Study of Longitudinal Neurodegeneration Biomarkers to Support the Early Diagnosis of Alzheimer's Disease

Author: Giovana Gavidia Bovadilla

Supervisor: Alexander Perera Lluna

This dissertation is submitted for the Degree of Ph.D.
Biomedical Engineering Ph.D. Program

Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial (ESAII)
Universitat Politècnica de Catalunya

Bioinformatics and Biomedical
Signals Laboratory, at Biomedical
Research Center (CREB)-UPC

Barcelona, 2018

To my sweet daughter Alessandra and my loving Jordi, for their infinite love and patience. I could not have done this without you.

To my beloved parents who have always encouraged me to believe in myself.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Author: Giovana Gavidia Bovadilla
Barcelona, 2018

Acknowledgements

I would never have been able to finish my thesis without the guidance of my supervisor and the committee members, help from friends, and support from my family.

First of all, I would like to thank my supervisor, Alexandre Perera, for his guidance and encouragement during the last few years. He introduced me to the wonderful world of Data Science and taught me valuable knowledge not only for this thesis but also for developing my background in Biomedical Engineering and Bioinformatics. I am also very grateful to Pere Caminal for introduced me to my supervisor. Thank you for being so kind at every moment.

Many thanks to my friends and fellow mates at the Bioinformatics and Biomedical Signals Laboratory, at Biomedical Research Center (CREB) and the IBEC, especially to Samir Kanaan for helping me to carry out an important part of this study but also for his continued support and encouragement. Similarly, many thanks to Francesc Fernández, María Maqueda, Patricia Díaz, Sergi Picart, Rudys Magrans, Leonardo Sarlabous and Puy Ruiz; who have been indispensable for addressing in the best comfortable way all these years of work.

I would also like to thank my colleagues at my workplace, the e-Health unit of Eurecat, for their caring and providing me with an excellent atmosphere for working and researching during these few last months. I would also like to thank Pere Ramón, who, as a good friend, helped me to disconnect and "to refuel" with engaging talks while we ran in our precious mountain.

Finally, I want to thank my family; my parents, brothers, sisters, my daughter Alessandra and my loving Jordi, for their unconditional support. My dear Jordi and Ale were always helping me with their words of encouragement, their immense love and the best cup of coffee during the long hours working on this thesis.

Abstract

Alzheimer's Disease (AD) is a progressive and neurodegenerative disorder characterised by pathological brain changes starting several years before clinical symptoms appear. Although at the moment there is no cure, earlier and accurate identification of those brain structures changes can help to improve diagnosis and monitoring, allowing that future treatments target the disease in its earliest stages, before irreversible brain damage or mental decline takes place. The brain of AD subjects shrinks significantly as the disease progress. Furthermore, ageing is the major risk factor for sporadic AD, older brains being more susceptible than young or middle-aged ones. However, seemingly healthy elderly brains lose matter in regions related to AD. Likewise, similar changes can also be found in subjects having mild cognitive impairment (MCI), which is a symptomatic pre-dementia phase of AD. Thus, in the research field, it is a challenging to distinguish brain changes in healthy elderly individuals from incipient AD; as well as in MCI subjects with risk to progress to AD. This work proposes two methods based on statistical learning methods, which are focused on characterising the ageing-related changes in brain structures of healthy elderly controls (HC), MCI and AD subjects, and addressing the estimation of current diagnosis (ECD) of HC, MCI and AD, as well as the prediction of future diagnosis (PFD) of these diagnostic groups mainly focused on the early diagnosis of conversion to AD. Analysis of longitudinal biomarkers based on Magnetic Resonance Imaging (MRI-based biomarkers) is the core of both methods. These biomarkers were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS), and corresponded to measures of cortical volume (CV) and thickness average of cortical regions, and measures of subcortical volume (SV). ADNI data include MRI biomarkers available at a 5-year follow up on HC, MCI and AD subjects, while OASIS data only include biomarkers measured at baseline on HC and AD.

In the first method, called M_{res}, in order to identify the variant (*vr*) and quasi-variant (*qvr*) brain regions over time, CV/TA/SV MRI-based biomarkers of HC subjects were characterised by using a Linear Mixed Effects (LME) approach on males and females, separately. All these LME models included the subject's age at each visitation (*age*) and years of education (*educ*) as covariates. We hypothesised that, by identifying both vr and qvr regions, it would

be possible to obtain an ageing-based null model, which would characterise their normal atrophy and growth patterns as well as the correlation between them. By using the null model on those subjects who had been clinically diagnosed as HC, MCI or AD, normal age-related changes were estimated, and then, deviation scores (residuals) from the observed MRI-based biomarkers were computed. The ECD for HC/MCI/AD subjects, as well as the PFD, were addressed through residual-based Support Vector Machines (SVM) modelling. SVM models were trained by using the pool of *vr* biomarkers plus the *age* as input. The advancement of early disease prediction was calculated as the average number of years advanced in the prediction of the future diagnosis of the subjects concerning the last known clinical diagnosis. We found reductions in most cortical volumes and thicknesses in HC (with evident gender differences) as well as in sub-cortical regions, including greater atrophy in the hippocampus.

The second method, called M_{raw}, continues along the same direction as the first method, addressing both ageing-related brain change understanding and the ECD and the PFD problems. However, unlike the M_{res}, this method is focused on directly analysing the raw MRI-based biomarkers values (without derived residuals) stratified by five-year age groups. Furthermore, instead of identifying variant biomarkers, M_{raw} includes a differential diagnosis-specific feature selection (FS) method, which is applied before classification. First, subjects with a stable last diagnostic, i.e., who remained as HC (sHC), MCI (sMCI) or AD (sAD) for the duration of the study were identified and labelled as such. To study the effect of ageing on brain structures, and to perform morphometric comparisons between sHC, sMCI and sAD, we fitted random intercept LME models for each MRI-based biomarker including the age as a covariate. After descriptive analysis, the differential diagnosis problem within each age group was addressed by carrying out three main experiments—sAD vs sHC, sMCI vs sHC, and sAD vs sMCI. For each, the most significant biomarkers were selected by sorting them according to their minimum redundancy and maximum relevance (mRMR) scores within diagnosis classes. After that, it was applied a feature selector based on an SVM wrapper. We built two-class SVM models to label subjects as HC, MCI or AD by using the pool of selected features plus age, gender and years of education as input. The regional analysis results confirm accelerated or reduced estimates of decline in all CV and TA measures with increasing age. Results confirm a frontotemporal pattern of atrophy in sHC subjects across the adult age-span, as well as in sMCI and sAD, which also support findings from the univariate analysis stage.

In addition, the reliability of both methods to correctly discriminate AD vs HC subjects was evaluated and compared by testing them on OASIS subjects observed at baseline. Several features were customised or excluded to facilitate such comparison. Models derived from

both methods were firstly re-trained from ADNI training data and then tested separately on remaining ADNI data, as well as, on OASIS data.

