9)$$

where k is the total number of parameters and LLR is the log likelihood ratio

- **F-test:** the F-test is a statistical test following an F-distribution under its null-hypothesis. In the context of this toolbox, it evaluates whether the variance of the full model (correctors and predictors) is significantly lower than the variance of the restricted model (only correctors). Under the null-hypothesis, the full-model does not provide any significantly better fit than the restricted model, resulting an F-statistic with $(df_{full}, df_{restricted})$ degrees of freedom and the corresponding p-value. Rejection of the null hypothesis is based upon the p-value.

$$f_{\text{score}} = \frac{\frac{SS_{restricted} - SS_{full}}{df_{full} - df_{restricted}}}{\frac{SS_{full}}{N - df_{full}}} \quad (3.10)$$

$$p_{\text{value}} = 1 - F(f_{\text{score}}, df_{restricted}, df_{full}) \quad (3.11)$$

where $SS_{restricted} = \|Y - f_c(X_c)\|^2$, $SS_{full} = \|Y - f_c(X_c) - f_p(X_p)\|^2$ and $F(x, d_1, d_2)$ is the F -distribution.

- **Penalized Residual Sum of Squares (PRSS), Variance-Normalized PRSS:** PRSS is introduced in this toolbox as another evaluation metric that accounts for the goodness of fit and penalizes for the model complexity. However, differently from other metrics, complexity is not computed with the degrees of freedom but using the curve shape itself. Hence, a complex model such as SVR with Gaussian

kernel that provides a linear curve will penalize as much as the GLM. VNPRSS is an adaptation of PRSS for data with high-variability, like medical images, and penalizes each error term by the inverse of the observations variance.

$$\text{PRSS} = \text{MSE} + \phi c_{\text{abruptness}}, \quad c_{\text{abruptness}} = \int f''(X) dx \quad (3.12)$$

$$\text{VN-PRSS} = \frac{\text{PRSS}}{c_{\text{variance}}}, \quad c_{\text{variance}} = \|f(X) - \bar{Y}\|^2 \quad (3.13)$$

3.2.4 Post-hoc analysis

The NeAT toolbox provides several functionalities for post-hoc analysis of the generated curves and statistical maps. Different model comparison strategies and a curve clustering algorithm are presented in what follows.

3.2.4.1 Model comparison

In order to compare K statistical maps generated using different fitting models we combine them into a single statistical map providing different information:

- **Diff-map** ($K = 2$): it provides the difference between maps, being useful for quantitative detection of differences between $K=2$ fitting models.
- **ABSdiff-map** ($K = 2$): it provides the absolute difference between maps, being useful for quantitative detection of differences between $K=2$ fitting models.
- **SE-map** ($K = 2$): it provides the squared difference between maps, being useful for quantitative detection of differences between $K=2$ fitting models.
- **RGB-map** ($K = 3$): it places each map in a different color channel. It might be useful to compare the intersection of several fitting models showing agreement and disagreement among them.
- **Best-map** ($K > 1$): it computes the best fitting model at each voxel. It might be useful for model localization in the brain.

3.2.4.2 Clustering

We incorporate a curve clustering functionality [201] for comparing the behavior of different brain regions. In that sense, we provide a scalable and non-parametric algorithm that is able to explore similarities and dissimilarities of the fitted curves across the brain and group them in a total of N_c clusters.

We adopted the hierarchical clustering framework [202] implemented in scikit-learn [196]. It is a bottom-up approach where initially each curve defines its own cluster. Next, pairs of clusters are successively merged according to a certain similarity metric and a linkage criterion. As a similarity metric, we use a weighted sum of distances:

$$S(x, y) = \sum_{i=0}^{N_D-1} w_n d_n(x, y) \quad (3.14)$$

where (x, y) are two different curves, d_n is the Euclidean distance between the n^{th} discrete derivative of each curve, w_n is the weight of each derivative to the total similarity metric and N_D is the total number of derivatives used. In our implementation, we fix $N_D = 3$ and $\mathbf{w}^T = [0.2, 0.8, 0.2]$. As a linkage criterion, we use the average distance between all possible pairs of elements of both clusters.

$$L(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} S(a, b) \quad (3.15)$$

where A, B are two different clusters, $|A|, |B|$ are the cardinalities of the clusters and a, b represent a curve from each cluster. Hence, at each step of the hierarchy, the two clusters that minimize the linkage criterion are combined. The algorithm stops when it reaches N_c clusters (a parameter predefined by the user).

3.2.5 Visualization

This toolbox provides *show-curves* and *show-data-distribution* functionalities and a graphical user interface (GUI) for visualization purposes. The *show-curves* is a command line functionality that reads either the voxel coordinates in mm (x,y,z) for voxel-based morphometry (VBM) analysis, or the vertex number (x) for surface-based morphometry (SBM) analysis, both referenced to the template specified in the configuration file. The *show-data-distribution* functionality allows the user to visualize the input data distribution (observations, residuals, covariates) using different types of plots: univariate and bivariate densities, boxplots and a categorical boxplots.

3.2.5.1 Graphical User Interface (GUI)

An interactive visualization GUI for 3D volumes (VBM) is provided for further analysis of the results. It allows to load 3D overlays over a template and visualize the generated curves for one or several fitting models of interest. Overlays must have the same extension as specified in the configuration file and can be either generated by the tool (e.g: statistical maps, model comparison maps or clustering maps) or external (e.g: brain structure atlases). The GUI simultaneously shows the three orthogonal planes (axial, coronal and sagittal) and the corresponding curve at any selected voxel. Inspection of the overall brain and associated curves can be done online using the cursor in an interactive way. Due to long rendering times, for visualization of 2D surfaces (2D) we recommend using other visualization software (e.g: FreeSurfer) in parallel with the *show-curves* functionality.

3.2.6 NeAT specifications

NeAT toolbox uses a configuration file to specify experiment related options such as input/output files or experiment parameters. The overall analysis pipeline (model estimation, statistical inference, visualization) is split into smaller steps using different scripts. A command line interface (CLI) is used for communication between the toolbox and the user, allowing to run the scripts and set specific parameters for the analysis (e.g. which fitting module to use as model estimator). The toolbox input files consists of covariates and images. Input covariates need to be stored in a spreadsheet either in .csv or .xls extension. Input images can be either preprocessed using voxel-based morphometry (VBM) or surface-based morphometry (SBM): nifti formats (.nii/.nii.gz), the Massachusetts General Hospital formats (.mgh/.mgz) and measurements of cortical

thickness (.thickness) and surface area (.area) can be used in the tool. Each analysis step of the global pipeline generates different output files saved under the directory specified in the configuration file. Statistical maps are saved using the same extension as input files allowing compatibility with other neuroimaging packages (e.g: visualization software). As a programming language, Python (version 3.6) is used due to its object-oriented programming paradigm that provides flexibility in toolboxes with increasing size and complexity. Moreover, Python is becoming progressively popular in the neuroimaging field with growing scientific libraries (e.g: [scipy \[203\]](#)), neuroimaging (e.g: [nibabel \[204\]](#)) or machine learning toolkits (e.g: [sklearn \[196\]](#)).

3.3 Case study 1: Atrophy patterns across the Alzheimer’s disease continuum

Nonlinear volumetric changes in grey matter across the Alzheimers disease (AD) continuum are shown to better fit some regions of the brain [[162](#)], suggesting that pathological processes may affect different brain locations at different times during disease progression. In [[162](#)] nonlinearity is modeled using GLM with a 3rd-order polynomial basis expansion, comparing the relevance of linear against higher-order predictors. In this work, we use NeAT to fit several nonlinear models to voxelwise grey matter volume (GMv) and statistically compare them.

Study participants were enrolled in a single-cohort study from the Alzheimer’s Disease and Other Cognitive Disorders Unit in Hospital Clinic of Barcelona (HCB). The cohort comprises 129 subjects (62 controls, 18 preclinical AD, 28 mild cognitive impairment (MCI) due to AD and 21 diagnosed AD) that underwent an MRI scan, registered to a common space, and a CSF lumbar puncture. The AD continuum is defined biologically by the AD-CSF index [[205](#)] which combines CSF biomarkers into a single indicator that determines the position of each subject along the AD continuum. For further details on both MRI processing and CSF acquisition, refer to [[162](#)].

Following the standard procedure of splitting covariates into confounding factors and predictors, we fit a corrector GLM model using gender and a second order polynomial expansion of age. We use AD-CSF index as the predictor variable fitting several models to the GMv corrected observations: (i) GLM with third order polynomial expansion, (ii) GAM using b-splines as smoothing function (iii) SVR using third order polynomial kernel and (iv) SVR using Gaussian kernel. We use an F-test to statistically compare all predictor models. Statistical significance was set to $p < 0.001$ uncorrected for multiple comparisons with a cluster-extent threshold of 100 voxels.

In [Figure 3.2](#) we show some examples of the visualization GUI using the *best-map* option to compare the aforementioned fitting methods. Results using GLM with polynomial basis expansion are coherent with the ones found in [[162](#)]. Moreover, using the toolbox we are able to detect regions that result in larger goodness-of-fit using nonlinear models even though they use greater number of degrees of freedom, indicating that may be better described using nonlinear rather than linear models. There is a high overlap between second order polynomial expansion of GLM, GAM with b-splines and SVR with polynomial and Gaussian kernels. Due to the low numbers of degrees of freedom used, GLM and GAM appear to be the most relevant models across the brain. On the other hand, using a Gaussian kernel on SVR employ higher number of degrees of freedom and its

relevance is restricted at the center of typical AD subcortical regions (e.g: hippocampus and amygdala).

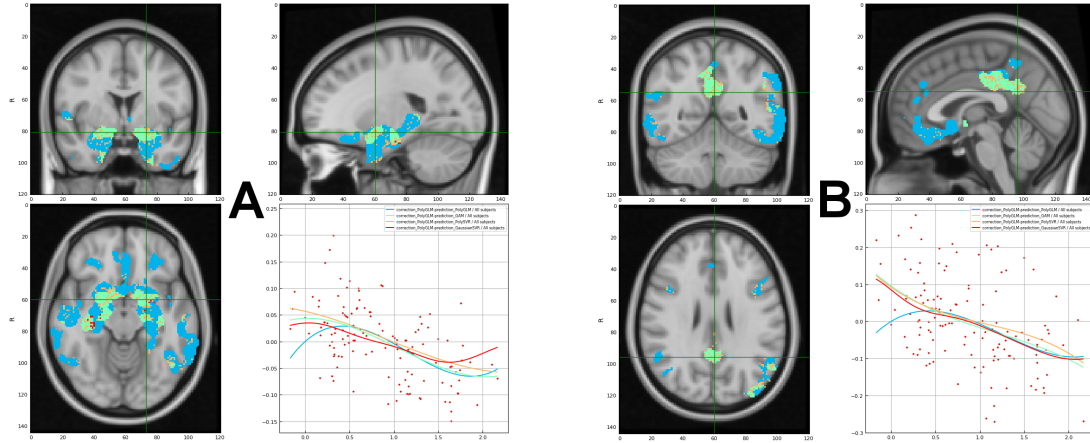


Figure 3.2: Comparison between different curve fitting models: third order polynomial expansion of GLM (blue), B-splines GAM (green), SVR with polynomial kernel (yellow) and SVR with Gaussian kernel (red). The *best-map* is used for statistical comparison, showing the best (in terms of F-test) model among all four models with statistical significance using uncorrected $p < 0.001$ separately for each model. Estimated curves show the variation of grey matter volume (y-axis) and AD-CSF index (x-axis). Based on CSF amyloid-beta and tau levels, the AD-CSF index measures biomarker progression using a single index normalized between 0 (no altered biomarkers) and 2 (full AD-like alteration) [205]. The figure on (A) corresponds to the left hippocampus and the figure on (B) corresponds to the right precuneus.

Further analysis of the results can be done using the clustering functionality of the tool. Using the GAM model, we look for regions with similar atrophy patterns along the AD continuum. In Figure 3.3, we show the result for $N_c = 6$ clusters. Symmetric patterns are found in both hemispheres and, interestingly, inner and outer parts of brain regions show similar patterns at different scales, showing this characteristic *onion-like* map.

3.4 Case study 2: Effects of $APOE-\epsilon 4$ in brain aging

The apolipoprotein E (ApoE) has an important role in regulating amyloid-beta ($A\beta$) plaques deposition in the brain, a key neuropathological AD biomarker that shows abnormal levels of accumulation in cognitively intact individuals, characterizing the pre-clinical phase of AD. The ApoE gene, $APOE$, is polymorphic and contains three different alleles referred as $APOE-\epsilon 2$, $-\epsilon 3$ and $-\epsilon 4$ coding three different isoforms and six different genotypes. The isoform $APOE-\epsilon 4$ has been shown to be less efficient in $A\beta$ clearance than other isoforms ($\epsilon 2$, $\epsilon 3$), related to loss of neuroprotective function [206], memory decline [207] on cognitively healthy subjects or grey matter hippocampal degeneration [208]. Hence, $APOE-\epsilon 4$ allele is considered the major genetic risk for late-onset AD.

In this work, we analyze the impact of $APOE-\epsilon 4$ allele load on brain morphology and its interaction with age. The ALFA (ALzheimer's and FAMilies) cohort presented in [128]

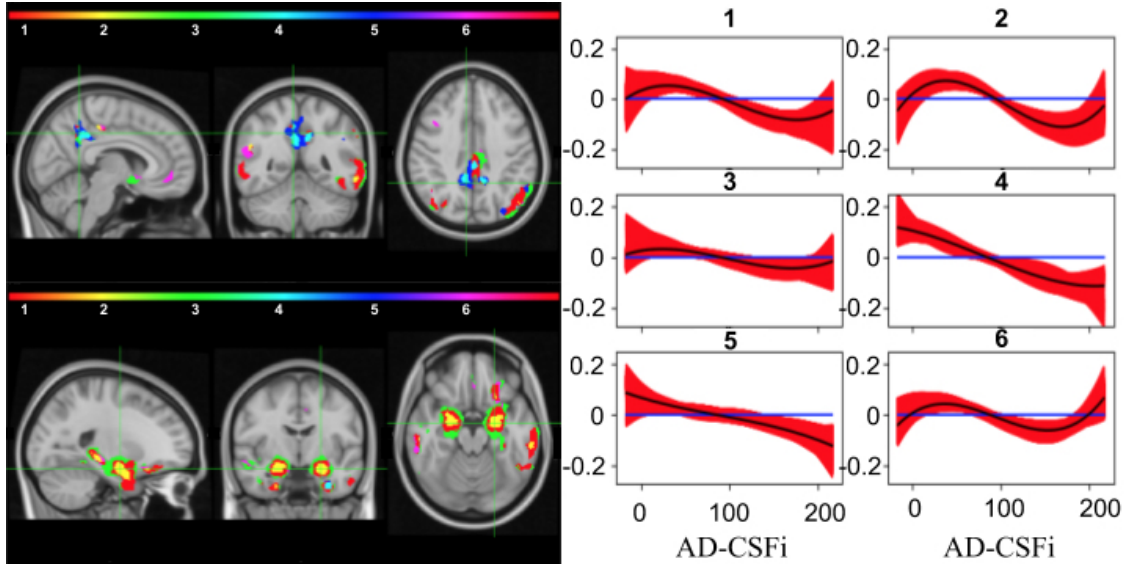


Figure 3.3: Curve clustering algorithm run on relevant atrophy patterns along the AD-CSF index using GAM fitting. The number of clusters is set to $N_{clusters} = 6$. On the left, we show the relevant voxels color-coded to describe the association of each voxel with each cluster. On the right, we show all curves associated to each cluster (red) and their respective centroid (black).

was used for this purpose, involving 533 subjects that underwent *APOE* genotyping and an MRI scan. For statistical analysis, participants were pooled according to the *APOE*- $\epsilon 4$ allele load: 65 homozygotes (HO) that have *APOE*-genotype with 2 copies of the *APOE*- $\epsilon 4$ allele, 207 heterozygotes (HE) with a single copy of the *APOE*- $\epsilon 4$ allele and 261 non-carriers (NC).

3.4.1 *APOE* genotype effects on brain morphology in normal aging population.

In this second case study, we begin replicating the results of [190] with respect to the *APOE* genotype effects on brain morphology and explore possible non-linear effects. The baseline model consists of three dummy variables characterizing each genotype (NC, HE, HO) defining the number of $\epsilon 4$ alleles. Sex, years of education, total intracranial volume and linear and quadratic expansions of age were included as covariates. Due to the reported interactions [209] of *APOE* status and age, we fit the model with the interaction terms *APOE* \times age and *APOE* \times age². We apply the contrast [-1,0,1] on dummy variables indicating *APOE*- $\epsilon 4$ count, defining an additive model that predicts incremental/decremental effects of *APOE* genotype. Results using the linear model are shown in Figure 3.4(A), replicating the findings in [190]. The use of the tool allowed us to study non-linear effects of the genotype. Concretely, in Figure 3.4(B) and Figure 3.4(C) we show results using GAM and SVR with polynomial kernel models, respectively. Smaller effects are observed and only relevant effects are found in regions such as bilateral thalamus, right hippocampus, right superior frontal and small cluster around the right caudate and the left middle occipital. Nonlinear modeling fails behind linear modeling of *APOE*- $\epsilon 4$ count, probably because it is a categorical ($C=3$) predictor. Hence, due to higher

degrees of freedom used in GAM and SVR, only larger significant values survive the used threshold.

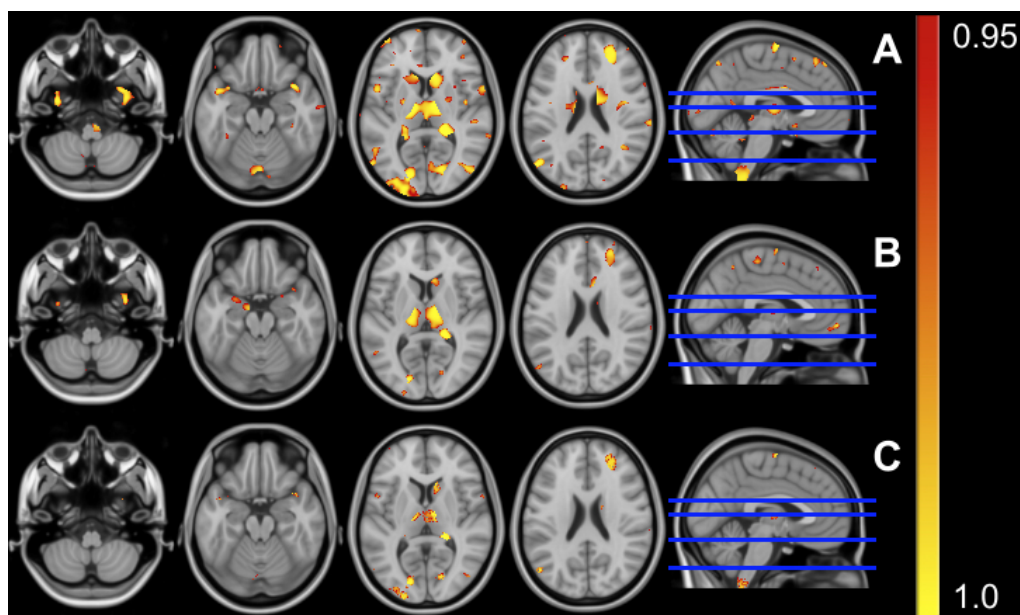


Figure 3.4: Statistical inference using volumetric data and different curve fitting modules: using GLM (A), using GAM (B) and using SVR with a polynomial kernel (C). For visualization purposes, statistical significance threshold is set to $p < 0.05$ uncorrected.

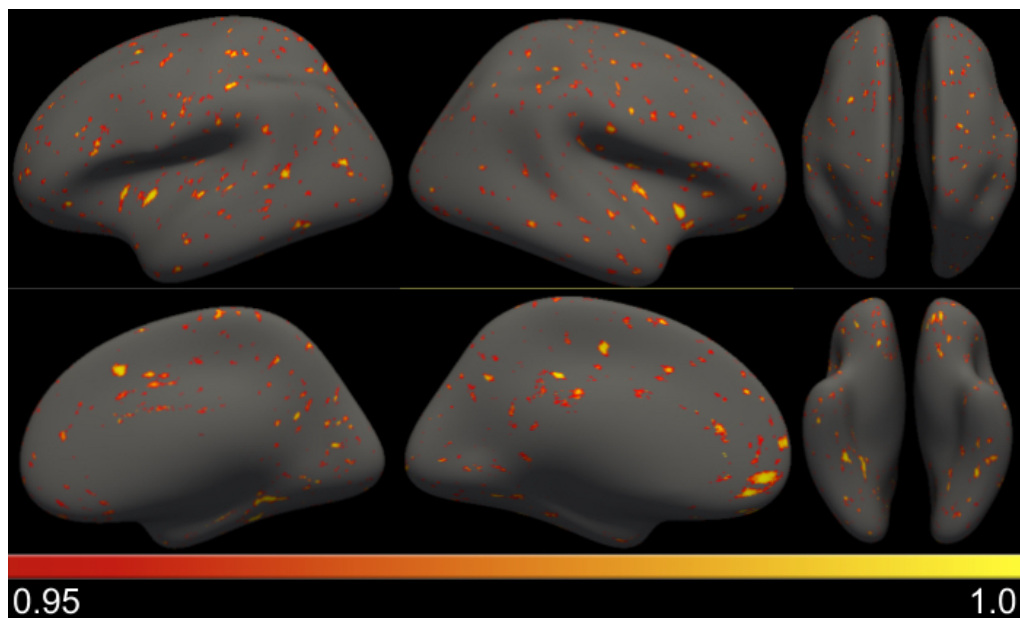


Figure 3.5: Statistical inference using cortical thickness data and GLM. For visualization purposes, statistical significance threshold is set to $p < 0.05$ uncorrected.

Using the tool, we could also study the *APOE* genotype effects in cortical thickness data. In Figure 3.5, we show the results on different surface views using the GLM model. In this case, even smaller effects are found being statistically relevant ($p < 0.05$) in small clusters across the brain, specially in regions such as the insular cortex and fusiform.

3.4.2 Interaction between *APOE* genotype and age in normal aging population.

In this second part, we investigate the interaction between *APOE* genotype and age on brain morphology. For this purpose, we model each *APOE* genotype separately to find their associated curves and generate a goodness-of-fit metric using the F-test. Statistical inference threshold is set to $p < 0.001$. We perform post-hoc analysis combining statistical maps into an RGB-map that sums up the results of all three *APOE*-genotype models: we place each model (NC, HE, HO) in each R, G, B channel, respectively. Volumetric and cortical thickness analyses were performed but no significant results were found with the latter.

In Figure 3.7, we show the RGB-map and the associated curves of regions corresponding to significant effect of age on brain morphology of homozygotes *APOE*-e4 carriers (see Section 3.4.1): right hippocampus, right caudate and right cerebellar crus. We present two different curve fitting models: using polynomial expansion of second order of the GLM on the left and B-splines GAM on the right. Clearly, relevant regions for the HO group show nonlinear relationship between age and voxel intensities. Statistical and *RGB-maps* present analogous results on polynomial expansion of GLM and GAM analysis. The right hippocampus and the right cerebellar crus follow a quadratic curve with age similar to GLM fitting. HO subjects show an earlier decreasing of GMv in both regions compared to NC and HE around their fifties with an initial volumetric increase on middle-aged individuals, more pronounced in the cerebellum, again replicating the results in [190]. On the other hand, GMv volume on the right caudate appears to decrease at the sixth decade for all *APOE* genotypes but decaying faster for HO subjects. Due to the non-quadratic behaviour of the right caudate, it appears to be better modeled with GAM, as shown in Figure 3.6.

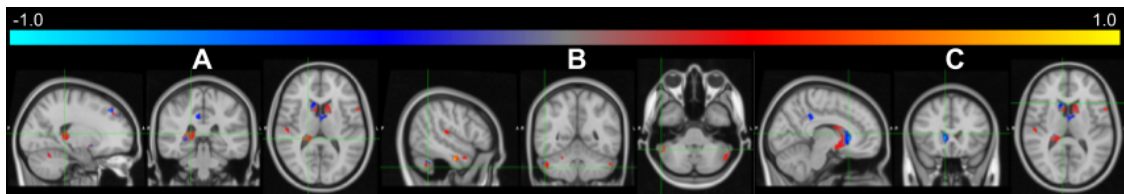


Figure 3.6: Differences between statistical maps of HE model using GLM and GAM at different brain ROIs: right hippocampus (A), right caudate (B) and right cerebellar crus (C). A positive (negative) value indicates that GAM (GLM) is statistically better using the f-test metric.

3.5 Conclusions

The standard neuroimaging software suites incorporate linear effects analysis in their pipelines. However, it has been shown that many factors describe a non-linear association with imaging modalities. NeAT arises as a tool for non-linear analysis of medical images. NeAT is a modular, flexible and user-friendly toolbox suited for the study of non-linear associations between imaging and non-imaging features. It provides several curve fitting methods for voxelwise and surface-based modeling and different metrics for statistical inference of the results. Several visualization features are available, such as an interactive GUI that shows statistical maps together with the resulting fitted curves.

Finally, post-hoc analysis functionalities such as model comparison (e.g: linear vs. non-linear) or a curve clustering algorithm that show similar fittings across the brain are available. Altogether defines a complementary tool to standard processing software that uncovers non-linear associations between neuroimaging and a set of factors (e.g: age, environmental factors, disease, genetics or demographics). We illustrate NeATs capabilities with two already published study cases for validation and show some features of the tool. Moreover, we stressed out the benefits of using NeAT, such as a comparison between linear and non-linear models.

In the first case of study, we analysed the nonlinear patterns of brain atrophy along the Alzheimer's continuum. We found that in many typical AD regions (e.g: hippocampus and precuneus) non-linear models better describe their evolution than linear methods and that using a linear model with polynomial basis expansion of the non-linear factors is limiting the analysis. However, due to the low-complexity of the linear model, it still becomes relevant at other parts of the brain. In the second case of study we replicate and extend the previous work from [190] analyzing the *APOE* – $\epsilon 4$ count effect on brain volume and surface for cognitively unimpaired individuals. Significant effects were found in already reported regions such that thalamus or hippocampus. Moreover, we use nonlinear methods to model the interaction between *APOE* – $\epsilon 4$ count and age, findings regions (e.g: caudates) where nonlinear effects were larger than linear ones.

Future work: the potential of NeAT is expected to expand as it will grow. At the short term, the expansion of the tool to ROI-based analysis is granted. Moreover, a longitudinal analysis module might be interesting due to the increasing number of cohorts with longitudinal follow-up visits. Seemingly, the integration of some imaging modalities (PET, DTI) should be easy and the implementation of others (fMRI) can be considered for future revisions of the tool. Other statistical methods, such as Partial Least Squares (PLS) or Canonical Correlation Analysis (CCA) can be incorporated for multivariate effects modeling. Finally, other curve fitting models (e.g: based on neural networks) can be designed and implemented.

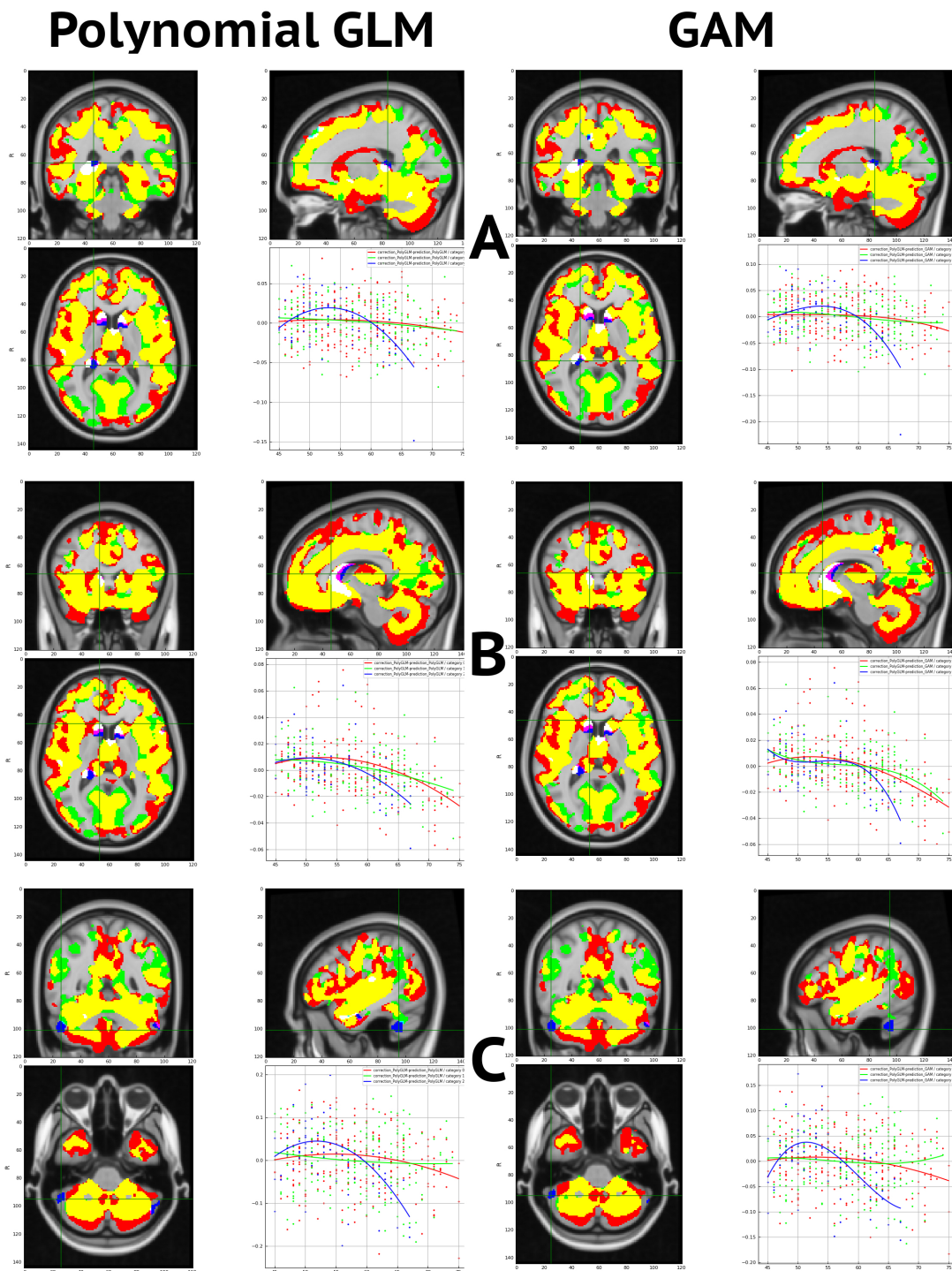


Figure 3.7: Interaction between age and the $APOE-\epsilon 4$ genotype using second order polynomial expansion of GLM (left) and B-splines GAM (right). Three different regions are shown at each row: (A) right hippocampus, (B) right caudate and (C) right cerebellar crus. Statistical analysis using F-test and uncorrected $p < 0.001$ threshold with cluster size of 100 voxels. Statistical comparison using the RGB map, where R corresponds to 0 copies of the allele, G to 1 copy and B to 2 copies. Other colors are any possible combination of them, meaning that are relevant for more than one $APOE$ genotype.

Brain structure in AD pathology

Alzheimer's disease is a multi-factorial disease that can be measured in-vivo using pathophysiological and topographical biomarkers. While the former reflect in-vivo AD pathology (amyloidosis and tauopathy), the latter provide evidence of brain damage and are useful to monitor disease progression and staging and support the diagnosis, even though they are not specific to AD. Magnetic resonance imaging (MRI) is an example of downstream topographical biomarker that provides information about brain tissue configuration and is used as a valid biomarker of neuronal loss. Relating both types of biomarkers is useful to study the whole Alzheimer's continuum and may provide better understanding of brain changes due to underlying AD pathology. In this chapter we study the relationship of cerebrospinal fluid (CSF) and MRI biomarkers in the whole continuum of AD with special focus on preclinical stages. We report the results of a univariate and a multivariate model of the brain concluding that different topographical patterns apply at different stages of the disease, specially at early stages.

4.1 Introduction

Human brains are constantly evolving throughout life, changing their neurobiological structure according to an uncountable number of factors ranging from genetics or hormonal to vascular factors [210]. Hence, we can define multiple interdependent processes occurring at the same time in each subject's brain. As an example, a *normal aging* process describes common changes in groups of subjects with similar age and can be characterized by subject's real age. Often, in elderly subjects, other processes related to *dementia* occur in parallel and eventually result in cognitive or memory decline. The etiology of most of those processes is still unknown but a large literature of promising research is being published, specially for the Alzheimer's disease (AD), the most common type of dementia. The work in [211] goes through several examples of recent findings and comments on the found evidences related to AD. In this line, cerebrospinal fluid (CSF) biomarkers, as a proxy measure of AD pathophysiological processes ($A\beta$ and tau deposition), show an acceptable sensitivity and specificity for diagnostic purposes and can be used as a measure of disease progression along the AD continuum [205]. Promising developments in PET imaging could provide us with more sensitive and specific pathophysiological biomarkers [60]. The incorporation of pathophysiological markers into diagnostic frameworks lead to the definition of the preclinical stage of AD (PC), where subjects present abnormal levels of $A\beta$ in the brain without cognitive decline or clinical symptoms [2].

On the other hand, brain cell integrity can be assessed by high-field MR imaging that provides detailed brain anatomical information, as explained in Section 2.2. MRI can be used to study neurodegenerative processes associated with brain development and aging [212], cognitive decline [213, 214] and amyloid positivity [166, 167, 215]. The inter-

relationship between AD pathophysiological markers, neurodegeneration and cognition is highly studied in the literature [216, 217]. Compared to pathophysiological markers (CSF and PET), MRI constitute a late but cheaper and non-invasive biomarker for disease diagnosis and progression. The goal of this chapter is to relate brain anatomical structure to underlying AD pathology to find characteristic patterns of brain structural variations due to AD. To this end, we consider two separate processes governing brain morphology along the AD continuum: a process driven by confounders that we call “cAD” (e.g. normal aging) and a process driven by disease pathology that we call “AD”. We acknowledge the presence of other processes (e.g. non-AD pathologies) specially in later stages of the disease, but these analyses are out of the scope of this work.

Structural brain changes are heterogeneous among subjects [218] and may not be uniform across the brain, specially if related to dementia. However, we hypothesize that these changes can be described by a small set of interdependent underlying processes that define common morphological patterns across all subjects. In this work, partial least squares (PLS) is used to jointly model the variation of brain morphology (measured using MRI biomarkers) and both confounders and AD pathology (measured through CSF biomarkers). We build two different models to find specific effects due to confounders factors and AD and couple them in a single brain structural model. Moreover, we constrain both models to be orthogonal to each other in order to disentangle confounders from AD processes on brain morphology.

We investigate two different modeling techniques of the brain: a univariate model that studies brain anatomical regions separately and a multivariate model that study all brain regions in a single model. For that purpose, we use PLS Correlation (PLSC) and PLS Regression (PLSR), respectively. We perform statistical inference on the models learned and derive a predictive model that uses MRI (cheap, non-invasive biomarker) to predict CSF biomarkers in the latter case.

4.1.1 Related work

Studies integrating both CSF and MRI information can be found in the literature in diagnostic [219] or prediction [220] contexts. The first work attempts to use cross-sectional MRI and CSF measurements to predict different clinical stages along the disease continuum, comparing the discriminative power of both biomarkers. In the latter, baseline measurements are used to predict future clinical change measured through Clinical Dementia Rating-sum of boxes (CDR-SB) and mini mental score examination (MMSE). Both works show that MRI outperforms CSF features in their tasks but they show complementary information when integrated in a joint model providing a boost in performance for classification and prediction. Other works study the impact of CSF biomarkers in brain integrity and clinical progression. Concretely, the work in [221] studies the relationship between CSF biomarkers with baseline values and longitudinal changes in brain morphology across different clinical AD categories. In a general linear model (GLM) framework, multiple comparisons between each CSF biomarker and each brain ROI are performed, concluding that CSF biomarkers can not explain group differences in baseline values and that they have a modest relationship to longitudinal brain atrophy. Finally, recent works from our group [215, 222] or others [166, 167] attempt to use MRI derived measurements to predict amyloid positivity in cognitively unimpaired subjects, with potential applications in clinical screening.

On the other hand, PLS modeling was first used in neuroimaging studies by McIntosh et al. in [223]. Since then, it has been widely used in neuroimaging analysis as indicated by the work in [224] that extensively reviews subsequent applications of PLS in the field. As an example, PLS was used to study the relationship between the cognitive profile and voxelwise MRI volumetric features in children and adolescents in [225]. In Alzheimer’s disease research, PLS has been used to relate MRI phenotypes with genetics, cognitive tests and disease categories. The work in [226] jointly models the variation of genotype and phenotype in Alzheimer’s disease, using single nucleotide polymorphisms (SNP) and 5 regional brain volumes as measurements, respectively. A white matter integrity analysis through diffusion tensor imaging (DTI) is analyzed in [227], where several diffusion parameters (axial diffusivity (AD), radial diffusivity (RD) and fractional anisotropy (FA)) are used to study group differences in white matter between different clinical AD stages. Similar to our work, the authors in [228] attempt to disentangle aging and disease processes using age and MMSE cognitive test as proxies, respectively. They show overlap between both processes using cross-sectional data while could disentangle both effects when using longitudinal data. However, MMSE may only characterize clinical AD stages and not the full continuum of the disease while using CSF biomarkers one could capture variability from a broader spectrum, specially at earlier stages, as we attempt to do in this chapter. Finally, PLS has also been used a feature extractor in a machine learning analysis. [229].

4.2 Methodology

4.2.1 Partial least squares

In general, partial least squares (PLS) methods (also known as *projection to latent structures*) analyze the relationship between two sets of measurements (X and Y) on the same observations. PLS methods also find a latent subspace built from linear combinations of the input measurements.

These methods were first introduced by Herman Wold in [230] for structural equation modeling (SEM) and have successfully been applied to chemometrics [231], econometrics [232] and psychology [233]. Two variants of this initial approach have been proposed in the literature: partial least squares regression (PLSR) and partial least squares correlation (PLSC). PLSR is used in the context of regression when the goal is to predict one measurement set (called response variables) from the other (called predictor variables) [234]. The model finds a latent subspace, derived from the predictor set, whose projections best predict the response variables. On the other hand, PLSC technique is used when the goal is to uncover shared information between the two sets of measurements [223]. PLSC finds two latent subspaces related to each of the measurement sets such that the latent projections have maximal covariance.

PLS techniques have been proven to work well when the number of observations is much smaller than the number of variables from the measurement sets as well as to cope well with correlated covariates.

4.2.1.1 Partial least squares Correlation

Partial Least Squares Correlation (PLSC) is a statistical method that describes the relationship between two sets of measurements, \mathbf{X} and \mathbf{Y} , on the same observations. This relationship is modeled as the covariance between both input spaces (\mathbf{X} , \mathbf{Y}) and the goal is to examine their shared information. The underlying assumption of PLSC modeling is that most of the joint variability between \mathbf{X} and \mathbf{Y} lies in a lower dimensional space, i.e. can be described by means of few latent patterns.

Let us assume we have N subjects with two sets of different measurements: K descriptive variables $\mathbf{X} \in \mathcal{R}^{N \times K}$ (e.g. brain structure using MRI) and M condition-related variables $\mathbf{Y} \in \mathcal{R}^{N \times M}$ (e.g. CSF biomarkers). Without loss of generality, we assume both variables to be mean-centered. Formally, PLSC is applied to identify two new sets of variables $\mathbf{T} \in \mathcal{R}^{N \times L}$, $\mathbf{U} \in \mathcal{R}^{N \times L}$, called latent variables, which are linear combinations of the original measurements \mathbf{X} and \mathbf{Y} , respectively. These new variables lie in lower and unobserved L -dimensional spaces derived by simultaneous decomposition of input variables trying to maximize their cross-covariance [235]. This idea is translated into finding the directions of maximum covariance between the original input spaces under the orthogonal constraint on the L projection vectors:

$$\begin{aligned} \text{maximize} \quad & \text{cov}(\mathbf{X}\mathbf{w}_l, \mathbf{Y}\mathbf{c}_l) = \text{cov}(\mathbf{t}_l, \mathbf{u}_l) \\ \text{s.t} \quad & \mathbf{w}_l^\top \mathbf{w}_{l'} = \delta(l - l'), \quad \mathbf{c}_l^\top \mathbf{c}_{l'} = \delta(l - l') \end{aligned} \quad (4.1)$$

where $\mathbf{w}_l \in \mathcal{R}^{K \times 1}$, $\mathbf{c}_l \in \mathcal{R}^{M \times 1}$ are the projection weight matrices from input to latent spaces. It follows from the properties of singular value decomposition (SVD) that $\mathbf{w}_l, \mathbf{c}_l$ are the left and right singular vectors of the covariance matrix $R = \mathbf{X}^\top \mathbf{Y}$ [236], respectively. Moreover, the covariance of the latent space at each dimension, i.e. $\text{cov}(\mathbf{t}_l, \mathbf{u}_l)$, is equal to the corresponding singular value. The final L -dimensional latent space is built by concatenating the corresponding latent variables:

$$\begin{aligned} \mathbf{T} &= [\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_{L-1}], \quad \mathbf{T} = \mathbf{X}\mathbf{W} \\ \mathbf{U} &= [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{L-1}], \quad \mathbf{U} = \mathbf{Y}\mathbf{C} \\ \text{where} \quad \mathbf{W} &= [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{L-1}], \quad \mathbf{C} = [\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{L-1}] \end{aligned} \quad (4.2)$$

4.2.1.2 Partial least squares Regression

Partial least squares regression (PLSR) is used when the goal is to predict one set of measurements, referred as response variables (\mathbf{Y}), from the other set called predictor variables (\mathbf{X}). PLSR is not only a multivariate method that can handle colinear predictors but can also predict multiple response variables and exploit the relation between them.

Formally, from a pool of N observations, PLSR finds a latent space ($\mathbf{T} \in \mathcal{R}^{N \times L}$) that both characterizes the predictor variables ($\mathbf{X} \in \mathcal{R}^{N \times K}$) and predict the response variables

($\mathbf{Y} \in \mathcal{R}^{N \times M}$):

$$\mathbf{X} = \mathbf{TP}^\top, \quad \hat{\mathbf{Y}} = \mathbf{XB} \quad (4.3)$$

where $\mathbf{P} \in \mathcal{R}^{K \times L}$ is the X-loading matrix and $\mathbf{B} \in \mathcal{R}^{K \times M}$ is the regression matrix whose computation is explained in the following. The number of predictors and response variables is K and M , respectively, while the latent subspace dimension is L .

Herein, we present the NIPALS solver solution [237] that attempts to iteratively solve the optimization problem by maximizing the covariance between input spaces (\mathbf{X}, \mathbf{Y}) at each iteration imposing orthogonality between consecutive latent dimensions ($\mathbf{t}_i \perp \mathbf{t}_{i+1}$). Other solvers (e.g. SIMPLS [238]) impose different orthogonality constraints between subspace dimensions resulting in different solutions. For a single latent dimension ($L = 1$) PLSR finds the underlying representations as linear combinations of input measurements $\mathbf{t}_0 = \mathbf{X}\mathbf{w}_0$, $\mathbf{u}_0 = \mathbf{Y}\mathbf{c}_0$ such that

$$\begin{aligned} \text{maximize} \quad & \text{cov}(\mathbf{X}\mathbf{w}_0, \mathbf{Y}\mathbf{c}_0) = \text{cov}(\mathbf{t}_0^\top \mathbf{u}_0) \\ \text{s.t.} \quad & \mathbf{w}_0^\top \mathbf{w}_0 = 1, \mathbf{c}_0^\top \mathbf{c}_0 = 1 \end{aligned} \quad (4.4)$$

Solving Equation 4.4 using Lagrange multipliers, weight vectors have the following analytic expression:

$$\mathbf{w}_0 = \frac{\mathbf{X}^\top \mathbf{u}_0}{\sqrt{\mathbf{u}_0^\top \mathbf{X}\mathbf{X}^\top \mathbf{u}_0}}, \quad \mathbf{c}_0 = \frac{\mathbf{Y}^\top \mathbf{t}_0}{\sqrt{\mathbf{t}_0^\top \mathbf{Y}\mathbf{Y}^\top \mathbf{t}_0}} \quad (4.5)$$

For a latent space of dimension $L > 1$, each subsequent latent factor is computed by iterating over the same optimization process (i.e. Equation 4.4 and Equation 4.5) and concatenating the results:

$$\begin{aligned} \mathbf{W} &= [\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_{L-1}] \in \mathcal{R}^{K \times L} \\ \mathbf{C} &= [\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{L-1}] \in \mathcal{R}^{M \times L} \\ \mathbf{T} &= [\mathbf{t}_0, \mathbf{t}_1, \dots, \mathbf{t}_{L-1}] \in \mathcal{R}^{N \times L} \\ \mathbf{U} &= [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{L-1}] \in \mathcal{R}^{N \times L} \end{aligned}$$

However, at each iteration, PLS uses deflated versions of input spaces (\mathbf{X} and \mathbf{Y}) from previous iterations as new predictor and response variables, forcing successive latent directions to be orthogonal to previous ones and hence, maximizing the input variance explained. The deflated version of \mathbf{X} and \mathbf{Y} at i -th iteration is:

$$\begin{aligned} \mathbf{X}_{i+1} &= \mathbf{X}_i - \mathbf{t}_i \mathbf{p}_i^\top = \left(\mathbf{I} - \frac{\mathbf{t}_i \cdot \mathbf{t}_i^\top}{\mathbf{t}_i^\top \mathbf{t}_i} \right) \mathbf{X}_i, \quad \mathbf{p}_i = \mathbf{X}_i^\top \mathbf{t}_i / (\mathbf{t}_i^\top \mathbf{t}_i) \\ \mathbf{Y}_{i+1} &= \mathbf{Y}_i - \mathbf{t}_i \mathbf{q}_i^\top = \left(\mathbf{I} - \frac{\mathbf{t}_i \cdot \mathbf{t}_i^\top}{\mathbf{t}_i^\top \mathbf{t}_i} \right) \mathbf{Y}_i, \quad \mathbf{q}_i = \mathbf{Y}_i^\top \mathbf{t}_i / (\mathbf{t}_i^\top \mathbf{t}_i) \end{aligned} \quad (4.6)$$

with $i = 0, \dots, L - 2$, $\mathbf{X}_0 = \mathbf{X}$ and $\mathbf{Y}_0 = \mathbf{Y}$. Due to continuous deflation of input spaces, weight matrices (\mathbf{W}, \mathbf{C}) do not directly relate input (\mathbf{X}, \mathbf{Y}) and latent (\mathbf{T}, \mathbf{U}) spaces. Accounting for that, rotation matrices, $\mathbf{R}_x, \mathbf{R}_y$, are defined:

$$\begin{aligned}\mathbf{R}_x &= \mathbf{W} \cdot (\mathbf{P}^\top \cdot \mathbf{W})^{-1}, & \mathbf{T} &= \mathbf{X} \cdot \mathbf{R}_x, \\ \mathbf{R}_y &= \mathbf{C} \cdot (\mathbf{Q}^\top \cdot \mathbf{C})^{-1}, & \mathbf{U} &= \mathbf{Y} \cdot \mathbf{R}_y\end{aligned}\quad (4.7)$$

where $\mathbf{P} = [\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{L-1}] \in \mathcal{R}^{K \times L}$, $\mathbf{Q} = [\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_{L-1}] \in \mathcal{R}^{M \times L}$ are the so-called loadings. Finally, from Eq. 4.6 arise the initial Eq. 4.3 for prediction:

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{R}_x\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{B} + \mathbf{E}\quad (4.8)$$

where $\mathbf{B} = \mathbf{R}_x\mathbf{Q}^\top$ is the regression matrix and $\mathbf{E} \in \mathcal{R}^{N \times M}$ models the regression error.

4.2.2 Partial least squares orthogonalization

A common assumption in neuroimaging studies is that the object of study (e.g: brain morphology) is affected at the same time by the condition of interest (e.g: dementia) and confounding variables (e.g: age, genetics). Hence, we need to control for confounding variables in any neuroimaging analysis in order to find meaningful results. The standard solution is to regress-out the unwanted factors on the condition of interest. In a PLS framework, it can be done estimating two separate models for confounders (M^C) and the variable of interest (M^I) and imposing orthogonality between both models. Each model is characterized by using two different Y-space variables ($\mathbf{Y}^C, \mathbf{Y}^I$) and the same X-space. Similarly to the work in [227], we introduce the orthogonality constraint in the optimization process forcing the associated latent subspace ($\mathbf{T}^C, \mathbf{T}^I$) to be orthogonal. Moreover, we add a coupling step in order to account for X-space related variance explained by the confounder variables \mathbf{Y}^C .

Firstly, the confounder model, M^C , is estimated using the data tuple $(\mathbf{X}, \mathbf{Y}^C)$ as predictor and response variables, respectively. The associated subspaces ($\mathbf{T}^C, \mathbf{U}^C$) are found by solving the regular expressions in Equation 4.1 and Equation 4.4 for an L^C -dimensional subspace with weights \mathbf{w}_i^C and \mathbf{c}_i^C for $i = 0 : L^C - 1$.

Secondly, the model of interest, M^I , is estimated using the data tuple $(\bar{\mathbf{X}}, \mathbf{Y}^I)$. To account for the variance explained by confounding variables, we define the new predictor variable $\bar{\mathbf{X}} = \mathbf{X} - \mathbf{T}^C \cdot (\mathbf{P}^C)^\top$ where we subtract the measurement variance explained by M^C . Moreover, we impose orthogonality between the estimated weights (\mathbf{w}_1) and the projection vectors of the confounder model (\mathbf{r}_1^C). The optimization process for M^I model is as follows:

$$\begin{aligned}\text{maximize} & \quad \text{cov}(\bar{\mathbf{X}}\mathbf{w}_1, \mathbf{Y}^I\mathbf{c}_1) \\ \text{s.t} & \quad \mathbf{w}_1^\top \mathbf{w}_{l'} = \delta(l - l'), \quad \mathbf{c}_1^\top \mathbf{c}_{l'} = \delta(l - l'), \quad \mathbf{w}_1^\top \mathbf{r}_1^C = \mathbf{0} \\ & \quad \forall i = 0, \dots, L^C - 1\end{aligned}\quad (4.9)$$

where for PLSC models r_i^C are the columns of the weight matrix \mathbf{W}^C and the deflation process between models to estimate $\bar{\mathbf{X}}$ ensures the orthogonality between subspaces. In PLSR models r_i^C are the columns of the rotation matrix R_X^C and $\mathbf{w}_l = \mathbf{w}_0$, $\mathbf{c}_l = \mathbf{c}_0$. For an L^I -dimensional subspace, the use of Lagrangian multipliers in the iterative solution results in a closed form solution for weight projections:

$$\begin{aligned} \mathbf{w}_1 &= \bar{\mathbf{X}}_1^\top \cdot \mathbf{u}_1 - \sum_{i=0}^{L^C-1} \frac{(\mathbf{r}_i^C)^\top \cdot \bar{\mathbf{X}}_1^\top \mathbf{u}_1}{(\mathbf{r}_i^C)^\top \cdot \mathbf{r}_i^C} \mathbf{r}_i^C \Rightarrow \mathbf{w}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} \\ \mathbf{c}_1 &= (\mathbf{Y}_1^I)^\top \cdot \mathbf{t}_1 \Rightarrow \mathbf{c}_1 = \frac{\mathbf{c}_1}{\|\mathbf{c}_1\|} \quad l = 0 : L^I - 1 \end{aligned} \quad (4.10)$$

where $\bar{\mathbf{X}}_l$ and \mathbf{Y}_l^I are the deflated versions of $\bar{\mathbf{X}}_0$ and \mathbf{Y}_0^I at l th iteration.

Finally, the full PLS model (M , L -dimensional) is built by concatenating latent scores and rotation matrices from both models: $\mathbf{T} = \text{concat}(\mathbf{t}^C_l, \mathbf{t}^I_r)$ and $\mathbf{R} = \text{concat}(\mathbf{R}_{Xl}^C, \mathbf{R}_{Xr}^I)$ where $l = 0, \dots, L^C - 1$ and $r = 0, \dots, L^I - 1$.

4.2.3 Statistical inference

The outcome measures for statistical inference of the estimated models are the *effect size* (ρ) and the *effect type* (ν_X, ν_Y). The effect size is a quantitative measure of the magnitude of a certain phenomenon while the effect type is defined over multivariate phenomena as the vector of proportions indicating relative effect sizes of each parameter. We use different definitions of both analytical metrics in PLSC and PLSR analyses.

4.2.3.1 Inference metrics for PLSC

Inherently, PLSC models have estimated two different latent subspaces each one related to X and Y input spaces, respectively. Hence, a good definition for the effect size is the covariance between both estimated subspaces at each dimension, while the effect types are defined as the vectors of projection to the associated subspace

$$\rho_l = \frac{1}{N-1} \mathbf{t}_l^\top \mathbf{u}_l, \quad \nu_{Xl} = \mathbf{w}_l, \quad \nu_{Yl} = \mathbf{c}_l \quad \forall l = 0, \dots, L-1 \quad (4.11)$$

4.2.3.2 Inference metrics for PLSR

On the other hand, even though PLSR models use projections from both X and Y input spaces in the learning process, only one associated latent subspace is estimated and used for prediction. The latent subspace is related to predictor variables (X) and it is characterized by their associated scores (\mathbf{T}) and projections (\mathbf{R}). Hence, we use S additional variables of interest (q_s) to evaluate their correlation with the estimated latent subspace and use it as the effect size. Accordingly, the effect type is the vector of projections from the input space to the associated latent subspace.

$$\begin{aligned}\rho_{l(s)} &= \frac{\text{cov}(\mathbf{t}_l, \mathbf{q}_s)}{\sigma_{\mathbf{t}_l} \cdot \sigma_{\mathbf{q}_s}} = \frac{\mathbf{t}_l^\top \cdot \mathbf{q}_s}{\sigma_{\mathbf{t}_l} \cdot \sigma_{\mathbf{q}_s}}, & \forall l = 0, \dots, L-1 \\ \nu_l &= r_l & \forall s = 0, \dots, S-1\end{aligned}\quad (4.12)$$

4.2.3.3 Permutation testing

Statistical hypothesis testing over the outcome measures of PLS models require to formulate a null hypothesis with an associated sampling distribution, referred as the null distribution. The null hypothesis states that the effect size is small/non-significant, i.e. there is no relationship between the two sets of measurements used to compute the effect size. The associated null-distribution can be estimated in various ways and in this work we use the non-parametric permutation testing framework [239]. It consists of randomly permuting subject indices in one measurement to break the initial relationship between both input sets and generating a new sample of unrelated variables. This process is repeated N_{perm} times, and for each permutation ($\pi(i)$, $\forall i=0, \dots, N_{\text{perm}} - 1$) a new PLS model is computed along with the associated effect size $\rho_l^{\pi(i)}$ at each dimension.

$$\rho_l^{\pi(i)} = \frac{1}{N-1} \mathbf{t}_{l, \pi(i)}^\top \cdot \mathbf{u}_l \quad (4.13)$$

$$(4.14)$$

where $\pi(i)$ is the i -th permutation without replacement of subject indices. The null distribution for the effect size is empirically built using $\rho_l^{\pi(i)}$. Statistical significance level (p-value) of the observed effect size at each dimension (ρ_l) is determined by the ratio of permutations that result in higher effect size.

$$p\text{-value}(\rho_l) = \frac{1}{N_{\text{perm}}} \|\rho_l^{\pi(i)} > \rho_l\|_0 \quad (4.15)$$

where $\|\cdot\|_0$ is the 0-norm operator, that counts the number of non-zero elements of a vector.

An important parameter for inference using permutation testing is the number of permutations (N_{perm}) to estimate the null distribution. Larger N_{perm} provides better estimations but increases the computational cost. Hence, there is a trade-off between computational complexity and the precision (P) of the $p\text{-value}(\rho_k)$ estimation. It is known that the Monte Carlo approximation of the p -value has a standard deviation of $\sqrt{\frac{p(1-p)}{N_{\text{perm}}}}$ involving the true real value of p . Since p is unknown, the authors in [240] suggest to use the upper bound $\frac{1}{\sqrt{N_{\text{perm}}}}$ and model $p\text{-value}(\rho_l)$ as approximately Gaussian with standard deviation P, referred to as the precision of the estimate. Then, for a desired precision P, the minimum number of permutations is

$$N_{\text{perm}} \geq \frac{1}{4 \cdot P^2} \quad (4.16)$$

4.3 Univariate brain structural effects

Magnetic resonance imaging (MRI) is useful for measuring brain soft-tissue (grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF)) and provides a good anatomical description of the brain. On the other hand, CSF biomarkers describe the biological state or condition of the brain and are used for staging the AD pathophysiological continuum. Jointly analyzing both sources of information may help to describe AD-related underlying processes that occur in the brain. In this section, we are interested in comparing regional effects of brain pathology in brain volume. For that purpose, we use the PLSC framework introduced in Section 4.2.1.1.

4.3.1 Material and methods

4.3.1.1 Data

In our experiments we use the publicly available dataset from the Alzheimer’s Disease Neuroimaging Initiative (ADNI, [241]) with imaging data preprocessed using FreeSurfer [242]. According to [1], we split the Alzheimer’s continuum ($N = 801$) in three stages: a total of $N_{\text{CU}} = 321$ cognitively unimpaired (CU) subjects, $N_{\text{MCI}} = 332$ subjects with mild cognitive impairment (MCI) and $N_{\text{AD-dementia}} = 148$ subjects diagnosed with dementia due to AD (AD-dementia). For symptomatic stages (MCI and AD-dementia), only amyloid positive subjects (CSF $A\beta < 192$ pg/mL, [170]) are considered. To study the presymptomatic stage of AD, we use subthresholds and build a separate data sample splitting the CU stage in three groups (T_0, T_1, T_2) based on CSF $A\beta$ tertiles: $A\beta_{T_0} \in (230, 295]$, $A\beta_{T_1} \in (175, 230]$, $A\beta_{T_2} \in (88, 175]$ with 104, 109, 108 subjects respectively. All subjects have cortical and subcortical grey matter volumetric information available for each of the $K = 84$ brain ROI [81, 243], age, sex and CSF biomarkers ($A\beta$, phosphorylated tau (p-tau), total tau protein (t-tau)) as condition-related variables. Both sets of features are mean-centered and scaled by the control (CU) group standard deviation.

4.3.1.2 Model definition

In neurodegenerative disease studies, it is normally assumed that there coexist many processes in the brain affecting its morphology and that they can be grossly classified into disease-related (variables of interest) and non-disease-related (confounders). Hence, according to the nomenclature introduced in Section 4.2.2, we define two different models: a confounder model ($M^C = M^{\text{cAD}}$) estimated using age and sex as response variables ($\mathbf{Y}^{\text{cAD}} \in \mathbb{R}^{N \times 2}$) and a pathological model ($M^I = M^{\text{AD}}$) that uses CSF biomarkers as response variables ($\mathbf{Y}^{\text{AD}} \in \mathbb{R}^{N \times 3}$). As predictor variables we use ROI-based volumetric features ($\mathbf{X}, \bar{\mathbf{X}} \in \mathbb{R}^{N \times K}$).

We set $L^{\text{cAD}} = 1$ to ease the interpretability yielding a multivariate model that describes the joint variability of regional brain volume and age and sex. Therefore the effect size and effect type read as follows:

$$\rho^{\text{cAD}} = \frac{1}{N-1} \mathbf{t}_0^\top \cdot \mathbf{u}_0, \quad \nu_X^{\text{cAD}} = \mathbf{w}_0, \quad \nu_Y^{\text{cAD}} = \mathbf{c}_0 \quad (4.17)$$

For the pathological model, M^{AD} , we create K univariate submodels (M_k^{AD} , $\forall k =$

$0, \dots, K - 1$) each one using the deflated version of regional volumetric features for the k -th ROI ($\bar{\mathbf{X}}_k$) as predictor variables. For each submodel we set $L_k^{AD} = 1$ yielding univariate structural model that describes regional effects of multivariate pathological effects. Therefore we define an effect size and effect type for each ROI:

$$\rho_k^{AD} = \frac{1}{N - 1} \mathbf{t}_{0_k}^\top \cdot \mathbf{u}_{0_k}, \quad \nu_{Y_k}^{AD} = \mathbf{c}_{0_k}, \quad \text{where } k = 0, \dots, K - 1 \quad (4.18)$$

The statistical significance of the pathological model is computed as stated in Section 4.2.3.3. The null distribution of the effect size for each submodel can be computed by permuting samples from the X-space:

$$\rho_k^{\pi(i)} = \frac{1}{N - 1} \mathbf{t}_{0, \pi(i)_k}^\top \cdot \mathbf{u}_{0_k} \quad (4.19)$$

For all experiments, we use a total of $N_{perm} = 10000$ permutations, corresponding to a precision of the p -value estimate of $P=0.005$. Statistical significance is set to p -value < 0.05 (uncorrected).

4.3.1.3 Clustering

Unsupervised clustering techniques allow to uncover common characteristics between sets of data without the need of specific labeling processes [244]. In this work we are interested in grouping brain ROI with similar CSF biomarker effects on their average volume. We use the k-means algorithm [244] on regional effect types $\nu_{Y_k}^{AD}$. We aim at finding a total of C clusters and their associated centroid in order to find the main modes (latent patterns) of volumetric variation. We choose $C = 4$ to study the full AD continuum and $C = 2$ for the analysis of the preclinical stage of AD.

4.3.2 Results

In this section we analyze the relationship between brain morphology and both confounders and markers of AD pathology along the Alzheimer's continuum. Using the PLSC framework presented in Section 4.2.1.1, we find morphological patterns associated with CSF biomarkers along the Alzheimer's continuum. We further split the data sample into clinical categories and perform the same analysis (cognitively unimpaired (CU) subjects, mild cognitive impairment (MCI) and dementia due to AD (AD-dementia)) to investigate specific effect sizes on different disease stages. Finally, we specifically study the presymptomatic stage of AD by splitting CU subjects into three tertiles (T0, T1, T2) regarding subject's CSF $A\beta$ burden.

4.3.2.1 Study of age association with brain morphometric features in normal aging

We first estimate the M^{cAD} model to regress-out the effect size of confounders (age and sex) on brain morphology in posterior analysis. In Figure 4.1 we show the confounders' effect type representing the regional contribution of age and sex variability to brain morphology. M^{cAD} model involves reduced whole brain volume with increasing age except for the choroid-plexus and the anterior cingulate. Regions that show higher decrease in volume with age are found in the temporal lobe, specially the hippocampus, while sex effect size is rather low in the whole brain.

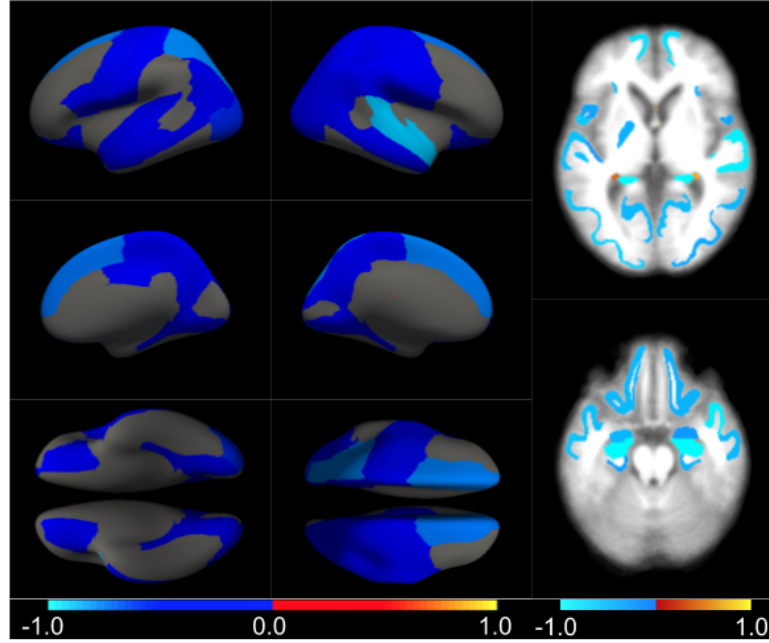


Figure 4.1: Brain morphometric effect type v_X^{cAD} related to confounders (age and sex) using M^{cAD} model. Color-code represents negative (blue) and positive (red) contributions on explaining confounder variability related to decreased and increased ROI volumes, respectively

4.3.2.2 Study of CSF biomarkers association with brain morphometric features along Alzheimer's continuum

AD pathological effects are thought to be spread non-uniformly across the brain [245] and may be only described by a small set of patterns along the disease continuum. To capture these characteristic patterns, we use the corrected M^{AD} model for regional analysis of CSF biomarkers association with brain volume at separate ROIs. CSF biomarkers pairwise relationship is shown in Figure 4.2 where CSF $A\beta$ show clear nonlinear aspect along the continuum with phosphorylated and total tau while tau proteins are shown to be highly correlated.

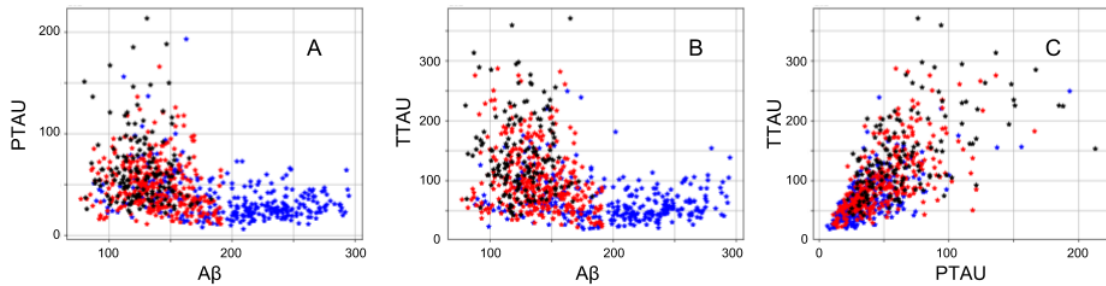


Figure 4.2: Pairwise CSF biomarker relationship along the Alzheimer's continuum: (A) CSF $A\beta$ vs. p-tau, (B) CSF $A\beta$ vs. t-tau and (C) CSF p-tau vs. t-tau. Different colors refer to different clinical categories: CU (blue), MCI (red) and AD-dementia (black)

All effect types are normalized to unit norm ($\|\nu_{Y_k}^{AD}\|_2 = 1$) and each value corresponds

to the relative contribution of each CSF biomarker on explaining the variance of ROI volumes. In Figure 4.3, we show the condition-related effect type ($\nu_{Y_k}^{AD}$) in statistically significant regions ($p < 0.05$). Each effect type value is splitted in different subfigures. A complete list of relevant regional AD effect sizes is found in Supplementary Table A.1.

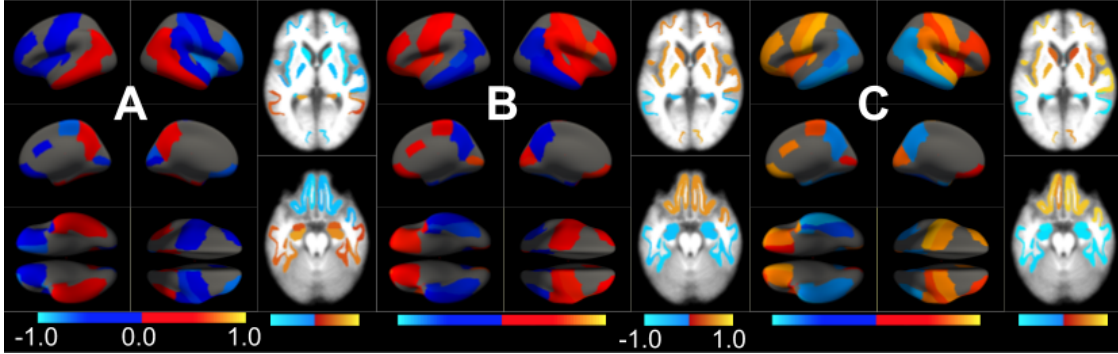


Figure 4.3: Effect type ($\nu_{Y_k}^{AD}$) of AD pathology on brain morphology for subjects along the Alzheimer's continuum. Each figure represents the contribution of CSF biomarkers in explaining ROI volume variability: (A) CSF A β , (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

We fit the k-means clustering algorithm to find a small set of representative CSF-related patterns that group all effect types ($\nu_{Y_k}^{AD}$) across the brain and along disease stages. It effectively results in four different clusters with the corresponding centroid representing patterns of AD pathology on brain morphology. Hence, we can define 4 different underlying processes governing brain morphology. In Table 4.1, we provide a list of the relevant ROIs associated to each process. CSF t-tau appears to explain most of the variability in many brain ROIs (clusters 0 and 1). Amyloid plaque deposition appears to favour the presence of tau protein in several temporal regions (e.g. hippocampus, middle temporal or amygdala) and reduce their volume (cluster 1). In contrast, several regions such as pallidum, precentral, lateral orbitofrontal or precentral appear to follow the opposite pattern once discounted the aging effect (cluster 0). On the other hand, CSF A β levels appear to highly affect regions such as insula, caudate and choroid plexus showing reduced volume with amyloid deposition once corrected by the aging process (clusters 3 and 4).

4.3.2.3 Study of CSF biomarkers association with brain morphometric features in different clinical AD stages.

In a second type of analysis, we consider that AD effects might be different along the disease continuum, as suggested by [245]. Hence, we repeat the analysis in separate clinical diagnostic groups (CU, MCI and AD-dementia) and provide a post-hoc comparison. At each stage, we use the corrected M^{AD} model to separately analyze regional CSF biomarkers effect size on each brain ROI volume. The statistically significant regions ($p < 0.05$) are listed in Supplementary Table A.2. The effect type is shown in Figures 4.4, 4.5 and 4.6 for CU, MCI and AD-dementia respectively. Each effect type value is displayed in separate subfigures. All effect types are normalized to unit norm ($\|\nu_{Y_k}^{AD}\|_2 = 1$) and each value corresponds to the relative contribution of each CSF biomarker on explaining the variance of each ROI volume.

Effect types in each ROI can be effectively clustered into four different characteristic

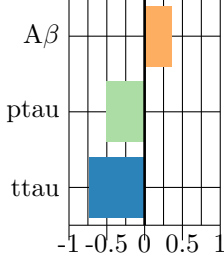
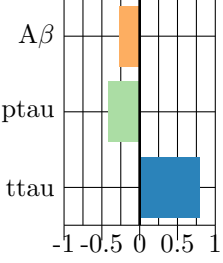
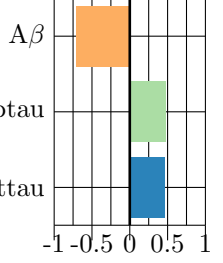
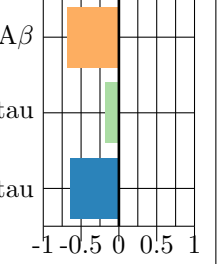
	Centroid 0	Centroid 1	Centroid 2	Centroid 3
				
ADc	Pallidum R Pallidum L Precentral L Precentral R Lateral Orbitofront. L Lateral Orbitofront. R Frontal Pole L Frontal Pole R Superior Temporal R Supramarginal R Caudal Ant. Cingul. R Medial Orbitofront. R ParsOrbitalis L Pericalcarine L Postcentral L Rostral Mid Frontal L Temporal Pole L Transverse Temporal L Accumbens Area L	Precuneus R Precuneus L Amygdala R Amygdala L Bankssts R Bankssts L Entorhinal R Entorhinal L Fusiform R Fusiform L Hippocampus R Hippocampus L Inferior Temporal R Inferior Temporal L Middle Temporal R Middle Temporal L Inferior Parietal R Inferior Parietal L	Insula R Insula L Caudate R Caudate L Paracentral R ParsOpercularis R Pericalcarine R Postcentral R Rostral Mid Front. R Cuneus L Medial Orbitofront. L	Choroid Plexus R Choroid Plexus L

Table 4.1: Unsupervised cluster analysis of CSF effect types on brain morphology along the Alzheimer’s continuum. Statistically relevant ($p < 0.05$) brain ROIs are clustered into four separate regions defined by their respective centroid (first row). Centroids show the relative contribution to the latent space of each CSF measure: CSF $A\beta$, CSF p-tau and CSF t-tau. A positive (negative) value corresponds to increased (decreased) ROI volume with each measure.

patterns. In Table 4.2, we provide a list of relevant ROIs associated to the four centroids describing each cluster. Among the four clusters, we find that increasing AD pathology (i.e. decreasing CSF $A\beta$ and increasing CSF t-tau/p-tau burden) are related to decreased ROI volume in clusters 0 and 2 and to increased ROI volume in clusters 1 and 3. Clusters 0 and 1 group regions whose variance is related to CSF t-tau/p-tau proteins variation while clusters 2 and 3 appeared to be described by CSF $A\beta$ load. Along the Alzheimer’s continuum, the choroid plexus appears to be relevant at each AD stage showing decreased ROI volume associated with an increment of CSF t-tau/p-tau values (cluster 0). The same pattern is followed by subcortical regions like amygdala, pallidum or putamen as well as the inferior temporal cortex at CU stage, the middle temporal at MCI stage or the inferior parietal region at AD-dementia. Interestingly, at early stages (CU) both caudate regions show increased volume with decreased CSF $A\beta$ values (cluster 3). Latent patterns 1 and 2 group most of the relevant regions at MCI stage. In cluster 1, subcortical regions such as the pallidum or thalamus as well as cortical regions like the precentral show increased mean ROI volumes with increasing CSF t-tau/p-tau values (cluster 1).

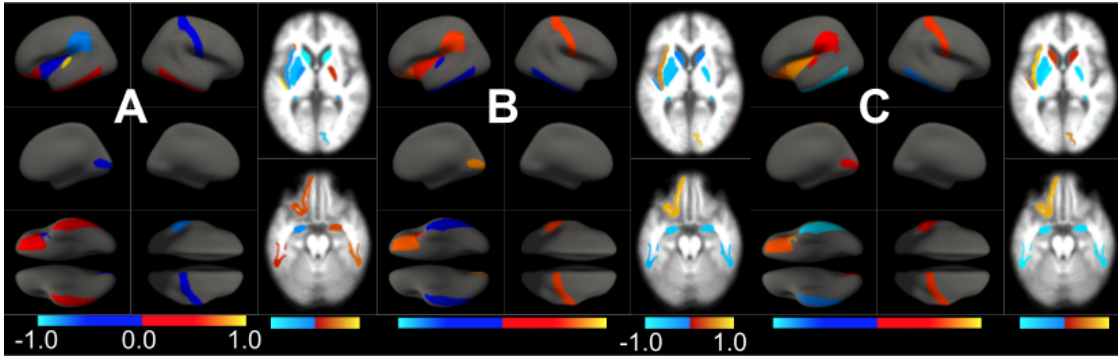


Figure 4.4: Effect type ($\nu_{Y_k}^{AD}$) of AD pathology on brain morphology for subjects in the CU stage. Each figure represents the contribution of CSF biomarkers in explaining ROI volume variability: (A) CSF $A\beta$, (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

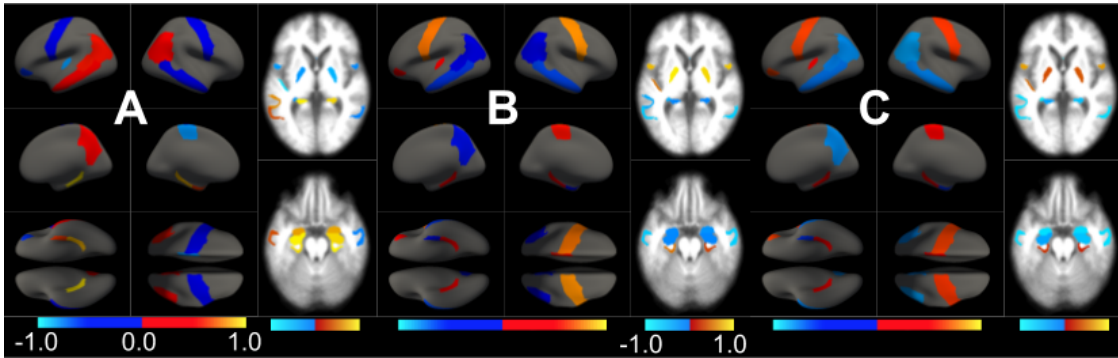


Figure 4.5: Effect type ($\nu_{Y_k}^{AD}$) of AD pathology on brain morphology for subjects in the MCI stage. Each figure represents the contribution of CSF biomarkers in explaining ROI volume variability: (A) CSF $A\beta$, (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

Typical AD-related regions within the temporal lobe such as the hippocampus, parahippocampus or the amygdala show decreased volumetric features with decreasing CSF $A\beta$ value (cluster 2). Other relevant regions for MCI stage are inferior parietal, superior temporal sulcus, precuneus or entorhinal showing decreased ROI volume associated to AD progression, either to increasing values of CSF p-tau/t-tau (cluster 0) or decreasing values of CSF $A\beta$ (cluster 2). Finally, in the AD-dementia stage, brain morphology can be explained mostly by CSF t-tau values. Already mentioned regions like the inferior parietal, choroid plexus, amygdala or entorhinal cortex show decreased mean ROI volume with increasing CSF t-tau/p-tau values (cluster 0), while subcortical pallidum shows the opposite effect.

4.3.2.4 Study of CSF biomarkers association with brain morphometric features at the asymptomatic stage of AD.

An interesting stage for prevention trials and clinical recruitment is the presymptomatic stage of AD, involving only cognitively unimpaired (CU) subjects. We study the use of subthresholds to find specific effect sizes at different phases of this presymptomatic

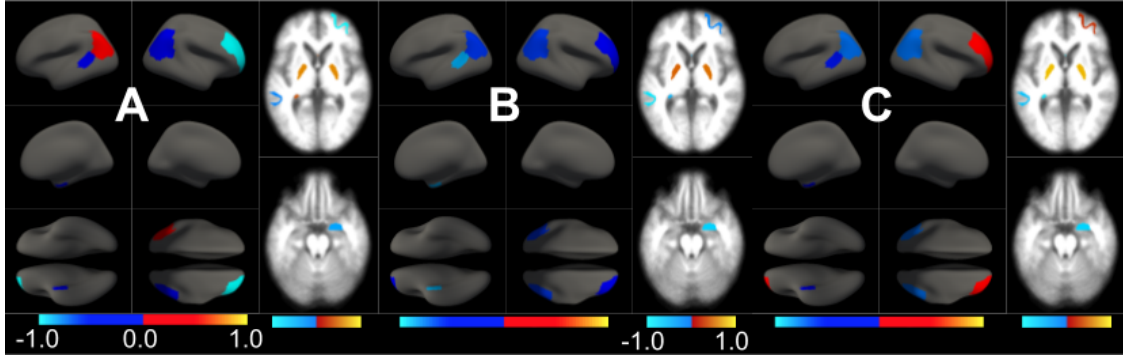


Figure 4.6: Effect type ($\nu_{Y_k}^{AD}$) of AD pathology on brain morphology for subjects in the AD-dementia stage. Each figure represents the contribution of CSF biomarkers in explaining ROI volume variability: (A) CSF $A\beta$, (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

stage. For this purpose, we split the sample into 3 balanced subsets divided using CSF $A\beta$ tertiles at thresholds 230 and 175 mg/mL. This categorization resulted in two groups (T0, T2) far above and below the standard amyloid-positive threshold at 192 mg/ml (used as a hallmark for preclinical AD stages), respectively, and a third group (T1) around it. Following the results in [246], nonlinear pairwise relationship between CSF $A\beta$ and tau proteins is shown in Figure 4.7, highlighting the three different groups.

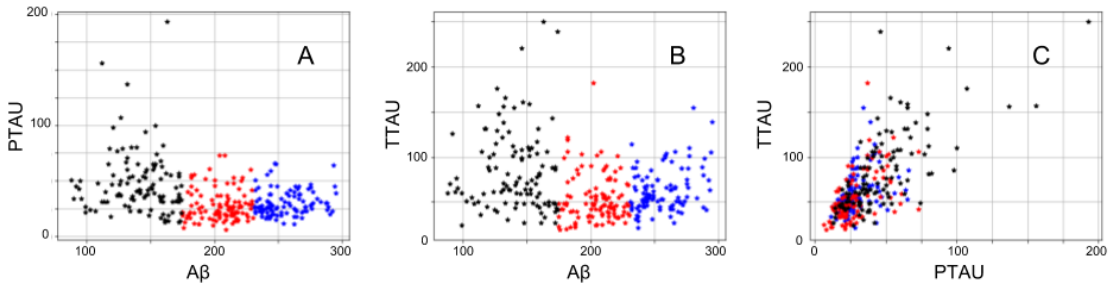


Figure 4.7: Pairwise CSF biomarker relationship in asymptomatic subjects: (A) CSF $A\beta$ vs. p-tau, (B) CSF $A\beta$ vs. t-tau and (C) CSF p-tau vs. t-tau. Different colors refer to different categories: T0 (blue), T1 (red) and T2 (black)

Following the same methodology as previous analysis, we use PLSC to define the M^{AD} model that relates corrected ROI-based volumetric features and CSF biomarkers. We perform statistical inference independently for each tertile and list the results in Supplementary Table A.3. CSF-related effect types of the statistically significant ROIs ($p < 0.05$) are shown in Figure 4.8 (T0), Figure 4.9 (T1) and Figure 4.10 (T2), where each subfigure is related to each CSF biomarker. Relevant regions are mostly found in temporal and parietal lobes. All patterns are normalized to unit norm ($\|\nu_{Y_k}^{AD}\|_2 = 1$) and correspond to the relative contribution of each CSF biomarker on explaining the variance of each ROI volume.

We use the clustering algorithm to find similar effect types at each ROI throughout all preclinical tertiles resulting in two different clusters (cluster 0 and cluster 1) associated (positively and negatively) to CSF p-tau/t-tau. In Table 4.3, we provide a list of relevant ROIs associated to the two centroids describing each cluster. The choroid plexus is the

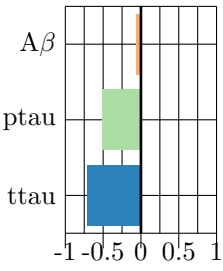
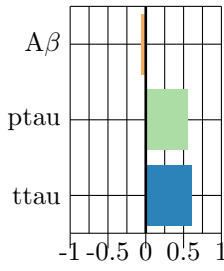
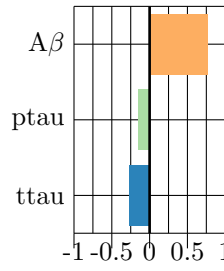
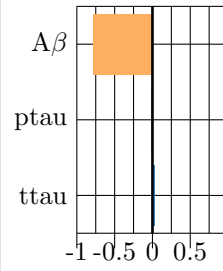
	Centroid 0	Centroid 1	Centroid 2	Centroid 3
				
CU	Pallidum R Pallidum L Amygdala L Amygdala R ChoroidPlexus L ChoroidPlexus R InfTemporal L InfTemporal R Putamen L	Pericalcarine R Postcentral R Insula L LatOrbitofront L	AccArea L TransvTempl L	Caudate L Caudate R Supramarg L
MCI	Precuneus R ChoroidPlexus L MidTemporal L MidTemporal R InferiorParietal R	Pallidum R Pallidum L Precentral R Precentral L Thalamus R Thalamus L ParsOrbitalis L	Parahipp. R Parahipp. L Amygdala R Amygdala L Hippocampus R Hippocampus L Bankssts L Entorhinal L InferiorParietal L	Paracentral L TransvTemp L
AD-dem	Bankssts L ChoroidPlexus L InferiorParietal L InferiorParietal R Amygdala R Entorhinal R	Pallidum L Pallidum R		RostMidFront R

Table 4.2: Unsupervised cluster analysis of CSF effect types on brain morphology along the Alzheimer’s continuum stratified by cognitive stage. Statistically relevant ($p < 0.05$) brain ROIs are clustered into four separate regions defined by their respective centroid (first row). Centroids show the relative contribution to the latent space of each CSF measure: CSF $A\beta$, CSF p-tau and CSF t-tau. A positive (negative) value corresponds to increased (decreased) ROI volume with each measure.

main exponent of cluster 0, showing decreased mean ROI with increasing CSF t-tau/p-tau values while examples of regions associated with cluster 1 are insula in T0, supramarginal and accumbens in T1 or pericalcarine in T2, showing increased mean ROI volume with increasing AD pathophysiological markers. Interestingly, this analysis of presymptomatic stages of AD using subthresholds shows that subcortical regions from the basal ganglia (amygdala or pallidum as well as the caudates), normally associated with Alzheimer’s disease, are not characteristic of any specific tertile and only appear to be relevant when studying the CU stage as a whole. On the other hand, the choroid plexus appears to be relevant at later tertiles (T1,T2). In Supplementary Table A.3, we provide the summary of regions across the brain related to each cluster for different preclinical tertiles.

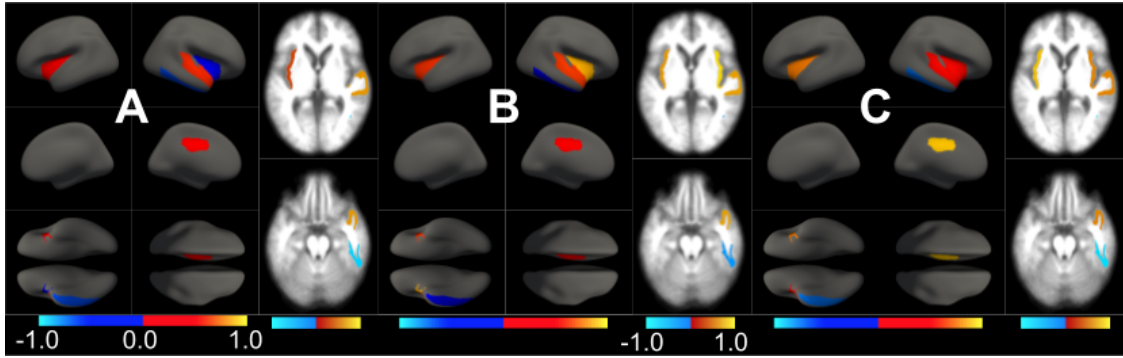


Figure 4.8: Effect type (ν_{Yk}^{AD}) of AD pathology on brain morphology for subjects in the T0 tertile. Each figure represents the contribution CSF biomarkers in explaining ROI volume variability: (A) CSF A β , (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

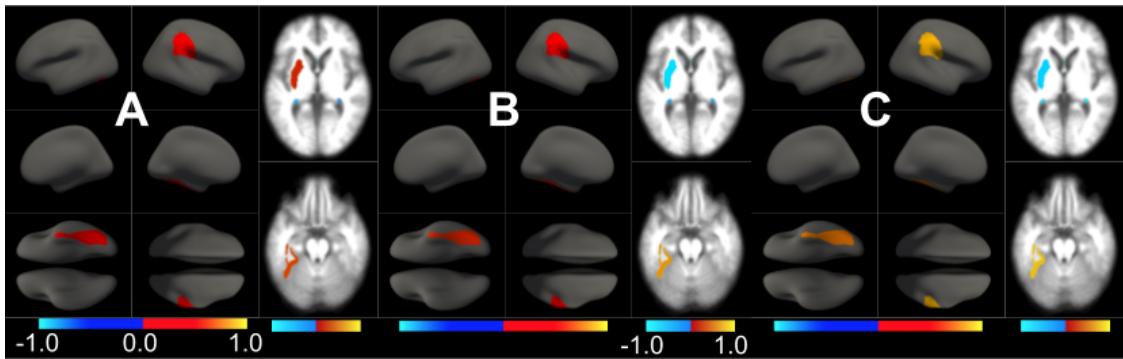


Figure 4.9: Effect type (ν_{Yk}^{AD}) of AD pathology on brain morphology for subjects in the T1 tertile. Each figure represents the contribution CSF biomarkers in explaining ROI volume variability: (A) CSF A β , (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

4.3.3 Discussion

In this section, we report the effect of normal aging and AD pathological processes on brain morphology. Age and sex are used as covariates in the confounder model mainly driven by age effects of global volume reduction [247] - except for the choroid plexus - that usually characterize cognitively unimpaired, amyloid-beta negative population. Its associated latent space is used as a confounding factor model to correct the disease model. AD pathological model is measured using CSF biomarkers and shows high effect sizes on brain morphology along the disease spectrum. Brain structure can be effectively described by a small set of underlying patterns correlated with CSF biomarkers. Comparing regional brain volume by whole brain marker of pathology show two different types of behavior once corrected by the confounder model: decreased and increased regional volume with biomarker progression showing patterns of faster and slower deterioration compared to the confounder model. Cerebral atrophy is defined as decreasing volumetric features with increasing CSF t-tau and p-tau and decreasing CSF A β and is found in typical-AD regions in the temporal lobe (hippocampus, superior, middle and inferior temporal, amygdala, fusiform and entorhinal cortex) as well as other regions such as precuneus and inferior parietal. Tau accumulation describes most of the volumetric

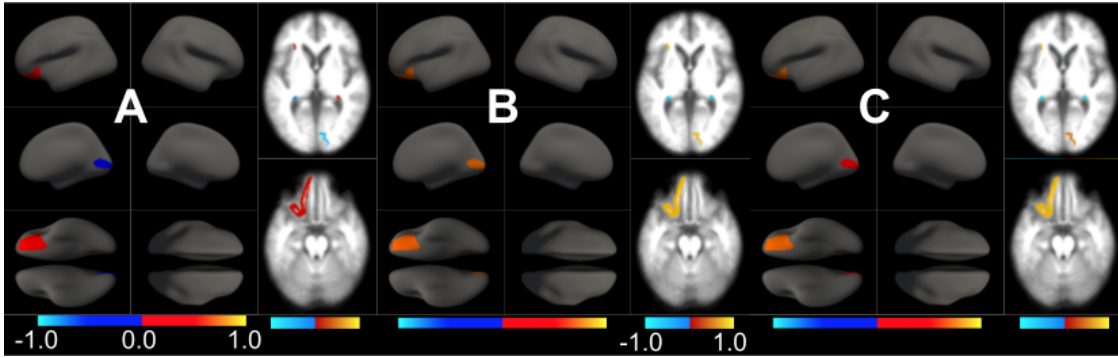


Figure 4.10: Effect type ($\nu_{Y_k}^{AD}$) of AD pathology on brain morphology for subjects in the T2 tertile. Each figure represents the contribution CSF biomarkers in explaining ROI volume variability: (A) CSF $A\beta$, (B) CSF p-tau, (C) CSF t-tau. Only brain ROIs with statistically significant effect size ($p < 0.05$) are shown.

	Centroid 0	Centroid 1
T0	InferiorTemporal R	SuperiorTemporal R PosteriorCingulate L Insula R Insula L
T1	ChoroidPlexus R ChoroidPlexus L Putamen L	Supramarginal R Fusiform L AccumbensArea R
T2	ChoroidPlexus R ChoroidPlexus L	Pericalcarine R LateralOrbitofrontal L

Table 4.3: Unsupervised cluster analysis of CSF effect types on brain morphology in the asymptomatic stage of AD. Statistically relevant ($p < 0.05$) brain ROIs are clustered into four separate regions defined by their respective centroid (first row). Centroids show the relative contribution to the latent space of each CSF measure: CSF $A\beta$, CSF p-tau and CSF t-tau. A positive (negative) value corresponds to increased (decreased) ROI volume with each measure.

variability on those regions associated with amyloid plaques deposition. On the other hand, increased mean ROI volume CSF biomarker progression is also present across the brain. This pattern is found in many regions, specially in the frontal lobe and in regions regions like pallidum, caudate nucleus and insular cortex. Even though it does not belong to the central nervous system (CNS), a well deserved remark for the choroid plexus, a region in the ventricles of the brain that produces CSF, that shows decreased volume with tau accumulation although seemed independent of CSF p-tau levels, indicating its non-specific character for AD.

Further analysis by using clinical diagnosis show different effect types along AD stages. At presymptomatic stage of AD (i.e. cognitively unimpaired (CU) subjects), volumetric variation of most brain ROI can be described by CSF t-tau/p-tau values. Subcortical regions such as pallidum, amygdala and putamen show decreased mean ROI volume with increasing tau protein concentrations. The same pattern is found in the cortical plexus or in cortical regions such as the inferior temporal. In contrast, pericalcarine, postcentral or the insular cortex show increasing their volume with increasing CSF p-tau/t-tau values. In order to further study the presymptomatic stage and with the aim to uncover specific CSF biomarkers association to brain morphology at different phases of early AD, we use subthresholds to split CU subjects into three tertiles (T0,T1,T2) based on CSF $A\beta$ burden. In the T0 tertile, considered as normal aging with high CSF $A\beta$ values, the insular cortex, cingulate and superior temporal show increased volume with CSF t-tau values. The most affected region is the choroid plexus and appears to decrease volume with increasing CSF tau proteins for subjects in later tertiles (T1 and T2). However, CSF effect size on subcortical regions becomes irrelevant when splitting the CU stage using subthresholds. More interestingly, the use of subthresholds hide CSF $A\beta$ -related patterns. To find variations in brain morphology associated to CSF $A\beta$ we need to analyze the CU stage as a whole and prominent effect sizes on the expansion of both caudates is found, related to expansion of lateral ventricles [163].

The relation between CSF biomarkers and brain morphology increases at MCI stage, becoming the clinical stage where most of the brain structure variation appears to be associated with CSF variation. Relevant effects of CSF on brain volume appear to be highly symmetrical. CSF $A\beta$ appears to be more relevant than in any other stage especially at subcortical regions. At MCI stage, typical AD regions in the temporal lobe such as hippocampus, parahippocampus and amygdala or middle temporal as well as the inferior parietal show decreased volume with markers of AD progression, confirming previous findings in the literature (e.g. [160]). Interestingly, central brain regions such as thalamus, precentral and pallidum show increased volume with pathology progression. At dementia stages, the influence of CSF biomarkers on brain morphology drastically diminishes, specially for CSF $A\beta$, which appears to be almost irrelevant. Only typical regions related to later stages of AD (inferior parietal, entorhinal cortex or pallidum) show significant effect sizes on the AD-dementia stage.

Nonetheless, there are some limitations in this work. First, we report preliminary results on the ADNI cohort but validation on an independent cohort would be required to assess the generalization of the results to other datasets as well as to the general population. Second, the oversimplification on the number of brain processes occurring in healthy adults brain, grossly split into AD and non-AD processes. Age and sex are used as the main confounders for non-AD effects while we acknowledge that many other factors (e.g: environmental, genetic) might be added to the model [69]. CSF biomarkers are used as a proxy for AD effects considering that amyloidosis, tauopathy and neurodegeneration are driving AD pathology. However, many other comorbidities might be found in AD subjects, specially at later stages [248]. Nonetheless, the work provides a methodology for the analysis of joint variation of imaging and non-imaging features and consistent results with the literature are found. New insights on brain morphology along the Alzheimer's continuum are also reported.