Results from both methods could confirm that the frontotemporal change corresponds to an inevitable process related to normal ageing and is not necessarily an AD-specific one. Furthermore, it is possible that the vulnerability of these areas to normal ageing-related decline contributes to their vulnerability to AD-related atrophy. Regarding the ECD problem, SVM models from both methods obtained better results than comparable methods in the literature, especially on AD vs HC experiment where most indicators rank in the first place (accuracy: 98.68%). Both methods also improve the PFD given the current clinical tests, both in prediction quality indicators as well as in the amount of time by which the diagnosis is advanced (up to 1.87 years earlier for subjects aged 80-84 years).

Resumen

La enfermedad de Alzheimer (AD) es un trastorno progresivo y neurodegenerativo caracterizado por cambios patológicos en el cerebro que comienzan varios años antes de aparecer los primeros síntomas clínicos. La identificación temprana y precisa de estos cambios ayuda a mejorar el diagnóstico y la monitorización, permitiendo que la enfermedad sea abordada en sus primeras etapas, antes de producirse un deterioro morfológico y mental irreversible. El cerebro de los sujetos con AD se reduce significativamente a medida que avanza la enfermedad, siendo el envejecimiento el principal factor de riesgo para la AD esporádica, donde los cerebros de la gente mayor son más susceptibles que los más jóvenes. Sin embargo, ha sido observado que los cerebros de los adultos mayores y de los sujetos en una fase anterior con deterioro cognitivo leve (MCI) pierden materia en regiones relacionadas con AD. Esta tesis propone dos métodos basados en métodos de aprendizaje estadísticos, que se centran en caracterizar los cambios relacionados con el envejecimiento en estructuras cerebrales de controles sanos de edad avanzada (HC), MCI y AD, y en abordar la estimación del diagnóstico actual (ECD) de estos grupos, así como la predicción de su diagnóstico futuro (PFD), principalmente en el diagnóstico precoz de la conversión a AD. Los datos utilizados corresponden a biomarcadores de neurodegeneración longitudinal obtenidas de imágenes de Resonancia Magnética (MRI). Estos biomarcadores se obtuvieron a partir de los estudios Alzheimer's Disease Neuroimaging Initiative (ADNI) y Open Access Series of Imaging Studies (OASIS). Los datos de ADNI incluyeron biomarcadores de MRI disponibles en un seguimiento de 5 años en sujetos HC, MCI y AD, mientras que los datos de OASIS solo incluyeron biomarcadores medidos al inicio del estudio en HC y AD. En el primer método, denominado M_{res}, los biomarcadores que cambiaron significativamente (vr) y los

que cambiaron en una reducida escala (qvr) fueron identificados en sujetos HC utilizando modelos lineales de efectos mixtos (LME). Asimismo, modelos nulos basados en el normal envejecimiento del cerebro fueron construidos para cada género. A través de estos ellos se buscó caracterizar la atrofia normal y los patrones de crecimiento de los biomarcadores vr y qvr, así como la correlación entre ellos. Estos modelos fueron utilizados en los sujetos HC, MCI y AD restantes para inferir los valores normales de los biomarcadores vr y luego calcular sus desviaciones (residuos) respecto a los biomarcadores observados. A diferencia de M_{res}, el segundo método denominado M_{raw}, se centra en el análisis de los valores directos de los biomarcadores MRI, estratificados por grupos de edad de cinco años. M_{raw} incluye un método de selección de características específicas del diagnóstico diferencial aplicado antes de la clasificación. En ambos métodos, se entrenaron máquinas soporte vectorial (SVM) para abordar tres experimentos: AD vs. HC, MCI vs. HC y AD vs. MCI. En M_{res}, los modelos SVM fueron entrenados a partir de los residuos calculados para los biomarcadores vr más la edad, mientras que en M_{raw}, se utilizó el grupo de características seleccionadas más la edad, el sexo y los años de educación. El avance de la predicción temprana de la enfermedad fue calculada como el promedio de años avanzados en el PFD con respecto al último diagnóstico clínico conocido. Los resultados confirman una reducción en todos los biomarcadores corticales a medida que la edad avanza, siendo el cambio de algunas regiones más acelerados que otras. Asimismo, se observó un patrón de atrofia frontotemporal en los tres grupos de sujetos. Con respecto al problema ECD, todos los modelos SVM obtuvieron mejor desempeño en la clasificación que los métodos comparables en la literatura, especialmente en AD vs. HC (Precisión: 98.68%). Ambos métodos también mejoraron la PFD, tanto en los indicadores de calidad de predicción como en el tiempo de avance en el diagnóstico (hasta 1.87 años antes en sujetos de 80-84 años).

Table of contents

List of figures

List of tables

Nomenclature

Roman Symbols

τ	Tau peptides.
$A\beta$	Beta-amyloid peptides.
$A\beta_{1-42}$	Aβ peptide extending to position 42.
AD	Alzheimer's Disease.
$AD-C$	AD-clinical process.
$AD-P$	AD-pathophysiological process.
$ADAS$	Alzheimer Disease's Assessment Scale.
$ADNI$	Alzheimer's Disease Neuroimaging Initiative.
AIC	Akaike Information Criteria.
$ANOVA$	Analysis of Variance.
APC	Annual Percent Change.
BIC	Bayesian Information Criteria.
CDR	Clinical Dementia Rating.
$CDRGLOBAL$	Clinical dementia rating global score.
$CDRSB$	Clinical Dementia Rating-Sum of Boxes.
CN	Control normal subjects at baseline labelled as such by ADNI study.
CSF	Cerebrospinal fluid.

ECD Estimation of current diagnosis.

FR Feature Ranking.

FS Feature Selection.

LDA Longitudinal Data Analysis.

LMCI Late MCI subjects at baseline labelled as such by ADNI study.

LME Linear Mixed-effects Models.

LOO Leave-one-out cross-validation method.

MCI Mild Cognitive Impairment.

MMSE Mini Mental State Examination.

MRI Magnetic Resonance Imaging.

mRMR Minimum Redundancy and Maximum Relevance.

MTL MTL. Medial Temporal Lobe.

OLS Ordinary Least Squares.

OPLS Orthogonal Partial Least Squares.

PFD Prediction of future diagnosis.

PLSR Partial Least Squares Regression.

ROI Brain regions of interest.

ROI Region of Interest.

RVM Relevance Vector Machine.

RVR RVM model for regression.

SVM Support Vector Machine.

P-τ Hyperphosphorylated τ.

T-τ Total tau.

Chapter 1

Introduction

1.1 Clinical Motivation

In 2015, an estimated 46.8 million people were living with dementia. This number is expected to increase to 50 million people in 2017. Alzheimer's Disease (AD) is the most common form of dementia, whose prevalence rises rapidly as the world population ages. In 2006, the worldwide prevalence of AD was 26.6 million. By 2050, a study based on mathematical model [11] predicted that prevalence would quadruple, by which time 1 in 85 persons worldwide will be living with the disease. Regarding European rates, a collaborative study carried out in the 1990's to determine the prevalence of dementia and major subtypes in several European population-based cohorts [78] found an AD's prevalence of 4.4% in 2346 demented individuals older than 65. Recently, a meta-analysis study based on literature research [91] have estimated the prevalence of AD in Europe at 5.05%, with woman's rate (7.13%) higher than men's rate (3.31%). The impact of the disease is huge, not just for the patient who loses the control of their life; but also for the family who cares for the patient; and for the society, that has to deal with the growing number of persons with AD in need of care.