4.4 Multivariate brain structural effects

In this section, we aim at predicting pathophysiological markers (i.e. CSF biomarkers) using brain morphological (i.e. MRI) multivariate features. Our modeling assumptions are that there coexist two main process in the brain driven by (i) aging and (ii) AD pathology) that can be described by means of few underlying structures related to brain morphometry. As opposed to Section 4.3 where we analysed pathological effects independently for each brain ROI, here we consider multivariate methods that account for inter- and intra-relations between brain regions. Besides descriptive analyses, these latent patterns are further used to predict AD pathophysiological markers. For this purpose, we use Partial Least Squares regression (PLSR) methodology introduced in Section 4.2.1.2.

4.4.1 Material and methods

4.4.1.1 Data

In our experiments we use the publicly available dataset from the Alzheimer’s Disease Neuroimaging Initiative (ADNI, [241]). We build a sample of $N = 802$ subjects split into diagnostic categories following recently published guidelines [16]: $N_{HC} = 185$ healthy controls (HC), $N_{PC} = 136$ preclinical (PC) subjects, $N_{MCI} = 330$ subjects labeled as mild cognitive impairment (MCI) and $N_{AD-dementia} = 148$ subjects diagnosed with dementia due to AD (AD-dementia). MCI and AD-dementia subjects are diagnosed following the standard criteria used in ADNI and reported in [38] while for PC subjects we select cognitively unimpaired (CU) subjects with positive amyloid-beta using the suggested threshold at [170] ($A\beta < 192$ pg/mL). All subjects have a T1-weighted MRI available, preprocessed using FreeSurfer, [242], and segmented according to [81] and [243] into regions of interest. Grey matter volume divided by the total intracranial volume (V) and cortical thickness (CT) averages are computed for each ROI and used as predictor variables. A total of $K_{CT} = 64$ and $K_V = 84$ cortical and subcortical ROIs are used, accordingly. Both types of features are mean-centered and scaled by the control (CU) group standard deviation. Age and CSF biomarkers ($A\beta$, p-tau, t-tau) are used as response variables. Together with age and CSF biomarkers, a normalized CSF index as diagnostic metric for disease progression (AD-CSF, [205]) is used for descriptive statistics of the derived latent model. The AD-CSF index is a combination of CSF biomarkers and has two different forms: (i) AD-CSF₁ that involves $A\beta$ and p-tau protein and (ii) AD-CSF₂ that involves $A\beta$ and t-tau.

4.4.1.2 Model definition

According to Section 4.2.2, we define two separate models for brain aging and AD effects ($M^C = M^{cAD}$, $M^I = M^{AD}$). The M^{cAD} model is built using brain morphometric features as predictor ($\mathbf{X}^{cAD} \in \mathbb{R}^{N \times K}$) and age as response ($\mathbf{Y}^{cAD} \in \mathbb{R}^{N \times 1}$) variables and fitted on HC subjects only. Thereafter, we use the methodology stated in Section 4.2.2 to estimate the M^{AD} model using a deflated version of brain morphometric features as predictor ($\overline{\mathbf{X}}^{cAD} \in \mathbb{R}^{N \times K}$) and CSF biomarkers ($A\beta$, p-tau, t-tau) as response variables ($\mathbf{Y}^{AD} \in \mathbb{R}^{N \times 3}$). To fit the M^{AD} model only subjects in the Alzheimer’s continuum (PC, MCI, AD-dementia) are considered. The linear regression model in Equation 4.8 is used to define the dimension of each latent space, (L^{cAD} , L^{AD}), by

evaluating the total mean absolute error (MAE) of the predicted response variables in a 5-fold cross-validation framework [125]. Accordingly, we set $L^{\text{cAD}} = 2$ and $L^{\text{AD}} = 4$. The final latent space, M with $L = 6$, is built by concatenating both models ($t_0 = t_0^{\text{cAD}}, t_1 = t_1^{\text{cAD}}, t_2 = t_0^{\text{AD}}, t_3 = t_1^{\text{AD}}, t_4 = t_2^{\text{AD}}, t_5 = t_3^{\text{AD}}$). As brain morphometric features, we study separately the effects on volumetric and cortical thickness data. We use t_i^V and t_i^{CT} to refer to the latent subspaces build using volumetric and cortical thickness features, respectively.

In order to model AD pathology, we adopt two different modeling techniques based on different priori hypothesis: (i) we consider that latent patterns (effect-type) governing AD pathology are linear across all Alzheimer’s continuum with possibly different effect sizes at different clinical stages and (ii) we consider separate models for each clinical stage of AD yielding a stage-wise linear model across the Alzheimer’s continuum. For the first case, we fit a single model using all subjects in the Alzheimer’s continuum, including preclinical and clinical stages, while for the second hypothesis we use 4 different models to analyze the effects of pathophysiological markers separately at each stage of the Alzheimer’s continuum (PC, MCI and AD-dementia) and in normal AD biomarker subjects (HC).

For statistical inference, we will use age, CSF biomarkers and AD-CSF indices as variables of interest ($q_s, \forall s = 0, \dots, 5$). The statistical relevance will be assessed using permutation testing, as stated in Section 4.2.3.3. For all experiments, we use a total of $N_{\text{perm}} = 10000$ permutations, corresponding to a precision of the p -value estimate of $P=0.005$. Statistical significance is set to p -value < 0.05 (uncorrected). Furthermore we study the predictive power of each model using a 5-fold cross-validation framework to report the mean absolute error (MAE) of each variable of interest.

4.4.2 Results

4.4.2.1 Single model for AD pathology

Using a single model for all disease continuum we assume that the effect type of AD pathology is preserved throughout the Alzheimer’s continuum even though the effect size might change at each stage. The effect size of brain condition markers on each latent factor ($\rho_{l(s)}$) is shown in Figure 4.11 and listed in Supplementary Tables A.6 and A.7 with the associated p -values for models using volumetric and cortical thickness features, respectively.

To better understand the model of each measurement (volumetric and cortical thickness), we first analyze the effect size related to the variables of interest at each latent dimension $\rho_{l(s)}$. In Figure 4.11, we show the absolute value of the correlation between each latent variable t_l and the respective indicators $s = \text{age, CSF biomarkers, AD-CSF}_1 \text{ and AD-CSF}_2$ evaluated separately on subjects with normal AD biomarkers (HC) and subjects at specific Alzheimer’s continuum stages (PC, MCI, AD-dementia).

Each effect type can be seen as a morphological pattern responding to certain conditions. According to the results in Supplementary Tables A.6 and A.7, we can distinguish three different types of relevant ($p < 0.05$) patterns along the Alzheimer’s continuum: (i) age-related patterns associated with high correlation with age, (ii) AD-related patterns that present high correlation with AD markers and (iii) mixed age and AD-related patterns

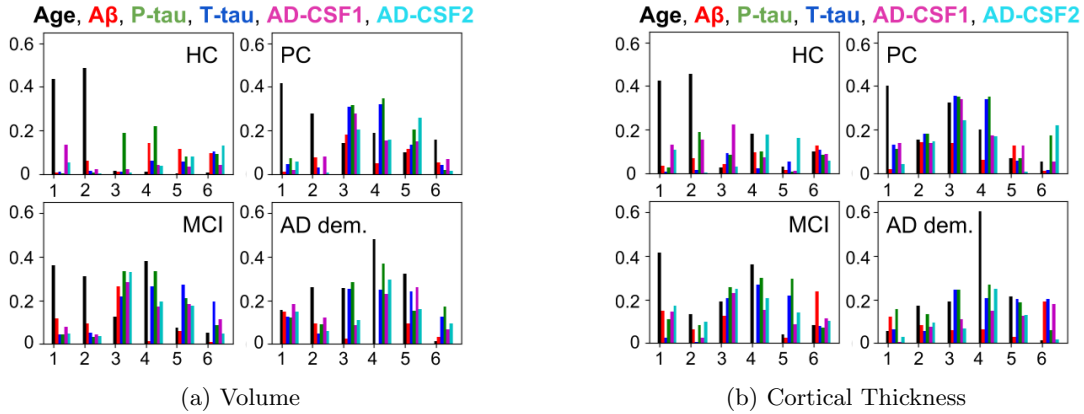


Figure 4.11: Absolute value of the effect size (single model) on several indicators evaluated at different stages: age (black), amyloid-beta (red), p-tau (green), t-tau (blue), AD-CSF₁ (magenta), AD-CSF₂ (cyan). Latent variables 0 and 1 correspond to brain aging model and latent variables 2, 3, 4 and 5 correspond to brain AD and are shown in the x-axis.

that show high correlation values with age and several indicators of AD progression. The first one is found relevant on normal aging subjects, while the other two coexists along the Alzheimer’s continuum, meaning that aging and AD processes might partially overlap across the brain. We show an example of each type of pattern for volumetric features in Figure 4.12 and for cortical thickness features in Figure 4.13, in line with the corresponding effect sizes at each stage.

In Figure 4.12a we show an age-related pattern with high effect size at pre-dementia stages (HC, PC and MCI) but low effect size at AD-dementia stage. This pattern is estimated using only HC subjects and M^{nAD} model and shows whole brain atrophy except for the choroid plexus regions. An AD-related pattern (t_2) is depicted in Figure 4.12b showing high correlation with AD pathophysiological markers both on subjects along the whole Alzheimer’s continuum and independently at all stages. It shows a highly symmetric pattern were positive and negative values indicate neurodegeneration and increased ROI volume, respectively, due to its positive correlation with CSF A β and negative correlation with CSF p-tau and CSF t-tau values. Cortical regions (parahippocampus, middle temporal, transverse temporal, inferior temporal) and subcortical regions (hippocampus, amygdala and fusiform) of the temporal lobe as well as regions such as the choroid plexus and, specially, the entorhinal cortex are the most affected regions positively correlated with t_2 , indicating neurodegeneration over the course of AD. Other regions such as the precentral or the pars orbitalis are negatively correlated with t_2 . Finally, a mixed age and AD-related pattern (t_3) is shown in Figure 4.12c, which is negatively correlated with age and positively correlated with CSF p-tau, CSF t-tau and their associated indices along the Alzheimer’s continuum and independently at each stage while it remains independent of CSF A β values. Among the regions driving this pattern, we found choroid plexus, precuneus (negatively correlated with t_3), temporal pole and parahippocampus (positively correlated with t_3) the most relevant ones, indicating that they are involved in both brain aging and brain AD processes.

Similarly, patterns associated with cortical thickness are shown in Figure 4.13 and are also split into age-related, AD-related and mixed age and AD-related effects. Firstly,

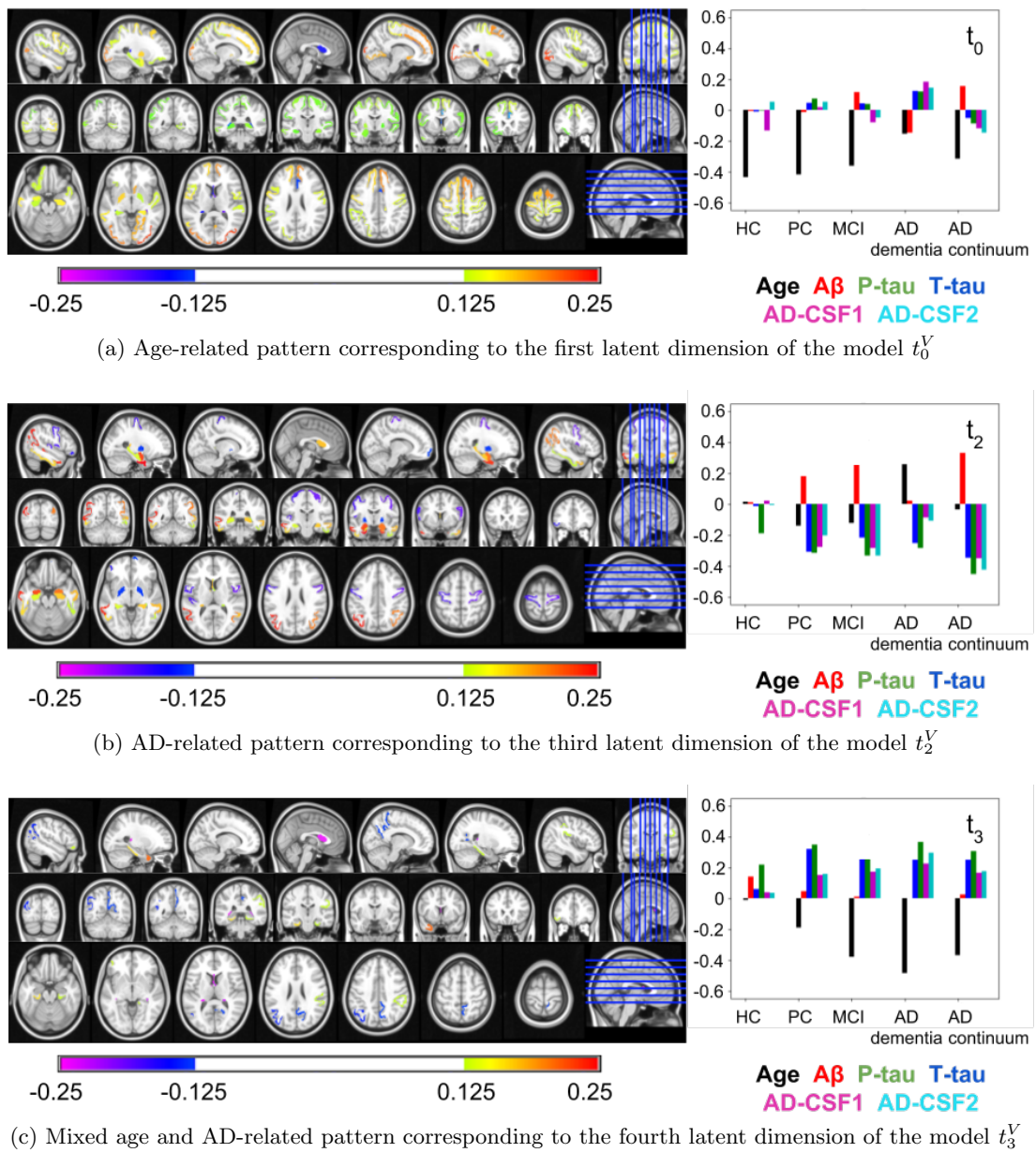


Figure 4.12: Examples of relevant latent brain morphological patterns (effect-type) using volumetric features. Values at each ROI represent their influence to the latent space. The corresponding effect size of each condition at different stages of the disease is shown on the right.

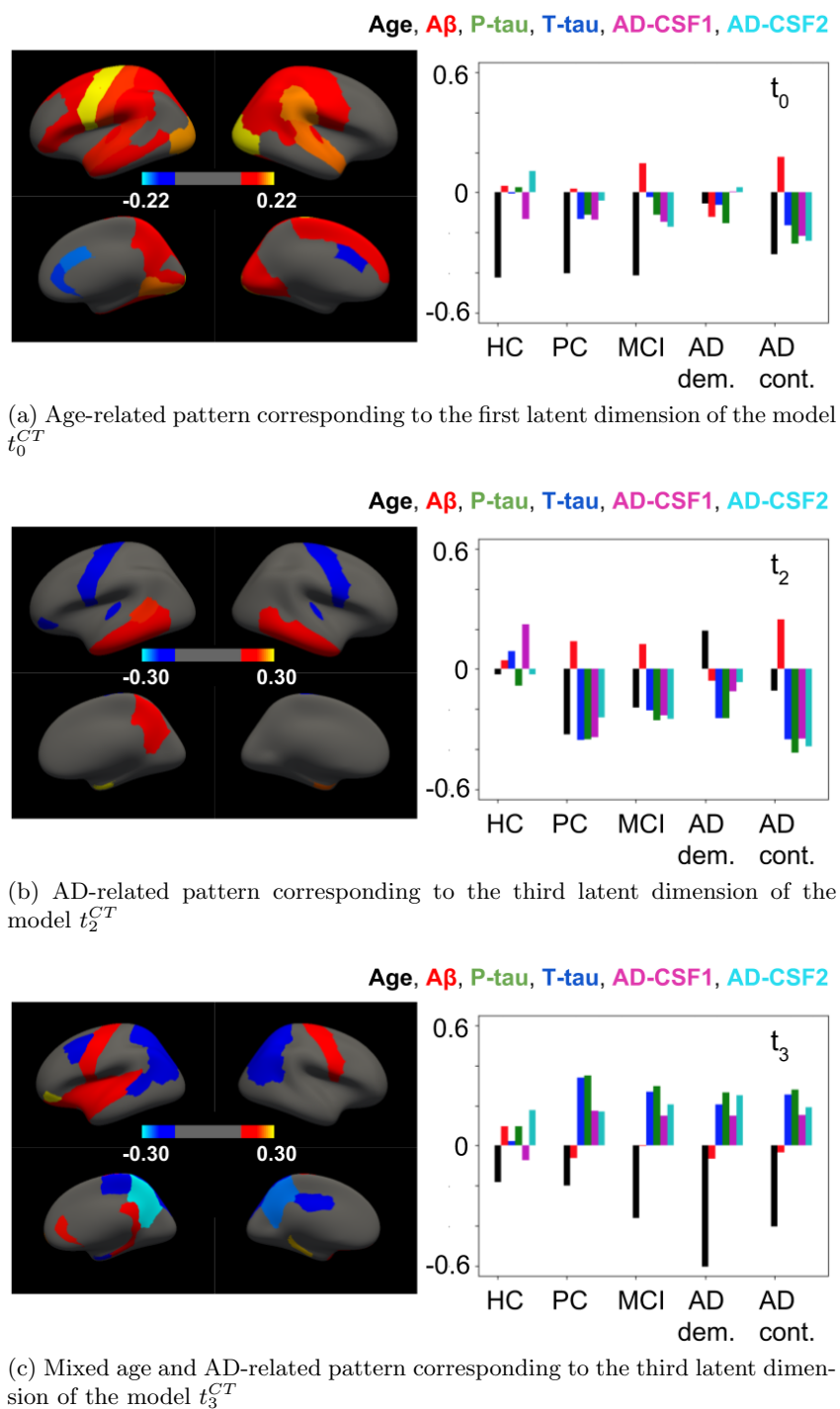


Figure 4.13: Examples of relevant latent brain morphology patterns (effect-type) using cortical thickness features. Values at each ROI represent their influence to the latent space. The corresponding effect size of each condition at different stages of the disease is shown on the right.

Figure 4.13a presents an age-related pattern involving overall cortical reduction with age in the normal aging model M^{nAD} except for the cingulate regions that show mild increased thickness. Secondly, an AD-related pattern is depicted in Figure 4.13b, involving cortical shrinkage of temporal regions (superior, middle, inferior) and the entorhinal cortex and increased thickness at regions such as precentral or transverse temporal. Finally, an age-AD mixed related pattern with opposite relationship between age (-) and CSF p-tau/t-tau (+) is shown in Figure 4.13c. It is associated with cortical shrinkage of parietal lobe regions and precuneus and increased thickness on parahippocampus, precentral, pars orbitalis, insula and superior temporal on the left hemisphere. Overall, similar findings are found for cortical thickness and volumetric data.

Finally, we evaluate the predictive power of the latent factors found for both age and CSF biomarkers at different disease stages. For better comparison, all indicators are globally standardized using z-score on all subjects (i.e. mean centered and normalized by variance). Table 4.4 shows the mean absolute error (MAE) for each class using cortical thickness and volumetric features and can be interpreted as the fraction of error per standard deviation. These results can be compared with standard linear regression method found in Supplementary Tables A.4 and A.5.

Features	Diagnosis	Age	CSF $A\beta$	CSF p-tau	CSF t-tau
Volume	HC	0.49	1.91	0.74	0.71
	PC	0.68	0.46	0.72	0.63
	MCI	0.67	0.4	0.70	0.72
	AD	0.85	0.37	0.88	0.94
Cortical Th.	HC	0.55	1.94	0.7	0.63
	PC	0.76	0.48	0.68	0.66
	MCI	0.77	0.42	0.69	0.72
	AD	0.91	0.38	0.87	0.88

Table 4.4: Predictive error of response variables for each diagnosis label and feature type using a single model for all Alzheimer’s continuum.

Age is best predicted using HC subjects and the prediction worsens at later stages at the Alzheimer’s continuum, showing that features affected by age are also affected by the disease. CSF $A\beta$ biomarker presents lower predictive error at Alzheimer’s continuum stages and high predictive error at HC stage probably due to the standardization process, where it presents low variance at Alzheimer’s continuum stages and higher variance at HC stage. On the other hand, CSF p-tau/t-tau seem not very predictable using brain morphometry measurements. Compared to standard linear regression, we found no clear gain of our methodology for the regression task even though the model provides relevant brain structural patterns related to AD pathophysiological markers.

4.4.2.2 Multiple models for AD pathology

Under the assumption that underlying processes governing brain structure along the Alzheimer’s continuum might differ between clinical and pathological conditions, we fit a brain AD model independently for each stage including (i) HC, (ii) PC, (iii) MCI and (iv) AD-dementia. Similarly to Section 4.4.2.1, we perform the analysis for volumetric and cortical thickness measurements and show the effect size of each brain latent model with respect to several variables of interest at their corresponding stage (Figure 4.14). A

complete list of all effect sizes with their associated p -values is listed in Supplementary Tables A.8 and A.9.

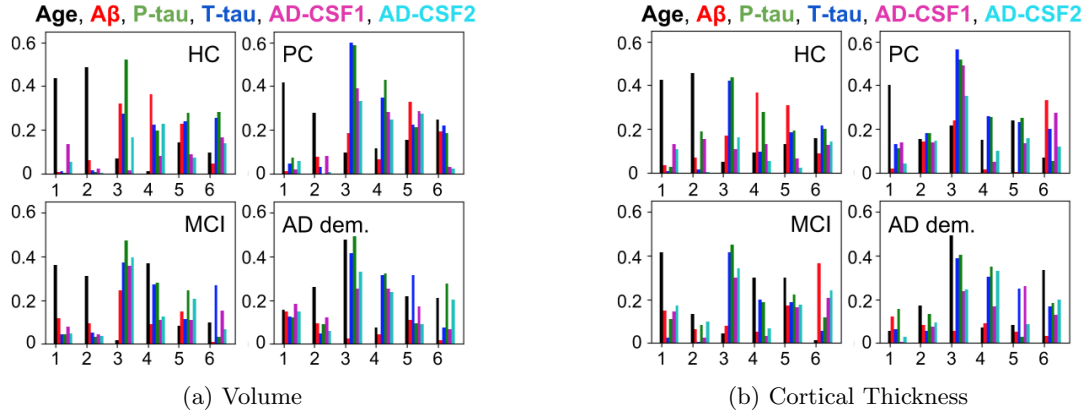


Figure 4.14: Absolute value of the effect size (multiple models) on several indicators evaluated at the respective disease stages: age (black), amyloid-beta (red), p-tau (green), t-tau (blue), AD-CSF-1 (magenta), AD-CSF-2 (cyan). Latent variables 0 and 1 correspond to brain aging model and latent variables 2, 3, 4 and 5 correspond to stagewise brain AD and are shown in the x-axis.

Underlying patterns from the brain aging model remain the same while latent factors related to brain AD model increase their effect size and statistical relevance at each stage. More specific patterns of brain AD are found for each Alzheimer’s continuum stage. Here, we focus on AD-related latent factors found in previous section (Section 4.4.2.1) to discover differences between brain morphological patterns related to AD at specific stages. Hence, in Figure 4.15 and Figure 4.16 we show the effect-type of the third latent dimension using volumetric (t_2^V) and cortical thickness features (t_2^{CT}), respectively. Both figures show the highest correlation with AD pathophysiological markers.

Here, we concretely study the third dimension of the model (t_2), an AD-related process that presents the highest correlation with AD pathophysiological markers with both cortical thickness and volumetric features. In Figure 4.15 we present the underlying pattern found at each stage of the disease using volumetric features. Compared to effect type found in previous analysis (Figure 4.12b), the effect type on MCI and AD-dementia stages appears to be very similar with the addition of the pallidum showing compensatory effect in the later. On the other hand, in the PC stage slightly differs from the global AD pattern yielding an specific preclinical signature. The neurodegeneration pattern involve choroid plexus and pallidum while pericalcarine show increased mean ROI volume. Again, similar results are found addressing volumetric and cortical thickness features.

For predictive analysis, all indicators are globally standardized using z-score on all subjects (i.e. mean centered and normalized by variance). In Table 4.5 we summarize the mean absolute error (MAE) per class that can be interpreted as the fraction of error per standard deviation. Even though stage-wise modeling provides a higher effect size and more relevant brain AD patterns, the predictive power of CSF biomarkers is slightly worse for late stages of the Alzheimer’s continuum, probably indicating a lower generalization power of the brain features found for the same number of latent dimensions. Nonetheless, PLSR has higher predictive power than the standard linear regression using

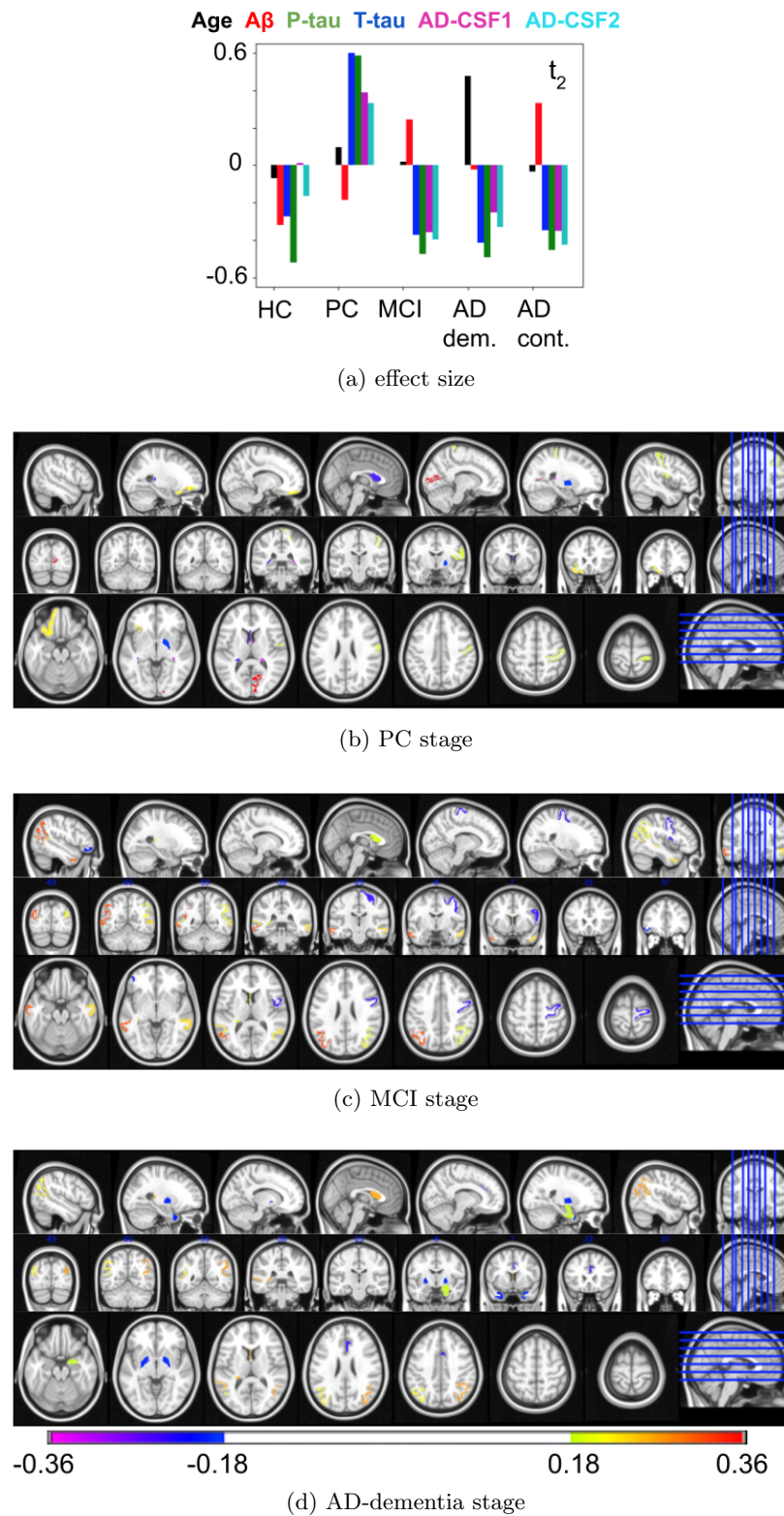


Figure 4.15: AD-related effect-type of the brain latent model using volumetric features. It corresponds to the third latent dimension of the model t_2^V at each stage of the Alzheimer's continuum. The effect size at each disease stage is shown on the top.

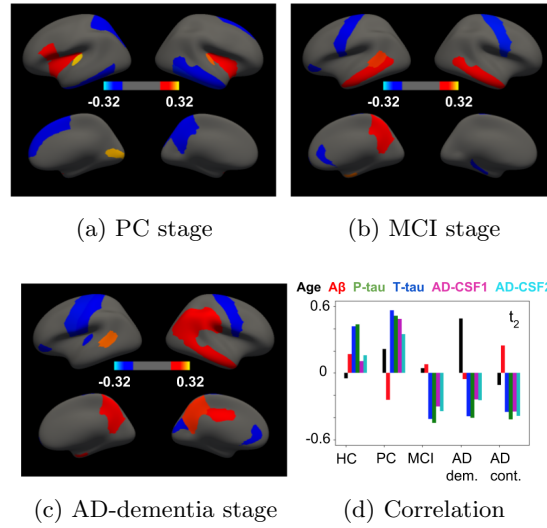


Figure 4.16: AD-related effect-type of the brain latent model using cortical thickness features. It corresponds to the third latent dimension of the model t_2^{CT} at each stage of the Alzheimer’s continuum. The effect size at each disease stage is shown on the bottom-right.

raw morphometric features (Supplementary Table A.5).

Features	Diagnosis	Age	CSF A β	CSF p-tau	CSF t-tau
Volume	HC	0.49	0.51	0.38	0.34
	PC	0.68	0.5	0.75	0.58
	MCI	0.67	0.41	0.72	0.75
	AD	0.85	0.33	1.07	1.0
Cortical Th.	HC	0.55	0.54	0.36	0.36
	PC	0.76	0.48	0.73	0.61
	MCI	0.77	0.43	0.7	0.77
	AD	0.91	0.34	0.98	1.05

Table 4.5: Predictive error of response variables for each diagnostic label and feature type using a separate models for subjects with normal AD biomarkers and for each Alzheimer’s continuum stage.

4.4.3 Discussion

In this work we aim at predicting AD pathophysiological markers from brain morphology in elderly adults. To this end, we disentangle aging and pathological processes by orthogonal projections in a lower-dimensional subspace describing brain morphology along the disease continuum. We provide two different approaches to model the Alzheimer’s continuum: linear and stage-wise linear models, assuming in the later that brain anatomy might differ between preclinical and clinical Alzheimer’s continuum stages. Statistical significance and predictive power of each model is assessed using either grey matter volume or cortical thickness measurements as brain morphological features. Joint modeling of both types of features do not reveal further insights since the overall effect and predictive

power remain similar indicating that they might not provide sufficient complementary information.

Both linear and stage-wise linear models find relevant patterns related to age, AD and a mixture of both, meaning that aging and disease processes might overlap, specially significant for later stages of the disease (MCI, AD-dementia). Using a single model for the whole Alzheimer's continuum, we find common structures that are affected over the course of the disease, delineating the AD signature supported in the literature [39]. On the other hand, fitting a model to each disease stage allows uncovering temporally specific patterns and increase the effect size of AD pathophysiological markers at each stage. While a single model finds relevant patterns involving typical AD-related regions in the temporal lobe, using multiple models allows to uncover significant structures at specific stages. Fitting a separate model for each stage along the Alzheimer's continuum increases the correlation between descriptive brain morphological patterns and AD pathophysiological markers. Concretely, we found that brain structure at MCI and AD-dementia stages retrieve morphological patterns overlapping with the previously found AD signature. However, at preclinical stage the joint variability of CSF biomarkers and brain structure follows a different and more specific pattern and thus, motivating the separate study of a preclinical AD signature, as several recent studies point out [163, 160]. Interestingly, the pallidum effect over the course of AD is only detected using stage-wise modelling since it shows early degeneration but then it remains unaffected through the course of the disease. Some other structures, such as pericalcarine at PC stage or pallidum at AD-dementia stage, showed increased volume/thickness once corrected by age, which might be interpreted as compensatory effects due to increasing brain activity at those regions [249, 250].

Moreover, we provide a confounder model orthogonal to the Alzheimer's continuum model that attempts to describe the brain aging process. It is estimated using only healthy control subjects and presents consistent results with the literature [251, 252, 253] for both volumetric and cortical thickness features involving global atrophy and cortical thinning, except for the choroid plexus region. This pattern appears to be highly correlated with pre-dementia stages (HC, PC, MCI) but has low effect size at the dementia stage, providing more evidence of a higher overlap between aging and disease processes at later stages.

Finally, we assess the predictive power of both models and compare it to a standard linear regression model. Using stage-wise modeling increases the overfitting at each stage, but the effect is mitigated using PLSR compared to linear regression. However, using a single model for the whole Alzheimer's continuum provides similar results for both methods. Overall, the predictive performance is moderate to low depending on the disease stage, being worse at later stages (AD-dementia).

Our results highlight the potential of using multivariate models to better detect and disentangle aging and AD pathological effects on brain structure instead of using univariate regression models and hypothesis testing studying each brain region independently [245]. Multivariate models are able to capture relevant and spatially distributed patterns of structural changes in the brain. In this work, we identify patterns of common variability between AD pathology, defined using CSF biomarkers, and brain structure, defined using either volumetric or thickness features. Throughout the literature, many multivariate approaches define AD pathology with binary or categorical variables (e.g: clinical diag-

nosis [219, 254] or amyloid status [215, 166]) and use machine learning as a multivariate model. However, the quantization of AD pathology in few categories reduces the modeling capacity of its variability and heterogeneity. To account for that, other works use single continuous clinical scores [228, 220], which is suited for modeling the later stages of the disease. Instead, CSF biomarkers, as used in our approach, are shown to better describe the full spectrum of AD [41]. Another recent multivariate approach to detect brain changes uses an event-based model (EBM) [175] that clusters subjects depending on spatially distributed and temporally aligned brain changes, being well-suited for e.g. patient staging. However, EBM makes a set of assumptions such that biomarkers vary independently and do not make use of underlying AD pathophysiological markers. Instead of finding a sequence of events, in this work we could identify few underlying patterns of brain variability that are maximally related to AD pathology finding three different types of patterns: age-related, AD-related and mixed age and AD related.

4.5 Conclusions

In this chapter, we study the relationship between brain morphology and AD pathology. We present a framework able to uncover latent patterns that capture most of the joint variability between both domains. Moreover, we extend it to embed a confounder model of brain morphology able to disentangle patterns of normal aging and AD pathology. Two different analysis of brain structure have been performed: (i) a univariate analysis, where each brain region is modeled independently and (ii) a multivariate analysis where we attempt to model inter- and intra-relationships between brain anatomical regions. In the univariate analysis we show that brain regions are affected differently by the underlying AD pathology with regions more related to amyloid plaques deposition and other to neurofibrillary tangles. Those brain regions are also affected differently throughout different clinical stages in Alzheimer’s disease. In the multivariate analysis we show that brain structural patterns of variation along the Alzheimer’s continuum can be age-related, AD-related and a mixture of both, indicating overlap between aging and disease processes. An AD-related pattern is found in the first mode of variation and can be seen as the *Alzheimer’s signature* also found throughout the literature. A different pattern of variation is found at preclinical stages fostering a specific *preclinical Alzheimer’s signature* that should be further investigated in future studies. An extension of this work might account for other processes present in the brain of subjects along the Alzheimer’s continuum. This would need the use of other biomarkers (e.g. DTI, FDG-PET, tau PET) and thus, an extension of this methodology to multi-domain approaches. In this line, the literature from the computer vision community on cross-modality retrieval using neural networks, autoencoders and different regularization strategies would be a good direction to take.

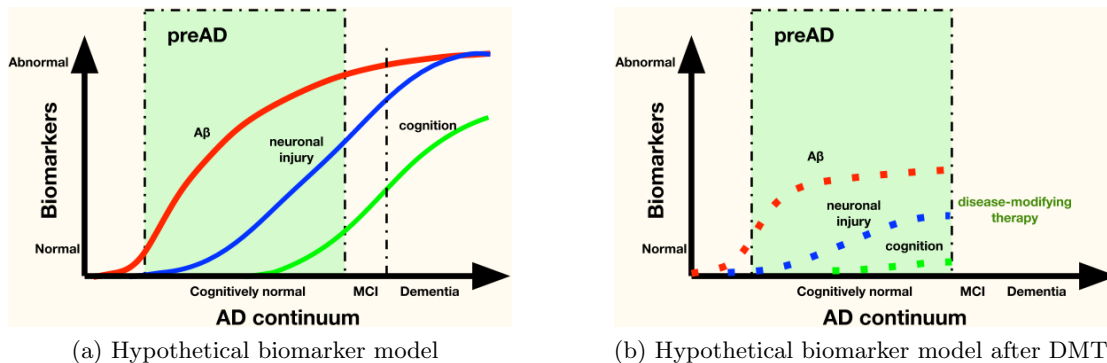
MRI and Machine Learning in Clinical Trials

The preclinical stage of Alzheimer’s disease has become a major focus in AD research as it constitutes an opportunity for disease modifying therapies. It is believed that early intervention and prevention have a better chance of success. Following the amyloid hypothesis, a promising research direction is to study anti-amyloid drugs either to stop amyloid production or to clear the amyloid from the brain in asymptomatic subjects with already elevated levels of amyloid. General population screening using gold standard tests becomes rapidly unfeasible when conducting large-scale studies due to invasiveness and cost. Hence, in this chapter we propose a protocol for clinical recruitment that includes a pre-screening layer using machine learning (ML) and magnetic resonance imaging (MRI) to detect subjects with abnormal levels of amyloid from a pool of cognitively unimpaired participants. We assess the predictive capacity of the methodology in cross-sectional and longitudinal studies and its utility in reducing the number of gold-standard tests and the cost burden in clinical trials.

This work has been done in collaboration with BarcelonaBeta Brain Research Center (BBRC), his Principal Investigator Dr. Juan Domingo Gispert and Dr. Paula Petrone, who proposed the longitudinal approach developed in Section 5.4.

5.1 Introduction

In Chapter 4, we showed some of the singular characteristics of the *preclinical Alzheimer’s signature* in structural MRI compared to the *AD signature*. It characterizes early changes in brain structure due to Alzheimer’s pathology and that might differ from those found in clinical stages of AD. Structural MRI, as a non-invasive and cheaper biomarker than invasive pathophysiological AD biomarkers, has great potential applications. Hence, the first question we attempt to answer in this chapter is whether preclinical Alzheimer’s pathology can be predicted from MR images. For that purpose, we use machine learning and cross-sectional data to evaluate the capacity of T1-weighted MR images to predict amyloid deposition in cognitively unimpaired individuals. We use a pattern recognition framework similar to the one introduced in Section 2.3. We further investigate whether other imaging modalities (diffusion and functional MRI) add some value to the final prediction, even though we rapidly discard functional MRI due to lack of sufficiently clean data. Secondly, we propose a new protocol for recruitment in clinical trials involving the previously derived methodology and we try to assess its potential as a pre-screening tool. The third question we formulate is whether volumetric rate of change between different acquisitions carry information about amyloid profile. We derive Jacobian determinant features as a measure of local tissue change between two acquisitions. A longitudinal study with follow-up times over 4 years is used to assess group-differences at different follow-up times. Finally, we attempt to find the minimum optimal time-span between



(a) Hypothetical biomarker model

(b) Hypothetical biomarker model after DMT

Figure 5.1: AD prevention hypothesis: applying disease modifying therapies (DMT) before neuronal loss may slow down both dead brain cells and cognitive decline progression. The figure (a) is based on the cascade hypothesis pictured in [40]. The figure (b) is mainly intended for visualization purposes and the curves are not based on any theoretical nor empirical model.

MRI acquisitions to capture differences due to amyloid depositions in the brain and evaluate the method in the clinical recruitment framework.

5.1.1 Motivation

As detailed in Section 2.1, nowadays Alzheimer’s disease has been established as a continuum preceded by a preclinical stage defined by cognitively healthy individuals with abnormal pathophysiological biomarkers. Following the amyloid hypothesis [40], we use the $A\beta$ status, measured using either CSF lumbar puncture or an amyloid PET scan, to define subjects at preclinical stage ($A\beta^-$) or healthy controls ($A\beta^+$). This preclinical stage is thought to last for 15-30 years before the onset of clinical and cognitive symptoms.

To date, agents aiming to slow down cognitive decline in symptomatic subjects have failed. So far, all phase III clinical trials conducted using agents targeting the amyloid cascade with the primary goal to halt or lessen disease progression have dropped out. Nonetheless, the anti-amyloid drugs tested do remove or stop the production of amyloid in the brain. A great deal of reasons may explain these failures, such as the validity of the amyloid hypothesis [255] or the heterogeneity of subjects with not enough evidence of amyloid [256]. The main reason, though, might be that these trials involved symptomatic subjects, a late stage in the disease continuum. From these failed results, it seems that once clinical symptoms appear, brain damage may be already irreversible and intervention at earlier stages of AD (pre-dementia stages) must be considered. Therefore, interventional studies and clinical trials have been shifting towards prevention of AD. Here, the preclinical stage is also a window of opportunity for disease modifying therapies hoping to revert or slow down the course of the disease before it is too late. We hypothetically represent this idea in Figure 5.1.

The challenge that early interventions are facing is how to identify cognitively healthy individuals at risk for AD dementia. The current gold standards for participant screening are amyloid PET and CSF- $A\beta$. PET scans use radioactive tracers and CSF lumbar puncture is invasive. Both tests are also expensive and many subjects have to be tested in order to find a few eligible ones, rising the overall cost of the trial and making it very

risky undermining the development of preventive therapies. Hence, these techniques are unsuited for the screening of the general population. Instead, our approach is to use less invasive and cost-effective tools for pre-screening. In this chapter, we propose to use machine learning and structural MRI information for prediction of preclinical Alzheimer's pathology (i.e. amyloid positivity) and use it for population screening in clinical trials.

5.2 Clinical trials

In this section, we provide an overview of some of the past clinical trials for drug development in Alzheimer's disease that failed and present some of the current clinical trials in late-stage development (phase II or III). The standard procedure in clinical trials is to test a treatment in a randomized and double-blinded study splitting the participants into at least two arms (experimental vs. placebo) and comparing several predefined outcomes among them. We focus on anti-amyloid therapies even though we show examples of others, as well. After that, the initial hypothesis of this work is formulated by proposing a modified version of the protocol used for population screening in clinical trials. A couple of models (simple and complex) are derived and used to test the hypothesis throughout the chapter.

5.2.1 Past clinical trials

The etiology of Alzheimer's disease is still unknown and hence, several hypothesis are under study. In earlier times, the neurotransmitter hypothesis was dominant, stating that the loss of cholinergic neurotransmitters was an indirect cause of cognitive impairment and that cholinergic drugs would improve cognitive functions [257]. In this line, donepezil and memantine were approved for medical use in the United States in late 90s and early 2000s. Currently, they are widely used for Alzheimer's disease interventions, being useful to relieve symptoms for short period of times, even though they are not able to delay the progression of AD. Since then, and for more than 15 years, no more drugs have been approved by the U.S. Food and Drug Administration (FDA) for Alzheimer's disease treatment.

At present times, the most prevalent hypothesis is the amyloid cascade, introduced by John Hardy and David Allsop in 1991 [258]. Anti-amyloid treatments represent almost a quarter (22.3%) of the total number of clinical trials (N=2173) up to 2019 [259]. Two different types of treatments targeting amyloid deposition can be distinguished: β -secretase amyloid precursor protein cleaving enzyme (BACE) inhibitors and anti-amyloid beta antibodies. The former attempts to stop the amyloid production while the latter aims at removing already produced amyloid from the brain. Many pharmaceutical companies are involved in drug development for Alzheimer's disease as it may represent huge revenues if approved for medical use. The most common practice for them is to sponsor clinical trials in public-private partnerships with institutions, universities, consortiums and initiatives.

Many anti-amyloid treatments have progressed to phase III because they were reported safe and show the ability to decrease the amount of amyloid in the brain in previous phases. Safety measures usually involve looking at amyloid related imaging abnormalities (ARIA) as well as side effects. However, to date, all of them show no significant effect on primary outcomes, i.e. related to cognition decline between experimental and placebo arms. Noteworthy, even some BACE inhibitors show worse cognition in participants that

Failed Phase III trials			
Study	Main study sponsors	Name of the drug	Condition
MISSION AD	Eisai	Elenbecestat	MCI due to/mild AD
EPOCH	Merck	Verubecestat	Mild to moderate AD
EXPEDITION-1/-2/-3/Pro	Eli Lilly	Solanezumab	Prodromal, mild and moderate AD
Marguerite RoAD	Hoffmann-La Roche	Gantenerumab	Mild AD
CREAD	Hoffmann-La Roche	Crenezumab	MCI/prodromal AD with amyloid pathology
EARLY	Janssen	Atabacestat	Asymptomatic at risk of Alzheimers dementia
ENGAGE/EMERGE	Eisai & Biogen	Aducanumab	MCI due to/mild AD
AMARANTH	AstraZeneca and Eli Lilly	Lanabecestat	MCI due to AD and mild AD
Generation S1/S2	Novartis and Amgen	Umibecestat	CU with two copies of APOE4 or one copy and amyloid positive

Table 5.1: Past phase III clinical trials that failed to show significant results in their primary outcomes.

undergo the treatment (e.g. AMARANTH and APECS studies) posing questions about the optimal dose, as high doses have been shown to interfere with synaptic plasticity and memory formation in mouse studies [260]. A promising BACE inhibitor was tested in the EARLY clinical trial, but dropped out due to serious liver toxicity and the sponsor (Janssen), in a press release, concluded that “the benefit-risk ratio is no longer favorable to continue development of atabacestat for people who have late-onset preclinical stage Alzheimer’s disease”. Nonetheless, BACE inhibitors remain the leading candidate for primary intervention. A list of some recent failed clinical trials is found in Table 5.1.

A systemic problem of failed clinical trials is that most of them have no publicly available data to analyze their failures and normally report that “internal futility studies” (which balance potential benefit vs. risk) “recommend to discontinue the trial”. Hence, no exact knowledge from several negative clinical trials avoiding to draw conclusions that could improve future trial designs. Moreover, these studies may account for risks other than patient health (e.g: economical) risking small potential benefits.