AD is a disease with both pathological brain processes and clinical decline occurring gradually, with dementia representing its last stage. Changes begin several years before the onset of clinical symptoms. At some point, neuronal dysfunction and neurodegeneration impair the ability to carry out essential human functions. Although at the moment there is no cure, earlier and accurate identification of those brain structures changes can help to improve diagnosis and monitoring, allowing that future treatments target the disease in its earliest stages, before irreversible brain damage or mental decline takes place. Thus, there is an urgent need for biomarker-based tests, which enable a more accurate and early diagnosis of AD [108] and the prediction of disease progression from Mild Cognitive Impairment

(MCI), which is a symptomatic predementia phase of AD [3]. Moreover, such tests could also improve monitoring AD progression, evaluation of new AD therapies, and enrichment of AD cohorts with specific subsets of AD subjects in clinical trials.

In the last few years, the potential of structural neuroimages based on Magnetic Resonance Imaging (MRI) is used to propose powerful neuroimaging biomarkers. MRI-based biomarkers quantify the degree of brain atrophy, allowing to measure the correlation between progression of cognitive impairment and atrophy rate. Despite cerebrospinal fluid (CSF) biomarkers including CSF-Aβ and CSF-τ have been suggested as the most informative AD biomarkers [108], many studies have also suggested their combination with other features types, such as MRI-based biomarkers, socio-demographic characteristics, among others, to increase early prediction accuracy [18]. On the other hand, despite older people are more susceptible than young or middle-aged ones for sporadic AD [45], studies have found that seemingly healthy elderly brains lose matter in regions related to AD. Likewise, similar changes can also be observed in subjects having MCI.

Thus, one of the most critical challenges in clinical AD research is to identify a set of reliable and robust AD features to distinguish brain changes in healthy elderly individuals from incipient AD; as well as in MCI subjects with risk to progress to AD. The research presented in this thesis aims to address the ageing-related brain change understanding, as well as the early diagnosis of MCI and AD subjects by using statistical learning methods to identify MRI-based biomarkers and socio-demographic features and build classifiers to evaluate them.

1.2 Clinical diagnosis.

The definitive diagnosis of AD requires the presence of high cognitive impairment and autopsy confirmation of the accumulation of both Aβ and τ. However, in clinical practice, the diagnosis of dementia due to AD and its previous stages are most often based on the criteria developed by the National Institute of Neurologic and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA). These criteria were initially established in 1984, and 27 years later, three research workgroups where formed by the National Institute on Aging (NIA) and the Alzheimer's Association to formulate a set of new guidelines and recommendations for AD diagnostic and research criteria for the continuum of AD [81, 3, 110]. As a result, in order to disambiguate the term "AD", researchers distinguished two AD process with three disease's stages contained within them [110]. The first process is named AD-pathophysiological (AD-P) process, which refers to the evidence of the underlying brain disease process at the preclinical stage. The second

process is named AD-clinical (AD-C) process, which refers to the clinical phase of illness, which includes not only AD dementia stage but also individuals within MCI stage due to AD-P. AD-P is thought to begin years before the emergence of AD-C. Figure 1.1 shows a model of the clinical trajectory, which is known as "The continuum of AD". That model represents the pathological-clinical continuum of AD where the stage of preclinical AD precedes both, MCI due to AD and AD dementia, but does not imply that all individuals with biomarker evidence of AD-P process will progress to the clinical phases of the illness. The three disease's stages within each of these processes are described following.

Preclinical AD. This stage includes individuals classified as "Not normal, not MCI". It also includes: (1) completely asymptomatic individuals with biomarker evidence suggestive of AD-P at risk for progression to AD dementia, (2) biomarker-positive individuals who are already demonstrating very subtle decline but not yet meeting standardized criteria for MCI, (3) individuals who carry one or more $ApoE\ \varepsilon 4$ alleles who are known to have an increased risk of developing AD dementia (AD-P biomarker-positive individuals), and (4) carriers of autosomal dominant mutations, who are in the pre-symptomatic biomarker-positive stage of their illness, and who will almost certainly manifest clinical symptoms and progress to dementia.

MCI due to AD. This stage refers to the symptomatic predementia phase of AD when sufficient damage has been accumulated to result in cognitive symptoms and impairment [3, 122]. Individuals in this stage are characterized as follow: (1) symptomatic but non-demented individuals who experience a gradually progressive cognitive decline with relative preservation of other cognitive domains and functional activities, and (2) do not meet the criteria for dementia. Studies showed that MCI patients progressed to AD at a yearly rate of 10% to 15%, and the probability of this progression depends on the interaction of genotypic-phenotypic-environmental factors.

Dementia due to AD. This stage refers to dementia in the clinical process. As a result of guidelines proposed by McKhann et al. [81], dementia due to AD was classified into three groups as follow.

1. Probable AD dementia, which includes typical clinical syndrome without histologic confirmation.

2. Possible AD dementia, which means atypical clinical features but no alternative diagnosis apparent and no histologic confirmation.

3. Probable or possible AD dementia with evidence of the AD pathophysiological process, which includes individuals who meet the core clinical criteria defined for types (1) or (2) with additional evidence of pathophysiological process detected through some AD biomarkers.

The first two classes should be used in all clinical settings. The third class should be used only for research purposes by taking into account both biomarkers for the pathophysiological process of AD and neuropsychological deficits, in the diagnostic criteria.

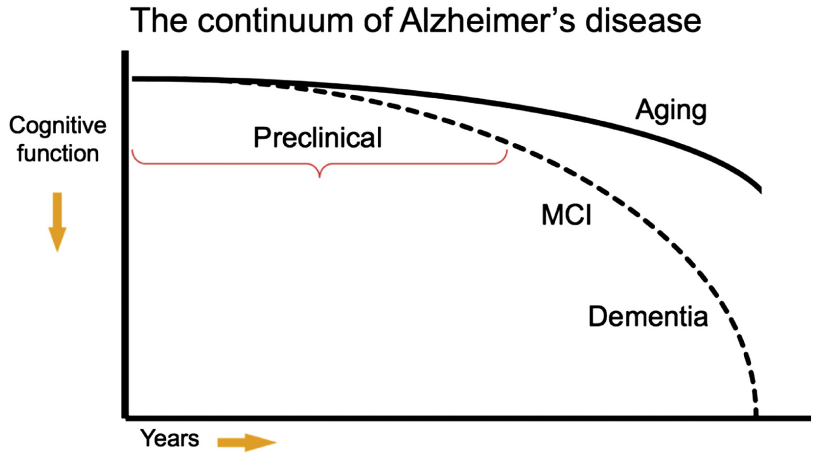


Fig. 1.1 Model of the clinical trajectory of AD. Source: Sperling et al. [110]

Regarding monitoring of AD in the clinical practice, neuropsychological tests such as the Clinical Dementia Rating (CDR) [86], the Mini-Mental Examination Score (MMSE) [46] and the Alzheimer's Disease Assessment Scale (ADAS) cognitive subscale [102] are used to monitor AD progression or treatment efficacy. However, although these tests unquestionably reflect an important aspect of disease progression (i.e. functional impairment), they also have several limitations such as relatively low specificity and reliability [87].