5.2.2 On-going clinical trials

After the aforementioned results, the community must formulate and try to answer several questions ranging from the validity of the amyloid hypothesis and the therapeutic effect of lowering brain amyloid depositions to clinical trial design. In this line, even though the amyloid hypothesis is still predominant in current phase II and phase III trials (specially using anti-amyloid beta antibodies), there are other hypotheses currently

Phase II trials			
Study	Main study sponsors	Name of the drug	Condition
Abvac40	Araclon Biotech S.L.	ABvac40	Amnestic mild cognitive impairment or very mild Alzheimer's disease
ACI-24-1801	AC Immune	ACI-24	Mild Alzheimer's disease
DESPIAD	University College London	CPHPC	Mild Alzheimer's disease
ETHERAL	Oryzon	ORY-2001	Mild-moderate Alzheimer's disease
LAURIET	Genentech	RO7105705	Moderate Alzheimer's disease
LM11A-31-BHS	Pharmatrophix Inc.	LM11A-31-BHS	Mild-moderate Alzheimer's disease
MAPT-CS1	Ionis Pharmaceuticals, Inc.	IONIS MAPTRx	Mild Alzheimer's disease
VALZ-PILOT	Hugo Lovheim	Valaciclovir	Early Alzheimer's disease

Table 5.2: Phase II trials conducted in Europe that are currently recruiting. Source: Clinical Trial Watch initiative from Alzheimer's Europe (last accessed: 23-09-2019)

being tested: targeting tau (e.g: RO7105705, IONIS MAPTRx, TRx0237), targeting neurotransmitters (e.g: the dopamine receptor Brexpiprazole), involving diet (e.g: OMEGA3 supplements) and targeting psychiatric disorders such as depression (e.g: Mirtazapine) or hyperactivity (e.g: Guanfacine), among others. A list of on-going european clinical trials is shown in Table 5.2 (phase II) and Table 5.3 (phase III).

A common characteristic of past clinical trials of amyloid-related treatments is that all targeted already symptomatic subjects (prodromal, moderate, mild AD). At these late stages of the Alzheimer's continuum is believed that brain damage is already irreversible yielding no improvement in cognition measures. Thus, amyloid treatments might be performed earlier in the disease continuum shifting towards prevention of AD. Some of the treatments that failed in earlier studies (BACE: elenbecestat, or antibodies: solanezumab and BAN2401) as well as new ones are used in trials targeting cognitively unimpaired individuals at risk for AD. This risk is evaluated by either amyloid status in sporadic AD (e.g: A4 study, A45) or by mutation carriers in DIAN-TU Next Generation or ADAD trial. A list of clinical trials in asymptomatic subjects is shown in Table 5.4.

5.2.3 Hypothesis

In this shift towards secondary and, eventually, primary prevention of AD, clinical trials face the same challenge: how to identify asymptomatic subjects at risk to develop AD dementia. In familial AD studies, a genetic verification of the autosomal dominant Alzheimer's disease mutations (APP, PSEN1, PSEN2) and clinical tests might suffice. However, for sporadic AD studies, in-vivo gold standard measures of Alzheimer's pathology are required for the initial characterization. Currently accepted pathophysiological

Phase III trials			
Study	Main study sponsors	Name of the drug	Condition
BREXPIPIRAZOLE	Otsuka Pharmaceutical	Brexpiprazole	Agitation in people with dementia of the Alzheimer's type
GRADUATE 1 & GRADUATE 2	Hoffmann-La Roche	Gantenerumab	Prodromal to mild Alzheimer's disease
LO-MAPT	University Hospital, Toulouse	Omega-3 treatment	Subjective memory complaints of family history of Alzheimer's disease
LUCIDITY	TauRx Therapeutics Ltd	TRx0237	Mild Alzheimer's disease
NorAD	Imperial College London	Guanfacine	Mild to moderate Alzheimer's disease
SYMBAD	University of Sussex	Mirtazapine	Alzheimer's disease and agitated behaviours
17-AVP-786-305	Avanir Pharmaceuticals	AVP-786	Agitation in people with dementia of the Alzheimer's type

Table 5.3: Phase III trials conducted in Europe that are currently recruiting. Source: Clinical Trial Watch initiative from Alzheimer's Europe (last accessed: 23-09-2019)

Study	Main study sponsors	Name of the drug	Condition
DIAN-TU Next Generation Trial	Washington University School of Medicine	Santenerumab or solanezumab	ADAD mutation carriers
A4 Trial Anti-Amyloid treatment in Asymptomatic Alzheimers Disease	Eli Lilly and Co.	Solanezumab	CU and $A\beta+$
The A3 Study: Ante-Amyloid Prevention of Alzheimer's disease	Eisai and NIA/NIH	Elenbecestat	CU just below amyloid positive threshold
The A45 Study	Eisai and NIA/NIH	BAN2401 and elenbecestat	CU and $A\beta+$
Autosomal Dominant Alzheimers Disease (ADAD) Colombia Trial	Genentech, Inc.	Crenezumab	Asymptomatic subjects who carry the PSEN1 E280A mutation

Table 5.4: Late-stage development trials (phase II/III or III) that attempt in cognitively unimpaired subjects.

biomarkers are either a lumbar puncture or a PET scan which are invasive and radioactive, respectively, and may hinder the initial recruitment process. Moreover, both gold standard biomarkers are rather expensive representing an important spending in the initial budget. As reviewed in section 2.1.2, the prevalence of Alzheimer’s pathology in cognitively unimpaired subjects might be rather low meaning that a great deal of participants (those who turn out to be amyloid-negative) would be unnecessarily scanned, as they won’t take part in the trial. Hence, an important amount of money is wasted in the screening phase for detecting subjects with Alzheimer’s pathology, which may unnecessarily increase the overall cost of the trial and so would the risk taken for the sponsors and collaborators. Therefore, we need less invasive and cost-effective techniques for Alzheimer’s pathology screening in general population.

In this chapter, we follow the amyloid hypothesis and thus, restrict the analysis on cognitively unimpaired subjects with evidence of amyloid positivity (which we refer to as preclinical subjects) as inclusion criteria. Our first hypothesis is that MRI can be used as a non-invasive and cheap biomarker for predicting amyloid positivity in asymptomatic subjects using machine learning techniques. If this holds true, the second hypothesis is that it can be used in clinical trials as a prescreening tool that may reduce the number of unnecessary gold standard tests and with that, the overall cost of the trial.

5.2.3.1 Standard vs. proposed clinical recruitment protocol

In clinical trials, the standard (S) recruitment protocol (Figure 5.2a) starts by a initial screening of the general population to build a initial pool of N_S cognitively healthy participants measured using standard clinical and cognitive tests. All N_S participants undergo a MRI scan to exclude subjects with brain injury and other unmet radiological criteria, leaving a total of M_S subjects. Finally, all M_S subjects are tested using gold standard measures (PET or CSF) looking for K amyloid positive subjects to enter the trial.

In the proposed (P) protocol (Figure 5.2b), we will add a layer of machine learning that will capitalize on already acquired images to pre-filter subjects which are more likely to be positive in the gold standard. If this approach is viable, this classifier would avoid unnecessary PET/CSF acquisitions. Since the pre-screening is not ideal, for the same number of K amyloid positive subjects we would require a larger initial pool of cognitively unimpaired subjects $N_P > N_S$. Assuming the same drop-out rate in the radiological exclusion criteria, we use all the available MRI scans for the remaining subjects to predict their amyloid status using machine learning. We then select those subjects most likely to be amyloid positive and, if our hypothesis holds true, a smaller number of subjects will remain to be tested using gold standard measures $M_P < M_S$ yielding the same number of K amyloid positive subjects in the end.

5.2.4 Savings

To study the viability of the proposed protocol, we analyze its performance in terms of overall savings that, at the same time, depends on the performance of the machine learning algorithm, the expected prevalence of amyloid positive subjects in the general population $\rho = 20\%$, the cost of the gold standard test ($C_{PET} = 3000\text{€}$), MRI acquisition ($C_{MRI} = 700\text{€}$) and clinical/cognitive assessment ($C_{COG} = 200\text{€}$).

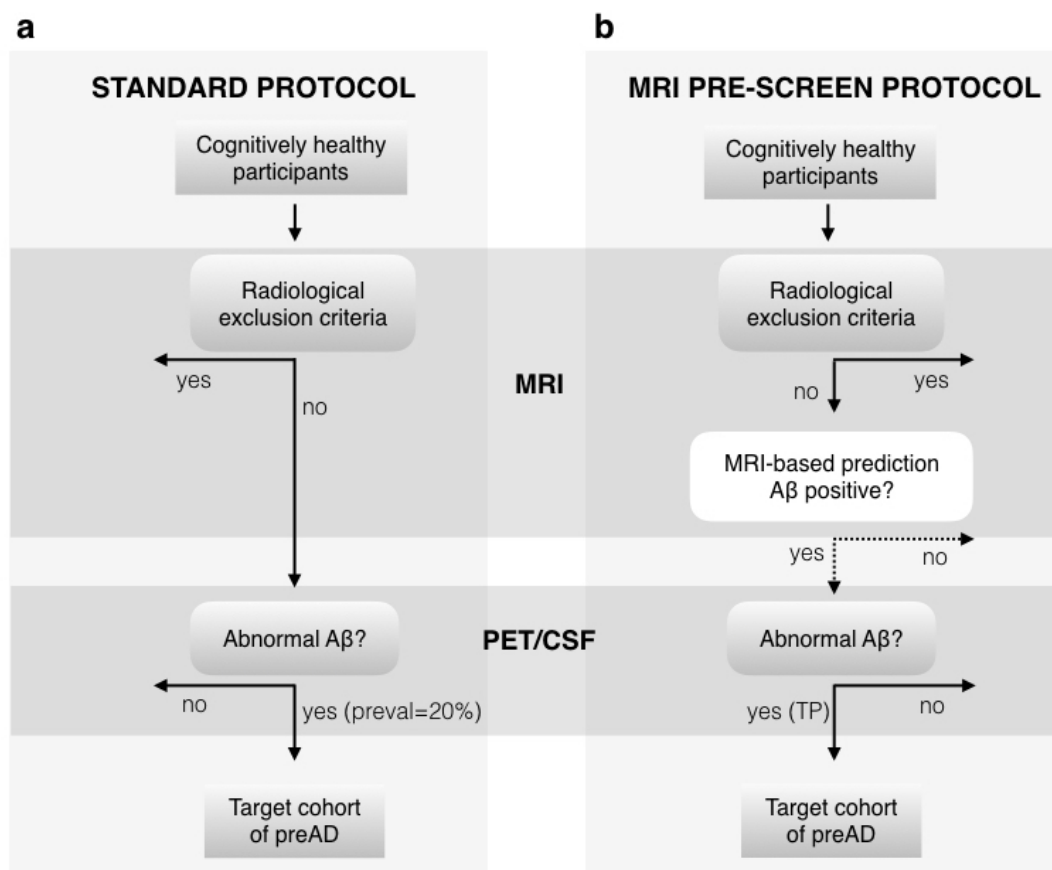


Figure 5.2: Outline of clinical recruitment protocol of subjects with preclinical (PC) Alzheimer’s pathology. (a) Standard recruitment protocol used in clinical trials. (b) Our proposed protocol with MRI-based machine learning. In the standard protocol, cognitively unimpaired participants undergo an MRI scan to identify radiological exclusion criteria (e.g. cerebrovascular disease). All cognitively unimpaired subjects with no brain injuries then participate in a PET/CSF test, and the fraction of $A\beta$ positives is determined only by the disease prevalence. The proposed protocol adds an exclusion layer in which automated MRI-based classification predicts a subset of amyloid positive subjects that are later subjected to PET/CSF acquisition. In this case, the target cohort is defined by the true positive rate ($TP\%$) of the classification algorithm. The initial participant pool needs to be larger for the proposed protocol due to classifier false negatives that may be excluded.

In statistical learning, to evaluate the performance of algorithms, the confusion matrix and its measures are often used. The confusion matrix is defined as follows:

$$C = \begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (5.1)$$

where TP are the true-positives, FP are false-positives, FN are false negatives and TN are true negatives. See that amyloid positive subjects correspond to positive samples ($P = TP + FN$) and are represented in the first row while healthy control subjects are negative examples and represented in the second row ($N = FP + TN$). Predicted amyloid positive ($TP + FP$) and predicted healthy controls ($FN + TN$) correspond to first and second columns, respectively. From the confusion matrix, other metrics arise:

$$\begin{aligned} \text{Precision (P)} &= \frac{TP}{TP + FP} \\ \text{Recall or sensitivity (R)} &= \frac{TP}{TP + FN} \\ \text{Specificiy (S)} &= \frac{TN}{TN + FP} \\ \text{Accuracy (A)} &= \frac{TP + TN}{TP + TN + FP + FN} \text{F1-score (A)} = 2 \frac{PR}{P + R} \end{aligned} \quad (5.2)$$

We develop two different models to represent a real clinical recruitment scenario: a simple model of a frictionless scenario (no drop-out) used to present preliminary results and a more complex model for further applications.

5.2.4.1 Simple model

In the standard protocol, cognitively healthy participants first undergo an MRI scan to identify radiological exclusion criteria (e.g. cerebrovascular disease). In this model we consider the ideal scenario where no subjects drop out at this stage ($M_S = N_S$) and we will tackle the non-ideal case afterwards. Hence, all subjects are then tested using either PET or CSF. The number of $A\beta$ positive subjects recruited, K , is given by the expected disease prevalence as $K = \rho \cdot N_S$, where N_S is the total number of individuals that will be screened. The overall economic cost in this scenario can be expressed as:

$$C_S = N_S \cdot (C_{PET} + C_{MRI}) = \frac{K}{\rho} \cdot (C_{PET} + C_{MRI}) \quad (5.3)$$

However, this procedure is costly and invasive and hampers large-scale studies. We propose another methodology that uses the available MR images and machine learning (ML) in the first place and provides an initial classification between healthy controls and subjects within the preclinical stage of AD. Then, only the subset of participants predicted as potential $A\beta$ positives subjects by the machine learning algorithm ($M_P = TP + FP$) undergo a verification step using a PET scan, where TP and FP are the

truly and falsely predicted positive respectively (Equation 5.1). The cost of the proposed methodology, C_P , is

$$C_P = N_P \cdot C_{MRI} + M_P \cdot C_{PET} \quad (5.4)$$

where N_P is the required number of initial participants in the proposed protocol with machine learning. The number of participants in the MRI-based screening is defined by the $A\beta$ positives ($TP + FN$) in the sample. Then, in the frictionless scenario assumed before where all subjects pass the radiological exclusion criteria ($N_P = M_P$), the total amount of participants is $N_P = (TP + FN)/\rho$, where $TP = K$ is the target number of $A\beta$ positives that need to be recruited. Using N_P , TP and the formulae for precision and recall (Equation 5.2), Equation 5.4 can be rewritten in terms of K , P (precision) and R (recall) as:

$$C_P = \frac{1}{\rho} \cdot (TP + FN) \cdot C_{MRI} + (TP + FP) \cdot C_{PET} = \frac{K}{\rho} \cdot \frac{C_{MRI}}{R} + \frac{K}{P} \cdot C_{PET} \quad (5.5)$$

$$C_P = \frac{K}{\rho} \cdot \left(\frac{C_{MRI}}{R} + \rho \frac{C_{PET}}{P} \right) \quad (5.6)$$

In addition, the number of individuals predicted $A\beta+$ that will undergo PET/CSF validation is :

$$M_P = TP + FP = \frac{TP}{P} = \frac{K}{P} \quad (5.7)$$

We can calculate the percentage savings as the differences in cost:

$$Savings_{COST} = \frac{C_S - C_P}{C_S} = 1 - \frac{1}{1 + C_{ratio}} \left(\frac{C_{ratio}}{R} + \frac{\rho}{P} \right) \quad (5.8)$$

where $C_{ratio} = \frac{C_{MRI}}{C_{PET}}$ is the cost ratio of the scans. Savings in terms of participant burden (i.e. avoided PET/CSF tests):

$$\begin{aligned} Savings_{CSF/PET} &= \frac{N_S - M_P}{N_S} \\ Savings_{CSF/PET} &= 1 - \frac{\rho}{P} \end{aligned} \quad (5.9)$$

While the number of participants that will avoid a CSF/PET due to the classifier layer depends on the classifier precision, the total number of participants that need to be screened depends on the recall: $N_P = K/R$. Thus, we use precision-recall (PR) curves to compute the savings achieved and assess the performance of the classification algorithm. We overlap the PR curves with a savings' heatmap to produce a full description of the system's performance and add several contour lines connecting points in the PR curve with constant savings.

5.2.4.2 Complex model

A more complex model is developed, taking into account a three layer of exclusion/inclusion tests that are applied sequentially: cognition/clinical assessment, MRI and PET/CSF acquisition. Cognition/clinical tests involve a visit with a doctor and cognitive tests that are designed to rule-out individuals with mental disorders that may interfere with the study outcomes (the rate of individuals excluded by this test is estimated at $e_{COG}=30\%$). An MRI scan is intended to identify incidental findings such as cerebrovascular disease. The amount of individuals excluded due to brain injury is estimated as $e_{MRI}=20\%$. Individuals who pass these requirements are considered cognitively healthy subjects with no brain injuries and they participate in a PET/CSF study to test for abnormal $A\beta$ levels (Figure 5.2a). Estimates of exam costs are:

$$C_S = N_{COG} \cdot C_{COG} + N_{MRI} \cdot C_{MRI} + N_{PET} \cdot C_{PET} \quad (5.10)$$

where C_S is the cost of standard protocol for a fixed number of K amyloid-positive subjects, $N_{PET} = M_S = \frac{K}{\rho}$ is the required number of subjects that are tested using gold standard measures, $N_{MRI} = \frac{K}{\rho \cdot (1 - e_{MRI})}$ is the required number of subjects that undergo an MRI scan taking into account a certain amount of dropout rate due to radiological exclusion criteria and $N_{COG} = \frac{K}{\rho \cdot (1 - e_{MRI}) \cdot (1 - e_{COG})}$ is the required pool size of subjects that are tested clinically taking into account a certain amount of dropout rate due to clinical exclusion criteria.

The factors $\left(\frac{1}{1 - e_{MRI}}\right)$ and $\left(\frac{1}{1 - e_{COG}}\right)$ increase the amount of subjects to be screened to compensate for the amount of subjects excluded by brain injury and abnormal cognition respectively. This can be factored in as an increment in the cost of the tests:

$$\hat{C}_{MRI} = \frac{1}{\rho \cdot (1 - e_{MRI})} \cdot C_{MRI}$$

$$\hat{C}_{COG} = \frac{1}{\rho \cdot (1 - e_{MRI}) \cdot (1 - e_{COG})} \cdot C_{COG}$$

Then:

$$C_S = \frac{K}{\rho} \left(\hat{C}_{COG} + \hat{C}_{MRI} + C_{PET} \right) \quad (5.11)$$

Our proposed protocol incorporates an exclusion layer in which automated MRI-based classification predicts a fraction of $A\beta+$ subjects that are likely to test positive in a CSF/PET scan (Figure 5.2b). The predictions amount to $M_P = TP + FP$. The following equations estimate the percentage of savings obtained by using the desired protocol instead of the standard protocol to recruit an identical number of $A\beta+$ individuals. Estimates of the exam costs take into account the machine learning classifier P (prediction) and R (recall):

$$C_P = N'_{COG} \cdot C_{COG} + N'_{MRI} \cdot C_{MRI} + N'_{PET} \cdot C_{PET} \quad (5.12)$$

where C_P is the cost of the proposed protocol and for a fixed number of K amyloid-positive subjects, $N'_{PET} = M_P = \frac{K}{P}$ is the required number of subjects that are tested

using gold standard measures, $N'_{MRI} = \frac{K}{\rho(1-e_{MRI})R}$ is the required number of subjects that undergo an MRI scan taking into account a certain amount of dropout rate due to radiological exclusion criteria and $N'_{COG} = \frac{K}{\rho(1-e_{MRI})(1-e_{COG})R}$ is the required pool size of subjects that are tested clinically taking into account a certain amount of dropout rate due to clinical exclusion criteria. Substituting those expressions, the overall cost of the proposed protocol is:

$$C_S = \frac{K}{\rho} \left(\frac{\hat{C}_{COG}}{R} + \frac{\hat{C}_{MRI}}{R} + \rho \frac{C_{PET}}{P} \right) \quad (5.13)$$

Hence, the fraction of savings can be computed as follows:

$$Savings_{COST} = \frac{C_S - C_P}{C_S} = 1 - \frac{1}{1 + C_{ratio}} \left(\frac{C_{ratio}}{R} + \frac{\rho}{P} \right) \quad (5.14)$$

where now, the cost ratio between cognitive and MRI costs and gold standard measures cost is $C_{ratio} = \frac{\hat{C}_{COG} + \hat{C}_{MRI}}{C_{PET}}$. Hence, we see that the reduction of CSF/PET tests is the same for either model, while for the savings related to cost the only difference is the cost ratio between pathological markers and other tests. In the simple model only MRI is considered, while in the more complex we account for cognitive testing and drop-out rate due to exclusion criteria, hence increasing the cost ratio.

5.3 Cross-sectional analysis

We begin by developing a proof-of-concept using cross-sectional data and study the viability of the proposed protocol using the simple savings models. We finally report the results using both savings models presented.

5.3.1 Material and methods

5.3.1.1 Subjects

CSF as well as T1 and DTI imaging data from 96 subjects of the ADNI cohort [241] has been downloaded. This subset of ADNI is selected to collect T1 and DTI images acquired on the same 3T scanner. In this sample, cognitively normal subjects are classified as PC if they had A β 42 CSF concentration below 192 pg/mL and as healthy control (HC) if else. This threshold for A β positivity (A β +) is selected as it optimally discriminates between controls and AD-dementia patients on the ADNI-independent autopsy-based AD CSF samples analysed with the multiplex xMAP Lumnex platform [170]. MCI and AD-dementia classification criteria applied in the ADNI project have been reported elsewhere [38] and are further refined by exclusively considering A β + individuals. This classification criteria results in 28 healthy controls (HC), 16 PC, 23 MCI due to AD (MCI), 29 dementia due to AD (AD-dementia) subjects.

Equivalent CSF, T1 and DTI imaging data is also acquired from a single-site cohort at the AD Unit Hospital Clinic of Barcelona (HCB) including 87 subjects (40 HC, 12 PC, 21 MCI due to AD (MCI), 14 dementia due to AD (AD-dementia)). Subjects are classified as PC if they were cognitively normal and A β +, defined according to recommended

	Age	Sex (M/F)	School years	APOE- ϵ 4 carriers	# Subjects
ADNI (All)	73.4 (5.3)	20/25	16.9 (2.5)	18	45
ADNI (HC)	72.0 (5.2)	14/15	17.2 (2.6)	8	29
ADNI (PC)	75.8 (4.6)	6/10	16.2 (2.1)	10	16
HCB (All)	64.6 (9.2)	33/19	11.6 (4.3)	19	52
HCB (HC)	61.1 (8.5)	26/14	11.9 (4.4)	13	40
HCB (PC)	72.7 (6.0)	7/5	10.6 (4.0)	6	12

Table 5.5: Demographics of the ADNI subset and HCB cohort used for this study

reference values with the INNOTEST ELISA (CSF A β 42 < 500 pg/mL) [164, 162, 90]. Criteria for MCI and AD-dementia have been previously reported [164] and, of note, only included A β + patients. The local ethics committee approved the study and all participants gave written informed consent.

5.3.1.2 Image acquisition and processing

HCB participants undergo MRI in the same 3T Siemens TrioTIM scanner providing 3D T1- weighted images with the following parameters: Repetition Time (TR) = 2300 ms, Echo Time (TE) = 2.98 ms, Inversion Time (TI) = 900 ms, Flip Angle = 9, voxel resolution of 111 mm³ on 240 sagittal slices; and two diffusion-weighted sequences: TR = 7600 ms, TE = 89 ms, Flip Angle = 90, 30 non-collinear directions (b = 1000 s/mm²) and 1 non-gradient volume (b = 0), voxel size of 2.05 x 2.05 x 2 mm³ on 60 axial slices. ADNI participants are scanned on 3T MRI devices, comprising 3D T1-weighted images: TR = 7.256 ms, TE = 2.988 ms, TI = 400 ms, Flip Angle = 11, voxels resolution of 1.02 x 1.02 x 1.2 mm³ on 196 sagittal slices; and a diffusion-weighted sequence: TR = 13000 ms, TE = 69.1 ms, Flip Angle = 90, 41 non-collinear directions (b = 1000 s/mm²) and 5 non-gradient images, voxel resolution 1.3672 x 1.3672 x 2.7 mm³ on 59 axial slices.

T1-weighted images are segmented with the Statistical Parametric Mapping (SPM) VBM8 toolbox [261] to obtain CSF, grey matter (GM), and white matter (WM) probabilistic maps. Then, a binary GM mask is formed by those voxels whose probability of belonging to GM is bigger than the probability of belonging to any other tissue. The AAL atlas [80] in the Montreal Neurological Institute (MNI) standard space is warped to each subjects T1 native space by means of ANTS [85] and using the DARTEL template obtained from the segmentation step. AAL maps are resliced to the T1 resolution using nearest neighbour interpolation and intersected with the GM mask. Volumes of all AAL regions are computed for every subject and normalized by total intracranial volume, calculated as the sum of the CSF, WM and GM masks.

Structural connectivity matrices (SCM) are also computed. To this end, AAL regions are superimposed to the DTI data using a similar protocol as the one described above. For tractography, to create a seed mask insensitive to grey matter volume changes, the interface between grey and white matter is dilated by one voxel. Then, probabilistic tractography is performed between pairs of AAL regions using FSL's ProbtrackX [86] (see further details in [90]).

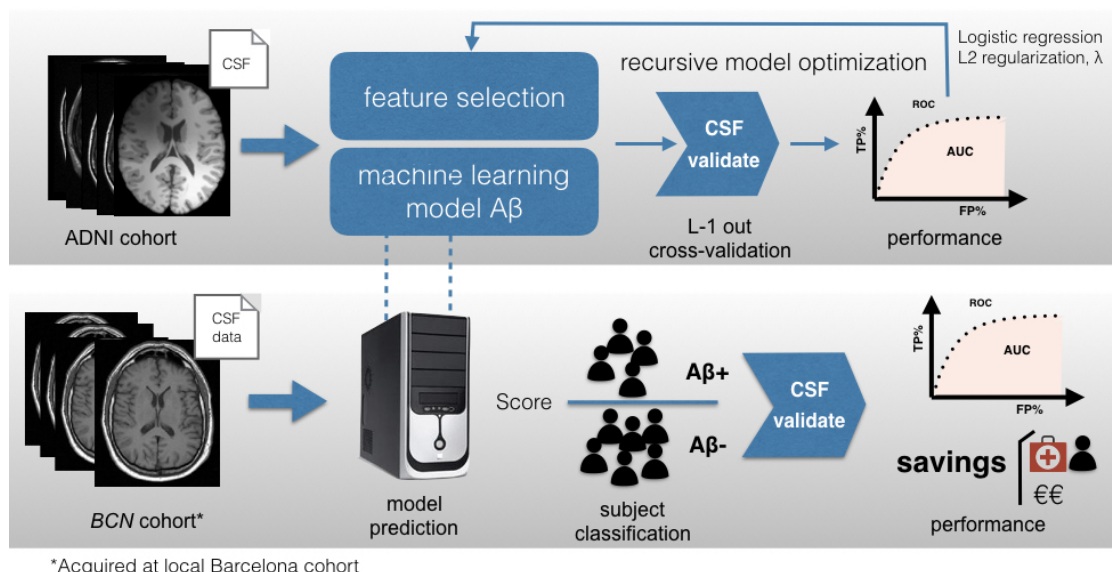


Figure 5.3: Core model training and validation for automated MRI-based classification of $A\beta$ positive ($A\beta+$) subjects using logistic regression. The ADNI dataset is used for model training. Feature extraction from GMVs of ROIs is followed by feature selection for dimensionality reduction. The parameters of the model are optimized iteratively based on the training data to maximize performance measured in terms of area under the receiver-operating characteristic curve (ROC). The final model is tested on the HBC dataset to predict $A\beta+$ subjects and validated using CSF biomarkers as ground truth. Model predictivity and performance are assessed in terms of area under the ROC and the overall savings produced in the alternative recruitment scenario that includes MRI-based classification.

5.3.1.3 Overview of machine learning methods

Our core machine learning pipeline is trained using the ADNI cohort dataset and tested with our local dataset from HBC (Figure 5.3). The performance of the overall system is validated using CSF biomarkers as gold-standard.

Feature extraction and selection. Feature selection techniques are used to retrieve a reduced set of informative features from high dimensional medical images. From each T1-weighted image, we extract a 90-dimensional vector corresponding to the grey matter volume (GMV) of each of the 90 regions of interest (ROIs) described in the AAL atlas. From each DTI image, we extract a 90x90 structural connectivity matrix (SCM) for atlas ROIs (SCM). Due to its symmetry, the upper triangular of SCM determines a 4005-feature vector.

The different acquisition protocols across medical centers introduce biases that are reflected in considerable variability in the distribution of imaging feature vectors. Thus, imaging features are normalized using z-scores computed independently for each cohort to make data comparable and ensure standardized feature weights.

Different filter feature selection methods have been compared for performance. We focus on balancing the relevance and the redundancy of the subset of features and three meth-

ods are evaluated for this purpose: F-test, mRMR-FCD and mRMR-FCQ [262]. The first two provided optimal performance in the selection of T1 and DTI imaging features respectively and were used throughout this study. We discard the use of wrapper and embedded feature selection methods since they are sensible to small sample sizes and result in high generalization error indicating severe overfitting of the training set.

Classification. Cognitively unimpaired individuals (HC and PC) in the two cohorts are used to train and test a classifier to predict dichotomous $A\beta$ status (either positive or negative). The ADNI cohort is used for training the classifier and the HCB data for test. In addition, two different classifiers are built, one based on T1 data alone and, another one, based on a combination of T1 and DTI data. Different classification schemes are tested out, including logistic regression (LR), support vector machine (SVM) with linear and gaussian kernels, random forests and k-nearest neighbours. Linear models (logistic regression and linear-SVM) are the most succesful in practice, giving similar results in practice. Finally, due to slightly better performance and easier interpretation of the model, a binary LR classifier is used with L2 regularization and a penalty term λ , optimized by a leave-one-out cross-validation strategy in the ADNI cohort. The LR classifier outputs a score for each subject that can be interpreted as the probability of being $A\beta+$ or $A\beta-$, as validated by the CSF biomarker readout. The optimized model is tested for performance in terms of the savings in cost of clinical trial recruitment. To this end, precision (fraction of positive prediction correctly classified), recall (also known as sensitivity, i.e. fraction of positive instances correctly classified) and specificity (fraction of negative instances correctly classified) are calculated using the dichotomized CSF $A\beta$ as gold standard. Precision-recall (PR) and receiver-operating characteristic (ROC) curves are also computed to evaluate method performance. For a target normal population of mean age 65, the prevalence of amyloid pathology is estimated to be 20% [55]. We employ a bootstrapping approach to simulate the required disease prevalence and iteratively repeat the classification protocol, yielding different splits of the test set at each run. For that purpose, at each iteration we randomly select N_{test} subjects form the test set with an HC/PC ratio fixed by the prevalence ρ . As a parallel analysis and for the sake of comparability with previously reported AD-dementia/MCI classifiers, an “optimized diagnostic” classifier is derived with an identical pipeline and using HC, MCI and AD-dementia subjects in the two cohorts to discriminate between HC-MCI and HC-AD-dementia (for further details see Appendix B.1.1). Finally, we also assess the performance of the classifier optimized to discriminate PC from HC to classify between of HC vs. clinical diagnostic categories (MCI and AD-dementia).

5.3.2 Results

First we report our standardization protocol to make data actionable across training and testing cohorts (Section 5.3.2.1). Then, we evaluate the performance of a classifier based on GMV from T1-weighted images only (Section 5.3.2.2) and an extended classifier based on a combination of T1 and DWI features (Section 5.3.2.3). Demographic characteristics of the studied samples are shown in Table 5.5.

5.3.2.1 Data normalization

Following normalization of the raw imaging features from the training cohort (ADNI) and test cohort (HCB), the resulting distributions of z-score features are confirmed as match-

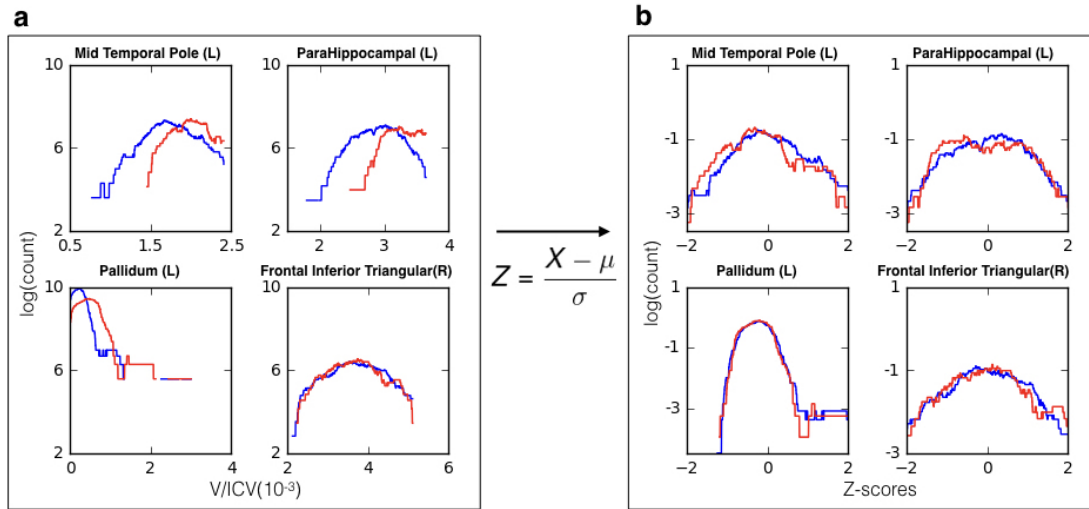


Figure 5.4: Examples of distributions of T1 features for three GMV ROIs (a) before and (b) after the z-score normalization. The ADNI and HBC cohort distributions are shown in red and blue respectively.

Features	HC vs. PC				
	AUC	Precision	Recall	Savings% (Cost)	Savings% (CSF/PET)
10	0.65	0.48	0.75	40.9	58.3
20	0.76	0.54	0.83	47.1	62.9
30	0.70	0.37	0.83	33.4	45.9
40	0.65	0.42	0.66	32.7	52.3
50	0.61	0.36	0.58	22.3	44.4

Table 5.6: Performance of the PC classification model according to the number of features selected (10, 20, 30, 40, 50). In the classification HC vs PC, optimal performance (AUC = 0.76) is achieved with 20 features and the optimal P-R threshold that produces the highest savings is highlighted.

ing distributions under the Kolmogorov-Smirnov (KS) test ($p \leq 0.05$) (Figure 5.4).

5.3.2.2 Classification of PC subjects using T1-weighted images

F-test selection provides the most discriminative features in the classification HC vs PC based on ADNI data (Figure 5.5, Appendix B.1.2). The classification algorithm provides a ranking and a weight for these features (Figure 5.5) based on their relevance. Maximal area under the ROC curve (AUC) and maximal clinical recruitment savings can be achieved for 20 T1 features (Table 5.6).

Using the complex model presented in 5.2.4.2, for a machine learning classifier with $P=0.54$, $R=0.83$, the savings in the cost of recruiting 300 subjects amounts to: 34%. In terms of amount of CSF/PET scans saved it is kept to 62.9%. as it only depends on the preclinical prevalence and the precision of the method.

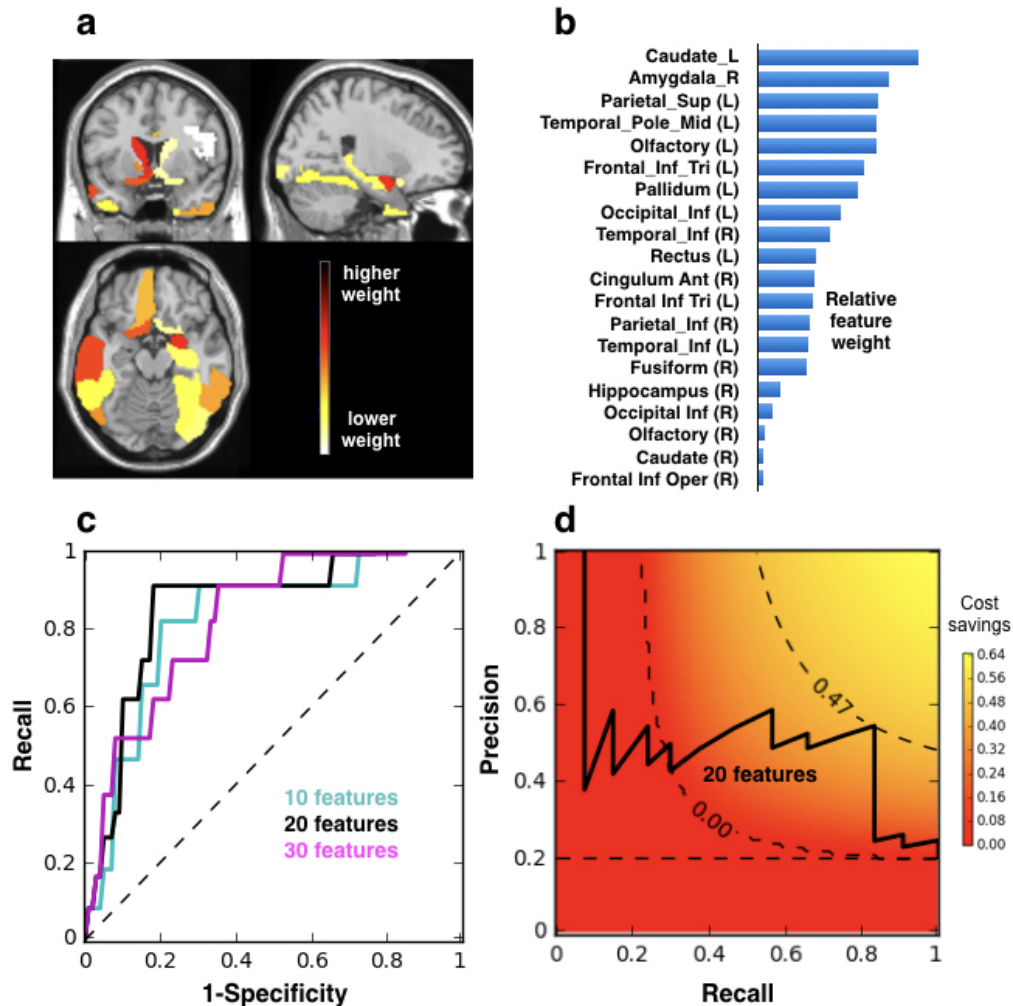


Figure 5.5: (a) ROI Features selected for the best performing method (F-test, 20 features). (b) Ranked list of most informative features according to the weights provided by the LR decision function (ADNI cohort). (c) ROC analysis of the method performance for 10, 20, and 30 top ranking features based on the validation cohort (HCB). (d) $Savings_{COST}$ as a function of precision-recall calculated based on the optimal number of features ($n = 20$) applied to the HCB cohort. Color scale and dashed contour lines indicate savings. In particular, a contour line with $savings = 0.001$ demarks the boundary of area of the precision-recall curve above which $savings > 0$. Optimal PR threshold provides the highest savings ($P = 0.54$, $R = 0.83$, $Savings = 47\%$). Horizontal dotted line at $P=0.2$ denotes savings of a random classifier, note that this precision value is equal to the prevalence of the PCs population.

Features	HC vs. PC				
	AUC	Precision	Recall	Savings% (Cost)	Savings% (CSF/PET)
10	0.70	0.41	0.68	32.6	51.2
20	0.65	0.60	0.59	40.9	66.6
30	0.70	0.54	0.59	37.9	62.9
40	0.69	0.83	0.51	43.3	75.9
45	0.69	0.85	0.60	49.3	76.5
50	0.70	0.66	0.60	43.8	69.7

Table 5.7: Performance of classification according to the number of features selected (10, 20, 30, 40, 45, 50). Optimal performance (AUC = 0.69) is provided by 45 features. The optimal PR threshold that produces the highest savings is listed ($Savings_{COST} = 49\%$, $Savings_{CSF/PET} = 76.5\%$).

5.3.2.3 Classification of PC subjects using multimodal feature vectors

Multimodal classification is performed combining imaging features from T1 and DTI. Z-score transformed features concatenated from both modalities yield 4095-dimensional feature vectors. The optimal PR threshold that yields the highest savings depends on the number of features used in the classifier (Table 5.7, Figure 5.6). Feature selection with the mRm-FCD [262] method results in a 45-dimensional feature vector (features listed in Appendix B.1.3). Notably, only 2 out of these 45 most informative features are T1 features (feature 18: Olfactory (L), feature 23: Pallidum), while 43 are DTI feature-pairs. This result suggests that DTI features may be more informative than T1 features for identifying preclinical subjects. Figure 5.6 shows the most DTI-connected regions, given by the weighted sum of the DTI-feature pairs and reflects their relevance in the identification of PC. For example, the Pallidum is highlighted as a ROI with multiple connections as well as by itself as a GMV feature. Remarkably, almost same performance (AUC) and cost savings are obtained for both T1 and the multimodal classifiers but precision is higher ($P = 0.85$) in the multimodal model than in the T1 case ($P = 0.54$) resulting in larger savings with regards to participant burden (76.5%). The description of the association between CSF A42 levels and tractography falls beyond the interest of this work as it has been described elsewhere [90]

Using the complex model presented in 5.2.4.2, for a machine learning classifier with $P=0.85$, $R=0.60$, the savings in the cost of recruiting 300 subjects amounts to: 36%. In terms of amount of CSF/PET scans saved it is kept to 76% as it only depends on the preclinical prevalence and the precision of the method.

5.3.3 Discussion

Our work investigates the automated screening of PC individuals using machine learning methods on features derived from structural MRI. We present the potential integration of such a technology as a pre-screening tool for the recruitment of subjects in secondary prevention trials and present a framework to quantify the benefits of such approach in economic and participant burden terms. Noteworthy, it is not proposed to replace the gold standard PET/CSF tests, but to provide a prediction of amyloid positivity in order to avoid unnecessary PET/CSF tests and, thus, reduce participant burden and economic

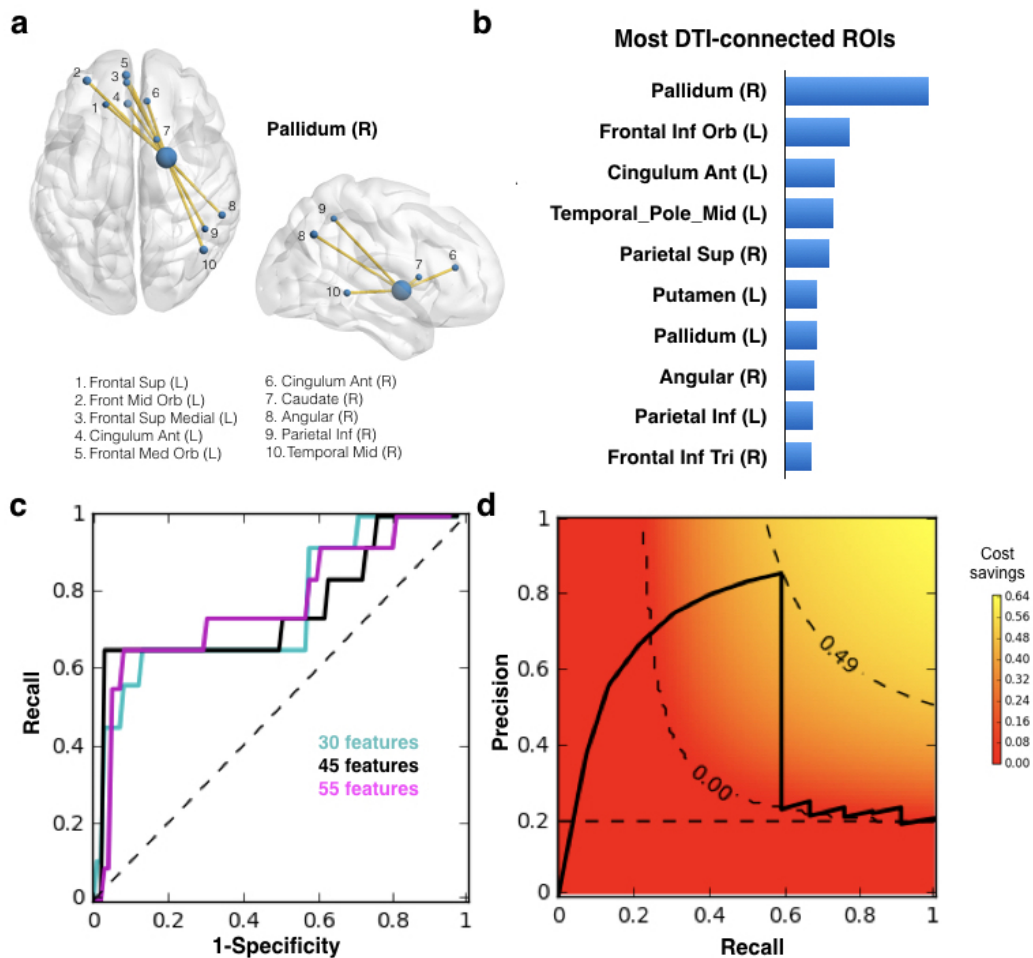


Figure 5.6: (a) Example ROI DTI connections to Pallidum as feature-pairs in the multimodal vector for machine learning. (b) Ranked list of most connected ROIs according to weighted sum of features from the SCM matrix based on the training set ADNI. Weights are provided by the LR decision function. (c) ROC analysis of the method performance for 30, 44, and 55 top ranking features for the multimodal T1+DTI feature vector based on validation dataset HCB. (d) SavingsCOST as a function of precision-recall calculated based on the optimal number of features ($n = 45$) calculated on the validation dataset HCB. Color scale and dashed contour lines indicate savings. Contour line with savings=0.001 demarks the boundary of the meaningful area of the precision-recall curve below which all savings=0. Optimal P vs R pairing provides the highest savings ($P = 0.85$, $R = 0.60$, $Savings_{COST} = 49\%$). Horizontal dotted line at $P=0.2$ denotes savings of a random classifier.

costs.

Validation across cohorts. We utilize public MRI and CSF data available from ADNI for feature selection and model training. Our independent HCB cohort serves for model testing, simulating a real trial recruitment scenario. The data normalization scheme enables validation overcoming the differences between data acquisition protocols and scanners in both institutions. Thus, the relevance of our models can be generalized to other population samples, with different age and education ranges. Of note, performance of the algorithm is stable when reversing the training and test datasets. Therefore, we foresee that the performance of our PC classification will hold when prospectively applied to independent cohorts with practical utility for PC screening purposes.

Feature selection. By design, our data-driven approach is agnostic to previously reported anatomical changes due to AD. This allows us to validate the selected features given their role in preclinical and prodromal stages of AD according to the literature. As expected, many of the features found informative in PC classification relate to the established AD signature. In the case of T1, characteristic AD affected regions include atlas ROIs such as the hippocampus and amygdala, temporal and parietal areas as well as the cingulum, thalamus, rectus and fusiform giri (Figure 5.5). Regions such as olfactory bulb pertain to well documented alterations of the olfactory system in the pathogenesis of AD and other dementias [209]. Other regions showing a high discriminative power such as the caudates, have been found altered in cognitively normal individuals at high genetic risk of developing AD [263, 264]. Finally, the role of other regions such the globus pallidus is not as clear in PC (Figure 5.6). One explanation would be that this atlas ROI includes the nucleus basalis of Meynert, which is an adjacent region not explicitly identified in the AAL parcellation. In a previous voxelwise analysis of the HCB cohort, the nucleus basalis showed alterations in PC [162] and, moreover, this region has been consistently reported to degenerate in early AD stages [128]. Interestingly, the features that classify between HC and PC yield an $AUC < 0.2$ to discriminate among HC and the diagnostic categories (i.e. MCI, AD-dementia). An AUC significantly departing from 0.5 is classifying de facto. However, being well below 0.5 suggests that some features that characterize PC from HC appear in the opposite direction as some of the features that classify HC vs. AD-dementia. This interpretation is further reinforced by the fact that classification models built using an identical pipeline but trained specifically to recognize diagnostic categories perform accordingly to what has been reported in the literature ($AUC=0.95$ for 20 T1 features, $P=1.0$, $R=0.76$), and are based mainly on hippocampal and temporal lobe volumes (Appendix B.1.1). Future work should improve the identification of discriminant brain regions using voxelwise approaches to fully characterize the preclinical AD signature using MRI and to identify overlaps with the established AD signature [264]. Overall, our results show that feature selection has prioritized DTI features over T1 features in a multimodal imaging vector. Thus, DTI feature-pairs may be more robust than T1 features, compensating for the lower image quality (resolution, noise, blur) and artifacts given by surrogate measures of brain anatomical connectivity in DTI. Also, DTI conveys complementary information from white matter connectivity. Our data raises the question of whether changes in connectivity may be more evident in early AD than volumetric changes [90]. Nevertheless, the relevance of DTI over T1 features observed in this study should be further confirmed in other datasets. Despite its similar performance, the multimodal model is more precise than the T1-based model, which translates into less false positives and therefore less unnecessary PET/CSF tests

	Simple protocol		Complex protocol	
	T1	T1 + DTI	T1	T1 + DTI
Savings in cost (€)	47%	49%	34%	36%
Savings in PET/CSF	63%	76%	63%	76%

Table 5.8: Savings summary using the simple and the complex model for population screening in clinical trials.

in exchange for more MRI scans given the low recall of the model.

Range of application and PC prevalence. In this proof-of-concept study, our MRI-based classifier thresholds are optimized to maximize the cost savings. The model can be tuned for other applications such as minimizing participant burden (Equation 5.9). The practical utility of our MRI-based screening protocol highly depends on the prevalence of PC subjects, which in turn, is intimately related with the age range and age distribution of the clinical study. At lower disease prevalence, savings rely on higher classifier precision to make the MRI-based screening layer worthwhile (Equation 5.8). This proposed recruitment protocol is most efficient for younger cohorts which are also best suited for secondary intervention trials.

Precision vs. recall. Higher classifier precision reduces the number of unnecessary PET/CSF acquisitions (less false-positives). Higher classifier recall reduces the size of the initial participant pool for which MRI images need to be acquired by limiting the amount of false negatives that are missed in the classification. Ensuring a cognitively healthy participant pool also incurs associated costs (e.g. additional cognitive/clinical tests) that scale up with the number of participants. While it is true that MRI-based screening relies on recruiting a larger initial cognitively healthy participant pool and therefore more MRI scans and cognition tests, we estimate that the proposed protocol is still more efficient in terms of saving economic and participant burden.

Example of application. In a standard clinical trial recruitment protocol, if the prevalence of PC is 20%, 1500 individuals need to undergo CSF/PET test in order to identify 300 $A\beta+$ subjects. First, a preliminary MRI scan is acquired to exclude participants displaying incidental findings or other neuroradiological exclusion criteria. Instead, a modest MRI-based classifier (0.54 precision/0.84 recall), requires 1807 MRI images to predict an identical pool of 300 $A\beta$ positive individuals. From these images, only 555 will be predicted positive and sent for CSF/PET validation, with 300 resulting to be true positive and 255 will be false positives. This second protocol would spare 945 PET/CSF tests, in exchange for the additional 307 MRIs required to compensate for the false negatives missed by the classifier. In economic terms, the savings provided by the MRI-screening layer can prevent up to 63% unnecessary PET/CSF scans and save 47% of the cost as compared to the standard protocol that relies only the gold standard CSF/PET measures. Smaller figures are reported if using the more complex model introduced in Section 5.2.4.2, as seen in Table 5.8

Comparison to other approaches to detecting $A\beta$. At present, blood-based approaches to detecting $A\beta+$ are being actively pursued [265, 266, 267]. These methods show a high concordance (AUCs 0.80-0.90) between plasma $A\beta$ species against PET imaging.

Unlike our PC group that only included cognitively normal subjects, the performance of these methods was estimated including MCI and/or AD-dementia patients which might have favoured them. Nevertheless, these works demonstrate a link between cerebral and peripheral $A\beta$ alterations and, when validated and automated, they will represent a major resource to screen the general population and implement secondary AD prevention strategies. On the other hand, our MRI-based approach might be particularly useful in the context of research studies in which CSF and or PET imaging is envisaged, as in this case MRI scans are typically prescribed for safety reasons.

Limitations. The main limitation in our study is the small sample size of the studied groups. This is due to the limitation of available T1 and DTI data of the same individuals in the two cohorts. Nevertheless, previous reports on standardizing the evaluation of algorithms for computed-aided diagnosis of dementia based on structural MRI recommend using small training sets ($N \approx 30$) to mimic the clinical setting where limited data is available and to avoid overtraining issues [268]. On the other hand, to maximize the generalizability of the developed methods, they prioritized applying the algorithms to independent datasets with different demographics and scanning parameters. Based on this rationale, we prioritised testing the algorithm in independent cohorts over the total sample size. Nevertheless, future studies will focus on validating similar approaches in larger test sets to better account for sample demographics and confounders. To this end, and given the results in this report, T1 data might solely achieve similar levels of classification accuracy. In addition, the small sample size has limited the complexity of the machine learning schemes that could be implemented robustly. Future studies with bigger samples sizes will also allow the investigation of potential improvements in classification performance of more complex classifying schemes. Nevertheless, we consider this work as a promising proof-of-concept as, even with a simple logistic regression model, robust performance can be achieved across two independent datasets with distinct demographic characteristics, potentially leading to significant savings of gold-standard procedures.

Another limitation of our experimental approach is that thresholds for $A\beta+$ might not be completely equivalent across the two cohorts. The respective CSF $A\beta$ cut-offs have been independently validated to yield optimal discrimination between HC and AD-dementia patients on the respective analytical platforms (INNOTEST ELISA for HCB and xMAP-Luminex for ADNI). While absolute values are definitely not comparable, the correlation between them has been shown to be very high ($r > 0.75$ [269, 270, 271]). Nevertheless, according to published conversion equations, the respective thresholds are within the interlaboratory reproducibility of the analytical platforms which fall in the range of 15-30% [272]. Based on this rationale, we prefer to use published and validated cut-offs for comparability with previous reports with the respective datasets. A similar limitation may arise from equating positivity based on CSF and PET data unless cut-offs are defined against modalities, like in [273].

5.4 Longitudinal analysis

In this section, we study how amyloid positivity affects rate of change in brain morphological features. Moreover, we aim at finding the minimal follow-up time to detect substantial changes in brain structure.

Number of visits	Number of subjects
2	295
3	63
4	27
5	15
6	3
Total	403 subjects and 841 Jacobian maps

Table 5.9: Distribution of the number of 3D-T1 MRI acquisitions per subject.

5.4.1 Material and methods

5.4.1.1 Subjects

Subjects for this study are selected from the ADNI database [241] provided that they have two or more longitudinal 3D-T1 MRI acquisitions and cerebrospinal fluid (CSF) biomarker data publicly available. Subjects are assigned biomarker-assisted diagnostic categories following recently published guidelines [16]. Subjects labelled as “Normal” in ADNI were classified as amyloid negative cognitively unimpaired healthy controls (HC) if CSF $A\beta$ was above 192 pg/mL and preclinical (PC) if CSF $A\beta$ was below 192 pg/mL. Subjects are categorized as MCI due to AD (MCI) or dementia due to AD (AD-dementia) according to the ADNI diagnostic categories reported in [38] and we select only those individuals with CSF $A\beta$ levels below 192 pg/mL to exclude subjects harboring non-AD pathological changes. At baseline, this diagnostic algorithm yield 79 HC, 50 PC and 274 MCI/AD-dementia, a total of 403 subjects with complete imaging and CSF data. As an additional inclusion criteria, in follow-up visits, all subjects must remain stable in the same diagnostic category. We exclude subjects that progress between diagnostic categories within the time span of the study due to small sample size (13 PC converters from HC, 13 MCI/AD-dementia converters from PC and 1 MCI/AD-dementia converter from HC).

5.4.1.2 MRI data

Structural 3D-T1 MRI images are acquired across different scanners and institutions. Each image is associated with a cognition score and a set of CSF biomarker values (amyloid-beta, total tau and phosphorylated tau). The date of the CSF extraction is selected to be within 90 days from the date of the MRI scan. Each subject has at least one follow-up visit with the corresponding T1-MRI image, cognition score and CSF biomarker values. The number of visits may differ across subjects (Table 5.10). The total number of MRI scans analyzed is 980. The time interval between visits is, at least, 6 months apart.

Image Analysis The SPM12 [261] neuroimaging software suit is used for every step of this longitudinal analysis pipeline. All image pairs corresponding to the same subject from the ADNI database are processed with longitudinal pairwise registration. Images in each pair are averaged and their respective Jacobian determinant is calculated, which reflects the regional cerebral volumetric changes between the respective time-points. DARTEL normalization [274] is applied onto average images to normalize Jacobian determi-

nant maps to MNI space [275] and allows comparison across subjects. The intensity of each voxel in the Jacobian image is normalized by the interval of time between reference and follow-up visits (i.e. Δt). The number of Jacobian determinant maps for each subjects diagnostic category is: 184 HC, 114 PC and 543 MCI/AD-dementia. On top of the voxelwise analysis, a regional analysis is also performed. To this end, regions of interest (ROIs) in the AAL atlas are masked by each subjects grey matter segmentation and the mean value of the remaining voxels intensity per region is computed [80].

5.4.1.3 Automated recognition of PC volumetric changes using machine learning

All Jacobian determinant maps from each subject are labeled using subjects label (i.e. PC, HC), leaving a study cohort of $N = 129$ ($N_{HC} = 79$, $N_{PC} = 50$). Importantly, as mentioned before, we only consider pairs of images for which no transitions have been observed across categories. This analysis is performed only on the PC and HC subjects.

Feature selection: Due to the limited sample size and high-dimensionality of the Jacobian determinant maps, we perform feature selection to keep an optimal percentage of the most relevant features. To this end, we use a filter feature selection method based on F-test, taking into consideration Jacobian features and subject labels. The F-test metric is used to create a ranking of all Jacobian features and finally a fixed percentage of the highly ranked features are used for classification [117].

Classification and performance evaluation: Ridge logistic regression with hyperparameter C [276] is used for binary classification of Jacobian features within the nested cross-validation (CV) framework [125] defined in Figure 5.7. It consists of an inner CV loop for model selection and an outer CV loop for assessing model performance. First, in the outer loop, subjects are randomly divided into 80% train set and 20% test set previously fixing a prevalence of interest (the percentage of samples of the amyloid positive class). For each subject in either set, all available Jacobian determinant maps are used for classification. The train set is used for feature selection and model optimization while the test set is left out for final model evaluation. The random split by subject ensures that there is no contamination of the test set with Jacobian determinants of the train set.

Feature selection is computed using only the train set. In the model optimization step, the train set is further split into sub-train (2/3) and validation (1/3) sets using a (k=3)-fold cross-validation. A grid search strategy is used to optimize the classifier hyperparameter C by maximizing the f1-score on the validation set. Finally, the model is estimated using the optimized hyperparameter C on the whole train set. Then, the model is applied to the test set to compute standard performance metrics (i.e. area under the receiver operating curve (AUC), accuracy, precision, recall/sensitivity, specificity and f1-score). We also report the reduction of economic cost (i.e. savings) of using this classification framework as a tool for AD screening. This procedure is repeated $n=100$ times and performance results are reported using the average and standard deviation.

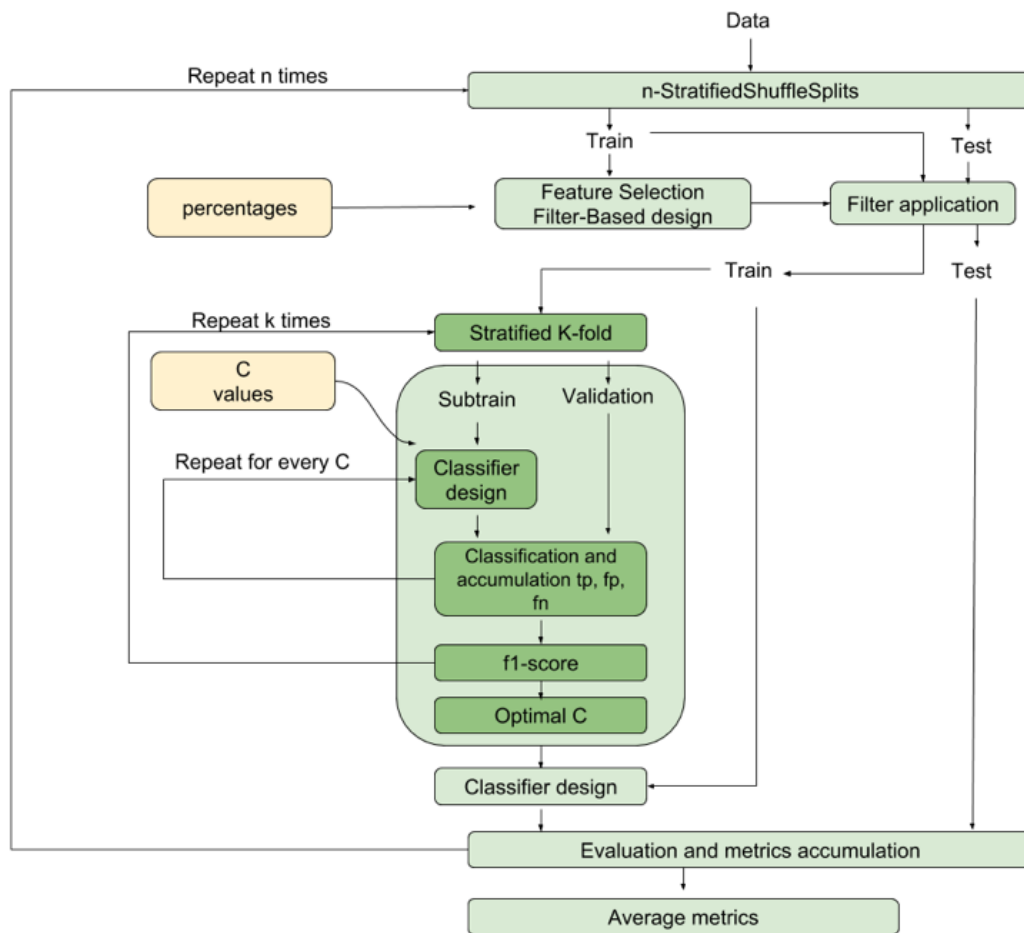


Figure 5.7: Workflow of the optimization and evaluation of the classification method. The performance of the final classifier is evaluated on a fresh test set that has not been used for training.

5.4.1.4 Statistical Analysis

The aim of the statistical analysis is to identify significant group differences in brain volumetric change rate between AD stages. We investigate the location of these stage-specific changes, and whether they represent a volume increase (positive changes) or decrease (negative changes). Every Jacobian determinant map is treated as an independent variable.

Two-sample t-test: Statistical analyses are performed by comparing any combination of two subject categories. The uncorrected threshold for statistical significance is $p < 0.005$. Spatial clustering of regions with statistically relevant voxels is applied to rule out false-positives, with a clustering threshold of $k > 100$ voxels under which voxel clusters with smaller sizes are discarded.

Data normalization: The effects of normal ageing on brain structural changes is considered as a confounder and regressed out [277]. Coefficients for linear regression on age are fitted using only HC (i.e. individuals that are amyloid negative, asymptomatic in

Category	HC ($A\beta^-$)	PC ($A\beta^+$)	MCI ($A\beta^+$)	AD-dementia ($A\beta^+$)	MCI/AD-dementia ($A\beta^+$)
Number of subjects	79	50	196	78	274
Age (years) at baseline (mean; std)	73.97 (5.97)	76.04 (6.25)	73.55 (6.55)	75.44 (7.37)	74.1 (6.85)
Sex (F/M)	37/42	21/29	79/117	33/45	112/162
Follow-up (years) period (mean; std)	2.48 (1.38)	2.32 (1.32)	2.2 (1.09)	1.4 (0.46)	1.97 (1.02)

Table 5.10: Dataset demographics at baseline.

all visits). The age corresponding to each Jacobian determinant is defined as the mean age between the two visits i.e. $age = (age_{reference} + age_{follow-up})/2$.

5.4.2 Results

5.4.2.1 Demographic and follow-up comparisons

We include a total of 430 subjects at baseline with at least one follow-up visit over three categories: HC (n=79), PC (n=50) and MCI/AD-dementia (n=274). Demographic data and follow-up period is presented in Table 5.10 split into different categories.

We denote as Δt the time interval between two follow-up visits (i.e. reference and target images). The distribution of the time interval (Δt) between follow-up visits on all subjects is given in Figure 5.8. The median of the distribution is 2.01 years. A subset of the cohort for which $3.5 > \Delta t > 2.5$ years is used for some of the machine learning studies, given that longer Δt account for more signal-to-noise ratio into disease progression. Demographics for this subset of subjects is provided in Table 5.11. Supplementary Table B.14 provides information about the research facility and type of scanner that were used for each of the subjects in this reduced cohort.

5.4.2.2 Machine learning

We use machine learning for voxelwise prediction of amyloid positive subjects (PC) among cognitively unimpaired subjects. A realistic prevalence for PC subjects on middle-age adults is 20% [55]. We use this prevalence to fix the ratio of PC subjects in the test set on all machine learning experiments. Another key parameter of the analysis is the temporal distance (Δt) between reference and target images used to compute the Jacobian determinant maps. In Figure 5.9, we report the performance of the classifier as a function of minimal Δt values in the test set. It is observed that even though we normalize each Jacobian determinant map with respect to the Δt parameter, the preclinical signature is within the detection range when visits are at least 2.5 years apart. In the case in which $\Delta t > 2.5$ years, the performance of the classifier based on structural changes is much better than a classifier trained on individual images as reported in the previous

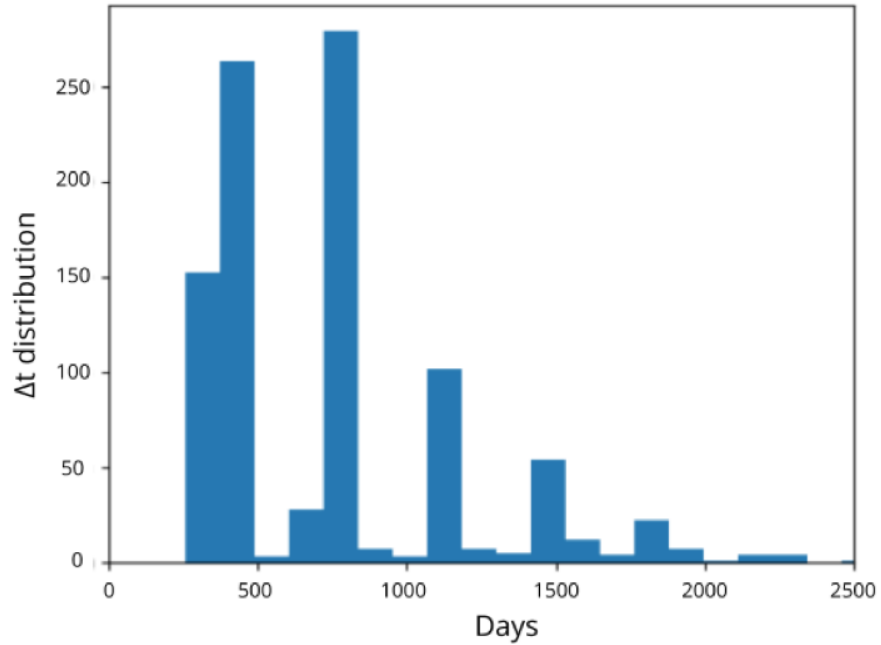


Figure 5.8: Distribution of the interval Δt between reference and follow-up visits across the whole dataset.

Category	HC ($A\beta^-$)	PC ($A\beta^+$)	MCI ($A\beta^+$)	AD-dementia ($A\beta^+$)	MCI/AD-dementia ($A\beta^+$)
Number of subjects	15	10	38	0	38
Age (years) at baseline (mean; std)	76.51 (6.18)	76.0 (3.97)	72.987 (6.11)	-	72.987 (6.11)
Sex (F/M)	8/7	5/5	27/11	-	27/11
Follow-up (years) period (mean; std)	4.17 (1.035)	4.21 (0.98)	3.91 (0.82)	-	3.91 (0.82)

Table 5.11: Demographics of the subset of the study cohort for which $\Delta t > 2.5$ used for machine learning classification

cross-sectional study that reports an $AUC = 0.76$ (Section 5.3). When using Jacobian determinant maps with smaller temporal distance ($\Delta t < 2.5$ years), the mean performance is worse than the cross-sectional analysis, probably due to the low signal to noise ratio between the changes due to normal brain aging and the changes due to amyloid positivity. The optimal temporal span in terms of AUC and savings between data acquisitions is $\Delta t > 2.5$ years. The number of subjects with follow-up visits between $2.5 < \Delta t < 3.5$ years from baseline is reduced to 15 HC, 10 PC and 38 MCI/AD-dementia subjects with 25, 16 and 52 Jacobian determinant maps respectively. In what follows, throughout the chapter, we use only Jacobian determinant maps within the optimal temporal span ($2.5 < \Delta t < 3.5$ years) for evaluation purposes. The use of jacobians within this temporal span ($2.5 < \Delta t < 3.5$ years) for training the system and evaluating it in all other cases has also been tested, with poor generalization (Supplementary Table B.17). Receiver operating characteristic curve (ROC) and precision-recall (PR) curves of the classifier are shown in Figure 5.10. A savings heatmap that responds to Equation 5.8 is overlaid on the PR curve, while the mean and standard deviation of the model performance are plotted against the random classifier on the ROC curve.

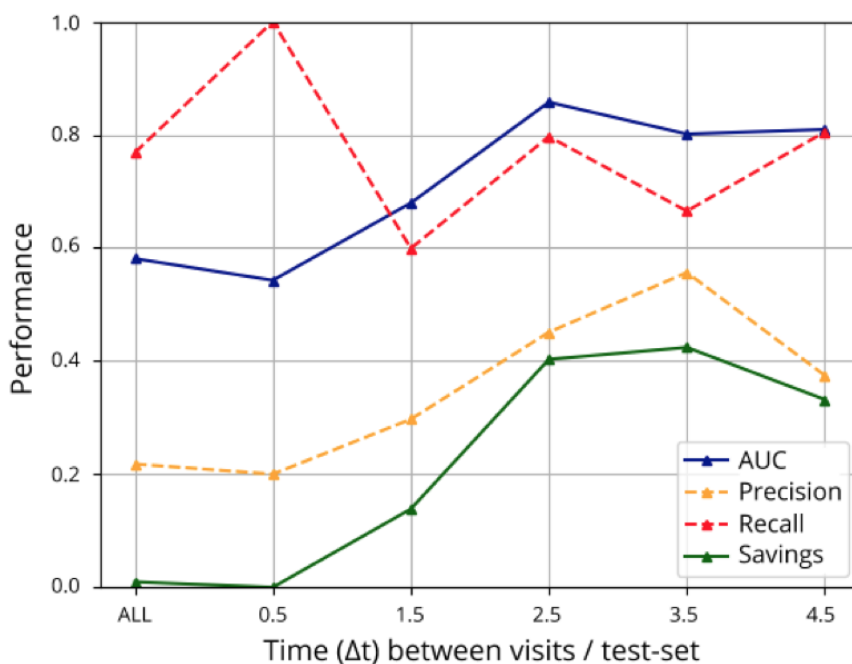


Figure 5.9: AUC and savings (blue, green) reported using Jacobian determinant maps with different time intervals (Δt) between reference and target and a fixed prevalence of $\rho = 20\%$ amyloid positive subjects on the test set. To compute savings we use optimal precision and recall values plotted in dashed orange and red lines, respectively using the cost function defined in Equation 5.8.

The impact of different number of features used to train our multivariate algorithm is presented in Table 5.12, evaluated on our dataset which is imbalanced (36% of preclinical subjects). Note that the prevalence of preclinical subjects on the test set is forced to 20% using permutations. When using a low number of features, the model underrepresents the preclinical signature, not being able to capture all data heterogeneity. In contrast, when using a high number of features, the model is not able to generalize results to unseen

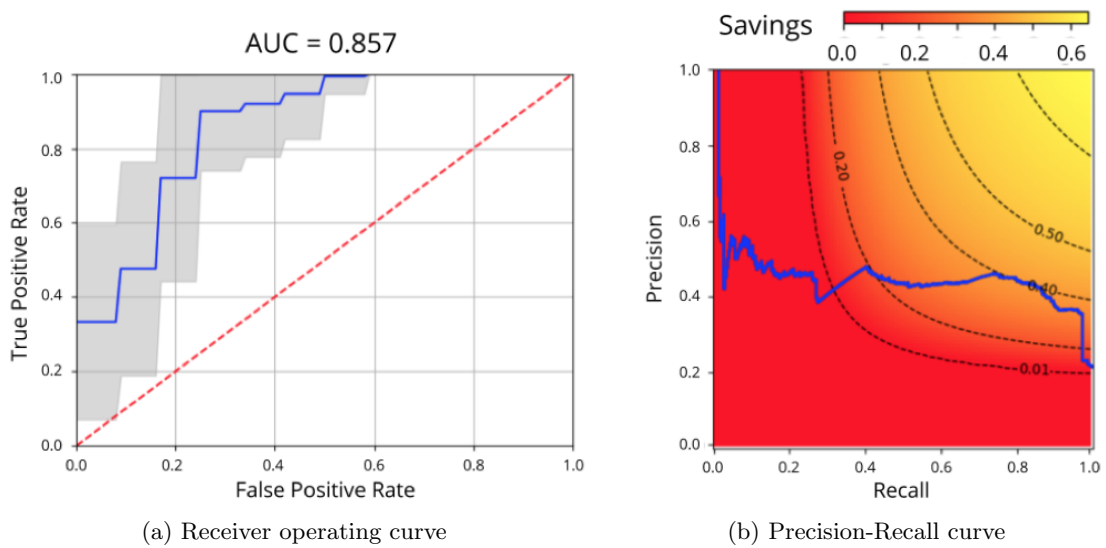


Figure 5.10: ROC and PR curves for Jacobian determinant maps with time spans in the range $2.5 < \Delta t < 3.5$ years using 0.5% of the features. On the left, the ROC curve is averaged across different development/test splits: the mean curve (blue) with the standard deviation (gray) and the curve of a random classifier (red). On the right, the PR curve of the classifier (blue) is overlaid on a savings heatmap (Equation 5.8). Black lines indicate points of equal savings.

Jacobian determinant maps, overfitting the development set. Hence, the best results are obtained using a moderate number of features that are able to both represent the preclinical signature and still generalize well to the test set. We also test an embedded, multivariate feature selection method based on ℓ_1 -norm minimization resulting in lower performance (Supplementary Table B.16).

An optimal compromised solution between several metrics is to design our model using a 0.5% of the total Jacobian features. In this case, after the 100 iterations of the nested cross-validation framework, a heatmap of selected features is shown in Figure 5.11. As expected, the top selected features correspond to typical regions affected by AD pathology like the caudates, fusiform or parahippocampal gyrus, presenting high overlap with the statistical analysis presented in the next section. This result shows that a machine learning classifier trained on changes in specific brain regions has the capacity to predict the presence of early amyloid pathology in asymptomatic individuals as measured by MRI.

5.4.2.3 Volumetric changes in preclinical stage of AD

In parallel to the machine learning classification model, we carry out a voxel-wise statistical analysis using the full dataset of Jacobian determinant maps to identify the regions of volumetric change that are statistically significant between the different categories HC, PC and MCI/AD-dementia (Figure 5.12). Stable PC individuals show significantly higher grey matter (GM) atrophy in the parahippocampal and fusiform gyri as compared to amyloid negative cognitively unimpaired subjects, as shown on the left hand side of Figure 5.12. Apparent mild GM increments are detected in the caudate heads, probably

# features (%)	AUC (95% CI)	B.acc (95% CI)	Acc. (95% CI)	Prec. (95% CI)	Sens. (95% CI)	Spec. (95% CI)	F-score (95% CI)
6 (0.00)	0.78 (0.50-0.99)	0.70 (0.375-0.875)	0.57 (0.20-0.80)	0.33 (0.13-0.50)	0.91 (0.33-1.00)	0.48 (0-0.75)	0.48 (0.19-0.67)
65 (0.01)	0.81 (0.60-0.97)	0.74 (0.478-0.834)	0.63 (0.27-0.73)	0.38 (0.18-0.43)	0.91 (0.33-1.00)	0.56 (0.12-0.75)	0.50 (0.28-0.6)
653 (0.10)	0.85 (0.67-1.0)	0.77 (0.60 - 0.88)	0.65 (0.53-0.8)	0.37 (0.27-0.50)	0.97 (0.67-1.00)	0.57 (0.42-0.75)	0.53 (0.38-0.67)
1633 (0.25)	0.86 (0.72-1.00)	0.77 (0.67-0.88)	0.65 (0.46-0.8)	0.37 (0.27-0.50)	0.98 (0.67-1.00)	0.53 (0.33-0.75)	0.53 (0.43-0.67)
3266 (0.50)	0.86 (0.71-0.97)	0.77 (0.60 - 0.88)	0.64 (0.46-0.8)	0.36 (0.26-0.50)	0.97 (0.67-1.00)	0.56 (0.33-0.75)	0.53 (0.38-0.67)
6532 (1.00)	0.87 (0.72-0.97)	0.76 (0.58 - 0.88)	0.65 (0.53-0.80)	0.36 (0.25-0.50)	0.933 (0.67-1.00)	0.58 (0.42-0.75)	0.52 (0.36-0.67)
13064 (2.00)	0.86 (0.64-1.00)	0.75 (0.50-0.88)	0.65 (0.46-0.80)	0.36 (0.20-0.50)	0.917 (0.67-1.00)	0.58 (0.33-0.75)	0.51 (0.31-0.67)
32661 (5.00)	0.80 (0.49-1.00)	0.67 (0.42 - 0.86)	0.57 (0.40-0.77)	0.30 (0.14-0.47)	0.837 (0.33-1.00)	0.50 (0.33-0.71)	0.44 (0.33-0.71)
65323 (10.00)	0.77 (0.40-1.00)	0.66 (0.42-0.86)	0.573 (0.4-0.77)	0.298 (0.14-0.47)	0.813 (0.33-1.00)	0.51 (0.33-0.75)	0.43 (0.33-0.75)

Table 5.12: Performance of the system using a different number of features evaluated on the interval $2.5 < \Delta t < 3.5$ years. Metrics used are: area under the curve (AUC), balanced accuracy (B.acc.), accuracy (Acc.), precision (Prec.), recall/sensitivity (Sens.), specificity (Spec.) and F-score.

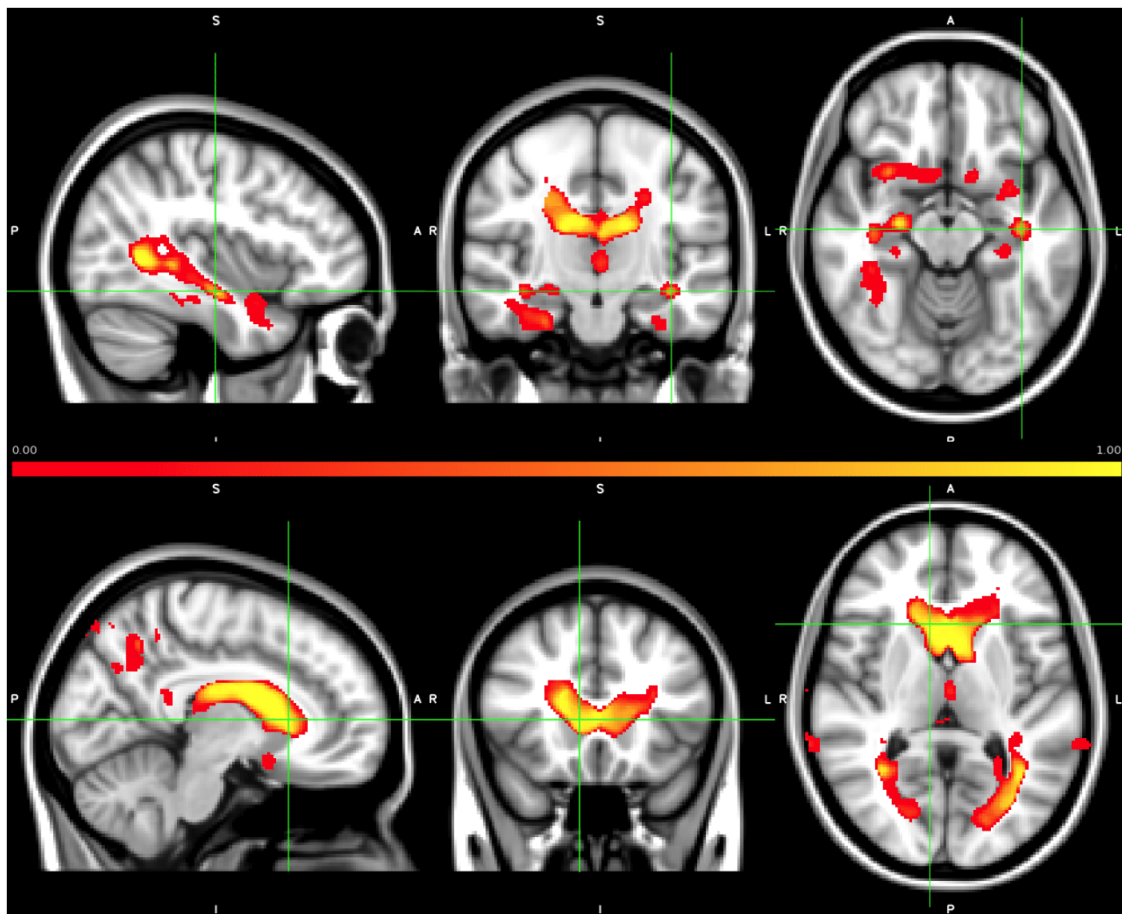


Figure 5.11: Normalized feature maps of the 0.5% of features selected during the 100 different splits of the development/test sets, representing the frequency of selection of each feature. Those features have optimal capacity to detect the presence of early amyloid pathology in asymptomatic individuals.

as a surrogate effect of ventricular expansion. Furthermore, comparison of longitudinal volumetric changes between amyloid negative cognitively unimpaired subjects and stable symptomatic ones (amyloid positive MCI or AD-dementia subjects) reveals the well-known AD signature involving temporo-parietal and posterior cingulate areas, as well as most of the basal ganglia [278], as shown on the right hand side of Figure 5.12. Of note, apparent GM increments are also detected in periventricular areas, including the caudates and medial thalamus.

5.4.3 Discussion

The goal of this work is to assess whether brain structural changes captured by subsequent magnetic resonance images can indicate the presence of abnormal amyloid levels in cognitively unimpaired subjects using machine learning-techniques. In addition, we also aim at characterizing the preclinical signature voxelwise using Jacobian determinant maps as a measure of volumetric rate of change.

A machine learning framework is implemented for the classification of amyloid positive subjects using Jacobian determinant maps as features for classification. The best

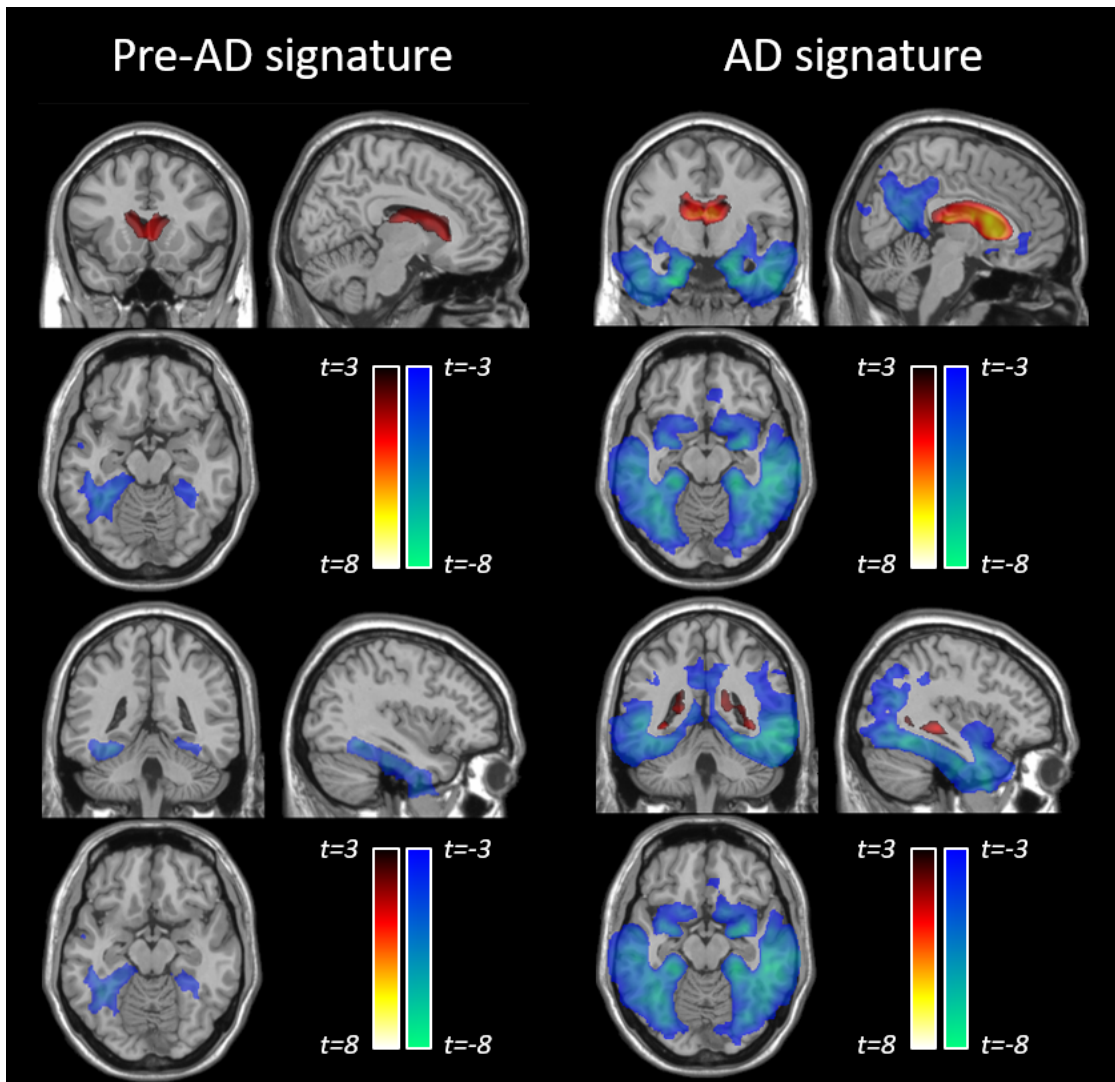


Figure 5.12: Statistical maps for group comparison between HC and PC (preclinical AD signature) and HC and MCI/AD-dementia (AD signature) subjects. Statistical significance was set to uncorrected p-value < 0.005 and minimum spatial extent $k > 100$.

achieved performance in our longitudinal classifier (AUC: 0.87) significantly improves the performance we previously reported for a cross-sectional classifier (AUC 0.76) in Section 5.3. This performance is significantly higher than what was reported in previous works that, on top of using MRI ROI data, built classifiers adding demographics (AUC: 0.63), demographics and genetics (AUC: 0.62-0.66) and demographics, neuropsychology and APOE (AUC: 0.74) [166, 167]. It is possible that adding complementary information to the MRI such as demographics and genetic risk factors may improve the performance of our machine learning classifier. While the field strength of the scanners is 1.5T for all subjects, there is large heterogeneity in the site ID, so we believe this has had small or no influence on the performance metrics of the classifier.

The increased performance of our classifier may be accounted for two factors. On the one hand and unlike similar previously reported classifiers, we use voxelwise data as features. Coupled with an efficient feature selection strategy, this allows the classifier to

select the most discriminant brain regions, independent of a-priori cortical parcellations. On the other hand, we use subsequent images that correspond to the same individuals, thus eliminating an important percentage of the between-subject variability present in cross-sectional setups. In this regard, we observe that our classifier works significantly better only when the pairs of MRI scans that are used for evaluation are acquired more than 2.5 years apart. This time period is likely related to the protracted evolution of the neuroanatomical changes in preclinical AD stages. At more advanced stages of the disease, more rapid evolution of brain structural changes is expected, and thus, the benefits of a longitudinal classifier would potentially be evident with shorter time intervals. It remains to be explored how these promising results would be affected by the use of different scanners. Still, a time gap of 2.5 years for resolving PC is within the timescale of relevance for AD screening or the follow up of subjects enrolled in secondary prevention clinical trials, which typically last a decade. In this context, this work and our earlier cross-sectional study on MRI using ML (Section 5.3) show that even though the performance of the ML classifier is not high, if implemented as a screening tool it can save resources in a clinical trial setting.

The main discriminative features between amyloid positive and healthy controls mostly include AD-related areas in the medial and inferior temporal lobe, as well as the lateral ventricles which can be considered as the preclinical AD signature. Increased expansion of the lateral and inferior lateral ventricles in cognitively unimpaired individuals with lower levels of CSF amyloid-beta has been shown previously, along with increased atrophy in the fusiform gyri as well as in middle temporal and posterior cingulate cortices [279, 280, 281, 282, 283]. In this regard, the preclinical AD signature found in our study does not significantly depart from published reports and, as can be seen in Figure 5.12, is very much in line with the expected pattern of atrophy in AD, though to a lesser magnitude and extent. In addition to (peri)ventricular regions, Figure 5.11 also shows the fusiform gyri and middle temporal regions to display significant discriminative capacity to discriminate amyloid positive vs negative CU individuals, as expected [280]. Additional detail on the brain areas contributing to such discriminative power is now provided in Supplementary Table B.15.

The predictive capacity achieved by this classifier does not place this method as substitute of gold-standard tests to detect amyloid abnormalities. Still, if used for triaging of subjects, e.g. clinical trial recruitment, we demonstrated that it could allow significant savings in terms of the number of costly gold-standard tests that would have to be performed to detect a fixed number of amyloid positive, cognitively healthy subjects. Used in this way, in a cognitively unimpaired population with a prevalence of amyloid positivity of 20%, the accuracy of the longitudinal classifier would allow a reduction of up to 55% of unnecessary PET or CSF tests, which translates to a 40% reduction of the total cost, according to the savings model (Section 5.2.4). Nevertheless, in a clinical trial recruitment setting, it can be more advantageous instead to optimize the sensitivity of the classifier to maximize the number of detected at-risk individuals, at the cost of a slightly poorer specificity which might decrease these cost savings.

Due to the limited sample size for training and the large inter-subject variability of cerebral morphology, we use a simple but effective model for prediction of amyloid positivity. Our method is fully automatic from feature extraction and signature learning to classification. However, the presence of high-dimensional and low informative features together with the overlap between normal aging and AD processes in the brain, reduces the overall

precision of the system. To account for that, future efforts will need larger longitudinal datasets and many initiatives are contributing to achieve this [166, 128].

We observe much higher sensitivity than specificity. This is likely given the limited size and imbalance of the cohort but also most likely due to the fact that we are imposing an imbalance on the test set to simulate the amyloid positivity prevalence of 20% typically found in a clinical trial setting. On top of this, given the limited sample size and the large amount of features used for classification (voxels), we might have incurred in an overfitting of the existing data, potentially resulting in an overestimation of the capacity of the classifier. Therefore, our results need to be validated on independent datasets, but the scarcity of longitudinal MRI datasets with CSF biomarker levels has prevented us to conduct such validation in this work. Still, in the previous ROI-based cross-sectional study (Section 5.3), we successfully validated a very similar classifier with two independent datasets without a major loss of the classifiers performance.

To further characterize the preclinical AD signature, a statistical analysis is conducted and we report longitudinal morphological changes in cognitively unimpaired subjects with abnormal amyloid CSF levels. This preclinical AD signature comprises atrophy of the parahippocampal and fusiform gyri and expansion of the lateral ventricles. This pattern is in line with previous reports of longitudinal volumetric changes associated to the presence of abnormal amyloid levels from ADNI participants that have been replicated in an independent cohort [163]. On the other hand, expansion of the caudate heads falls beyond this known pattern. Being in the proximity of the lateral ventricles, it may be questioned whether the detected increase in the volume of the caudates is an actual feature associated to preclinical AD stages or an artifact of the processing methodology to detect volumetric changes. By smoothing spatially continuous Jacobian determinant maps, it could be considered that the observed increase in caudate volumes could be a side effect of the “spillover of” the Jacobian determinant maps due to the expansion of the ventricles. To address this question, we performed a post-hoc analysis of the caudate volumes between the HC and PC groups, but using the longitudinal Freesurfer pipeline to compute change in caudate volumes. Since the subcortical segmentation implemented in Freesurfer uses an ROI-approach based on a probabilistic atlas [81], it can be considered to be virtually free from the potential spillover effect of continuous Jacobian determinant maps. Results show that the changes in caudate volumes is not significantly different between HC and PC individuals ($p > 0.3$) and, thus, it can be concluded that the observed caudate head expansion is artifactual and secondary to ventricular expansion. Still, this signal might contribute to the detection of the presence of amyloid-burden in cognitively unimpaired individuals.

This study has some limitations. Even though data comes from a heterogeneous sample with different sites, and MRI scanners, the MRI acquisition is harmonized according the ADNI protocol. Therefore, the performance of our method when applied to MRI samples using different acquisition protocols may deviate from what is here reported. Actually, the ultimate validation of the generalizability of the results here reported can only be accomplished by applying the method here developed to an independent sample. In the previous section, the performance of a similar cross-sectional classifier was kept stable when derived and validated in two independent cohorts. Therefore, it can be expected the same behaviour in this longitudinal extension of the classifier. Our study relies on the ADNI cohort which is well-known for its data quality and unique in having corresponding MRI and CSF data and a longitudinal aspect required for a study using

Jacobian determinants. The low amount of subjects with MRIs acquired with more than 2.5 years needed for a good signal to noise ratio certainly impose a limitation to our results and encourage future validation efforts. For example, one misclassification error has a huge impact in the performance metrics. To mitigate this effect, we repeat the workflow 100 times in order to report mean performance metrics. Nevertheless, the effect of misclassification can still be observed in the large confidence intervals that are found for each one of the metrics.

Finally, we use CSF amyloid as the gold-standard for amyloid positivity and not PET imaging. It could be argued that the performance of the classifier could be sensitive to the selection of the gold-standard method. However, the agreement between CSF and PET determinations of amyloid is very high, particularly in the intermediate ranges where thresholds for positivity typically lie.

Unfortunately, only a subset of 13 subjects respond to these criteria, from these only 2 subjects undergo this transition within a time frame of $\Delta t < 2.5$ years between consecutive scans. The sample size is therefore too small for a machine learning workflow. Nevertheless, the prediction of a transition from normal to preclinical AD stages is a question of utmost importance to research (e.g. observational studies) and clinical practice (e.g: clinical trials) and a natural follow-up to the present study.

5.5 Conclusions

In this last chapter, we provide a brief overview about clinical trials for Alzheimer's disease treatments, with focus in the recent shift to study anti-amyloid therapies in pre-clinical stages of the disease. We hypothesize that if preclinical subjects could be detected using MRI in combination with machine learning algorithms it can have potential benefits when used for triaging in recruitment protocols for clinical trials.