1.3 Risk factors

The single major risk factor for developing AD is age. Older brains are more susceptible than young or middle-aged ones for sporadic AD [45]. Most cases of AD are seen in order adults, after of 65 years.

Regarding genetic risks, a genetic pattern of inheritance is not observed in sporadic AD. However, there exists a gene found on chromosome 19 associated with AD risk, which is responsible for the production of a protein that carries cholesterol in the blood, called

apolipoprotein E ($ApoE$). From the three main types of this protein, carriers of the $ApoE$-$\varepsilon 4$ allele have more risk of developing this disease, being sufficient the presence of at least one allele to advance the age of clinical onset of AD dementia.

Another risk factor to consider is the influence of socio-demographic characteristics. Both higher levels of education or occupational attainment may act as a protective factor. For example, studies have found that higher levels of education may increase the brain reserve, leading to larger brain structures, which can help to counter the effects of brain atrophy [64, 76].

Finally, coexisting health problems, such as cardiovascular diseases and Type 2 diabetes can increase the risk of developing AD. Cardiovascular diseases, such as stroke, atrial fibrillation, coronary heart disease, and heart failure are very prevalent in elderly individuals and have regularly been linked to AD [26]. This association might be due to shared risk factors between those and AD, and the fact that cardiac disease causes hypoperfusion and microemboli, which have been implicated in the aetiology of AD. For example, heart failure causes damage to blood vessels in the brain, meaning less blood flow and possible neuronal death. On the other hand, many studies have confirmed that the risk of dementia and AD is higher in individuals with type 2 diabetes mellitus, even in a pre-diabetes stage [26]. This tendency is because the increase in glucose and insulin levels causes a direct neurotoxicity.

1.4 Neuropathology of AD.

AD is characterized by pathological changes in the brain at molecular and morphological levels. Based on causative factors triggered those changes, AD subjects can be grouped into two groups including the familiar AD and sporadic AD [36]. The inheritable AD is the smallest group, which is affected by mutations in the genes of presenilin 1 (PS1) and presenilin 2 (PS2). This group is typically associated with an early onset, before the age of 65 years. However, even if several members of a family have in the past been diagnosed as having AD, this does not mean that another member of the family will necessarily develop it. On the other hand, as mentioned in Section 1.3, the sporadic AD subjects are not affected by the presence of such known genetic causes, but age is their most significant risk factor for developing AD.

1.4.1 Molecular changes.

Molecular changes in the brain begin several years before the onset of clinical symptoms. These changes include accumulation of beta-amyloid (Aβ) plaques and neurofibrillary

tangles composed of τ amyloid fibrils, and these proteins are associated with neuronal dysfunction and neurodegeneration causing cognitive impairment. In 1992, Hardy and Higgins [57] proposed the "Amyloid Cascade Hypothesis", which affirms that $A\beta$ production and accumulation is the causative agent of AD pathogenesis. Figure 1.2, presented by Cummings [20], shows graphically that hypothesis. In that figure, it is observed the amyloid precursor protein (APP) causes the generation of the $A\beta$ peptide, which is followed by multiple secondary steps, to cell death, neurotransmitter deficit and cognitive and behavioural abnormalities. The amyloid hypothesis was the core of many investigations focused on developing agents for disease-modification. However, a few years ago, the total validity of the amyloid hypothesis has been questioned because there was no validation based on treatment of that hypothesis [20]. Despite that fact, currently, Aβ deposition is the first detectable change in AD, but the initial cause of its appearance in the brain is still undefined.

1.4.2 Morphological changes.

Brain atrophy is one of the initial consequence of neuronal death due to AD. The pattern of atrophy in AD is not random, but usually, it evolves slow changes [45]. Firstly, this pattern involves the medial temporal lobe (MTL) including mainly the entorhinal cortex, and the hippocampus. After that, change continues through association areas in medial parietal, lateral temporal and frontal regions, eventually affecting all regions of cortex [45, 99]. Expansions of the ventricular and sulcal CSF regions also are observed [90, 35]. Once the patient has reached a diagnosis of AD, neurodegeneration is usually found throughout the neocortex and subcortical regions, with significant atrophy of the temporal, parietal, and frontal cortices Risacher and Saykin [99]. The main brain structures changes identified for each AD stage are briefly described below.

Pre-clinical stage. In this stage, there is evidence that the early $A\beta$ deposition within the brain is associated with the grey matter atrophy, where the main affected regions are the neocortex and the parietal and frontal lobules, including higher atrophy degree within the hippocampus and the cingulate gyrus [36]. However, even these findings, a complete understanding of the brain deposition of $A\beta$ remains an important study field.

MCI stage Many studies have shown that the atrophy degree in MCI subjects is intermediate between controls and AD demented subjects. Global grey matter volume in MCI subjects has been observed smaller than in AD, but higher than in healthy control (HC) subjects, with significant local reductions in grey matter in the MTL, insula, thalamus, temporal neocortex, parahippocampal cortex, orbitofrontal and inferior parietal areas [66, 13]. However, when

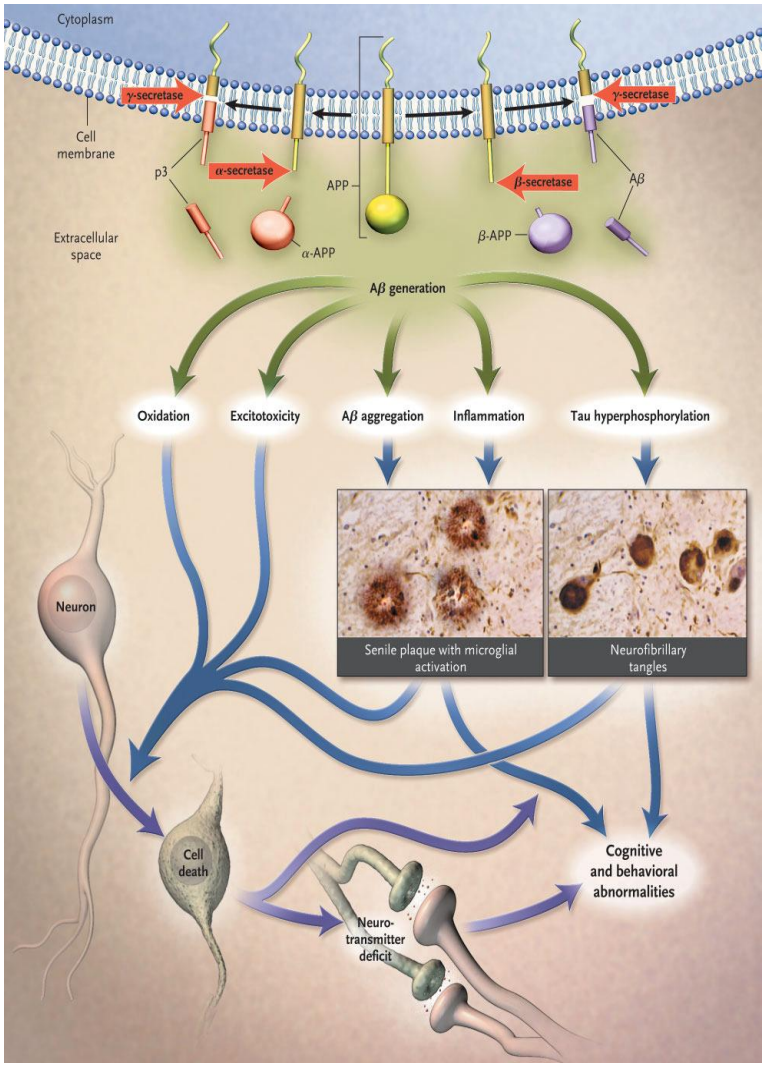


Fig. 1.2 Graphic representation of the "*Amyloid cascade hypothesis*"of AD proposed by Cummings [21].