We begin by presenting a proof-of-concept study establishing that machine learning can detect amyloid pathology in cognitively normal populations using cross-sectional (T1-weighted alone and together with DTI MRI) and longitudinal data (Jacobian determinant maps derived from T1-weighted MRI). The performance of the cross-sectional classifier achieved a modest AUC of 0.76 while it improved in the longitudinal study (AUC= 0.87). Even though the performance of the classifier does not allow for it to substitute gold-standard methods to determine the presence of amyloid pathology, results show that this method could be used prospectively for preclinical subjects' screening in future cohorts, following appropriate data standardization. Such an approach can lead to significant reduction in participant burden and economic costs.

Hence, we propose a modified clinical recruitment protocol that introduces a pre-screening layer using machine learning layer that capitalizes the already acquired MRI scan for safety reasons and assess its viability in terms of reduction of the gold standard tests' reduction and monetary savings. We provide a theoretical framework to evaluate the impact of the proposed method using two different models trying to mimic real clinical recruitment scenarios. The potential benefits of the proposed method depend on the algorithm performance as well as the impact of expected prevalence of amyloid pathology over the general population and the cost of biomarker acquisitions. We report the results in a cross-sectional study using T1-weighted features (reduction of 60% of unnecessary gold-standard test and 47% in the recruitment costs) and T1+DTI features (reduction

of 76% of unnecessary gold-standard test and 49% in the recruitment costs).

Finally, we use longitudinal data to assess the minimum follow-up time in order to detect morphologic changes due to amyloid pathology. We conclude that a time span of 2.5 years is required to detect significant changes in brain structure. When use it as triaging tool, it would lead to significant reductions of 55% of unnecessary gold-standard tests and of 40% of the cost to detect a fixed number of cognitively healthy individuals in preclinical AD stages.

The main limitations of this work from the clinical perspective are that (i) we miss amyloid positive subjects at the prescreening step and this could be prejudicial for those individuals and (ii) not all amyloid positive subjects will develop Alzheimer's disease in their life span. Both limitations should be addressed in the future, for example, following the discarded subjects for, at least, 2.5 years (as suggested by our results) or incorporating tau pathological biomarkers in later stages of the trial. Future works should also focus on studying large multi-center samples that allows more complex modeling of amyloid pathology using MRI as well as validate the methodology in operational environments. Other complementary biomarkers in the triaging step can be easily incorporated due to the flexibility of the theoretical framework developed.

Conclusions and future work

Summary

The goal of this thesis is to study the Alzheimer's pathology signature present in MR images. MRI is a valid topographical biomarker of disease progression that measures synaptic loss and dead neurons. It is highly correlated with clinical phenotypes and can be used for patient staging. In this thesis, we use MRI to model brain structure along the Alzheimer's continuum which may have applications in interventional studies and clinical practice. As from recent updates to diagnostic frameworks in the literature, we define Alzheimer's disease as a clinicobiological entity characterized by early biological changes that may result in cognitive decline. We use in-vivo pathophysiological biomarkers (cerebrospinal fluid, concretely) to describe AD multi-pathological continuum accounting for amyloidosis, tauopathy and neurodegeneration.

Our analysis is focused in the understanding and modeling of brain structure at preclinical stages of AD: the *preclinical AD* signature. MRI biomarkers (e.g: hippocampal volume) are thought to become abnormal just before the onset of clinical symptoms even though brain damage begin much before that, according to the amyloid hypothesis [40]. Besides regional biomarkers of brain atrophy, studying brain structure as a whole uncovers hidden patterns of neurodegeneration as seen in Chapter 4. We use statistical learning and pattern recognition methods to develop descriptive (Chapters 3, 4) and predictive (Chapters 4, 5) models of Alzheimer's pathology.

We have developed an open-source toolbox for the analysis of nonlinear effects on medical images in Chapter 3. We provide the community a tool for statistical analysis which is flexible, modular and compatible with standard processing softwares. It helps to uncover regions with relevant nonlinear effects that could be hidden in linear analysis. We experiment with two cases of study: firstly, we analyze neurodegeneration along the Alzheimer's continuum and secondly we examine the interaction between APOE4 and age on brain atrophy. In both studies, we show several nonlinear effects on brain regional volume and cortical thickness.

The framework developed in Chapter 4 disentangles morphometric patterns associated to aging and Alzheimer's disease. It can be used to study other conditions or related-factors that may influence brain tissue configuration. The methodology is based on latent modeling (partial least squares) assuming that underlying brain processes lie in a low dimensional space. The work in Chapter 4 reflects that brain regions might be differently affected by the underlying pathology (amyloidosis or tauopathy) along the Alzheimer's continuum. Moreover, each region is differently related to pathophysiological markers depending on cognitive severity. Multivariate structural patterns reveal different brain structure at preclinical and clinical stages of AD, supporting the idea of an existing

preclinical AD signature in line with the reported literature in Chapter 2.

Finally, in Chapter 5 we study the capacity to detect amyloid positive subjects from a pool of cognitively unimpaired individuals. We propose a new recruitment protocol for clinical trials that uses machine learning and MRI as a triaging tool for preclinical subjects. Using this methodology, we develop a framework to study the impact of the proposed protocol in reducing the number of unnecessary gold-standard tests and monetary savings for clinical trials. A proof-of-concept study for clinical recruitment that capitalizes the already taken MRI for safety measures is presented and we report benefits in reducing the number of unnecessary gold-standard tests (76%) and monetary savings (49%) in clinical trials. In a longitudinal setting, we found that a minimum follow-up time of 2.5 years is required to detect significant changes due to amyloid positivity which remains within the range of clinical trial duration.

Future work

As a result of this thesis, several on-going, planned and suggested works can be derived. We hope that the toolbox developed in Chapter 3 could be used by the neuroimaging community to model nonlinear effects in brain imaging in a wide range of applications that haven't been tested. A natural extension of Chapter 4 would be to use multimodal (two or more measurements) modeling of biomedical data of the same observations. Latent models from the autoencoder literature (noisy autoencoder or variational autoencoders) constitute a natural generalization of the framework developed in this thesis. Adversarial training techniques can also be used in order to learn common distributions across different data modalities to uncover similarities between different sources of information. These methods must account for subjects with uncomplete data for which all the aforementioned frameworks result a suitable solution.

As for Chapter 5, there is work in progress to validate the framework in an operational environment: the ALFA project from Barcelona Beta Brain Research Center¹ [128]. Our study is restricted to a limited subject sample and therefore our models can only recognize a limited aspect of the PC signature. We expect an enhancement in performance when using larger training sets helping to develop more complex models and including demographic, genetic and cognitive data in the model. Another line of improvement is to predict the actual CSF A β 42 levels in order to dichotomize the result at a later stage and accommodate for optimal A β 42 thresholds in different applications. Future evaluation in a clinical trial setting, with the required regulatory measures, would report the overall potential of the framework. The utility of this methodology for disease-modifying trials in other mental disorders is still not explored and may widen the range of clinical applications of this work. With regards to the longitudinal study, one interesting area for further exploration is the classification of subjects that undergo a transition between normal and preclinical amyloid biomarkers within the timeframe of two consecutive scans. In principle, one could hypothesize that this category of transitioning subjects will not necessarily follow the same pattern of brain volumetric change as either the normal or the preclinical group.

¹<https://www.barcelonabeta.org>

Publications

Publications derived from this thesis

Submitted

- **Adrià Casamitjana**, Verónica Vilaplana, Santi Puch, Asier Aduriz, Carlos Lpez, Grégory Operto, Raffaele Cacciaglia, Carles Falcon, José Luis Molinuevo and Juan Domingo Gispert. NeAT: a nonlinear analysis toolbox for neuroimaging. *Neuroinformatics*, 2019. Submitted
- **Adrià Casamitjana**, Paula Petrone, José Luis Molinuevo, Juan Domingo Gispert and Verónica Vilaplana. Projection to Latent Spaces disentangles pathological effects on brain morphology in the asymptomatic phase of Alzheimer’s disease. *Frontiers in Neurology*, 2019. Submitted

Journals

- **Adrià Casamitjana***, Paula Petrone*, Alan Tucholka, Carles Falcon, Stavros Skouras, José Luis Molinuevo, Verónica Vilaplana and Juan Domingo Gispert. MRI-based screening of preclinical Alzheimer’s disease for prevention clinical trials. *Journal of Alzheimer’s Disease*, 2018. [215]
- Paula Petrone*, **Adrià Casamitjana***, Carles Falcon, Miquel Artigues, Grégory Operto, Raffaele Cacciaglia, José Luis Molinuevo, Verónica Vilaplana and Juan Domingo Gispert. Prediction of amyloid pathology in cognitively unimpaired individuals using voxel-wise analysis of longitudinal structural brain MRI. *Alzheimer’s Research & Therapy*, 2019. [222]
- **Adrià Casamitjana**, Verónica Vilaplana, Paula Petrone, José Luis Molinuevo and Juan Domingo Gispert. Shared latent structures between imaging features and biomarkers in early stages of Alzheimer’s disease: a predictive study. *IEEE Journal of Biomedical and Health Informatics*, 2019. [284]

Abstracts

- Paula Petrone, **Adrià Casamitjana**, Carles Falcon, Miquel Artigues, Grégory Operto, Stavros Skouras, Raffaele Cacciaglia, José Luis Molinuevo, Verónica Vilaplana, Juan Domingo Gispert and Gemma Salvadó. Characteristic brain volumetric changes in the AD preclinical signature. *Alzheimer’s Association International Conference*, 2018. [285]

- **Adrià Casamitjana**, Paula Petrone, Miquel Artigues, José Luis Molinuevo, Juan Domingo Gispert and Verónica Vilaplana. Projection to latent spaces disentangles specific cerebral morphometric patterns associated to aging and preclinical AD. *Alzheimer's Association International Conference*, 2018. [286]
- Paula Petrone, Verónica Vilaplana, **Adrià Casamitjana**, Dalila Sánchez Escobedo, Alan Tucholka, Raffaele Cacciaglia, Grégory Operto, Stavros Skouras Carles Falcon, José Luis Molinuevo and Juan Domingo Gispert. Magnetic resonance imaging and machine learning make a valuable combined tool for the screening of preclinical AD. *Alzheimer's Association International Conference*, 2017. [287]

In Proceedings

- **Adrià Casamitjana**, Verónica Vilaplana, Paula Petrone, José Luis Molinuevo and Juan Domingo Gispert. Shared Latent Structures Between Imaging Features and Biomarkers in Early Stages of Alzheimer's Disease. *International Workshop on Predictive Intelligence In MEdicine*, 2018. [288]
- Santi Puch, Asier Aduriz, **Adrià Casamitjana**, Veronica Vilaplana, Paula Petrone, Grgory Operto, Raffaele Cacciaglia, Stavros Skouras, Carles Falcon, Jos Luis Molinuevo and Juan Domingo Gispert. Voxelwise nonlinear regression toolbox for neuroimage analysis: Application to aging and neurodegenerative disease modeling. *Machine Learning for Health, NIPS Workshop*, 2016. [289]

Other publications

- **Adrià Casamitjana**, Santi Puch, Asier Aduriz, Elisa Sayrol, Verónica Vilaplana. 3D Convolutional Neural Networks for Brain Tumor Segmentation: a comparison of multi-resolution architectures. *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2016. [290]
- Li Wang, Dong Nie, Guannan Li, lodie Puybareau, Jose Dolz, Qian Zhang, Fan Wang, Jing Xia, Zhengwang Wu, Jiawei Chen, Kim-Han Thung, Toan Duc Bui, Jitae Shin, Guodong Zeng, Guoyan Zheng, Vladimir S Fonov, Andrew Doyle, Yongchao Xu, Pim Moeskops, Josien PW Pluim, Christian Desrosiers, Ismail Ben Ayed, Gerard Sanroma, Oualid M Benkarim, **Adrià Casamitjana**, Verónica Vilaplana, Weili Lin, Gang Li, Dinggang Shen. Benchmark on Automatic 6-month-old Infant Brain Segmentation Algorithms: The iSeg-2017 Challenge. *IEEE Transactions on Medical Imaging*, 2019. [291].
- Hugo J Kuijf, J Matthijs Biesbroek, Jeroen de Bresser, Rutger Heinen, Simon Andermatt, Mariana Bento, Matt Berseth, Mikhail Belyaev, M Jorge Cardoso, **Adrià Casamitjana**, D Louis Collins, Mahsa Dadar, Achilleas Georgiou, Mohsen Ghafourian, Dakai Jin, April Khademi, Jesse Knight, Hongwei Li, Xavier Llad, Miguel Luna, Qaiser Mahmood, Richard McKinley, Alireza Mehrdash, Sbastien Ourselin, Bo-yong Park, Hyunjin Park, Sang Hyun Park, Simon Pezold, Elodie Puybareau, Leticia Rittner, Carole H Sudre, Sergi Valverde, Verónica Vilaplana, Roland Wiest, Yongchao Xu, Ziyue Xu, Guodong Zeng, Jianguo Zhang, Guoyan Zheng, Christopher Chen, Wiesje van der Flier, Frederik Barkhof, Max A Viergever, Geert Jan

Biessels. Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities; Results of the WMH Segmentation Challenge. *IEEE Transactions on Medical Imaging*, 2019. [72]

- **Adrià Casamitjana**, Marcel Catà , Irina Sanchez Muriana, Marc Combalia, Verónica Vilaplana. Cascaded V-Net Using ROI Masks for Brain Tumor Segmentation. *International MICCAI Brainlesion Workshop*, 2017. [292]

Bibliography

- [1] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine, *et al.*, “Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 7, no. 3, pp. 280–292, 2011. [1](#), [10](#), [12](#), [49](#)
- [2] B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, *et al.*, “Preclinical Alzheimer’s disease: definition, natural history, and diagnostic criteria,” *Alzheimer’s & Dementia*, vol. 12, no. 3, pp. 292–323, 2016. [1](#), [12](#), [13](#), [41](#)
- [3] D. Rueckert, B. Glocker, and B. Kainz, “Learning clinically useful information from images: past, present and future,” 2016. [1](#), [19](#)
- [4] W. H. Organisation, “Dementia fact sheet December 2017”.” <https://www.who.int/en/news-room/fact-sheets/detail/dementia>, 2017. [1](#), [2](#)
- [5] W. H. Organisation, “World Health Statistics in 2014”.” <https://www.who.int/mediacentre/news/releases/2014/world-health-statistics-2014/en/>, 2018. [1](#)
- [6] M. Prince, A. Wimo, M. Guerchet, G.-C. Ali, Y.-T. Wu, and M. Prina, “World Alzheimers Report 2015.The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends. Alzheimers Disease International,” 2015. [2](#)
- [7] B. Winblad, P. Amouyel, S. Andrieu, C. Ballard, C. Brayne, H. Brodaty, A. Cedazo-Minguez, B. Dubois, D. Edvardsson, H. Feldman, *et al.*, “Defeating Alzheimer’s disease and other dementias: a priority for European science and society,” *The Lancet Neurology*, vol. 15, no. 5, pp. 455–532, 2016. [2](#)
- [8] C. Reitz and R. Mayeux, “Alzheimer’s disease: epidemiology, diagnostic criteria, risk factors and biomarkers,” *Biochemical pharmacology*, vol. 88, no. 4, pp. 640–651, 2014. [2](#)
- [9] J.-C. Chen, M. A. Espeland, R. L. Brunner, L. C. Lovato, R. B. Wallace, X. Leng, L. S. Phillips, J. G. Robinson, J. M. Kotchen, K. C. Johnson, *et al.*, “Sleep duration, cognitive decline, and dementia risk in older women,” *Alzheimer’s & Dementia*, vol. 12, no. 1, pp. 21–33, 2016. [2](#)
- [10] L. Mosconi, A. Rahman, I. Diaz, X. Wu, O. Scheyer, H. W. Hristov, S. Vallabhajosula, R. S. Isaacson, M. J. de Leon, and R. D. Brinton, “Increased Alzheimer’s

- risk during the menopause transition: A 3-year longitudinal brain imaging study,” *PloS one*, vol. 13, no. 12, p. e0207885, 2018. [2](#)
- [11] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, *et al.*, “The diagnosis of dementia due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 7, no. 3, pp. 263–269, 2011. [3](#), [10](#), [12](#)
- [12] D. S. Knopman, S. T. DeKosky, J. Cummings, H. Chui, J. Corey-Bloom, N. Relkin, G. Small, B. Miller, and J. Stevens, “Practice parameter: diagnosis of dementia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology,” *Neurology*, vol. 56, no. 9, pp. 1143–1153, 2001. [3](#), [8](#)
- [13] T. G. Beach, S. E. Monsell, L. E. Phillips, and W. Kukull, “Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010,” *Journal of neuropathology and experimental neurology*, vol. 71, no. 4, pp. 266–273, 2012. [3](#), [8](#)
- [14] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, “Clinical diagnosis of Alzheimer’s disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease,” *Neurology*, vol. 34, no. 7, pp. 939–939, 1984. [3](#), [8](#), [12](#)
- [15] B. Dubois, H. H. Feldman, C. Jacova, H. Hampel, J. L. Molinuevo, K. Blennow, S. T. DeKosky, S. Gauthier, D. Selkoe, R. Bateman, *et al.*, “Advancing research diagnostic criteria for Alzheimer’s disease: the IWG-2 criteria,” *The Lancet Neurology*, vol. 13, no. 6, pp. 614–629, 2014. [4](#), [10](#), [12](#), [13](#)
- [16] C. R. Jack Jr, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, *et al.*, “NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 14, no. 4, pp. 535–562, 2018. [4](#), [5](#), [11](#), [12](#), [60](#), [93](#)
- [17] E. Karran, M. Mercken, and B. De Strooper, “The amyloid cascade hypothesis for Alzheimer’s disease: an appraisal for the development of therapeutics,” *Nature reviews Drug discovery*, vol. 10, no. 9, p. 698, 2011. [4](#)
- [18] K. M. Langa and J. F. Burke, “Preclinical Alzheimer Disease Early Diagnosis or Overdiagnosis?,” *JAMA internal medicine*, 2019. [4](#)
- [19] K. A. Jobst, L. P. Barnetson, and B. J. Shepstone, “Accurate prediction of histologically confirmed Alzheimer’s disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, X-ray CT, and Apo E4 in medial temporal lobe dementias,” *International Psychogeriatrics*, vol. 9, no. S1, pp. 191–222, 1997. [7](#)
- [20] S. Karantzoulis, J. E. Galvin, Braak, Salmon, Mckhann, McKhann, Jack, Sperling, Hodges, Butters, *et al.*, “Distinguishing Alzheimer’s disease from other major forms

- of dementia,” *Expert review of neurotherapeutics*, vol. 11, no. 11, pp. 1579–1591, 2011. 7
- [21] D. S. Knopman, J. E. Parisi, A. Salviati, M. Floriach-Robert, B. F. Boeve, R. J. Ivnik, G. E. Smith, D. W. Dickson, K. Johnson, L. Petersen, *et al.*, “Neuropathology of cognitively normal elderly,” *Journal of Neuropathology & Experimental Neurology*, vol. 62, no. 11, pp. 1087–1095, 2003. 7
- [22] D. Bennett, J. Schneider, Z. Arvanitakis, J. Kelly, N. Aggarwal, R. Shah, and R. Wilson, “Neuropathology of older persons without cognitive impairment from two community-based studies,” *Neurology*, vol. 66, no. 12, pp. 1837–1844, 2006. 7
- [23] A. Alzheimer, “Uber eine eigenartige Erkrankung der Hirnrinde,” *Zentralbl. Nervenhe. Psych.*, vol. 18, pp. 177–179, 1907. 7
- [24] R. Katzman, “The prevalence and malignancy of Alzheimer disease: a major killer,” *Archives of neurology*, vol. 33, no. 4, pp. 217–218, 1976. 7
- [25] P. J. Whitehouse, D. L. Price, R. G. Struble, A. W. Clark, J. T. Coyle, and M. R. Delon, “Alzheimer’s disease and senile dementia: loss of neurons in the basal forebrain,” *Science*, vol. 215, no. 4537, pp. 1237–1239, 1982. 7
- [26] R. D. Terry and R. K. Katzman, “Senile dementia of the Alzheimer type,” *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 14, no. 5, pp. 497–506, 1983. 7
- [27] J. C. Morris, D. McKeel, M. Storandt, E. H. Rubin, J. L. Price, E. A. Grant, M. J. Ball, and L. Berg, “Very mild Alzheimer’s disease: informant-based clinical, psychometric, and pathologic distinction from normal aging,” *Neurology*, vol. 41, no. 4, pp. 469–469, 1991. 8
- [28] D. N. Levine, J. M. Lee, and C. Fisher, “The visual variant of Alzheimer’s disease: a clinicopathologic case study,” *Neurology*, vol. 43, no. 2, pp. 305–305, 1993. 8
- [29] C. Flicker, S. H. Ferris, and B. Reisberg, “Mild cognitive impairment in the elderly: predictors of dementia,” *Neurology*, vol. 41, no. 7, pp. 1006–1006, 1991. 8
- [30] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, “Mild cognitive impairment: clinical characterization and outcome,” *Archives of neurology*, vol. 56, no. 3, pp. 303–308, 1999. 8
- [31] J. C. Morris, M. Storandt, J. P. Miller, D. W. McKeel, J. L. Price, E. H. Rubin, and L. Berg, “Mild cognitive impairment represents early-stage Alzheimer disease,” *Archives of neurology*, vol. 58, no. 3, pp. 397–405, 2001. 8
- [32] R. C. Petersen, “Mild cognitive impairment as a diagnostic entity,” *Journal of internal medicine*, vol. 256, no. 3, pp. 183–194, 2004. 8
- [33] B. Dubois and M. L. Albert, “Amnestic MCI or prodromal Alzheimer’s disease?,” *The Lancet Neurology*, vol. 3, no. 4, pp. 246–248, 2004. 8
- [34] G. A. Jicha, J. E. Parisi, D. W. Dickson, K. Johnson, R. Cha, R. J. Ivnik, E. G. Tangalos, B. F. Boeve, D. S. Knopman, H. Braak, *et al.*, “Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia,” *Archives of neurology*, vol. 63, no. 5, pp. 674–681, 2006. 8

- [35] R. C. Petersen, J. E. Parisi, D. W. Dickson, K. A. Johnson, D. S. Knopman, B. F. Boeve, G. A. Jicha, R. J. Ivnik, G. E. Smith, E. G. Tangalos, *et al.*, “Neuropathologic features of amnesic mild cognitive impairment,” *Archives of neurology*, vol. 63, no. 5, pp. 665–672, 2006. [8](#)
- [36] B. Dubois, H. H. Feldman, C. Jacova, S. T. DeKosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, *et al.*, “Research criteria for the diagnosis of Alzheimer’s disease: revising the NINCDS–ADRDA criteria,” *The Lancet Neurology*, vol. 6, no. 8, pp. 734–746, 2007. [9](#)
- [37] B. Dubois, H. H. Feldman, C. Jacova, J. L. Cummings, S. T. DeKosky, P. Barberger-Gateau, A. Delacourte, G. Frisoni, N. C. Fox, D. Galasko, *et al.*, “Revising the definition of Alzheimer’s disease: a new lexicon,” *The Lancet Neurology*, vol. 9, no. 11, pp. 1118–1127, 2010. [9](#)
- [38] C. R. Jack Jr, M. Albert, D. S. Knopman, G. M. McKhann, R. A. Sperling, M. Carrillo, W. Thies, and C. H. Phelps, “Introduction to revised criteria for the diagnosis of Alzheimers disease: National Institute on Aging and the Alzheimer Association Workgroups,” *Alzheimer’s & dementia: the journal of the Alzheimer’s Association*, vol. 7, no. 3, p. 257, 2011. [10](#), [13](#), [60](#), [82](#), [93](#)
- [39] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, *et al.*, “The diagnosis of mild cognitive impairment due to Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 7, no. 3, pp. 270–279, 2011. [10](#), [12](#), [22](#), [69](#)
- [40] C. R. Jack Jr, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski, “Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade,” *The Lancet Neurology*, vol. 9, no. 1, pp. 119–128, 2010. [10](#), [13](#), [15](#), [23](#), [72](#), [107](#)
- [41] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, H. H. Feldman, G. B. Frisoni, H. Hampel, W. J. Jagust, K. A. Johnson, D. S. Knopman, *et al.*, “A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers,” *Neurology*, vol. 87, no. 5, pp. 539–547, 2016. [11](#), [70](#)
- [42] J. A. Hardy and G. A. Higgins, “Alzheimer’s disease: the amyloid cascade hypothesis,” *Science*, vol. 256, no. 5054, pp. 184–186, 1992. [11](#)
- [43] E. Mohandas, V. Rajmohan, and B. Raghunath, “Neurobiology of Alzheimer’s disease,” *Indian journal of psychiatry*, vol. 51, no. 1, p. 55, 2009. [11](#)
- [44] E. Portelius, H. Zetterberg, T. Skillbäck, U. Törnqvist, U. Andreasson, J. Q. Trojanowski, M. W. Weiner, L. M. Shaw, N. Mattsson, and K. Blennow, “Cerebrospinal fluid neurogranin: relation to cognition and neurodegeneration in Alzheimer’s disease,” *Brain*, vol. 138, no. 11, pp. 3373–3385, 2015. [11](#)
- [45] H. Wellington, R. W. Paterson, E. Portelius, U. Törnqvist, N. Magdalinou, N. C. Fox, K. Blennow, J. M. Schott, and H. Zetterberg, “Increased CSF neurogranin concentration is specific to Alzheimer disease,” *Neurology*, vol. 86, no. 9, pp. 829–835, 2016. [11](#)

- [46] B. D. James, R. S. Wilson, P. A. Boyle, J. Q. Trojanowski, D. A. Bennett, and J. A. Schneider, “TDP-43 stage, mixed pathologies, and clinical Alzheimer’s-type dementia,” *Brain*, vol. 139, no. 11, pp. 2983–2993, 2016. [11](#)
- [47] A. Vergallo, R.-S. Bun, N. Toschi, F. Baldacci, H. Zetterberg, K. Blennow, E. Cavedo, F. Lamari, M.-O. Habert, B. Dubois, *et al.*, “Association of cerebrospinal fluid α -synuclein with total and phospho-tau181 protein concentrations and brain amyloid load in cognitively normal subjective memory complainers stratified by Alzheimer’s disease biomarkers,” *Alzheimer’s & Dementia*, vol. 14, no. 12, pp. 1623–1631, 2018. [11](#)
- [48] H. M. Snyder, R. A. Corriveau, S. Craft, J. E. Faber, S. M. Greenberg, D. Knopman, B. T. Lamb, T. J. Montine, M. Nedergaard, C. B. Schaffer, *et al.*, “Vascular contributions to cognitive impairment and dementia including Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 11, no. 6, pp. 710–717, 2015. [11](#)
- [49] A. Brugulat-Serrat, G. Salvadó, C. H. Sudre, O. Grau-Rivera, M. Suárez-Calvet, C. Falcon, G. Sánchez-Benavides, N. Gramunt, K. Fauria, M. J. Cardoso, *et al.*, “Patterns of white matter hyperintensities associated with cognition in middle-aged cognitively healthy individuals,” *Brain imaging and behavior*, pp. 1–12, 2019. [11](#)
- [50] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, “Alzheimer’s Disease Neuroimaging Initiative (ADNI),” *Neurology*, vol. 74, no. 3, pp. 201–209, 2010. [12](#)
- [51] K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff, *et al.*, “The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease,” *International Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009. [12](#), [21](#)
- [52] B. Dunn, “Early Alzheimer’s Disease: Developing Drugs for Treatment; Draft Guidance for Industry,” *Services USDoHaH, Administration FaD, (CDER) CfDEaR, (CBER) CfBEaR*, eds. Silver Spring, MD: Office of the Federal Register, National Archives and Records Administration, pp. 7060–1, 2018. [12](#)
- [53] J. Cummings and N. Fox, “Defining disease modifying therapy for Alzheimer’s disease,” *The journal of prevention of Alzheimer’s disease*, vol. 4, no. 2, p. 109, 2017. [13](#)
- [54] M. Crous-Bou, C. Minguillón, N. Gramunt, and J. L. Molinuevo, “Alzheimer’s disease prevention: from risk factors to early intervention,” *Alzheimer’s research & therapy*, vol. 9, no. 1, p. 71, 2017. [13](#)
- [55] W. J. Jansen, R. Ossenkoppele, D. L. Knol, B. M. Tijms, P. Scheltens, F. R. Verhey, P. J. Visser, P. Aalten, D. Aarsland, D. Alcolea, *et al.*, “Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis,” *Jama*, vol. 313, no. 19, pp. 1924–1938, 2015. [13](#), [14](#), [85](#), [96](#)
- [56] B. Dubois, S. Epelbaum, F. Nyasse, H. Bakardjian, G. Gagliardi, O. Uspenskaya, M. Houot, S. Lista, F. Cacciamani, M.-C. Potier, *et al.*, “Cognitive and neuroimaging features and brain β -amyloidosis in individuals at risk of Alzheimer’s disease

- (insight-read): a longitudinal observational study,” *The Lancet Neurology*, vol. 17, no. 4, pp. 335–346, 2018. [13](#), [21](#)
- [57] M. C. Donohue, R. A. Sperling, R. Petersen, C.-K. Sun, M. W. Weiner, and P. S. Aisen, “Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons,” *Jama*, vol. 317, no. 22, pp. 2305–2316, 2017. [14](#)
- [58] S. Kern, H. Zetterberg, J. Kern, A. Zettergren, M. Waern, K. Höglund, U. Andreasson, H. Wetterberg, A. Börjesson-Hanson, K. Blennow, *et al.*, “Prevalence of preclinical Alzheimer disease: Comparison of current classification systems,” *Neurology*, vol. 90, no. 19, pp. e1682–e1691, 2018. [14](#)
- [59] G. Livingston, A. Sommerlad, V. Orgeta, S. G. Costafreda, J. Huntley, D. Ames, C. Ballard, S. Banerjee, A. Burns, J. Cohen-Mansfield, *et al.*, “Dementia prevention, intervention, and care,” *The Lancet*, vol. 390, no. 10113, pp. 2673–2734, 2017. [14](#)
- [60] R. Ossenkoppele, G. D. Rabinovici, R. Smith, H. Cho, M. Schöll, O. Strandberg, S. Palmqvist, N. Mattsson, S. Janelidze, A. Santillo, *et al.*, “Discriminative accuracy of [18F] flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders,” *Jama*, vol. 320, no. 11, pp. 1151–1162, 2018. [14](#), [41](#)
- [61] R. Ossenkoppele, D. R. Schonhaut, M. Schöll, S. N. Lockhart, N. Ayakta, S. L. Baker, J. P. O’Neil, M. Janabi, A. Lazaris, A. Cantwell, *et al.*, “Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer’s disease,” *Brain*, vol. 139, no. 5, pp. 1551–1567, 2016. [14](#), [23](#)
- [62] C. R. Jack Jr, H. J. Wiste, S. D. Weigand, T. M. Therneau, D. S. Knopman, V. Lowe, P. Vemuri, M. M. Mielke, R. O. Roberts, M. M. Machulda, *et al.*, “Age-specific and sex-specific prevalence of cerebral β -amyloidosis, tauopathy, and neurodegeneration in cognitively unimpaired individuals aged 50–95 years: a cross-sectional study,” *The Lancet Neurology*, vol. 16, no. 6, pp. 435–444, 2017. [14](#)
- [63] C. Westbrook and J. Talbot, *MRI in Practice*. John Wiley & Sons, 2018. [14](#)
- [64] R. Felix, W. Schörner, M. Laniado, H. Niendorf, C. Claussen, W. Fiegler, and U. Speck, “Brain tumors: MR imaging with gadolinium-DTPA,” *Radiology*, vol. 156, no. 3, pp. 681–688, 1985. [15](#)
- [65] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, *et al.*, “MR pulse sequences: what every radiologist wants to know but is afraid to ask,” *Radiographics*, vol. 26, no. 2, pp. 513–537, 2006. [15](#)
- [66] J. L. Whitwell, D. W. Dickson, M. E. Murray, S. D. Weigand, N. Tosakulwong, M. L. Senjem, D. S. Knopman, B. F. Boeve, J. E. Parisi, R. C. Petersen, *et al.*, “Neuroimaging correlates of pathologically defined subtypes of Alzheimer’s disease: a case-control study,” *The Lancet Neurology*, vol. 11, no. 10, pp. 868–877, 2012. [15](#)
- [67] M. E. Murray, N. R. Graff-Radford, O. A. Ross, R. C. Petersen, R. Duara, and D. W. Dickson, “Neuropathologically defined subtypes of Alzheimer’s disease with

- distinct clinical characteristics: a retrospective study,” *The Lancet Neurology*, vol. 10, no. 9, pp. 785–796, 2011. [15](#)
- [68] A. Dong, J. B. Toledo, N. Honnorat, J. Doshi, E. Varol, A. Sotiras, D. Wolk, J. Q. Trojanowski, C. Davatzikos, and A. D. N. Initiative, “Heterogeneity of neuroanatomical patterns in prodromal Alzheimer’s disease: links to cognition, progression and biomarkers,” *Brain*, vol. 140, no. 3, pp. 735–747, 2016. [15](#)
- [69] Y. Stern, “Cognitive reserve in ageing and alzheimer’s disease,” *The Lancet Neurology*, vol. 11, no. 11, pp. 1006–1012, 2012. [15](#), [59](#)
- [70] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014. [16](#)
- [71] P. Pretorius and G. Quaghebeur, “The role of MRI in the diagnosis of MS,” *Clinical radiology*, vol. 58, no. 6, pp. 434–448, 2003. [16](#)
- [72] H. J. Kuijf, J. M. Biesbroek, J. de Bresser, R. Heinen, S. Andermatt, M. Bento, M. Berseth, M. Belyaev, M. J. Cardoso, A. Casamitjana, *et al.*, “Standardized assessment of automatic segmentation of white matter hyperintensities; results of the wmh segmentation challenge,” *IEEE transactions on medical imaging*, 2019. [16](#), [111](#)
- [73] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, “MRI denoising using non-local means,” *Medical image analysis*, vol. 12, no. 4, pp. 514–523, 2008. [16](#)
- [74] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, “N4ITK: improved N3 bias correction,” *IEEE transactions on medical imaging*, vol. 29, no. 6, p. 1310, 2010. [16](#)
- [75] U. Vovk, F. Pernus, and B. Likar, “A review of methods for correction of intensity inhomogeneity in MRI,” *IEEE transactions on medical imaging*, vol. 26, no. 3, pp. 405–421, 2007. [16](#)
- [76] J. Ashburner and K. J. Friston, “Unified segmentation,” *Neuroimage*, vol. 26, no. 3, pp. 839–851, 2005. [16](#)
- [77] J. Ashburner and K. J. Friston, “Voxel-based morphometrythe methods,” *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000. [16](#)
- [78] X. Hua, A. D. Leow, N. Parikshak, S. Lee, M.-C. Chiang, A. W. Toga, C. R. Jack Jr, M. W. Weiner, P. M. Thompson, A. D. N. Initiative, *et al.*, “Tensor-based morphometry as a neuroimaging biomarker for Alzheimer’s disease: an MRI study of 676 AD, MCI, and normal subjects,” *Neuroimage*, vol. 43, no. 3, pp. 458–469, 2008. [16](#), [19](#), [23](#)
- [79] M. J. Clarkson, M. J. Cardoso, G. R. Ridgway, M. Modat, K. K. Leung, J. D. Rohrer, N. C. Fox, and S. Ourselin, “A comparison of voxel and surface based cortical thickness estimation methods,” *Neuroimage*, vol. 57, no. 3, pp. 856–865, 2011. [16](#)

- [80] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002. [16](#), [83](#), [94](#)
- [81] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, *et al.*, “Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain,” *Neuron*, vol. 33, no. 3, pp. 341–355, 2002. [16](#), [49](#), [60](#), [104](#)
- [82] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et al.*, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006. [16](#)
- [83] B. Fischl, “FreeSurfer,” *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012. [16](#), [25](#)
- [84] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011. [16](#), [25](#)
- [85] B. B. Avants, N. Tustison, and G. Song, “Advanced normalization tools (ANTs),” *Insight j*, vol. 2, pp. 1–35, 2009. [16](#), [83](#)
- [86] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, “FSL,” *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012. [16](#), [83](#)
- [87] G. Stebbins and C. Murphy, “Diffusion tensor imaging in Alzheimer’s disease and mild cognitive impairment,” *Behavioural neurology*, vol. 21, no. 1, 2, pp. 39–49, 2009. [17](#)
- [88] J. W. Prescott, A. Guidon, P. M. Doraiswamy, K. Roy Choudhury, C. Liu, J. R. Petrella, and A. D. N. Initiative, “The Alzheimer structural connectome: changes in cortical network topology with increased amyloid plaque burden,” *Radiology*, vol. 273, no. 1, pp. 175–184, 2014. [17](#)
- [89] J. Acosta-Cabronero and P. J. Nestor, “Diffusion tensor imaging in Alzheimer’s disease: insights into the limbic-diencephalic network and methodological considerations,” *Frontiers in aging neuroscience*, vol. 6, p. 266, 2014. [17](#)
- [90] A. Tucholka, O. Grau-Rivera, C. Falcon, L. Rami, R. Sanchez-Valle, A. Lladó, J. D. Gispert, J. L. Molinuevo, A. D. N. Initiative, *et al.*, “Structural connectivity alterations along the Alzheimer’s Disease continuum: reproducibility across two independent samples and correlation with cerebrospinal fluid Amyloid- β and Tau,” *Journal of Alzheimer’s Disease*, vol. 61, no. 4, pp. 1575–1587, 2018. [17](#), [83](#), [88](#), [90](#)
- [91] E. L. Dennis and P. M. Thompson, “Functional brain connectivity using fMRI in aging and Alzheimer’s disease,” *Neuropsychology review*, vol. 24, no. 1, pp. 49–62, 2014. [17](#)
- [92] P. Vemuri, D. T. Jones, and C. R. Jack, “Resting state functional MRI in Alzheimer’s Disease,” *Alzheimer’s research & therapy*, vol. 4, no. 1, p. 2, 2012. [17](#), [23](#)

- [93] F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G. B. Frisoni, and M. Filippi, “Resting state fMRI in Alzheimer’s disease: beyond the default mode network,” *Neurobiology of aging*, vol. 33, no. 8, pp. 1564–1578, 2012. 17
- [94] A. Badhwar, A. Tam, C. Dansereau, P. Orban, F. Hoffstaedter, and P. Bellec, “Resting-state network dysfunction in Alzheimer’s disease: a systematic review and meta-analysis,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 8, pp. 73–85, 2017. 17
- [95] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, “Statistical parametric maps in functional imaging: a general linear approach,” *Human brain mapping*, vol. 2, no. 4, pp. 189–210, 1994. 18, 25
- [96] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006. 18, 19, 20
- [97] C. Rasmussen and C. Williams, “Gaussian Processes for Machine Learning the MIT Press,” 2006. 18
- [98] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001. 18, 20
- [99] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, “Single subject prediction of brain disorders in neuroimaging: promises and pitfalls,” *Neuroimage*, vol. 145, pp. 137–165, 2017. 18
- [100] R. V. Marinescu, N. P. Oxtoby, A. L. Young, E. E. Bron, A. W. Toga, M. W. Weiner, F. Barkhof, N. C. Fox, S. Klein, D. C. Alexander, *et al.*, “TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer’s Disease,” *arXiv preprint arXiv:1805.03909*, 2018. 18
- [101] M. de Bruijne, “Machine learning approaches in medical image analysis: From detection to diagnosis,” 2016. 18
- [102] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, “Machine learning for neuroimaging with scikit-learn,” *Frontiers in neuroinformatics*, vol. 8, p. 14, 2014. 19
- [103] C. Misra, Y. Fan, and C. Davatzikos, “Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI,” *Neuroimage*, vol. 44, no. 4, pp. 1415–1422, 2009. 19
- [104] B. Fischl and A. M. Dale, “Measuring the thickness of the human cerebral cortex from magnetic resonance images,” *Proceedings of the National Academy of Sciences*, vol. 97, no. 20, pp. 11050–11055, 2000. 19
- [105] A. M. Winkler, P. Kochunov, J. Blangero, L. Almasy, K. Zilles, P. T. Fox, R. Duggirala, and D. C. Glahn, “Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies,” *Neuroimage*, vol. 53, no. 3, pp. 1135–1146, 2010. 19
- [106] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *NeuroImage*, vol. 15, pp. 273–279, January 2002. 20

- [107] X. Zhu, H.-I. Suk, K.-H. Thung, Y. Zhu, G. Wu, and D. Shen, “Joint Discriminative and Representative Feature Selection for Alzheimers Disease Diagnosis,” in *International Workshop on Machine Learning in Medical Imaging*, pp. 77–85, Springer, 2016. [20](#)
- [108] Y. Zhu, X. Zhu, M. Kim, D. Shen, and G. Wu, “Early Diagnosis of Alzheimers Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 264–272, Springer, 2016. [20](#)
- [109] L. Sørensen, A. Pai, C. Anker, I. Balas, M. Lillholm, C. Igel, and M. Nielsen, “Dementia diagnosis using MRI cortical thickness, shape, texture, and volumetry,” in *Proc MICCAI workshop challenge on computer-aided diagnosis of dementia based on structural MRI data*, pp. 111–118, 2014. [20](#)
- [110] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, “COMPARE: classification of morphological patterns using adaptive regional elements,” *IEEE transactions on medical imaging*, vol. 26, no. 1, pp. 93–105, 2006. [20](#)
- [111] K. Van Leemput, A. Bakkour, T. Benner, G. Wiggins, L. L. Wald, J. Augustinack, B. C. Dickerson, P. Golland, and B. Fischl, “Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI,” *Hippocampus*, vol. 19, no. 6, pp. 549–557, 2009. [20](#)
- [112] T. W. Way, B. Sahiner, L. M. Hadjiiski, and H.-P. Chan, “Effect of finite sample size on feature selection and classification: a simulation study,” *Medical physics*, vol. 37, no. 2, pp. 907–920, 2010. [20](#)
- [113] B. Mwangi, T. S. Tian, and J. C. Soares, “A review of feature reduction techniques in neuroimaging,” *Neuroinformatics*, vol. 12, no. 2, pp. 229–244, 2014. [20](#)
- [114] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative, *et al.*, “Multimodal classification of Alzheimer’s disease and mild cognitive impairment,” *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011. [20](#)
- [115] Y. Li, Y. Wang, G. Wu, F. Shi, L. Zhou, W. Lin, D. Shen, A. D. N. Initiative, *et al.*, “Discriminant analysis of longitudinal cortical thickness changes in Alzheimer’s disease using dynamic and network features,” *Neurobiology of aging*, vol. 33, no. 2, pp. 427–e15, 2012. [20](#)
- [116] E. E. Bron, M. Smits, W. J. Niessen, and S. Klein, “Feature Selection Based on the SVM Weight Vector for Classification of Dementia,” *IEEE journal of biomedical and health informatics*, vol. 19, no. 5, pp. 1617–1626, 2015. [20](#)
- [117] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007. [20](#), [94](#)
- [118] M. López, J. Ramírez, J. M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, R. Chaves, P. Padilla, M. Gómez-Río, A. D. N. Initiative, *et al.*, “Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer’s disease,” *Neurocomputing*, vol. 74, no. 8, pp. 1260–1271, 2011. [20](#)

- [119] M. Barker and W. Rayens, “Partial least squares for discrimination,” *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 3, pp. 166–173, 2003. [20](#)
- [120] R. Wolz, P. Aljabar, J. V. Hajnal, and D. Rueckert, “Manifold learning for biomarker discovery in MR imaging,” in *International Workshop on Machine Learning in Medical Imaging*, pp. 116–123, Springer, 2010. [20](#)
- [121] H. Park *et al.*, “ISOMAP induced manifold embedding and its application to Alzheimer’s disease and mild cognitive impairment,” *Neuroscience Letters*, vol. 513, no. 2, pp. 141–145, 2012. [20](#)
- [122] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002. [20](#)
- [123] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. [20](#)
- [124] M. Liu, D. Zhang, D. Shen, A. D. N. Initiative, *et al.*, “Ensemble sparse classification of Alzheimer’s disease,” *NeuroImage*, vol. 60, no. 2, pp. 1106–1116, 2012. [20](#)
- [125] G. Varoquaux, P. R. Raamana, D. A. Engemann, A. Hoyos-Idrobo, Y. Schwartz, and B. Thirion, “Assessing and tuning brain decoders: cross-validation, caveats, and guidelines,” *NeuroImage*, vol. 145, pp. 166–179, 2017. [21](#), [61](#), [94](#)
- [126] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, “The Alzheimer’s disease neuroimaging initiative,” *Neuroimaging Clinics*, vol. 15, no. 4, pp. 869–877, 2005. [21](#)
- [127] R. J. Bateman, C. Xiong, T. L. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, *et al.*, “Clinical and biomarker changes in dominantly inherited Alzheimer’s disease,” *New England Journal of Medicine*, vol. 367, no. 9, pp. 795–804, 2012. [21](#), [25](#)
- [128] J. L. Molinuevo, N. Gramunt, J. D. Gispert, K. Fauria, M. Esteller, C. Minguillon, G. Sánchez-Benavides, G. Huesa, S. Morán, R. Dal-Ré, *et al.*, “The ALFA project: a research platform to identify early pathophysiological features of Alzheimer’s disease,” *Alzheimer’s & Dementia: Translational Research & Clinical Interventions*, vol. 2, no. 2, pp. 82–92, 2016. [21](#), [34](#), [90](#), [104](#), [108](#)
- [129] G. D. Rabinovici, C. Gatsonis, C. Apgar, K. Chaudhary, I. Gareen, L. Hanna, J. Hendrix, B. E. Hillner, C. Olson, O. H. Lesman-Segev, *et al.*, “Association of amyloid positron emission tomography with subsequent change in clinical management among Medicare beneficiaries with mild cognitive impairment or dementia,” *Jama*, vol. 321, no. 13, pp. 1286–1294, 2019. [21](#)
- [130] I. Bos, S. Vos, R. Vandenberghe, P. Scheltens, S. Engelborghs, G. Frisoni, J. L. Molinuevo, A. Wallin, A. Lleó, J. Popp, *et al.*, “The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics,” *Alzheimer’s research & therapy*, vol. 10, no. 1, p. 64, 2018. [21](#)
- [131] A. W. Toga, S. C. Neu, P. Bhatt, K. L. Crawford, and N. Ashish, “The global Alzheimer’s association interactive network,” *Alzheimer’s & Dementia*, vol. 12, no. 1, pp. 49–54, 2016. [21](#)