comparing MCI subjects who did subsequently converted to AD (AD_c) with the MCI who remain stable (MCI_s), the first group have shown accelerated atrophy than MCI_s.

Dementia due to AD. As neuronal degeneration progresses, atrophy in specific areas typical of AD becomes readily detectable and measurable by using neuroimages. Studies have found early pathological changes in MTL regions, such as reduction of hippocampus volume and the entorhinal cortex; then as disease progress, changes are gradually extended to temporoparietal cortical regions.

Morphological changes are assessed by neuroimaging biomarkers, which are described below in the Section 1.6.2.

1.5 Brain morphological changes due to ageing.

The effects of ageing on the brain are widespread and include atrophy caused by dendritic pruning, as well as a loss of synapses and neurons. Studies have demonstrated that seemingly healthy subjects lose brain matter over time and brain age-related changes and function are not uniform across the whole brain or over subjects. The volume of the brain decreases with age at a rate of approximately 0.2–0.5% per year [43], and this rate might be even greater over the age of 70 [93]. Regionally, ageing-related atrophy has been observed across many of the cortical regions [104, 64, 43] with a prominent decline in the prefrontal cortex and the slight decline of the temporal cortex and parahippocampal cortex [104]. Fjell et al. [43], Jiang et al. [64] found reductions in several subcortical structures including the caudate nucleus, amygdala, cerebellum, and hippocampus, the latter being the most studied structure, with annual atrophy rate of about 2.0% [43]. Studies on cross-sectional and longitudinal MRI have found significant correlations between gender and cortical and subcortical regions, but there are inconsistencies between the results [64, 104].

However, many of these age-related changes are shared by neurodegenerative diseases. Some studies have found that cognitively normal elderly have Aβ deposition in the brain with similar levels of the substances observed in subjects with dementia due to AD [85]. τ protein values in AD is higher than in HC, and this protein is also found in other neurodegenerative diseases referred to as tauopathies [125]. Furthermore, at the macroscopic level, part of ageing-related atrophy occurs in areas vulnerable to AD, while other changes are observed in areas less characteristic of early-stage AD [43, 64, 104]. Due to the shared biochemical and morphological characteristics, it is a complex task to discriminate some of these ageing-related changes in healthy elderly subjects from subjects affected by early stage of AD.

1.6 Biomarkers for AD.

Biomarkers of AD are physiological, biochemical and anatomical indicators that can be measured in vivo and that indicate specific features of disease-related pathological changes [61]. These biomarkers can improve both the diagnostic and prognostic accuracy of AD and its differentiation from other neurodegenerative diseases [115]. They enable physicians to identify individuals at risk for the disease, even before symptoms appear, and possibly prevent or slow down the progression to dementia.

Based on the current literature, the following five AD biomarkers are the most widely studied:

- Decreased CSF Aβ42,
- Increased CSF tau,
- Decreased [F-18]-fluorodeoxyglucose uptake on positron emission tomography (FDG-PET),
- PET amyloid imaging, and
- Structural MRI measures of cerebral atrophy

Because AD biomarkers play an important role both as outcome measures and as inclusion criteria, researchers have dedicated efforts to measure these biomarkers in cross-sectional and longitudinal observational studies and to establish the correct ordering of the relevant biomarkers and their relationships to clinical symptoms. In 2010, Jack et al. [61] proposed a framework of AD development where the five AD biomarkers do not reach abnormal levels simultaneously but do so in an ordered manner. The framework represents a hypothetical model of the time-dependent ordering of onset and maxima of those five AD biomarkers, where the disease stages appear related to these biomarkers. The five biomarkers are grouped according to two groups: (1) Biomarkers of Aβ-plaque deposition, where are both CSF Aβ42 and amyloid PET imaging; and (2) Biomarkers of degeneration including the CSF tau, FDG-PET, and MRI biomarkers.

One year later, as result of the AD biomarkers revision carried out in 2011 by the NIA and the Alzheimer's Association (see Section 1.2), Sperling et al. [110] proposed a new biomarker model adapted from the original model presented by Jack et al. [61]. In this model, the pre-clinical phase of AD is expanded, and biomarkers are considered as dynamic, i.e., each biomarker changes over time and follows a non-linear time course. Figure 1.3 shows the main statements of the model, in which the researchers highlight the following findings:

- Brain $A\beta$ accumulation biomarkers are identified by CSF-Aβ42 assay or PET amyloid imaging. These biomarkers become abnormal first, possibly with a lag greater than a decade until clinical symptoms could be quantified. The model highlight that brain $A\beta$ accumulation is *necessary but not sufficient* to produce the clinical symptoms of MCI and dementia,
- Biomarkers of synaptic dysfunction obtained from FDG-PET and functional MRI may demonstrate abnormalities in very early stage, particularly in *APOE* gene ε4 allele carriers.
- Elevated CSF-τ is a biomarker of neuronal injury, and it is measured in the CSF.
- Brain structures are markers of neuronal loss obtained from structural MRI images. These biomarkers seem to become abnormal a bit later in comparison with the previous biomarkers, following a pattern of change affected by disease progression.
- Either cognitive and clinical function decline retain a close relationship with brain structure atrophy in both MCI and dementia stages.