- [132] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults,” *Journal of cognitive neuroscience*, vol. 22, no. 12, pp. 2677–2684, 2010. [21](#)
- [133] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, and C. Davatzikos, “A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages,” *NeuroImage*, vol. 155, pp. 530–548, 2017. [21](#), [22](#)
- [134] L. deToledo Morrell, T. Stoub, M. Bulgakova, R. Wilson, D. Bennett, S. Leurgans, J. Wu, and D. Turner, “MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD,” *Neurobiology of aging*, vol. 25, no. 9, pp. 1197–1203, 2004. [21](#)
- [135] W. W. Seeley, R. K. Crawford, J. Zhou, B. L. Miller, and M. D. Greicius, “Neurodegenerative diseases target large-scale human brain networks,” *Neuron*, vol. 62, no. 1, pp. 42–52, 2009. [22](#)
- [136] P. Vemuri and C. R. Jack, “Role of structural MRI in Alzheimer’s disease,” *Alzheimer’s research & therapy*, vol. 2, no. 4, p. 23, 2010. [22](#)
- [137] G. B. Frisoni, N. C. Fox, C. R. Jack Jr, P. Scheltens, and P. M. Thompson, “The clinical use of structural MRI in Alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, p. 67, 2010. [22](#)
- [138] B. C. Dickerson, I. Goncharova, M. Sullivan, C. Forchetti, R. Wilson, D. Bennett, L. A. Beckett, and L. deToledo Morrell, “MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer’s disease,” *Neurobiology of aging*, vol. 22, no. 5, pp. 747–754, 2001. [22](#)
- [139] F. Falahati, E. Westman, and A. Simmons, “Multivariate data analysis and machine learning in Alzheimer’s disease with a focus on structural magnetic resonance imaging,” *Journal of Alzheimer’s disease*, vol. 41, no. 3, pp. 685–708, 2014. [22](#)
- [140] M. R. Sabuncu, E. Konukoglu, A. D. N. Initiative, *et al.*, “Clinical prediction from structural brain MRI scans: a large-scale empirical study,” *Neuroinformatics*, vol. 13, no. 1, pp. 31–46, 2015. [22](#)
- [141] S. Klöppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, *et al.*, “Accuracy of dementia diagnosis direct comparison between radiologists and a computerized method,” *Brain*, vol. 131, no. 11, pp. 2969–2974, 2008. [22](#)
- [142] K. Nho, L. Shen, S. Kim, S. L. Risacher, J. D. West, T. Foroud, C. R. Jack Jr, M. W. Weiner, and A. J. Saykin, “Automatic prediction of conversion from mild cognitive impairment to probable Alzheimer’s disease using structural magnetic resonance imaging,” in *AMIA Annual Symposium Proceedings*, vol. 2010, p. 542, American Medical Informatics Association, 2010. [22](#)
- [143] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. B. Frisoni, and D. L. Collins, “MRI-based automated computer classification of probable AD versus normal controls,” *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 509–520, 2008. [22](#)

- [144] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Schill, J. D. Rohrer, N. C. Fox, C. R. Jack Jr, J. Ashburner, and R. S. Frackowiak, “Automatic classification of MR scans in Alzheimer’s disease,” *Brain*, vol. 131, no. 3, pp. 681–689, 2008. [22](#)
- [145] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack Jr, “Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies,” *Neuroimage*, vol. 39, no. 3, pp. 1186–1197, 2008. [22](#)
- [146] C. Aguilar, E. Westman, J.-S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, *et al.*, “Different multivariate techniques for automated classification of MRI data in Alzheimers disease and mild cognitive impairment,” *Psychiatry Research: Neuroimaging*, vol. 212, no. 2, pp. 89–98, 2013. [22](#)
- [147] S. G. Costafreda, I. D. Dinov, Z. Tu, Y. Shi, C.-Y. Liu, I. Kloszewska, P. Mecocci, H. Soininen, M. Tsolaki, B. Vellas, *et al.*, “Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment,” *Neuroimage*, vol. 56, no. 1, pp. 212–219, 2011. [22](#)
- [148] S. F. Eskildsen, P. Coupé, D. García-Lorenzo, V. Fonov, J. C. Pruessner, D. L. Collins, A. D. N. Initiative, *et al.*, “Prediction of Alzheimer’s disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning,” *Neuroimage*, vol. 65, pp. 511–521, 2013. [22](#)
- [149] C. Davatzikos, P. Bhatt, L. M. Shaw, K. N. Batmanghelich, and J. Q. Trojanowski, “Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification,” *Neurobiology of aging*, vol. 32, no. 12, pp. 2322–e19, 2011. [22](#)
- [150] M. Sjöbeck, M. Haglund, and E. Englund, “Decreasing myelin density reflected increasing white matter pathology in Alzheimer’s diseasea neuropathological study,” *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, vol. 20, no. 10, pp. 919–926, 2005. [23](#)
- [151] R. Migliaccio, F. Agosta, K. L. Possin, G. D. Rabinovici, B. L. Miller, and M. L. Gorno-Tempini, “White matter atrophy in Alzheimer’s disease variants,” *Alzheimer’s & Dementia*, vol. 8, no. 5, pp. S78–S87, 2012. [23](#)
- [152] B. Zhang, Y. Xu, B. Zhu, and K. Kantarci, “The Role of Diffusion Tensor Imaging in Detecting Microstructural Changes in Prodromal Alzheimer’s Disease,” *CNS neuroscience & therapeutics*, vol. 20, no. 1, pp. 3–9, 2014. [23](#)
- [153] T. M. Nir, J. E. Villalon-Reina, G. Prasad, N. Jahanshad, S. H. Joshi, A. W. Toga, M. A. Bernstein, C. R. Jack Jr, M. W. Weiner, P. M. Thompson, *et al.*, “Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer’s disease,” *Neurobiology of aging*, vol. 36, pp. S132–S140, 2015. [23](#)
- [154] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Identifying patients with Alzheimer’s disease using resting-state fMRI and graph theory,” *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2132–2141, 2015. [23](#)

- [155] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaei, A. Babajani-Feremi, A. D. N. Initiative, *et al.*, “Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM,” *Journal of neuroscience methods*, vol. 282, pp. 69–80, 2017. [23](#)
- [156] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, “Gaussian process classification of Alzheimer’s disease and mild cognitive impairment from resting-state fMRI,” *NeuroImage*, vol. 112, pp. 232–243, 2015. [23](#)
- [157] D. Zhang, D. Shen, A. D. N. Initiative, *et al.*, “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease,” *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012. [23](#)
- [158] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, *et al.*, “Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis,” *NeuroImage*, vol. 101, pp. 569–582, 2014. [23](#)
- [159] Z. Wang, X. Zhu, E. Adeli, Y. Zhu, F. Nie, B. Munsell, G. Wu, *et al.*, “Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning,” *Medical image analysis*, vol. 39, pp. 218–230, 2017. [23](#)
- [160] C. Fennema-Notestine, D. J. Hagler Jr, L. K. McEvoy, A. S. Fleisher, E. H. Wu, D. S. Karow, and A. M. Dale, “Structural MRI biomarkers for preclinical and mild Alzheimer’s disease,” *Human brain mapping*, vol. 30, no. 10, pp. 3238–3253, 2009. [23](#), [59](#), [69](#)
- [161] R. Killiany, B. Hyman, T. Gomez-Isla, M. Moss, R. Kikinis, F. Jolesz, R. Tanzi, K. Jones, and M. Albert, “MRI measures of entorhinal cortex vs hippocampus in preclinical AD,” *Neurology*, vol. 58, no. 8, pp. 1188–1196, 2002. [23](#)
- [162] J. D. Gispert, L. Rami, G. Sánchez-Benavides, C. Falcon, A. Tucholka, S. Rojas, and J. L. Molinuevo, “Nonlinear cerebral atrophy patterns across the Alzheimer’s disease continuum: impact of APOE4 genotype,” *Neurobiology of aging*, vol. 36, no. 10, pp. 2687–2701, 2015. [23](#), [25](#), [26](#), [33](#), [83](#), [90](#)
- [163] C. Falcon, A. Tucholka, G. C. Monté-Rubio, R. Cacciaglia, G. Operto, L. Rami, J. D. Gispert, J. L. Molinuevo, A. D. N. Initiative, *et al.*, “Longitudinal structural cerebral changes related to core CSF biomarkers in preclinical Alzheimer’s disease: a study of two independent datasets,” *NeuroImage: Clinical*, vol. 19, pp. 190–201, 2018. [23](#), [59](#), [69](#), [104](#)
- [164] J. L. Molinuevo, P. Ripolles, M. Simó, A. Lladó, J. Olives, M. Balasa, A. Antonell, A. Rodriguez-Fornells, and L. Rami, “White matter changes in preclinical Alzheimer’s disease: a magnetic resonance imaging-diffusion tensor imaging study on cognitively normal older people with positive amyloid β protein 42 levels,” *Neurobiology of aging*, vol. 35, no. 12, pp. 2671–2680, 2014. [23](#), [83](#)
- [165] J. Pegueroles, E. Vilaplana, V. Montal, F. Sampedro, D. Alcolea, M. Carmona-Iragui, J. Clarimon, R. Blesa, A. Lleó, J. Fortea, *et al.*, “Longitudinal brain structural changes in preclinical Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 13, no. 5, pp. 499–509, 2017. [23](#)

- [166] M. Ten Kate, A. Redolfi, E. Peira, I. Bos, S. J. Vos, R. Vandenberghe, S. Gabel, J. Schaevebeke, P. Scheltens, O. Blin, *et al.*, “MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study,” *Alzheimer’s research & therapy*, vol. 10, no. 1, p. 100, 2018. [23](#), [41](#), [42](#), [70](#), [102](#), [104](#)
- [167] M. Ansart, S. Epelbaum, G. Gagliardi, O. Colliot, D. Dormont, B. Dubois, H. Hampel, S. Durrleman, A. D. N. Initiative*, and the INSIGHT-preAD study, “Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis,” *Statistical methods in medical research*, p. 0962280218823036, 2019. [23](#), [41](#), [42](#), [102](#)
- [168] C. M. Clark, J. A. Schneider, B. J. Bedell, T. G. Beach, W. B. Bilker, M. A. Mintun, M. J. Pontecorvo, F. Hefti, A. P. Carpenter, M. L. Flitter, *et al.*, “Use of florbetapir-PET for imaging β -amyloid pathology,” *Jama*, vol. 305, no. 3, pp. 275–283, 2011. [23](#)
- [169] G. E. Alexander, K. Chen, P. Pietrini, S. I. Rapoport, and E. M. Reiman, “Longitudinal PET evaluation of cerebral metabolic decline in dementia: a potential outcome measure in Alzheimers disease treatment studies,” *American Journal of Psychiatry*, vol. 159, no. 5, pp. 738–745, 2002. [23](#)
- [170] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, *et al.*, “Cerebrospinal fluid biomarker signature in Alzheimer’s disease neuroimaging initiative subjects,” *Annals of neurology*, vol. 65, no. 4, pp. 403–413, 2009. [23](#), [49](#), [60](#), [82](#)
- [171] X. Liu, K. Chen, T. Wu, D. Weidman, F. Lure, and J. Li, “Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer’s disease,” *Translational Research*, vol. 194, pp. 56–67, 2018. [23](#)
- [172] M. Lorenzi, M. Filippone, G. B. Frisoni, D. C. Alexander, S. Ourselin, A. D. N. Initiative, *et al.*, “Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer’s disease,” *NeuroImage*, 2017. [23](#)
- [173] A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, and D. C. Alexander, “A data-driven model of biomarker changes in sporadic Alzheimer’s disease,” *Brain*, vol. 137, no. 9, pp. 2564–2577, 2014. [23](#)
- [174] V. Venkatraghavan, E. E. Bron, W. J. Niessen, S. Klein, A. D. N. Initiative, *et al.*, “Disease progression timeline estimation for Alzheimer’s disease using discriminative event based modeling,” *NeuroImage*, vol. 186, pp. 518–532, 2019. [23](#)
- [175] A. L. Young, R. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. C. Firth, D. M. Cash, D. L. Thomas, K. M. Dick, J. Cardoso, *et al.*, “Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference,” *Nature communications*, vol. 9, no. 1, p. 4273, 2018. [23](#), [70](#)
- [176] P. A. Freeborough and N. C. Fox, “The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI,” *IEEE transactions on medical imaging*, vol. 16, no. 5, pp. 623–629, 1997. [25](#)

- [177] M. N. Samtani, M. Farnum, V. Lobanov, E. Yang, N. Raghavan, A. DiBernardo, V. Narayan, and A. D. N. Initiative, “An improved model for disease progression in patients from the Alzheimer’s disease neuroimaging initiative,” *The Journal of Clinical Pharmacology*, vol. 52, no. 5, pp. 629–644, 2012. [25](#)
- [178] M. S. Mendiondo, J. W. Ashford, R. J. Kryscio, and F. A. Schmitt, “Modelling mini mental state examination changes in Alzheimer’s disease,” *Statistics in medicine*, vol. 19, no. 11-12, pp. 1607–1616, 2000. [25](#)
- [179] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoek, S. L. Macaulay, R. Martins, P. Maruff, *et al.*, “Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer’s disease: a prospective cohort study,” *The Lancet Neurology*, vol. 12, no. 4, pp. 357–367, 2013. [25](#)
- [180] P. S. Insel, R. Ossenkoppele, D. Gessert, W. Jagust, S. Landau, O. Hansson, M. W. Weiner, N. Mattsson, A. D. N. Initiative, *et al.*, “Time to amyloid positivity and preclinical changes in brain metabolism, atrophy, and cognition: evidence for emerging amyloid pathology in Alzheimer’s disease,” *Frontiers in neuroscience*, vol. 11, p. 281, 2017. [25](#)
- [181] P. S. Insel, N. Mattsson, M. C. Donohue, R. S. Mackin, P. S. Aisen, C. R. Jack Jr, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, A. D. N. Initiative, *et al.*, “The transitional association between β -amyloid pathology and regional brain atrophy,” *Alzheimer’s & Dementia*, vol. 11, no. 10, pp. 1171–1179, 2015. [25](#)
- [182] M. R. Sabuncu, R. S. Desikan, J. Sepulcre, B. T. T. Yeo, H. Liu, N. J. Schmansky, M. Reuter, M. W. Weiner, R. L. Buckner, R. A. Sperling, *et al.*, “The dynamics of cortical and hippocampal atrophy in Alzheimer disease,” *Archives of neurology*, vol. 68, no. 8, pp. 1040–1048, 2011. [25](#)
- [183] N. Schuff, D. Tosun, P. S. Insel, G. C. Chiang, D. Truran, P. S. Aisen, C. R. Jack Jr, M. W. Weiner, A. D. N. Initiative, *et al.*, “Nonlinear time course of brain volume loss in cognitively normal and impaired elders,” *Neurobiology of aging*, vol. 33, no. 5, pp. 845–855, 2012. [25](#)
- [184] J. Kornak, J. A. Fields, S. Farmer, B. F. Boeve, H. J. Rosen, A. L. Boxer, J. Bove, D. Brushaber, G. Coppola, C. Dheel, *et al.*, “Nonlinear N-score estimation for establishing cognitive norms from the National Alzheimer’s Coordinating Center (NACC) dataset,” *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, vol. 14, no. 7, pp. P390–P391, 2018. [25](#)
- [185] A. M. Fjell, L. T. Westlye, H. Grydeland, I. Amlien, T. Espeseth, I. Reinvang, N. Raz, D. Holland, A. M. Dale, K. B. Walhovd, *et al.*, “Critical ages in the life course of the adult brain: nonlinear subcortical aging,” *Neurobiology of aging*, vol. 34, no. 10, pp. 2239–2247, 2013. [25](#)
- [186] A. M. Fjell, K. B. Walhovd, L. T. Westlye, Y. Østby, C. K. Tamnes, T. L. Jernigan, A. Gamst, and A. M. Dale, “When does brain aging accelerate? Dangers of quadratic fits in cross-sectional studies,” *Neuroimage*, vol. 50, no. 4, pp. 1376–1383, 2010. [25](#)

- [187] E. J. Vinke, M. De Groot, V. Venkatraghavan, S. Klein, W. J. Niessen, M. A. Ikram, and M. W. Vernooij, “Trajectories of imaging markers in brain aging: the Rotterdam Study,” *Neurobiology of aging*, vol. 71, pp. 32–40, 2018. [25](#)
- [188] G. Ziegler, R. Dahnke, and C. Gaser, “Models of the aging brain structure and individual decline,” *Frontiers in neuroinformatics*, vol. 6, p. 3, 2012. [25](#)
- [189] J.-B. Poline, J. L. Breeze, S. S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, K. G. Helmer, D. S. Marcus, R. A. Poldrack, Y. Schwartz, *et al.*, “Data sharing in neuroimaging research,” *Frontiers in neuroinformatics*, vol. 6, p. 9, 2012. [25](#)
- [190] R. Cacciaglia, J. L. Molinuevo, C. Falcón, A. Brugulat-Serrat, G. Sánchez-Benavides, N. Gramunt, M. Esteller, S. Morán, C. Minguillón, K. Fauria, *et al.*, “Effects of APOE- ϵ 4 allele load on brain morphology in a cohort of middle-aged healthy individuals with enriched genetic risk for Alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 14, no. 7, pp. 902–912, 2018. [26](#), [35](#), [37](#), [38](#)
- [191] R. Henson and W. Penny, “ANOVAs and SPM,” *Wellcome Department of Imaging Neuroscience, London, UK*, 2003. [27](#)
- [192] R. Christensen, *Plane answers to complex questions: the theory of linear models*. Springer Science & Business Media, 2011. [28](#)
- [193] T. J. Hastie, “Generalized additive models,” in *Statistical models in S*, pp. 249–307, Routledge, 2017. [28](#)
- [194] L. Breiman and J. H. Friedman, “Estimating optimal transformations for multiple regression and correlation,” *Journal of the American statistical Association*, vol. 80, no. 391, pp. 580–598, 1985. [28](#)
- [195] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, “Support vector regression machines,” in *Advances in neural information processing systems*, pp. 155–161, 1997. [29](#)
- [196] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011. [29](#), [31](#), [33](#)
- [197] F. Dinuzzo, M. Neve, G. D. Nicolao, and U. P. Gianazza, “On the representer theorem and equivalent degrees of freedom of SVR,” *Journal of Machine Learning Research*, vol. 8, no. Oct, pp. 2467–2495, 2007. [29](#)
- [198] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, “A practical guide to support vector classification,” 2003. [29](#)
- [199] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112. Springer, 2013. [29](#)
- [200] Y. Sakamoto, M. Ishiguro, and G. Kitagawa, “Akaike information criterion statistics,” *Dordrecht, The Netherlands: D. Reidel*, vol. 81, 1986. [30](#)
- [201] J. Jacques and C. Preda, “Functional data clustering: a survey,” *Advances in Data Analysis and Classification*, vol. 8, no. 3, pp. 231–255, 2014. [31](#)

- [202] F. Murtagh and P. Legendre, “Wards hierarchical agglomerative clustering method: which algorithms implement Wards criterion?,” *Journal of classification*, vol. 31, no. 3, pp. 274–295, 2014. [31](#)
- [203] E. Jones, T. Oliphant, and P. Peterson, “SciPy: open source scientific tools for Python,” 2014. [33](#)
- [204] M. Brett, M. Hanke, B. Cipollini, M.-A. Côté, C. Markiewicz, S. Gerhard, E. Larson, Y. Halchenko, C. Cheng, *et al.*, “nibabel: 2.3.3,” *Zenodo*, 2019. [33](#)
- [205] J. L. Molinuevo, J. D. Gispert, B. Dubois, M. T. Heneka, A. Lleó, S. Engelborghs, J. Pujol, L. C. de Souza, D. Alcolea, F. Jessen, *et al.*, “The AD-CSF-index discriminates Alzheimer’s disease patients from healthy controls: a validation study,” *Journal of Alzheimer’s Disease*, vol. 36, no. 1, pp. 67–77, 2013. [33](#), [34](#), [41](#), [60](#)
- [206] C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu, “Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy,” *Nature Reviews Neurology*, vol. 9, no. 2, p. 106, 2013. [34](#)
- [207] R. J. Caselli, A. C. Dueck, D. Osborne, M. N. Sabbagh, D. J. Connor, G. L. Ahern, L. C. Baxter, S. Z. Rapcsak, J. Shi, B. K. Woodruff, *et al.*, “Longitudinal modeling of age-related memory decline and the APOE ϵ 4 effect,” *New England Journal of Medicine*, vol. 361, no. 3, pp. 255–263, 2009. [34](#)
- [208] N. Filippini, A. Rao, S. Wetten, R. A. Gibson, M. Borrie, D. Guzman, A. Kertesz, I. Loy-English, J. Williams, T. Nichols, *et al.*, “Anatomically-distinct genetic associations of APOE 4 allele load with regional cortical atrophy in Alzheimer’s disease,” *Neuroimage*, vol. 44, no. 3, pp. 724–728, 2009. [34](#)
- [209] M. ten Kate, E. J. Sanz-Arigitá, B. M. Tijms, A. M. Wink, M. Clerigue, M. Garcia-Sebastian, A. Izagirre, M. Ecay-Torres, A. Estanga, J. Villanua, *et al.*, “Impact of APOE-4 and family history of dementia on gray matter atrophy in cognitively healthy middle-aged adults,” *Neurobiology of aging*, vol. 38, pp. 14–20, 2016. [35](#), [90](#)
- [210] R. Peters, “Ageing and the brain,” *Postgraduate medical journal*, vol. 82, no. 964, pp. 84–88, 2006. [41](#)
- [211] P. S. Aisen, J. Cummings, C. R. Jack, J. C. Morris, R. Sperling, L. Frölich, R. W. Jones, S. A. Dowsett, B. R. Matthews, J. Raskin, *et al.*, “On the path to 2025: understanding the Alzheimer’s disease continuum,” *Alzheimer’s research & therapy*, vol. 9, no. 1, p. 60, 2017. [41](#)
- [212] A. M. Fjell, L. McEvoy, D. Holland, A. M. Dale, K. B. Walhovd, A. D. N. Initiative, *et al.*, “What is normal in normal aging? Effects of aging, amyloid and Alzheimer’s disease on the cerebral cortex and the hippocampus,” *Progress in neurobiology*, vol. 117, pp. 20–40, 2014. [41](#)
- [213] M. Wirth, S. Villeneuve, C. M. Haase, C. M. Madison, H. Oh, S. M. Landau, G. D. Rabinovici, and W. J. Jagust, “Associations between Alzheimer disease biomarkers, neurodegeneration, and cognition in cognitively normal older people,” *JAMA neurology*, vol. 70, no. 12, pp. 1512–1519, 2013. [41](#)

- [214] D. Mungas, B. R. Reed, W. Jagust, C. DeCarli, W. Mack, J. Kramer, M. Weiner, N. Schuff, and H. Chui, “Volumetric MRI predicts rate of cognitive decline related to AD and cerebrovascular disease,” *Neurology*, vol. 59, no. 6, pp. 867–873, 2002. [41](#)
- [215] A. Casamitjana, P. Petrone, A. Tucholka, C. Falcon, S. Skouras, J. L. Molinuevo, V. Vilaplana, J. D. Gispert, A. D. N. Initiative, *et al.*, “MRI-based screening of preclinical Alzheimer’s disease for prevention clinical trials,” *Journal of Alzheimer’s Disease*, no. Preprint, pp. 1–14. [41](#), [42](#), [70](#), [109](#)
- [216] L. Iaccarino, G. Tammewar, N. Ayakta, S. L. Baker, A. Bejanin, A. L. Boxer, M. L. Gorno-Tempini, M. Janabi, J. H. Kramer, A. Lazaris, *et al.*, “Local and distant relationships between amyloid, tau and neurodegeneration in Alzheimer’s Disease,” *NeuroImage: Clinical*, vol. 17, pp. 452–464, 2018. [42](#)
- [217] R. Ossenkoppele, R. Smith, T. Ohlsson, O. Strandberg, N. Mattsson, P. S. Insel, S. Palmqvist, and O. Hansson, “Associations between tau, $A\beta$, and cortical thickness with cognition in Alzheimer disease,” *Neurology*, vol. 92, no. 6, pp. e601–e612, 2019. [42](#)
- [218] R. Kanai and G. Rees, “The structural basis of inter-individual differences in human behaviour and cognition,” *Nature Reviews Neuroscience*, vol. 12, no. 4, p. 231, 2011. [42](#)
- [219] P. Vemuri, H. Wiste, S. Weigand, L. Shaw, J. Trojanowski, M. Weiner, D. S. Knopman, R. C. Petersen, C. Jack, *et al.*, “MRI and CSF biomarkers in normal, MCI, and AD subjects: diagnostic discrimination and cognitive correlations,” *Neurology*, vol. 73, no. 4, pp. 287–293, 2009. [42](#), [70](#)
- [220] P. Vemuri, H. Wiste, S. Weigand, L. Shaw, J. Trojanowski, M. Weiner, D. S. Knopman, R. C. Petersen, C. Jack, *et al.*, “MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change,” *Neurology*, vol. 73, no. 4, pp. 294–301, 2009. [42](#), [70](#)
- [221] A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, J. B. Brewer, A. M. Dale, A. D. N. Initiative, *et al.*, “CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer’s disease,” *Journal of Neuroscience*, vol. 30, no. 6, pp. 2088–2101, 2010. [42](#)
- [222] P. M. Petrone, A. Casamitjana, C. Falcon, M. Artigues, G. Operto, R. Cacciaglia, J. L. Molinuevo, V. Vilaplana, and J. D. Gispert, “Prediction of amyloid pathology in cognitively unimpaired individuals using voxel-wise analysis of longitudinal structural brain MRI,” *Alzheimer’s research & therapy*, vol. 11, no. 1, pp. 1–13, 2019. [42](#), [109](#)
- [223] A. McIntosh, F. Bookstein, J. V. Haxby, and C. Grady, “Spatial pattern analysis of functional brain images using partial least squares,” *Neuroimage*, vol. 3, no. 3, pp. 143–157, 1996. [43](#)
- [224] A. Krishnan, L. J. Williams, A. R. McIntosh, and H. Abdi, “Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review,” *Neuroimage*, vol. 56, no. 2, pp. 455–475, 2011. [43](#)

- [225] G. Ziegler, R. Dahnke, A. Winkler, and C. Gaser, “Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents,” *NeuroImage*, vol. 82, pp. 284–294, 2013. [43](#)
- [226] M. Lorenzi, B. Gutman, D. P. Hibar, A. Altmann, N. Jahanshad, P. M. Thompson, and S. Ourselin, “Partial least squares modelling for imaging-genetics in Alzheimer’s disease: Plausibility and generalization,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 838–841, IEEE, 2016. [43](#)
- [227] E. Konukoglu, J.-P. Coutu, D. H. Salat, B. Fischl, A. D. N. I. (ADNI, *et al.*, “Multivariate statistical analysis of diffusion imaging parameters using partial least squares: Application to white matter variations in Alzheimer’s disease,” *Neuroimage*, vol. 134, pp. 573–586, 2016. [43](#), [46](#)
- [228] C. Wachinger, A. Rieckmann, and M. Reuter, “Latent processes governing neuroanatomical change in Aging and Dementia,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 30–37, Springer, 2017. [43](#), [70](#)
- [229] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, F. Segovia, A. s Disease Neuroimaging Initiative, *et al.*, “Early diagnosis of Alzheimer s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images,” *Neurocomputing*, vol. 151, pp. 139–150, 2015. [43](#)
- [230] H. Wold, “Path models with latent variables: The NIPALS approach,” in *Quantitative sociology*, pp. 307–357, Elsevier, 1975. [43](#)
- [231] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: a basic tool of chemometrics,” *Chemometrics and intelligent laboratory systems*, vol. 58, no. 2, pp. 109–130, 2001. [43](#)
- [232] J. Henseler, C. M. Ringle, and R. R. Sinkovics, “The use of partial least squares path modeling in international marketing,” in *New challenges to international marketing*, pp. 277–319, Emerald Group Publishing Limited, 2009. [43](#)
- [233] M. Rönkkö, C. N. McIntosh, and J. Antonakis, “On the adoption of partial least squares in psychological research: Caveat emptor,” *Personality and Individual Differences*, vol. 87, pp. 76–84, 2015. [43](#)
- [234] S. Wold, A. Ruhe, H. Wold, and W. Dunn, III, “The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses,” *SIAM Journal on Scientific and Statistical Computing*, vol. 5, no. 3, pp. 735–743, 1984. [43](#)
- [235] H. Abdi, “Partial least squares regression and projection on latent structure regression (PLS Regression),” *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 1, pp. 97–106, 2010. [44](#)
- [236] H. Abdi and L. J. Williams, “Principal component analysis,” *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010. [44](#)
- [237] H. Wold, “Nonlinear iterative partial least squares (NIPALS) modelling: some current developments,” in *Multivariate Analysis-III*, pp. 383–407, Elsevier, 1973. [45](#)

- [238] S. De Jong, “SIMPLS: an alternative approach to partial least squares regression,” *Chemometrics and intelligent laboratory systems*, vol. 18, no. 3, pp. 251–263, 1993. 45
- [239] P. I. Good, *Permutation, Parametric, and Bootstrap Tests of Hypotheses (Springer Series in Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2004. 48
- [240] M. Ojala and G. C. Garriga, “Permutation tests for studying classifier performance,” *Journal of Machine Learning Research*, vol. 11, no. Jun, pp. 1833–1863, 2010. 48
- [241] “Alzheimer’s Disease Neuroimaging Initiative (ADNI).” <http://adni.loni.usc.edu/>. Accessed: 2019-09-27. 49, 60, 82, 93
- [242] “FreeSurfer (FS).” <https://surfer.nmr.mgh.harvard.edu/>. Accessed: 2019-09-27. 49, 60
- [243] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, *et al.*, “Automatically parcelating the human cerebral cortex,” *Cerebral cortex*, vol. 14, no. 1, pp. 11–22, 2004. 49, 60
- [244] D. J. MacKay, “An Example Inference Task: Clustering,” in *Information theory, inference and learning algorithms*, ch. 20, pp. 284–292, Cambridge university press, 2003. 50
- [245] “Relationship between CSF biomarkers of Alzheimer’s disease and rates of regional cortical thinning in ADNI data, author=Tosun, Duygu and Schuff, Norbert and Shaw, Leslie M and Trojanowski, John Q and Weiner, Michael W and Alzheimer’s Disease NeuroImaging Initiative and others, journal=Journal of Alzheimer’s Disease, volume=26, number=s3, pages=77–90, year=2011, publisher=IOS Press,” 51, 52, 69
- [246] M. J. de Leon, E. Pirraglia, R. S. Osorio, L. Glodzik, L. Saint-Louis, H.-J. Kim, J. Fortea, S. Fossati, E. Laska, C. Siegel, *et al.*, “The nonlinear relationship between cerebrospinal fluid A β 42 and tau in preclinical Alzheimer’s disease,” *PloS one*, vol. 13, no. 2, p. e0191240, 2018. 55
- [247] H. Lemaitre, A. L. Goldman, F. Sambataro, B. A. Verchinski, A. Meyer-Lindenberg, D. R. Weinberger, and V. S. Mattay, “Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume?,” *Neurobiology of aging*, vol. 33, no. 3, pp. 617–e1, 2012. 57
- [248] P. M. Doraiswamy, J. Leon, J. L. Cummings, D. Marin, and P. J. Neumann, “Prevalence and impact of medical comorbidity in Alzheimer’s disease,” *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 57, no. 3, pp. M173–M177, 2002. 59
- [249] M. W. Bondi, W. S. Houston, L. T. Eyler, and G. G. Brown, “fMRI evidence of compensatory mechanisms in older adults at genetic risk for Alzheimer disease,” *Neurology*, vol. 64, no. 3, pp. 501–508, 2005. 69

- [250] Z. Qi, X. Wu, Z. Wang, N. Zhang, H. Dong, L. Yao, and K. Li, "Impairment and compensation coexist in amnesic MCI default mode network," *Neuroimage*, vol. 50, no. 1, pp. 48–55, 2010. 69
- [251] C. Hutton, B. Draganski, J. Ashburner, and N. Weiskopf, "A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging," *Neuroimage*, vol. 48, no. 2, pp. 371–380, 2009. 69
- [252] D. H. Salat, R. L. Buckner, A. Z. Snyder, D. N. Greve, R. S. Desikan, E. Busa, J. C. Morris, A. M. Dale, and B. Fischl, "Thinning of the cerebral cortex in aging," *Cerebral cortex*, vol. 14, no. 7, pp. 721–730, 2004. 69
- [253] M. Thambisetty, J. Wan, A. Carass, Y. An, J. L. Prince, and S. M. Resnick, "Longitudinal changes in cortical thickness associated with normal aging," *Neuroimage*, vol. 52, no. 4, pp. 1215–1223, 2010. 69
- [254] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A. D. N. Initiative, *et al.*, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects," *Neuroimage*, vol. 104, pp. 398–412, 2015. 70
- [255] G. P. Morris, I. A. Clark, and B. Vissel, "Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease," *Acta neuropathologica communications*, vol. 2, no. 1, p. 135, 2014. 72
- [256] S. Salloway, R. Sperling, N. C. Fox, K. Blennow, W. Klunk, M. Raskind, M. Sabbagh, L. S. Honig, A. P. Porsteinsson, S. Ferris, *et al.*, "Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease," *New England Journal of Medicine*, vol. 370, no. 4, pp. 322–333, 2014. 72
- [257] P. T. Francis, A. M. Palmer, M. Snape, and G. K. Wilcock, "The cholinergic hypothesis of Alzheimer's disease: a review of progress," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 66, no. 2, pp. 137–147, 1999. 73
- [258] J. Hardy and D. Allsop, "Amyloid deposition as the central event in the aetiology of Alzheimer's disease," *Trends in pharmacological sciences*, vol. 12, pp. 383–388, 1991. 73
- [259] L. Pei-Pei, Y. Xie, M. Xiao-Yan, and K. Jian-Sheng, "History and progress of hypotheses and clinical trials for Alzheimer's disease," *Signal Transduction and Targeted Therapy*, vol. 4, pp. 1–22, 2019. 73
- [260] S. Filser, S. V. Ovsepiyan, M. Masana, L. Blazquez-Llorca, A. B. Elvang, C. Volbracht, M. B. Müller, C. K. Jung, and J. Herms, "Pharmacological inhibition of BACE1 impairs synaptic plasticity and cognitive functions," *Biological psychiatry*, vol. 77, no. 8, pp. 729–739, 2015. 74
- [261] "Statistical Parametric Mapping (SPM)." <http://www.fil.ion.ucl.ac.uk/spm/>. Accessed: 2019-09-27. 83, 93
- [262] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 8, pp. 1226–1238, 2005. 85, 88

- [263] M. Grothe, H. Heinsen, and S. J. Teipel, "Atrophy of the cholinergic basal forebrain over the adult age range and in early stages of Alzheimer's disease," *Biological psychiatry*, vol. 71, no. 9, pp. 805–813, 2012. [90](#)
- [264] B. C. Dickerson, A. Bakkour, D. H. Salat, E. Feczko, J. Pacheco, D. N. Greve, F. Grodstein, C. I. Wright, D. Blacker, H. D. Rosas, *et al.*, "The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals," *Cerebral cortex*, vol. 19, no. 3, pp. 497–510, 2008. [90](#)
- [265] A. Nakamura, N. Kaneko, V. L. Villemagne, T. Kato, J. Doecke, V. Doré, C. Fowler, Q.-X. Li, R. Martins, C. Rowe, *et al.*, "High performance plasma amyloid- β biomarkers for Alzheimer's disease," *Nature*, vol. 554, no. 7691, p. 249, 2018. [91](#)
- [266] V. Ovod, K. N. Ramsey, K. G. Mawuenyega, J. G. Bollinger, T. Hicks, T. Schneider, M. Sullivan, K. Paumier, D. M. Holtzman, J. C. Morris, *et al.*, "Amyloid β concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis," *Alzheimer's & Dementia*, vol. 13, no. 8, pp. 841–849, 2017. [91](#)
- [267] A. Nabers, L. Perna, J. Lange, U. Mons, J. Schartner, J. Guldnhaupt, K.-U. Saum, S. Janelidze, B. Holczek, D. Rujescu, *et al.*, "Amyloid blood biomarker detects Alzheimer's disease," *EMBO molecular medicine*, vol. 10, no. 5, 2018. [91](#)
- [268] E. E. Bron, M. Smits, W. M. Van Der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J. M. Papma, R. M. Steketee, C. M. Orellana, R. Meijboom, *et al.*, "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge," *NeuroImage*, vol. 111, pp. 562–579, 2015. [92](#)
- [269] D. Irwin, C. McMillan, J. Toledo, S. Arnold, L. Shaw, L.-S. Wang, V. Lee, J. Trojanowski, and M. Grossman, "Comparison of Cerebrospinal Fluid Levels of Tau and A β 1-42 in Alzheimer's Disease and Frontotemporal Degeneration Using Two Analytical Platforms (P02. 055)," 2012. [92](#)
- [270] L.-S. Wang, Y. Y. Leung, S.-K. Chang, S. Leight, M. Knapik-Czajka, Y. Baek, L. M. Shaw, V. M.-Y. Lee, J. Q. Trojanowski, and C. M. Clark, "Comparison of xMAP and ELISA assays for detecting cerebrospinal fluid biomarkers of Alzheimer's disease," *Journal of Alzheimer's disease*, vol. 31, no. 2, pp. 439–445, 2012. [92](#)
- [271] A. M. Fagan, L. M. Shaw, C. Xiong, H. Vanderstichele, M. A. Mintun, J. Q. Trojanowski, E. Coart, J. C. Morris, and D. M. Holtzman, "Comparison of analytical platforms for cerebrospinal fluid measures of β -amyloid 1-42, total tau, and p-tau181 for identifying Alzheimer disease amyloid plaque pathology," *Archives of neurology*, vol. 68, no. 9, pp. 1137–1144, 2011. [92](#)
- [272] J.-H. Kang, M. Korecka, J. B. Toledo, J. Q. Trojanowski, and L. M. Shaw, "Clinical utility and analytical challenges in measurement of cerebrospinal fluid amyloid- β 1-42 and τ proteins as Alzheimer disease biomarkers," *Clinical chemistry*, vol. 59, no. 6, pp. 903–916, 2013. [92](#)

- [273] O. Hansson, J. Seibyl, E. Stomrud, H. Zetterberg, J. Q. Trojanowski, T. Bittner, V. Lifke, V. Corradini, U. Eichenlaub, R. Batrla, *et al.*, “CSF biomarkers of Alzheimer’s disease concord with amyloid- β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts,” *Alzheimer’s & Dementia*, vol. 14, no. 11, pp. 1470–1481, 2018. [92](#)
- [274] J. Ashburner, “A fast diffeomorphic image registration algorithm,” *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007. [93](#)
- [275] A. C. Evans, D. L. Collins, S. Mills, E. Brown, R. Kelly, and T. M. Peters, “3D statistical neuroanatomical models from 305 MRI volumes,” in *1993 IEEE conference record nuclear science symposium and medical imaging conference*, pp. 1813–1817, IEEE, 1993. [94](#)
- [276] S. Le Cessie and J. C. Van Houwelingen, “Ridge estimators in logistic regression,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 41, no. 1, pp. 191–201, 1992. [94](#)
- [277] J. Dukart, M. L. Schroeter, K. Mueller, A. D. N. Initiative, *et al.*, “Age correction in dementia—matching to a healthy brain,” *PLoS one*, vol. 6, no. 7, p. e22193, 2011. [95](#)
- [278] S. L. Risacher and A. J. Saykin, “Neuroimaging biomarkers of neurodegenerative diseases and dementia,” in *Seminars in neurology*, vol. 33, pp. 386–416, Thieme Medical Publishers, 2013. [101](#)
- [279] M. Storandt, M. A. Mintun, D. Head, and J. C. Morris, “Cognitive decline and brain volume loss as signatures of cerebral amyloid- β peptide deposition identified with Pittsburgh compound B: cognitive decline associated with A β deposition,” *Archives of neurology*, vol. 66, no. 12, pp. 1476–1481, 2009. [103](#)
- [280] A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, K. Blennow, J. B. Brewer, A. M. Dale, and A. D. N. Initiative, “Brain atrophy in healthy aging is related to CSF levels of A β 1-42,” *Cerebral Cortex*, vol. 20, no. 9, pp. 2069–2079, 2010. [103](#)
- [281] D. Tosun, N. Schuff, D. Truran-Sacrey, L. M. Shaw, J. Q. Trojanowski, P. Aisen, R. Peterson, M. W. Weiner, A. D. N. Initiative, *et al.*, “Relations between brain tissue loss, csf biomarkers, and the apoe genetic profile: a longitudinal mri study,” *Neurobiology of aging*, vol. 31, no. 8, pp. 1340–1354, 2010. [103](#)
- [282] J. A. Becker, T. Hedden, J. Carmasin, J. Maye, D. M. Rentz, D. Putcha, B. Fischl, D. N. Greve, G. A. Marshall, S. Salloway, *et al.*, “Amyloid- β associated cortical thinning in clinically normal elderly,” *Annals of neurology*, vol. 69, no. 6, pp. 1032–1042, 2011. [103](#)
- [283] E. M. Arenaza-Urquijo, J.-L. Molinuevo, R. Sala-Llonch, C. Solé-Padullés, M. Balasa, B. Bosch, J. Olives, A. Antonell, A. Lladó, R. Sánchez-Valle, *et al.*, “Cognitive reserve proxies relate to gray matter loss in cognitively healthy elderly with abnormal cerebrospinal fluid amyloid- β levels,” *Journal of Alzheimer’s Disease*, vol. 35, no. 4, pp. 715–726, 2013. [103](#)

- [284] A. Casamitjana, V. Vilaplana, P. Petrone, J. L. Molinuevo, and J. D. Gispert, “Shared latent structures between imaging features and biomarkers in early stages of Alzheimer’s disease: a predictive study,” *IEEE journal of biomedical and health informatics*, 2019. [109](#)
- [285] P. M. Petrone, A. Casamitjana, C. Falcon, M. Artigues, G. Operto, S. Skouras, R. Cacciaglia, J. L. Molinuevo, V. Vilaplana, J. D. Gispert, *et al.*, “Characteristic brain volumetric changes in the AD preclinical signature,” *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, vol. 14, no. 7, p. P1235, 2018. [109](#)
- [286] A. Casamitjana Díaz, P. Petrone, M. Artigues, J. L. Molinuevo, J. D. Gispert, and V. Vilaplana Besler, “Projection to latent spaces disentangles specific cerebral morphometric patterns associated to aging and preclinical AD,” *Alzheimer’s & dementia*, vol. 14, no. 7, pp. 869–870, 2018. [110](#)
- [287] P. M. Petrone, V. Vilaplana, A. Casamitjana, D. S. Escobedo, A. Tucholka, R. Cacciaglia, G. Operto, S. Skouras, C. Falcon, J. L. Molinuevo, *et al.*, “Magnetic resonance imaging and machine learning make a valuable combined tool for the screening of preclinical AD,” *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, vol. 13, no. 7, p. P1245, 2017. [110](#)
- [288] A. Casamitjana, V. Vilaplana, P. Petrone, J. L. Molinuevo, and J. D. Gispert, “Shared Latent Structures Between Imaging Features and Biomarkers in Early Stages of Alzheimer’s Disease,” in *International Workshop on Predictive Intelligence In MEdicine*, pp. 60–67, Springer, 2018. [110](#)
- [289] S. Puch, A. Aduriz, A. Casamitjana, V. Vilaplana, P. Petrone, G. Operto, R. Cacciaglia, S. Skouras, C. Falcon, J. L. Molinuevo, *et al.*, “Voxelwise nonlinear regression toolbox for neuroimage analysis: Application to aging and neurodegenerative disease modeling,” *arXiv preprint arXiv:1612.00667*, 2016. [110](#)
- [290] A. Casamitjana, S. Puch, A. Aduriz, and V. Vilaplana, “3D Convolutional Neural Networks for Brain Tumor Segmentation: a comparison of multi-resolution architectures,” in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 150–161, Springer, 2016. [110](#)
- [291] L. Wang, D. Nie, G. Li, É. Puybureau, J. Dolz, Q. Zhang, F. Wang, J. Xia, Z. Wu, J. Chen, *et al.*, “Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge,” *IEEE transactions on medical imaging*, 2019. [110](#)
- [292] A. Casamitjana, M. Catà, I. Sánchez, M. Combalia, and V. Vilaplana, “Cascaded V-Net using ROI masks for brain tumor segmentation,” in *International MICCAI Brainlesion Workshop*, pp. 381–391, Springer, 2017. [111](#)
- [293] Y. Sasaki *et al.*, “The truth of the F-measure,” *Teach Tutor mater*, vol. 1, no. 5, pp. 1–5, 2007. [150](#), [151](#)

Appendices

Appendix A Chapter 3

Appendix A.1 List of relevant regions along the AD continuum

Relevant regions list	
ROI name	AD continuum
Pallidum R	0.0005
Pallidum L	0.0014
Paracentral R	0.0112
Paracentral L	
Parahippocampal R	-
Parahippocampal L	-
ParsOperuclaris R	0.001
ParsOperuclaris L	-
ParsOrbitalis R	-
ParsOrbitalis L	0.0
ParsTriangularis R	-
ParsTriangularis L	-
Pericalcarine R	0.0031
Pericalcarine L	0.0
Postcentral R	0.0012
Postcentral L	0.0425
PosteriorCingulate R	-
PosteriorCingulate L	-
Precentral R	0.0
Precentral L	0.0
Precuneus R	0.0001
Precuneus L	0.0034
Putamen R	-
Putamen L	-
RostralAnteriorCingulate R	-
RostralAnteriorCingulate L	-
RostralMiddleFrontal R	0.0384
RostralMiddleFrontal L	0.0411
SuperiorFrontal R	-
SuperiorFrontal L	-
SuperiorParietal R	-
SuperiorParietal L	-
SuperiorTemporal R	0.0064

Relevant regions list	
ROI name	AD continuum
SuperiorTemporal L	-
Supramarginal R	0.023
Supramarginal L	-
TemporalPole R	-
TemporalPole L	0.0042
Thalamus R	-
Thalamus L	-
TransverseTemporal R	-
TransverseTemporal L	0.0008
Insula R	0.0281
Insula L	0.0002
AccumbensArea R	-
AccumbensArea L	0.0302
Amygdala R	0.0
Amygdala L	0.0
Bankssts R	0.0045
Bankssts L	0.0
CaudalAnteriorcingulate R	0.0211
CaudalAnteriorcingulate L	-
CaudalMiddleFrontal R	-
CaudalMiddleFrontal L	-
Caudate R	-
Caudate L	0.0029
ChoroidPlexus R	0.0028
ChoroidPlexus L	0.0013
Cuneus R	0.0028
Cuneus L	0.0027
Entorhinal R	0.0
Entorhinal L	0.0
FrontalPole R	0.0006
FrontalPole L	0.0233
Fusiform R	0.0084
Fusiform L	0.0093
Hippocampus R	0.0
Hippocampus L	0.0
InferiorParietal R	0.0
InferiorParietal L	0.0
InferiorTemporal R	0.0006
InferiorTemporal L	0.0
IsthmusCingulate R	-
IsthmusCingulate L	-
LateralOccipital R	-
LateralOccipital L	-
LateralOrbitofrontal R	0.0
LateralOrbitofrontal L	0.0
Lingual R	-
Lingual L	-

Relevant regions list	
ROI name	AD continuum
MedialOrbitofrontal R	0.007
MedialOrbitofrontal L	0.0161
MiddleTemporal R	0.0
MiddleTemporal L	0.0

Table A.1: List of statistically relevant ($p < 0.05$) brain ROIs and their associated p-value in the AD continuum analysis of the CSF effect on mean ROI volume

Appendix A.2 List of relevant regions in different clinical AD stages.