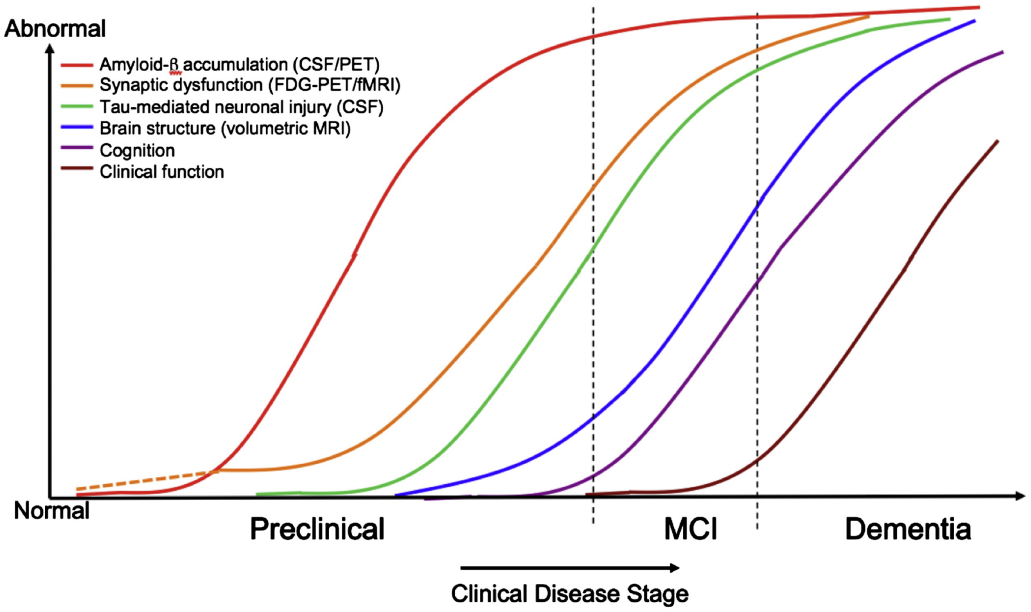


Fig. 1.3 Hypothetical model of dynamic biomarkers of the AD. Source: Sperling et al. [110]

1.6.1 CSF-based Biomarkers.

CSF is a clear fluid largely produced by the choroid plexus and found in the ventricular system and subarachnoid spaces surrounding the brain and spinal cord. Due to CSF directly interacts with the extracellular space in the brain reflecting the associated biochemical/pathologic changes, it is an ideal source for developing biomarkers in AD [56]. Many studies suggest that changes in CSF levels of *tau* and Aβ occur in early stages of AD. A CSF profile consisting of decreased concentrations of the Aβ peptide extending to position 42 (Aβ_{1-42}) and increased concentrations of total tau (T-τ) and hyperphosphorylated τ (P-τ) have been associated with AD signature in the CSF [108, 117, 61, 122]. Probably, damaged and dying neurons that harbour dystrophic tau neurites and tangles release tau levels in CSF and as a consequence, its levels are increased. On the other hand, levels of Aβ_{1-42} may be reduced due to it is accumulated into insoluble plaques in the AD brain and is not available in a diffusible form. However, despite the utility of these biomarkers for diagnosis of AD, there exist several issues that researchers have to face. First, CSF-Aβ and CSF-τ are also found in healthy elderly people and MCI, thus, it is necessary to discriminate AD-related CSF abnormalities from the CSF profile in normal subjects due to ageing and MCI. Additionally, the analysis of these proteins levels in the same sample often varies significantly from institution to institution due to not being applied a standard procedure protocol.

1.6.2 Neuroimaging biomarkers

Neuroimaging is one of the most promising areas of research focused on early detection of AD. Currently, image-based diagnostic techniques, such as MRI and computed tomography, are considered part of the standard workup for AD, being used to exclude other types of brain diseases, whose symptoms could be confused with AD (e.g., brain tumours or bruising epidural). Furthermore, with the increasing potential of these techniques and the advent of new modalities in medical imaging, its use has not only been focused on improving the accuracy of clinical disease diagnosis, but it has also benefited the potential of these images to monitor the progression of diseases and effects of treatments. Neuroimaging methods are capable of detecting early substantial brain changes not only in AD subjects but also in MCI and in cognitively normal subjects who may be in the preclinical stage of AD [36].

By using neuroimages, Ewers et al. [36] were able to reconstruct the AD trajectory in the living brain and create a hypothetical model of the temporal profile of neuroimaging biomarkers, see Figure 1.4. That model shows a set of imaging modality-specific changes within the AD stages, where biomarkers measured in CSF, PET, functional MRI and structural MRI could detect early brain changes even in the preclinical stage of AD. However, it is

relevant to take notice of that trajectory could be not uniform and may be affected by several factors including risk-associated and protective genes, comorbidities, brain reserve, among others [35].

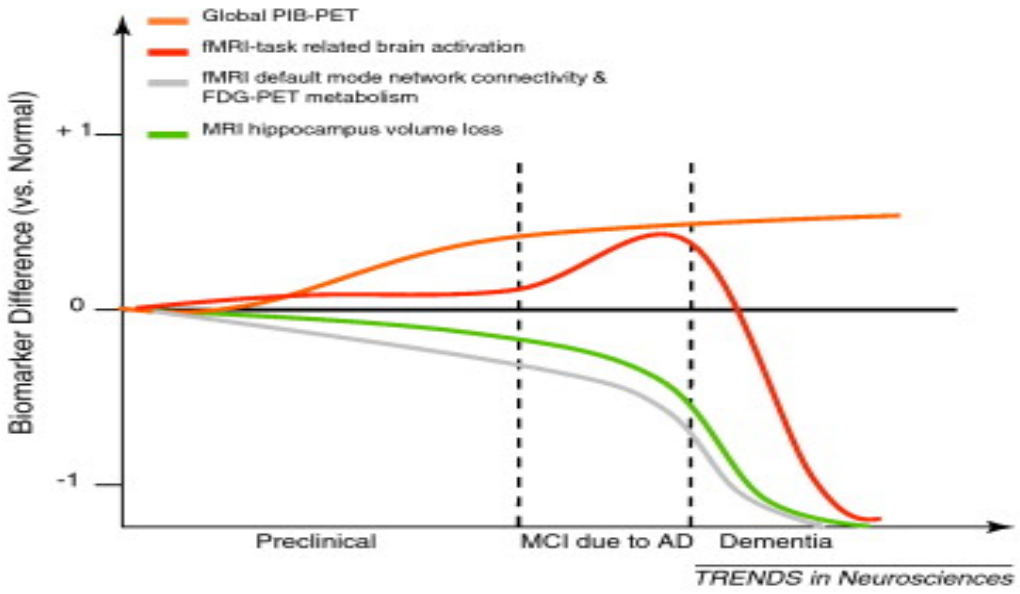


Fig. 1.4 Hypothetical model of main meuroimaging biomarkers applied along AD stages. Fuente: Ewers et al. [36].

Magnetic Resonance Imaging (MRI). Magnetic Resonance (MR) images is a type of structural image which produces clear reproductions of the brain, and it is beneficial in ruling out AD-related changes and other causes of dementia, such as tumours or strokes. These images give a macroscopic visualisation of atrophy caused by neuronal death and allow clinicians for multiple tracking over time to compare the stuff lost since the first images to evaluate the applied treatments. However, at the moment, there is no agreement on standardised values for brain volume, which would help to determine a significant brain shrinkage for any individual at a single point in time. In this study, we have focused on analysing biomarkers based on MRI. They are introduced briefly in Section 1.7 and details about its development, analysis, and application on AD are described in the Chapter 2.

Positron Emission Tomography (PET). PET is a type of functional image, which provides information about reductions of the brain cell activity in certain regions. Neurodegeneration is accompanied by the decline in synaptic function, which is assessed by FDG-PET images. These images obtained from patients with AD show a specific topographic pattern of

decreased glucose uptake in brain regions critical for memory, learning and problem-solving. PET-based biomarkers are widely used in studies focused on the early diagnosis of AD [36, 99, 122, 19, 129, 114, 65, 73]. However, currently, this method is more invasive and expensive, and less accessible. Thus, it has a limited application in everyday clinical practice.

Molecular imaging. Since a few years ago, molecular imaging tracers are applied on patients to evaluate possible AD or other causes of cognitive decline. Tracers, such as the florbetapir F-18 and the flutametamol F-18, are molecules that bind to and reveals amyloid plaques in the brain. Because these plaques are labelled with a radioactive tracer, it can be visualised during a PET brain scan. However, as mentioned previously (see Section 1.6.1), many people have amyloid plaques in the brain but have no present symptoms of cognitive decline or AD. Thus, at the moment, amyloid imaging is not recommended for routine use in patients suspected of having AD.

1.7 MRI-based biomarkers.

One of the main challenges of AD neuroimaging-related studies is the combination of MR images processing techniques with data-driven statistical approaches to obtain measures from brain regions of interest (ROI), which are called MRI-based biomarkers.

1.7.1 Characteristics and advantages.

Studies carried out by Clark et al. [16], Hampel et al. [54], Fjell and Walhovd [42], Hampel et al. [55], Weiner et al. [122] have presented a detailed review of the recent studies focused on the use of AD biomarkers, including the ones based on structural MRI and other sources. In summary, those studies described some characteristics expected from them to be considered as reliable measurements:

- Its replication and generalizability must be demonstrated to support the early diagnosis and prognosis at the individual level.
- Be able to predict the cognitive impairment.
- Be able to distinguish healthy elderly from subjects with prodromal AD.
- Be able to distinguish MCI subjects with risk of progression to AD.

In addition, some of the main proven advantages from using these biomarkers could be summarised as follow:

- MRI-based biomarkers describe and quantify certain patterns in the brain even for stages where the cognitive impairment is mild and possibly reversible helping to the early AD diagnosis as well as the prediction of disease advancement [67, 29, 36, 54, 55].
- Have shown to reach higher specificity and reliability than clinical tests, such as the CDR and the MMSE scores, even when the earliest symptoms are not visible [88].
- Are more effective than CSF-based biomarkers in detecting the cognitive decline [42].
- Allow distinguishing the ageing-related atrophy from the one related to the disease in order to understand the ageing effect on brain structures [104, 64, 43].

Regarding the role of MRI-based biomarkers along AD stages, they provide a dynamic and powerful approach for understanding the spectrum of AD-related brain changes with applications in clinical trials, screening, diagnosis and prognosis. As described in Section 1.6, as neuronal degeneration progresses, atrophy in certain areas typical of AD could be identified by MRI-based biomarkers (see Figure 1.3). Likewise, the hypothetical model represented in Figure 1.4 suggested that MRI-detected grey matter atrophy starts primarily, though not exclusively, in the hippocampus (green line), and continues to decline throughout the progression of the disease.

1.8 Longitudinal Data

Longitudinal studies play an important role in the health as well as in other areas (social, education, biological and agricultural sciences, education, economics, marketing, etc.). By contrast to the cross-sectional approach, the longitudinal design can provide increased statistical power by reducing the confounding effect of between-subject variability and be used to study of time-related change not only of the outcome but also of the covariates [41]. For example, Figure 1.5 shows the individual trajectories of hippocampal volume changes over time obtained from a longitudinal study carried out by Schuff et al. [105]. Hippocampus is the most studied region due to be strongly associated with declining cognitive function. In the figure, observations over time allowed them to identify that AD subjects had on average a greater volume reduction over than HC subjects, and MCI subjects had intermediate values between HC and AD subjects.

Longitudinal data analysis (LDA) applied on AD have become increasingly widespread over the last decades with the main focus on studying the within-subject and between-subject changes over time and identifying statistically significant biomarkers in order to propose

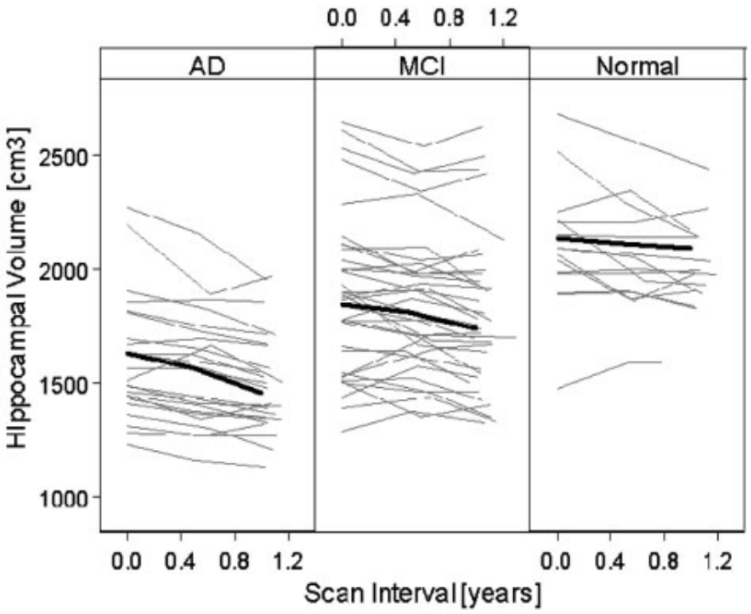


Fig. 1.5 Comparison of individual trajectories of hippocampal volume change over time from HC, MCI and AD subjects. The thick black lines indicate the mean trajectory of each group: Image Source: Schuff et al. [105]

better ways to accurately early diagnose AD and track its advancement. Longitudinal studies have specific features and challenges, which determine the main aspects of their complicated analysis. The most relevant are described below.

Covariance structure. Heterogeneous variability and correlation are common characteristics of repeated measurements on one individual. Heterogeneous variability refers to the variance of a feature over time. Correlation refers to know a priori the likely feature value in a specific time from previous values. As Fitzmaurice et al. [40] described, correlated observations are a positive feature of longitudinal data; they provide more precise estimates of the rate of change or the effect of covariates on that rate of change that would be obtained from an equal number of independent observations of different individuals. Both longitudinal data characteristics violate the fundamental assumptions of independence and homogeneity of variance of many standard statistical techniques, such as Analysis of Variance (ANOVA) and multiple linear regression. Accordingly, these data should be analysed with adequate statistical methods.

Balanced versus unbalanced designs. A study is defined as "balanced" over time when all subjects have the same number of repeated measurements obtained at a common set of

occasions (or timepoints). On the contrary, a study is defined as "unbalanced" over time when the repeated measurements are not obtained at a common set of timepoints due to mistimed measurements (observations made before or after the scheduled time).

Missing data. Missing data are the common problem in longitudinal studies, mostly in health sciences. A subject can be observed at baseline, missing at one follow-up time and then measured again at one of the next, resulting in an unbalanced study with a missing data pattern. However, to difference unbalanced studies due to missing data from those with repeated measurements unequally separated in time, the first studies are often referred as "incomplete" studies.

Time-varying and time-invariant covariates. Longitudinal studies permit repeated measures not only of the outcome but also of the covariates. The incorporation of covariates that change stochastically over time poses many intricate and complex analytic issues.