Relevant regions list			
ROI name	CU	MCI	AD
Pallidum R	0.0078	0.0124	0.0254
Pallidum L	0.0224	0.0086	0.0167
Paracentral R	-	-	-
Paracentral L	-	0.039	-
Parahippocampal R	-	0.0287	-
Parahippocampal L	-	0.02	-
ParsOperuclaris R	-	-	-
ParsOperuclaris L	-	-	-
ParsOrbitalis R	-	-	-
ParsOrbitalis L	-	0.0021	-
ParsTriangularis R	-	-	-
ParsTriangularis L	-	-	-
Pericalcarine R	0.0006	-	-
Pericalcarine L	-	-	-
Postcentral R	0.0121	0.0	-
Postcentral L	-	-	-
PosteriorCingulate R	-	-	-
PosteriorCingulate L	-	-	-
Precentral R	-	-	-
Precentral L	-	0.0037	-
Precuneus R	-	0.0208	-
Precuneus L	-	-	-
Putamen R	-	-	-
Putamen L	0.0298	-	-
RostralAnteriorCingulate R	-	-	-
RostralAnteriorCingulate L	-	-	-
RostralMiddleFrontal R	-	-	0.0465
RostralMiddleFrontal L	-	-	-
SuperiorFrontal R	-	-	-
SuperiorFrontal L	-	-	-
SuperiorParietal R	-	-	-
SuperiorParietal L	-	-	-
SuperiorTemporal R	-	-	-
SuperiorTemporal L	-	-	-
Supramarginal R	-	-	-

Relevant regions list			
ROI name	CU	MCI	AD
Supramarginal L	0.0203	-	-
TemporalPole R	-	-	-
TemporalPole L	-	-	-
Thalamus R	-	0.0421	-
Thalamus L	-	0.0389	-
TransverseTemporal R	-	-	-
TransverseTemporal L	0.0424	0.0308	-
Insula R	-	-	-
Insula L	0.0176	-	-
AccumbensArea R	-	-	-
AccumbensArea L	0.0018	-	-
Amygdala R	0.0197	0.0289	0.0271
Amygdala L	0.0096	0.0381	-
Bankssts R	-	-	-
Bankssts L	-	0.0006	0.0068
CaudalAnteriorcingulate R	-	-	-
CaudalAnteriorcingulate L	-	-	-
CaudalMiddleFrontal R	-	-	-
CaudalMiddleFrontal L	-	-	-
Caudate R	0.0014	-	-
Caudate L	0.0012	-	-
ChoroidPlexus R	0.0	-	-
ChoroidPlexus L	0.002	0.0079	0.0088
Cuneus R	-	-	-
Cuneus L	-	-	-
Entorhinal R	-	-	0.0163
Entorhinal L	-	0.0304	-
FrontalPole R	-	-	-
FrontalPole L	-	-	-
Fusiform R	-	-	-
Fusiform L	-	-	-
Hippocampus R	-	0.0023	-
Hippocampus L	-	0.0016	-
InferiorParietal R	-	0.0003	0.0024
InferiorParietal L	-	0.0	0.0316
InferiorTemporal R	0.0035	-	-
InferiorTemporal L	0.0218	-	-
IsthmusCingulate R	-	-	-
IsthmusCingulate L	-	-	0
LateralOccipital R	-	-	-
LateralOccipital L	0.0094	-	-
LateralOrbitofrontal R	-	-	-
LateralOrbitofrontal L	-	-	-
Lingual R	-	-	-
Lingual L	-	-	-
MedialOrbitofrontal R	-	-	-
MedialOrbitofrontal L	-	-	-

Relevant regions list			
ROI name	CU	MCI	AD
MiddleTemporal R	-	0.0004	-
MiddleTemporal L	-	0.0	-

Table A.2: List of statistically relevant ($p < 0.05$) brain ROIs and their associated p-value of the CSF effect on mean ROI volume. The analysis is performed independently on different cognitive categories along the AD continuum.

Appendix A.3 List of relevant regions in presymptomatic AD

Relevant regions list			
ROI name	T0	T1	T2
Pallidum R	-	-	-
Pallidum L	-	-	-
Paracentral R	-	-	-
Paracentral L	-	-	-
Parahippocampal R	-	-	-
Parahippocampal L	-	-	-
ParsOperuclaris R	-	-	-
ParsOperuclaris L	-	-	-
ParsOrbitalis R	-	-	-
ParsOrbitalis L	-	-	-
ParsTriangularis R	-	-	-
ParsTriangularis L	-	-	-
Pericalcarine R	-	-	0.0002
Pericalcarine L	-	-	-
Pastcentral R	-	-	-
Pastcentral L	-	-	-
PosteriorCingulate R	-	-	-
PosteriorCingulate L	0.0282	-	-
Precentral R	-	-	-
Precentral L	-	-	-
Precuneus R	-	-	-
Precuneus L	-	-	-
Putamen R	-	-	-
Putamen L	-	0.0194	-
RostralAnteriorCingulate R	-	-	-
RostralAnteriorCingulate L	-	-	-
RostralMiddleFrontal R	-	-	-
RostralMiddleFrontal L	-	-	-
SuperiorFrontal R	-	-	-
SuperiorFrontal L	-	-	-
SuperiorParietal R	-	-	-
SuperiorParietal L	-	-	-
SuperiorTemporal R	0.0293	-	-
SuperiorTemporal L	-	-	-
Supramarginal R	-	0.039	-
Supramarginal L	-	-	-

Relevant regions list			
ROI name	T0	T1	T2
TemporalPole R	-	-	-
TemporalPole L	-	-	-
Thalamus R	-	-	-
Thalamus L	-	-	-
TransverseTemporal R	-	-	-
TransverseTemporal L	-	-	-
Insula R	0.0227	-	-
Insula L	0.0284	-	-
AccumbensArea R	-	0.0209	-
AccumbensArea L	-	-	-
Amygdala R	-	-	-
Amygdala L	-	-	-
Bankssts R	-	-	-
Bankssts L	-	-	-
CaudalAnteriorcingulate R	-	-	-
CaudalAnteriorcingulate L	-	-	-
CaudalMiddleFrontal R	-	-	-
CaudalMiddleFrontal L	-	-	-
Caudate R	-	-	-
Caudate L	-	-	-
ChoroidPlexus R	-	0.0005	0.0
ChoroidPlexus L	-	0.0096	0.0083
Cuneus R	-	-	-
Cuneus L	-	-	-
Entorhinal R	-	-	-
Entorhinal L	-	-	-
FrontalPole R	-	-	-
FrontalPole L	-	-	-
Fusiform R	-	-	-
Fusiform L	-	0.0209	-
Hippocampus R	-	-	-
Hippocampus L	-	-	-
InferiorParietal R	-	-	-
InferiorParietal L	-	-	-
InferiorTemporal R	0.0135	-	-
InferiorTemporal L	-	-	-
IsthmusCingulate R	-	-	-
IsthmusCingulate L	-	-	-
LateralOccipital R	-	-	-
LateralOccipital L	-	-	-
LateralOrbitofrontal R	-	-	-
LateralOrbitofrontal L	-	-	0.033
Lingual R	-	-	-
Lingual L	-	-	-
MedialOrbitofrontal R	-	-	-
MedialOrbitofrontal L	-	-	-
MiddleTemporal R	-	-	-

Relevant regions list			
ROI name	T0	T1	T2
MiddleTemporal L	-	-	-

Table A.3: List of statistically relevant ($p < 0.05$) brain ROIs and their associated p-value in the preclinical analysis of the CSF effect on mean ROI volume

Appendix A.4 Linear regression for prediction

We use linear regression with raw brain morphometric features as a baseline method for predictive analysis. Linear regression uses raw features to predict AD pathological markers and age. Tables A.4 and A.5 summarize the results using a single model for the whole AD continuum and a model for each AD stage, respectively.

Features	Diagnosis	Age	CSF- $A\beta$	CSF-ptau	CSF-ttau
Volume	HC	0.5	1.91	0.8	0.72
	PC	0.69	0.51	0.74	0.71
	MCI	0.68	0.43	0.74	0.73
	AD	0.85	0.41	0.93	0.89
Cortical Th.	HC	0.53	1.89	0.77	0.76
	PC	0.72	0.51	0.73	0.78
	MCI	0.74	0.43	0.71	0.78
	AD	0.88	0.4	0.96	0.99

Table A.4: Predictive error of response variables for each diagnosis label and feature type using a single linear regression model Alzheimer’s pathology.

Features	Diagnosis	Age	CSF- $A\beta$	CSF-ptau	CSF-ttau
Volume	HC	0.5	0.67	0.58	0.49
	PC	0.69	0.88	1.4	0.86
	MCI	0.68	0.47	0.8	0.8
	AD	0.85	0.63	1.82	1.7
Cortical Th.	HC	0.53	0.67	0.46	0.47
	PC	0.72	0.78	1.41	0.96
	MCI	0.74	0.46	0.76	0.83
	AD	0.88	0.58	1.4	1.35

Table A.5: Predictive error of response variables for each diagnosis label and feature type using linear regression fitting a specific model for each Alzheimer’s cognitive stage.

Appendix A.5 Statistical inference tables

We provide numbers of effect-strength and associated p-value of AD pathological markers and age on the brain morphological latent model at different AD stages in Tables A.6, A.7, A.8, A.9.

AD stage	Marker	Latent dimension					
		First	Second	Third	Fourth	Fifth	Sixth
HC	Age	-0.44 (0.003)	0.49 (<0.001)	0.01 (0.789)	-0.01 (0.89)	-0.0 (0.983)	0.0 (0.941)
	CSF A β	-0.01 (0.899)	0.06 (0.311)	0.01 (0.903)	0.14 (0.072)	0.12 (0.205)	0.09 (0.245)
	CSF p-tau	-0.01 (0.807)	-0.01 (0.808)	-0.01 (0.894)	0.06 (0.422)	0.06 (0.508)	0.1 (0.216)
	CSF t-tau	-0.0 (0.996)	-0.01 (0.937)	-0.19 (0.076)	0.22 (0.016)	0.08 (0.485)	0.09 (0.418)
	AD-CSF1	-0.13 (0.197)	0.02 (0.893)	0.02 (0.776)	0.04 (0.601)	-0.03 (0.609)	-0.04 (0.638)
	AD-CSF1	0.05 (0.497)	-0.0 (0.976)	-0.01 (0.958)	0.04 (0.672)	0.08 (0.539)	-0.13 (0.224)
	PC	Age	-0.42 (0.031)	0.28 (0.465)	-0.14 (0.31)	-0.19 (0.211)	-0.1 (0.462)
CSF A β		-0.01 (0.844)	-0.07 (0.282)	0.18 (0.064)	0.05 (0.734)	0.11 (0.314)	0.05 (0.572)
CSF p-tau		0.05 (0.61)	0.03 (0.769)	-0.31 (0.011)	0.32 (0.003)	-0.13 (0.349)	-0.04 (0.754)
CSF t-tau		0.07 (0.452)	-0.0 (0.999)	-0.32 (0.006)	0.35 (<0.001)	-0.2 (0.082)	0.02 (0.891)
AD-CSF1		0.02 (0.767)	0.08 (0.238)	-0.28 (0.005)	0.15 (0.148)	-0.15 (0.202)	-0.07 (0.466)
AD-CSF1		0.05 (0.426)	0.01 (0.921)	-0.2 (0.015)	0.16 (0.074)	-0.26 (0.002)	-0.01 (0.836)
MCI		Age	-0.36 (0.072)	0.31 (0.196)	-0.12 (0.524)	-0.38 (<0.001)	-0.08 (0.693)
	CSF A β	0.12 (0.205)	-0.09 (0.427)	0.26 (0.001)	0.01 (0.982)	0.06 (0.693)	-0.01 (0.929)
	CSF p-tau	0.04 (0.387)	-0.05 (0.364)	-0.22 (0.006)	0.26 (<0.001)	-0.27 (<0.001)	-0.19 (0.01)
	CSF t-tau	0.04 (0.468)	-0.03 (0.616)	-0.33 (<0.001)	0.33 (<0.001)	-0.21 (0.048)	0.08 (0.394)
	AD-CSF1	-0.08 (0.436)	-0.04 (0.712)	-0.28 (<0.001)	0.17 (0.068)	-0.18 (0.065)	-0.11 (0.179)
	AD-CSF1	-0.05 (0.412)	0.03 (0.623)	-0.33 (<0.001)	0.19 (0.069)	-0.18 (0.114)	0.05 (0.601)
	AD	Age	-0.15 (0.25)	0.26 (0.092)	0.26 (0.181)	-0.48 (<0.001)	-0.32 (0.122)
CSF A β		-0.15 (0.38)	-0.1 (0.576)	0.02 (0.771)	0.0 (0.977)	-0.09 (0.294)	-0.03 (0.725)
CSF p-tau		0.12 (0.336)	0.05 (0.769)	-0.25 (0.003)	0.25 (0.004)	-0.24 (0.003)	-0.12 (0.164)
CSF t-tau		0.12 (0.229)	-0.09 (0.399)	-0.28 (0.003)	0.37 (<0.001)	-0.15 (0.191)	0.17 (0.093)
AD-CSF1		0.18 (0.336)	0.12 (0.576)	-0.09 (0.213)	0.23 (<0.001)	-0.26 (<0.001)	-0.07 (0.386)
AD-CSF1		0.15 (0.238)	-0.06 (0.717)	-0.11 (0.304)	0.3 (0.002)	-0.16 (0.104)	0.09 (0.384)
AD continuum		Age	-0.31 (0.079)	0.29 (0.136)	-0.03 (0.82)	-0.37 (<0.001)	-0.15 (0.312)
	CSF A β	0.15 (0.189)	-0.12 (0.385)	0.33 (<0.001)	0.02 (0.958)	0.04 (0.893)	0.0 (0.976)
	CSF p-tau	-0.05 (0.43)	0.03 (0.723)	-0.35 (<0.001)	0.25 (0.001)	-0.22 (0.058)	-0.14 (0.153)
	CSF t-tau	-0.09 (0.431)	0.01 (0.94)	-0.45 (<0.001)	0.31 (0.005)	-0.18 (0.314)	0.08 (0.518)
	AD-CSF1	-0.12 (0.243)	0.06 (0.679)	-0.35 (<0.001)	0.17 (0.351)	-0.19 (0.175)	-0.1 (0.335)
	AD-CSF1	-0.15 (0.246)	0.06 (0.709)	-0.43 (<0.001)	0.18 (0.517)	-0.18 (0.326)	0.03 (0.804)

Table A.6: Effect size (p-value) of each associated marker of interest on on average brain ROI volume using a single PLS model for Alzheimer's pathology.

AD stage	Marker	Latent dimension					
		First	Second	Third	Fourth	Fifth	Sixth
HC	Age	-0.42 (0.026)	0.45 (0.011)	-0.03 (0.85)	-0.18 (0.033)	-0.03 (0.752)	-0.1 (0.216)
	CSF A β	0.03 (0.739)	0.07 (0.335)	0.04 (0.611)	0.1 (0.24)	0.02 (0.849)	-0.13 (0.06)
	CSF p-tau	-0.01 (0.864)	0.01 (0.865)	0.09 (0.308)	0.02 (0.795)	-0.05 (0.512)	0.11 (0.252)
	CSF t-tau	0.03 (0.892)	0.19 (0.064)	-0.08 (0.292)	0.1 (0.179)	-0.01 (0.941)	0.08 (0.414)
	AD-CSF1	-0.13 (0.16)	0.15 (0.114)	0.22 (0.005)	-0.07 (0.493)	0.01 (0.927)	0.09 (0.217)
	AD-CSF1	0.11 (0.267)	-0.0 (0.98)	-0.03 (0.754)	0.18 (0.015)	0.16 (0.03)	0.06 (0.561)
PC	Age	-0.4 (0.055)	0.15 (0.676)	-0.33 (0.147)	-0.2 (0.588)	-0.07 (0.668)	-0.05 (0.701)
	CSF A β	0.02 (0.672)	-0.14 (0.087)	0.14 (0.157)	-0.06 (0.492)	0.13 (0.318)	-0.01 (0.937)
	CSF p-tau	-0.13 (0.2)	0.18 (0.166)	-0.35 (0.011)	0.34 (0.003)	-0.06 (0.8)	-0.02 (0.923)
	CSF t-tau	-0.11 (0.254)	0.18 (0.162)	-0.35 (0.011)	0.35 (<0.001)	0.07 (0.732)	0.17 (0.268)
	AD-CSF1	-0.14 (0.174)	0.14 (0.304)	-0.34 (0.008)	0.17 (0.245)	-0.13 (0.55)	0.05 (0.742)
	AD-CSF1	-0.04 (0.46)	0.14 (0.151)	-0.24 (0.025)	0.17 (0.095)	-0.01 (0.977)	0.22 (0.059)
MCI	Age	-0.41 (0.026)	0.13 (0.723)	-0.19 (0.551)	-0.36 (0.03)	0.04 (0.815)	-0.08 (0.394)
	CSF A β	0.15 (0.063)	-0.06 (0.584)	0.12 (0.095)	-0.0 (0.988)	0.02 (0.749)	0.24 (<0.001)
	CSF p-tau	-0.02 (0.551)	-0.01 (0.894)	-0.21 (0.006)	0.27 (<0.001)	0.22 (0.009)	-0.08 (0.422)
	CSF t-tau	-0.11 (0.184)	0.08 (0.401)	-0.26 (0.001)	0.3 (<0.001)	0.3 (<0.001)	-0.07 (0.455)
	AD-CSF1	-0.14 (0.129)	0.02 (0.858)	-0.23 (<0.001)	0.15 (0.065)	0.09 (0.435)	-0.11 (0.152)
	AD-CSF1	-0.17 (0.102)	0.1 (0.433)	-0.25 (0.002)	0.21 (0.009)	0.14 (0.193)	-0.1 (0.244)
AD	Age	-0.06 (0.5)	0.17 (0.318)	0.19 (0.556)	-0.6 (<0.001)	0.21 (0.568)	0.01 (0.942)
	CSF A β	-0.12 (0.49)	-0.08 (0.482)	-0.06 (0.508)	-0.06 (0.508)	0.03 (0.733)	0.19 (0.01)
	CSF p-tau	-0.06 (0.541)	-0.05 (0.555)	-0.25 (0.004)	0.21 (0.022)	0.2 (0.066)	-0.2 (0.017)
	CSF t-tau	-0.15 (0.533)	-0.13 (0.31)	-0.24 (0.003)	0.27 (0.002)	0.19 (0.112)	0.06 (0.569)
	AD-CSF1	0.0 (0.911)	-0.08 (0.182)	-0.11 (0.075)	0.15 (0.017)	0.12 (0.133)	-0.18 (0.016)
	AD-CSF1	0.03 (0.422)	-0.09 (0.181)	-0.07 (0.551)	0.25 (<0.001)	0.13 (0.255)	-0.02 (0.88)
AD continuum	Age	-0.31 (0.053)	0.15 (0.609)	-0.11 (0.723)	-0.4 (<0.001)	0.07 (0.733)	-0.04 (0.667)
	CSF A β	0.18 (0.068)	-0.08 (0.644)	0.25 (0.003)	-0.03 (0.924)	0.02 (0.883)	0.21 (<0.001)
	CSF p-tau	-0.17 (0.187)	0.02 (0.912)	-0.35 (<0.001)	0.25 (0.016)	0.17 (0.316)	-0.14 (0.174)
	CSF t-tau	-0.25 (0.136)	0.03 (0.912)	-0.42 (<0.001)	0.28 (0.046)	0.23 (0.161)	-0.05 (0.692)
	AD-CSF1	-0.22 (0.117)	0.03 (0.904)	-0.35 (<0.001)	0.15 (0.511)	0.07 (0.716)	-0.13 (0.145)
	AD-CSF1	-0.24 (0.099)	0.07 (0.794)	-0.39 (<0.001)	0.19 (0.356)	0.13 (0.507)	-0.07 (0.499)

Table A.7: Effect size (p-value) of each associated marker of interest on on average brain ROI cortical thickness using a single PLS model for Alzheimer's pathology.

AD stage	Marker	Latent dimension					
		First	Second	Third	Fourth	Fifth	Sixth
HC	Age	-0.44 (0.01)	0.49 (<0.001)	-0.07 (0.22)	-0.01 (0.88)	-0.14 (0.01)	0.1 (0.06)
	CSF A β	-0.01 (0.9)	0.06 (0.29)	-0.32 (<0.001)	0.36 (<0.001)	-0.23 (<0.001)	-0.05 (0.59)
	CSF p-tau	-0.01 (0.81)	-0.01 (0.8)	-0.27 (<0.001)	-0.22 (<0.001)	-0.24 (<0.001)	-0.25 (<0.001)
	CSF t-tau	-0.0 (0.99)	-0.01 (0.93)	-0.52 (<0.001)	-0.2 (0.07)	-0.28 (<0.001)	0.28 (<0.001)
	AD-CSF1	-0.13 (0.2)	0.02 (0.88)	0.01 (0.87)	-0.08 (0.29)	-0.09 (0.19)	-0.16 (0.02)
	AD-CSF1	0.05 (0.5)	-0.0 (0.98)	-0.16 (0.13)	-0.23 (0.02)	0.07 (0.55)	0.14 (0.21)
PC	Age	-0.42 (0.04)	0.28 (0.45)	0.1 (0.5)	0.12 (0.44)	-0.15 (0.28)	0.25 (0.04)
	CSF A β	-0.01 (0.85)	-0.07 (0.28)	-0.18 (0.04)	-0.06 (0.61)	0.33 (<0.001)	-0.19 (0.03)
	CSF p-tau	0.05 (0.61)	0.03 (0.8)	0.6 (<0.001)	0.35 (<0.001)	-0.22 (0.08)	-0.22 (0.09)
	CSF t-tau	0.07 (0.48)	-0.0 (1.0)	0.58 (<0.001)	0.43 (<0.001)	-0.21 (0.1)	-0.18 (0.14)
	AD-CSF1	0.02 (0.79)	0.08 (0.26)	0.39 (<0.001)	0.28 (<0.001)	-0.29 (<0.001)	0.03 (0.8)
	AD-CSF1	0.05 (0.47)	0.01 (0.92)	0.33 (<0.001)	0.25 (0.01)	-0.27 (<0.001)	0.02 (0.82)
MCI	Age	-0.36 (0.07)	0.31 (0.18)	0.02 (0.94)	0.37 (<0.001)	0.08 (0.61)	0.1 (0.52)
	CSF A β	0.12 (0.21)	-0.09 (0.4)	0.25 (0.01)	-0.09 (0.6)	-0.15 (0.14)	0.01 (0.94)
	CSF p-tau	0.04 (0.39)	-0.05 (0.36)	-0.37 (<0.001)	-0.27 (<0.001)	0.11 (0.16)	0.27 (<0.001)
	CSF t-tau	0.04 (0.5)	-0.03 (0.64)	-0.47 (<0.001)	-0.28 (<0.001)	0.24 (<0.001)	0.03 (0.75)
	AD-CSF1	-0.08 (0.43)	-0.04 (0.73)	-0.36 (<0.001)	-0.11 (0.31)	0.11 (0.23)	0.15 (0.08)
	AD-CSF1	-0.05 (0.44)	0.03 (0.61)	-0.4 (<0.001)	-0.12 (0.27)	0.21 (0.01)	0.07 (0.48)
AD	Age	-0.15 (0.27)	0.26 (0.08)	0.48 (<0.001)	-0.07 (0.73)	-0.22 (0.27)	-0.21 (0.28)
	CSF A β	-0.15 (0.36)	-0.1 (0.62)	-0.02 (0.81)	-0.04 (0.59)	-0.11 (0.23)	0.02 (0.85)
	CSF p-tau	0.12 (0.32)	0.05 (0.79)	-0.41 (<0.001)	-0.31 (<0.001)	-0.32 (<0.001)	0.07 (0.4)
	CSF t-tau	0.12 (0.23)	-0.09 (0.42)	-0.49 (<0.001)	-0.32 (<0.001)	0.09 (0.4)	-0.27 (<0.001)
	AD-CSF1	0.18 (0.34)	0.12 (0.58)	-0.25 (<0.001)	-0.25 (<0.001)	-0.17 (0.02)	0.06 (0.41)
	AD-CSF1	0.15 (0.25)	-0.06 (0.73)	-0.33 (<0.001)	-0.24 (0.02)	0.09 (0.34)	-0.2 (0.05)

Table A.8: Effect size (p-value) of each associated marker of interest on on average brain ROI volume using a multiple PLS model for Alzheimer's pathology.

AD stage	Marker	Latent dimension					
		First	Second	Third	Fourth	Fifth	Sixth
HC	Age	-0.42 (0.02)	0.45 (0.01)	-0.05 (0.64)	-0.09 (0.43)	-0.13 (0.14)	-0.16 (0.06)
	CSF A β	0.03 (0.74)	0.07 (0.33)	0.17 (0.01)	-0.37 (<0.001)	0.31 (<0.001)	-0.09 (0.23)
	CSF p-tau	-0.01 (0.83)	0.01 (0.84)	0.42 (<0.001)	-0.1 (0.33)	-0.18 (0.04)	-0.22 (0.01)
	CSF t-tau	0.03 (0.88)	0.19 (0.06)	0.44 (<0.001)	-0.28 (<0.001)	-0.19 (0.02)	0.2 (0.02)
	AD-CSF1	-0.13 (0.14)	0.15 (0.12)	0.11 (0.22)	0.13 (0.14)	-0.06 (0.45)	-0.12 (0.12)
	AD-CSF1	0.11 (0.28)	-0.0 (0.98)	0.16 (0.05)	-0.05 (0.63)	0.02 (0.8)	0.14 (0.09)
PC	Age	-0.4 (0.06)	0.15 (0.65)	0.22 (0.39)	0.15 (0.81)	-0.24 (0.26)	-0.07 (0.66)
	CSF A β	0.02 (0.67)	-0.14 (0.08)	-0.24 (0.02)	-0.01 (0.89)	-0.0 (0.98)	-0.33 (<0.001)
	CSF p-tau	-0.13 (0.17)	0.18 (0.16)	0.56 (<0.001)	-0.26 (0.03)	-0.23 (0.2)	0.2 (0.23)
	CSF t-tau	-0.11 (0.23)	0.18 (0.17)	0.52 (<0.001)	-0.25 (0.03)	-0.25 (0.11)	-0.05 (0.76)
	AD-CSF1	-0.14 (0.19)	0.14 (0.31)	0.49 (<0.001)	-0.05 (0.87)	-0.14 (0.51)	0.27 (0.05)
	AD-CSF1	-0.04 (0.44)	0.14 (0.16)	0.35 (<0.001)	-0.1 (0.37)	-0.16 (0.23)	0.12 (0.39)
MCI	Age	-0.41 (0.03)	0.13 (0.7)	0.04 (0.91)	-0.3 (0.28)	-0.3 (0.07)	0.01 (0.88)
	CSF A β	0.15 (0.07)	-0.06 (0.57)	0.08 (0.31)	0.05 (0.54)	0.17 (<0.001)	-0.36 (<0.001)
	CSF p-tau	-0.02 (0.56)	-0.01 (0.91)	-0.42 (<0.001)	0.2 (<0.001)	-0.18 (0.02)	0.05 (0.57)
	CSF t-tau	-0.11 (0.19)	0.08 (0.41)	-0.45 (<0.001)	0.19 (<0.001)	-0.22 (0.01)	0.12 (0.2)
	AD-CSF1	-0.14 (0.14)	0.02 (0.86)	-0.3 (<0.001)	0.03 (0.77)	-0.16 (0.05)	0.21 (<0.001)
	AD-CSF1	-0.17 (0.1)	0.1 (0.45)	-0.34 (<0.001)	0.06 (0.4)	-0.18 (0.05)	0.24 (<0.001)
AD	Age	-0.06 (0.49)	0.17 (0.32)	0.49 (0.02)	-0.07 (0.94)	0.08 (0.78)	-0.33 (0.06)
	CSF A β	-0.12 (0.5)	-0.08 (0.5)	-0.06 (0.53)	-0.09 (0.34)	0.05 (0.55)	-0.03 (0.66)
	CSF p-tau	-0.06 (0.58)	-0.05 (0.52)	-0.39 (<0.001)	-0.3 (<0.001)	-0.25 (0.01)	-0.17 (0.06)
	CSF t-tau	-0.15 (0.57)	-0.13 (0.31)	-0.4 (<0.001)	-0.35 (<0.001)	-0.03 (0.78)	0.18 (0.08)
	AD-CSF1	0.0 (0.91)	-0.08 (0.2)	-0.24 (<0.001)	-0.17 (0.01)	-0.26 (<0.001)	-0.13 (0.08)
	AD-CSF1	0.03 (0.4)	-0.09 (0.17)	-0.25 (<0.001)	-0.33 (<0.001)	-0.08 (0.4)	0.2 (0.01)

Table A.9: Effect size (p-value) of each associated marker of interest on on average brain ROI cortical thickness using a multiple PLS models for Alzheimer's pathology.

Appendix B Chapter 5

Appendix B.1 Cross-sectional analysis

Appendix B.1.1 Multiclass classification

Feature selection was carried out in the ADNI dataset for the classification of (i) healthy controls (HC) vs MCI due to AD and (ii) HC vs dementia due to AD (Figure B.1) following identical workflow to the one utilized in the classification of HC vs PC listed in the main text. T1 derived features in the HC vs MCI and HC vs AD-dementia models are highly overlapping. Notably, all discriminants of diagnostic categories have a negative weight indicating a loss of ROI volume for AD-dementia patients.

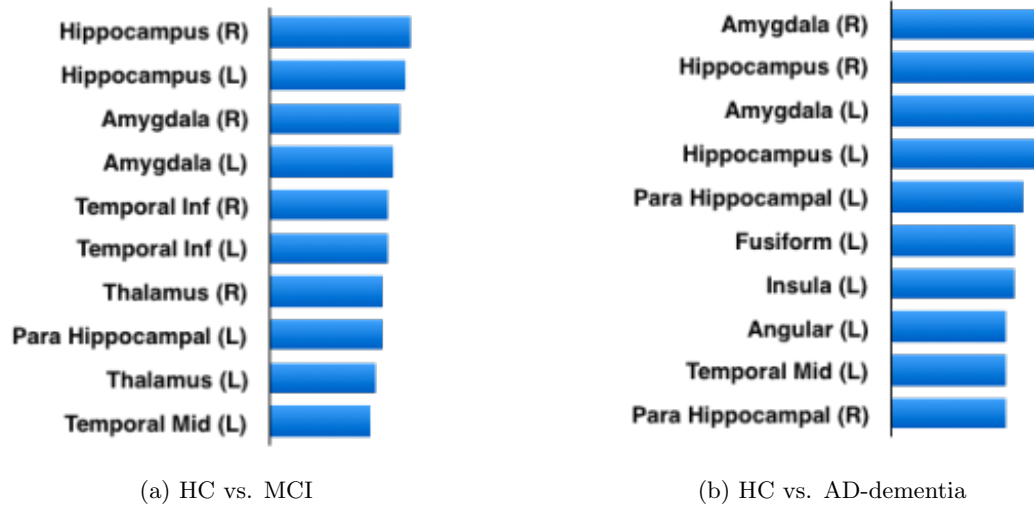


Figure B.1: Top 10 most relevant features for (a) HC vs MCI and (b) HC vs AD-dementia models. All features have a negative sign, indicating decreased volumes in MRI of AD-dementia patients with respect to HC.

The classification performance is shown in Tables B.10 and B.11 using T1 and T1+DTI derived features.

	HC vs MCI (a)			HC vs. AD-dementia (b)		
	AUC	Prec	Rec	AUC	Prec	Rec
10	0.968	0.89	0.92	0.931	0.709	0.895
20	0.956	0.885	0.85	0.92	0.70	0.84
30	0.945	0.821	0.852	0.92	0.7	0.84
40	0.947	0.774	0.889	0.91	0.61	0.89
50	0.926	0.75	0.889	0.892	0.57	0.89

Table B.10: T1 features have been trained on the ADNI cohort for the classification of (a) HC vs MCI and (b) HC vs AD-dementia. Preclinical subjects are excluded from the training and test sets. Performance (AUC) is evaluated on the HCB cohort for the classification of HC vs MCI and HC vs AD-dementia. Precision and recall are reported for the threshold that maximizes F1-score [293]

	HC vs MCI (a)			HC vs. AD-dementia (b)		
	AUC	Prec	Rec	AUC	Prec	Rec
10	0.932	0.85	0.81	0.96	0.86	0.86
20	0.947	1.0	0.762	0.94	0.8	0.86
30	0.93	0.78	0.86	0.95	0.76	0.93
40	0.94	1.0	0.71	0.94	0.8	0.86
50	0.93	1.0	0.71	0.91	0.71	0.86

Table B.11: T1+DTI features have been trained on the ADNI cohort for the classification of (a) HC vs MCI and (b) HC vs AD-dementia. Preclinical subjects have been excluded from the training and test sets. Performance (AUC) is evaluated on the HCB cohort for the classification of HC vs MCI and HC vs AD-dementia. Precision and recall are reported for the threshold that maximizes F1-score [293]

Appendix B.1.2 Feature selection: T1-derived features

In Table B.12, we show the top 20 features selected by the F-test method, ranked by logistic regression (LR) weight.

Ranking order	Brain ROI	LR weight
1	Caudate (L)	1.05
2	Amygdala (R)	0.86
3	Parietal Sup (L)	0.79
4	Temporal Pole Mid (L)	0.78
5	Olfactory (L)	0.78
6	Frontal Inf Tri (R)	0.70
7	Pallidum (L)	0.65
8	Occipital Inf (L)	0.54
9	Temporal Inf (R)	0.47
10	Rectus (L)	0.38
11	Cingulum Ant (R)	0.37
12	Frontal Inf Tri (L)	0.36
13	Parietal Inf (R)	0.34
14	Temporal Inf (L)	0.33
15	Fusiform (R)	0.32
16	Hippocampus (R)	0.15
17	Occipital Inf (R)	0.10
18	Olfactory (R)	0.05
19	Caudate (R)	0.04
20	Frontal Inf Oper (R)	0.04

Table B.12: Features derived from T1 images as normalized GMV from ROIs. Top-20 features selected using F-test are ranked according to the logistic regression (LR) absolute weights estimated on the training set (ADNI cohort).

Appendix B.1.3 Feature selection: Multimodal-derived features

In Table B.13, we show the top 45 features selected by mrm-FCD feature selection method.

Ranking order	Brain ROI	LR weight
1	Frontal Inf Orb (L), Temporal Pole Mid (L)	0.71
2	Occipital Inf (L), Parietal Inf (L)	0.46
3	Hippocampus (L), Frontal Inf Tri (R)	0.43
4	Angular (R), Pallidum (R)	0.39
5	Frontal Mid Orb (L), Pallidum (R)	0.39
6	Frontal Mid Orb (R), Thalamus (R)	0.39
7	Frontal Sup Medial (R), Amygdala (R)	0.38
8	Occipital Inf (L), Angular (L)	0.37
9	Frontal Inf Orb (L), Rectus (L)	0.37
10	Precentral (L), Hippocampus (L)	0.37
11	Cingulum Ant (L), Pallidum (L)	0.36
12	Frontal Med Orb (L), Pallidum (R)	0.35
13	Caudate (L), Pallidum (L)	0.34
14	Parietal Sup (R), Precuneus (R)	0.33
15	Occipital Inf (L), Cingulum Post (R)	0.31
16	Cingulum Ant (L), Pallidum (R)	0.31
17	Cingulum Post (R), Calcarine (R)	0.30
18	Olfactory (L)	0.27
19	Cingulum Mid (L), Temporal Pole Sup (R)	0.25
20	Putamen (L), Supp Motor Area (R)	0.24
21	Pallidum (R), Temporal Mid (R)	0.23
22	Pallidum (L), Parietal Sup (R)	0.22
23	Pallidum (L)	0.20
24	Frontal Sup Medial (L), Pallidum (R)	0.20
25	Cingulum Post (L), Fusiform (R)	0.20
26	Frontal Med Orb (L), Temporal Pole Mid (L)	0.20
27	Frontal Sup Medial (L), Temporal Sup (L)	0.19
28	Cuneus (L), Fusiform (R)	0.19
29	Cingulum Ant (R), Pallidum (R)	0.19
30	Occipital Inf (R), Temporal Mid (R)	0.17
31	Cingulum Post (R), Parietal Sup (R)	0.17
32	Frontal Sup (R), Temporal Pole Sup (R)	0.17
33	Cingulum Ant (L), Cingulum Post (L)	0.17
34	Putamen (L), Temporal Pole Mid (L)	0.14
35	Caudate (R), Pallidum (R)	0.14
36	Occipital Sup (L), Putamen (L)	0.13
37	Parietal Inf (R), Pallidum (R)	0.13
38	Temporal Pole Mid (L), Caudate (R)	0.12
39	Frontal Sup Medial (L), Insula (L)	0.10
40	Pallidum (L), SupraMarginal (R)	0.09
41	Fusiform (R), Angular (R)	0.08
42	Frontal Sup (L), Pallidum (R)	0.08

Ranking order	Brain ROI	LR weight
43	Cingulum Post (R), Occipital Inf (R)	0.05
44	Frontal Sup (R), Parietal Sup (R)	0.03
45	Caudate (R), Putamen (R)	0.00

Table B.13: Multimodal (T1+DTI) top ranking features as ordered by the logistic regression (LR) absolute weights estimated on the training set (ADNI cohort). Only two of these highly informative features are GMV from the T1 modality (Olfactory (L) and Pallidum (R)).

Appendix B.2 Longitudinal analysis

Appendix B.2.1 Selected subjects

In Table B.14, we show the acquisition sites and field strength of the selected subjects with available Jacobians that are, at least, 2.5 years apart.

RID	Diagnosis	Site ID	Field strength
96	Control	10/16 (m60,m72)	1.5T
118	Control	13/19(m60,m72)	1.5T
260	Control	114/47 (m60,m72, m84, 96, 108)	1.5T
352	Control	45/39 (m60,m72, m84, 96)	1.5T
498	Control	103/4 (m48,m60,m72)	1.5T
519	Control	110/11 (m48,m60,m72)	1.5T
559	Control	101	1.5T
619	Control	101	1.5T
637	Control	108	1.5T
672	Control	17/25(m48)	1.5T
677	Control	16/22(m60,m72,m84)	1.5T
685	Control	101	1.5T
923	Control	52/23(m48,m60,m72,m84,m96)	1.5T
1169	Control	16/22(m60,m72,m84)	1.5T
1250	Control	57/29(m48,m72,m84)	1.5T
257	PC	32	1.5T
259	PC	114/47 (m60,m72, m84, 96)	1.5T
295	PC	101	1.5T
479	PC	16/2 (m60,m72)	1.5T
403	PC	13	1.5T
520	PC	110/11 (m48,m60,m72)	1.5T
555	PC	6/24 (m48,m60,m72, m96)	1.5T
984	PC	9/15(m48,m60,m72,m84,m96)	1.5T
1098	PC	52/23(m48,m60,m72,m84,m96,m120)	1.5T
1276	PC	109	1.5T

Table B.14: Acquisition characteristics of the selected subjects on the interval $3.5 > \Delta t > 2.5$ years.

Appendix B.2.2 F-test feature selection

In Table B.15, we show the percentage of each region that best discriminates amyloid positivity in a pool of cognitively unimpaired subjects.

ROI number	ROI description	% of relevant voxels	minimum relevant voxels per iterations
1	Left Precentral	0	0
2	Right Precentral	3.19E-03	7.65E-05
3	Left Superior Frontal	0	0
4	Right Superior Frontal	0	0
5	Left Superior Frontal Orbital	0	0
6	Right Superior Frontal Orbital	9.33E-04	3.26E-05
7	Left Middle Frontal	0	0
8	Right Middle Frontal	0	0
9	Left Middle Frontal Orbital	0	0
10	Right Middle Frontal Orbital	0	0
11	Left Inferior Frontal Operculum	0	0
12	Right Inferior Frontal Operculum	2.96E-03	1.15E-04
13	Left Inferior Frontal	0	0
14	Right Inferior Frontal	0	0
15	Left Inferior Frontal Orbital	0	0
16	Right Inferior Frontal Orbital	0	0
17	Left Rolandic Operculum	0	0
18	Right Rolandic Operculum	3.97E-03	7.28E-05
19	Left Superior Motor	0	0
20	Right Superior Motor	0	0

ROI number	ROI description	% of relevant voxels	minimum relevant voxels per iterations
21	Left Olfactory	1.23E-01	8.82E-03
22	Right Olfactory	2.04E-01	9.78E-03
23	Left Superior Medial Frontal	0	0
24	Right Superior Medial Frontal	0	0
25	Left Medial Frontal Orbital	0	0
26	Right Medial Frontal Orbital	0	0
27	Left Rectus	1.37E-02	4.42E-04
28	Right Rectus	3.40E-02	1.60E-03
29	Left Insula	3.99E-03	6.64E-05
30	Right Insula	6.17E-02	6.86E-03
31	Left Anterior Cingulum	0	0
32	Right Anterior Cingulum	0	0
33	Left Middle Cingulum	1.23E-02	8.40E-04
34	Right Middle Cingulum	7.98E-04	1.20E-05
35	Left Cingulum	4.34E-02	3.96E-03
36	Right Cingulum	4.08E-02	1.15E-03
37	Left Hippocampus	1.43E-01	2.16E-02
38	Right Hippocampus	2.25E-01	3.97E-02
39	Left Parahippocampus	5.96E-02	1.62E-03
40	Right Parahippocampus	1.17E-01	1.10E-02
41	Left Amygdala	6.48E-02	1.84E-03
42	Right Amygdala	1.03E-02	4.54E-04
43	Left Calcarine	7.78E-02	1.52E-02
44	Right Calcarine	4.65E-02	3.93E-03
45	Left Cuneus	4.14E-02	1.38E-03
46	Right Cuneus	3.12E-02	4.97E-04
47	Left Lingual	1.03E-02	3.10E-04
48	Right Lingual	9.21E-03	5.08E-04
49	Left Superior Occipital	1.39E-01	8.83E-03
50	Right Superior Occipital	1.18E-01	7.87E-03

ROI number	ROI description	% of relevant voxels	minimum relevant voxels per iterations
51	Left Middle Occipital	4.07E-02	2.16E-03
52	Right Middle Occipital	7.65E-02	3.94E-03
53	Left Inferior Occipital	0	0
54	Right Inferior Occipital	8.79E-04	8.79E-06
55	Left Fusiform	2.03E-02	2.98E-04
56	Right Fusiform	3.40E-01	6.58E-02
57	Left Postcentral	0	0
58	Right Postcentral	6.50E-03	8.87E-05
59	Left Superior Parietal	4.49E-02	2.14E-03
60	Right Superior Parietal	7.61E-02	4.36E-03
61	Left Inferior Parietal	1.28E-02	2.70E-04
62	Right Inferior Parietal	1.41E-02	2.58E-04
63	Left Supramarginal	1.74E-02	2.86E-04
64	Right Supramarginal	8.86E-03	1.56E-04
65	Left Angular	7.68E-04	7.68E-06
66	Right Angular	4.39E-02	2.30E-03
67	Left Precuneus	1.59E-01	1.58E-02
68	Right Precuneus	7.09E-02	6.82E-03
69	Left Central Paracentral Lobule	0	0
70	Right Central Paracentral Lobule	0	0
71	Left Caudate	6.03E-01	3.99E-01
72	Right Caudate	7.50E-01	6.04E-01
73	Left Putamen	0	0
74	Right Putamen	2.86E-03	4.77E-05
75	Left Pallidum	0	0
76	Right Pallidum	0	0
77	Left Thalamus	1.49E-01	7.09E-02
78	Right Thalamus	1.89E-01	6.16E-02
79	Left Heschl	0	0
80	Right Heschl	0	0

ROI number	ROI description	% of relevant voxels	minimum relevant voxels per iterations
81	Left Superior Temporal	1.96E-02	7.83E-04
82	Right Superior Temporal	4.93E-03	1.27E-04
83	Left Superior Temporal Pole	5.55E-02	3.92E-03
84	Right Superior Temporal Pole	2.08E-03	3.28E-05
85	Left Middle Temporal	7.12E-03	4.07E-04
86	Right Middle Temporal	2.20E-02	4.87E-04
87	Left Middle Temporal Pole	7.08E-03	3.12E-04
88	Right Middle Temporal Pole	2.18E-01	1.53E-02
89	Left Inferior Temporal	9.15E-04	1.07E-05
90	Right Inferior Temporal	1.63E-01	1.64E-02
91	Left Cerebellum	1.99E-04	1.99E-06
92	Right Cerebellum	0	0
93	Left Cerebellum 3	0	0
94	Right Cerebellum 3	0	0
95	Left Cerebellum 4-5	0	0
96	Right Cerebellum 4-5	0	0
97	Left Cerebellum 6	0	0
98	Right Cerebellum 6	8.68E-04	2.89E-05
99	Left Cerebellum 7b	0	0
100	Right Cerebellum 7b	0	0
101	Left Cerebellum 8	0	0
102	Right Cerebellum 8	7.14E-03	1.63E-04
103	Left Cerebellum 9	0	0

ROI number	ROI description	% of relevant voxels	minimum relevant voxels per iterations
104	Right Cerebellum 9	1.09E-03	1.09E-05
105	Left Cerebellum 10	0	0
106	Right Cerebellum 10	0	0
107	Left Vermis 1-2	0	0
108	Right Vermis 1-2	0	0
109	Left Vermis 3	0	0
110	Right Vermis 3	0	0
111	Left Vermis 4-5	0	0
112	Right Vermis 4-5	0	0
113	Left Vermis 6	0	0
114	Right Vermis 6	0	0
115	Left Vermis	0	0
116	Right Vermis	0	0
117	Left Vermis 8	0	0
118	Right Vermis 8	4.48E-03	4.48E-05
119	Left Vermis 9	0	0
120	Right Vermis 9	0	0
121	Left Vermis 10 0		0
122	Right Vermis 10	0	0

Table B.15: Percentage of discriminant voxels that correspond to each of the brain regions of interest (ROIs)

Appendix B.2.3 L1-feature selection

Performance of the voxelwise classifier using a feature selection method based on l1-penalization in the optimization step is shown in Table B.16

Appendix B.2.4 Classification: training on a close Δt interval

In contrast to the approach used in the main text (Chapter 5) we used a voxelwise classifier trained on a restricted interval ($3.5 > \Delta t > 2.5$ years) and tested in all other cases. Results are listed in Table B.17.

Reg. parameter (mean # features)	AUC	B.acc	Acc.	Prec.	Sens.	Spec.	F-score
20000 (205.31)	0.748	0.651	0.575	0.303	0.777	0.525	0.429
50000 (576.17)	0.742	0.628	0.534	0.277	0.783	0.472	0.4046
100000 (1286.95)	0.8	0.634	0.537	0.28	0.8	0.4725	0.41
500000 (3814.73)	0.768	0.64875	0.541	0.286	0.817	0.48	0.42
1000000 (5130.59)	0.77	0.65	0.54	0.283	0.823	0.47	0.42

Table B.16: Performance of the system using a different feature selection method (l1-norm selection) and evaluated on the interval $3.5 > \Delta t > 2.5$ years. The number of features depend on the l1-norm regularization parameter and on the training data. We report the average number of features used per iteration of the evaluation loop. Metrics used are: area under the curve (AUC), balanced accuracy (B.acc.), accuracy (Acc.), precision (Prec.), sensitivity (Sens.), specificity (Spec.) and F-score.

# features (%)	AUC	B.acc	Acc.	Prec.	Sens.	Spec.	F-score
6 (0.001)	0.57	0.52	0.44	0.66	0.21	0.39	0.31
65 (0.01)	0.59	0.54	0.46	0.68	0.23	0.4	0.33
653 (0.1)	0.63	0.56	0.5	0.66	0.24	0.46	0.34
1633 (0.25)	0.62	0.55	0.52	0.61	0.24	0.49	0.33
3266 (0.5)	0.62	0.55	0.52	0.59	0.24	0.5	0.33
6532 (1)	0.62	0.55	0.51	0.61	0.24	0.49	0.33
13064 (2)	0.64	0.56	0.52	0.63	0.24	0.49	0.34
32661 (5)	0.65	0.58	0.5	0.71	0.25	0.44	0.36
65323 (10)	0.64	0.59	0.51	0.73	0.26	0.45	0.37

Table B.17: Performance of the system trained on the interval $3.5 > \Delta t > 2.5$ years and evaluated in all other cases. Metrics used are: area under the curve (AUC), balanced accuracy (B.acc.), accuracy (Acc.), precision (Prec.), sensitivity (Sens.), specificity (Spec.) and F-score.