Continuous and discrete responses. Longitudinal studies vary in the types of outcomes of interest. Outcomes in AD studies can be continuous variables, e.g., hippocampus volume, CSF-$A\beta$ range; or discrete variables, e.g., clinical disease status (HC, MCI or demented due to AD). Depending on the type of outcome, there exist adequate approaches for its statistical analysis.

Challenges. The benefits of a longitudinal design are not without cost. There are several challenges posed:

- Analysis of correlated data; where are required methods that can properly handle the intra-subject correlation of responses. If correlation is ignored, inferences such as statistical test or confidence intervals can be invalid.
- Participant follow-up; because in this type of studies there exist the risk of bias due to incomplete follow-up or drop out of subjects. In that case, statistical analysis may provide summaries that are not representative of the study population.
- Cost of longitudinal studies; these studies are expensive, suffer from subject-drop out over time and often span a relatively short period of follow up [35].

Both cross-sectional and longitudinal studies go hand in hand with statistical learning (SL) methods. These methods are valuable for summarising and obtaining assumptions about data. Several SL methods are described in more detail in Chapter 2.

1.9 Thesis outline

Relevant concepts of developing of neurodegeneration biomarkers, as well as their analysis using statistical learning methods, are described in Chapter 2. Regarding neurodegeneration biomarkers, this chapter provides an introduction to main processes related to brain MR images analysis addressing stages from image reading to measuring neurodegeneration biomarkers. Also, some of the most used tools to MRI-based biomarkers development and visualisation are briefly described. Then, it is presented an overview of several statistical learning methods applied to cross-sectional and longitudinal clinical data, mainly focused on methods related to this work. It includes methods for regression and classification, and common methods with which to assess their performance. Chapter 3 describes the main and secondary objectives, as well as the expected contributions of this thesis. In Chapters 4 and 5 are presented the two methods proposed in this work, which are focused on addressing both ageing-related brain change understanding, and the early prediction and diagnostic prediction advancement of MCI and AD. Univariate and multivariate statistical learning methods are applied together on MRI-based biomarkers from the ADNI longitudinal data. Chapter 6 presents the results of comparing the two previous methods by applying them on a new cohort integrated by MRI-based biomarkers from the OASIS study. In Chapter 7, we discuss the findings from the previous chapters and evaluates the performance of the proposed methods, relating the results with the objectives defined for this study. We also compare our results to the performance achieved in related work. Chapter 8 presents our conclusions, as well as points out some potential avenues of investigation. Finally, in Chapter 9 are listed the scientific papers in indexed journals produced during the development of this thesis, as well as the publications in conferences.

Chapter 2

Background

2.1 Development of MRI-based biomarker

Due to high complexity data arising from brain MR images, it is difficult to analyse them directly. In fact, the brain tissues are characterised by a varied, complex and often overlapping morphology, thus obtaining MRI-based biomarkers is not an easy and trivial task. Nowadays, the existence of powerful and user-friendly tools for image processing have overcome that issues, allowing to obtain biomarkers of neurodegeneration automatically. This chapter describes state-of-art methods applied to obtain MRI-based biomarkers. MR image analysis stages including pre-processing, segmentation, visualisation and biomarker quantification are reviewed in Section 2.1.1. Then, a brief description of automatic tools used directly or indirectly in this study to obtain these biomarkers is presented in Section 2.1.3.

2.1.1 Brain MR images analysis

MR image analysis deals with the development of problem-specific approaches for the enhancement of raw images, segmentation, quantitative measurements, and visualisation for further analysis. Steps applied for that purpose could be grouped into four main stages: (1) 3D image reconstruction, (2) Pre-processing, (3) Segmentation, and (4) Quantification of biomarkers.

Reading and 3D image reconstruction

Separate DICOM files are read into an *mxn* matrix and then assembled in parallel to create a 3D matrix of *mxnxz* dimensions. Here, *z* represent the number of slices (or DICOM files) available for each subject's MR image collection, see Figure 2.1. Every is integrated by a

finite set of image elements called pixels in 2D space or voxels in 3D space [31], see Figure 2.2. The values of each pixel or voxel are intensities representing the radio-frequency signals that are emitted by the brain tissue during the image acquisition process. These values are typically represented by a grey value $0, ..., 255$ in an MR image.

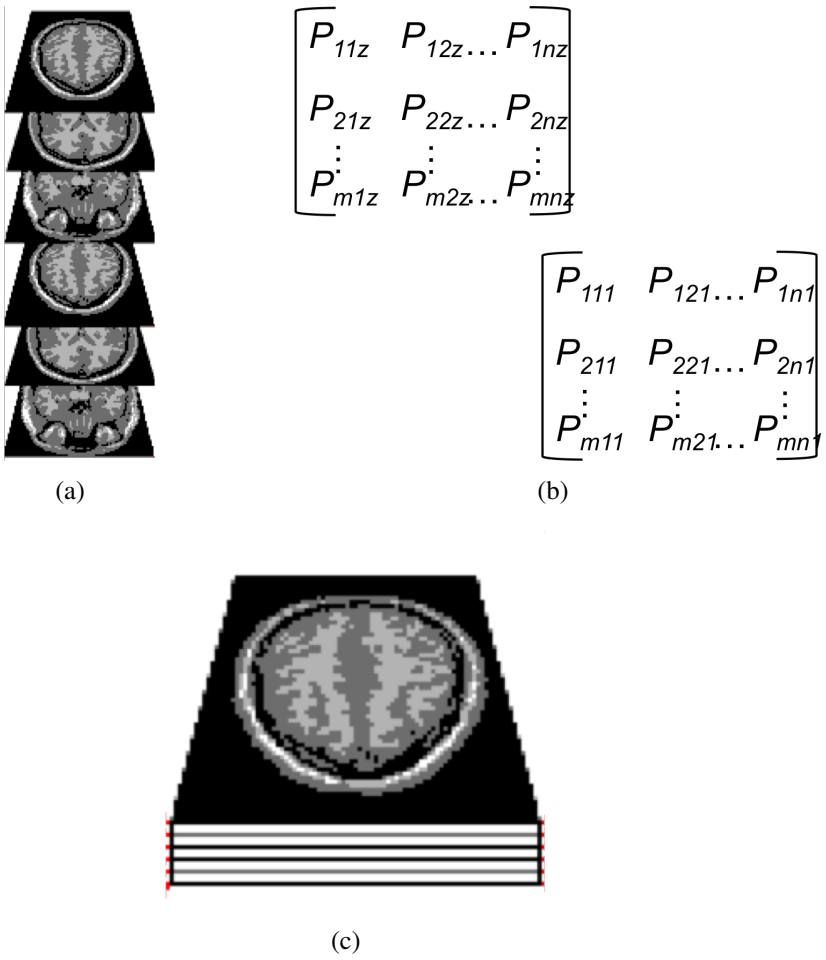


Fig. 2.1 3D brain image reconstruction.

Pre-processing

Defacing. Once obtained the 3D image reconstruction, the next step in many brain MRI analysis is the removal of extra-brain tissues. Defacing consists of identifying the facial features, such as eyes, nose and teeth, and removing it without damaging the brain tissue. This process ensures the individual anonymisation.

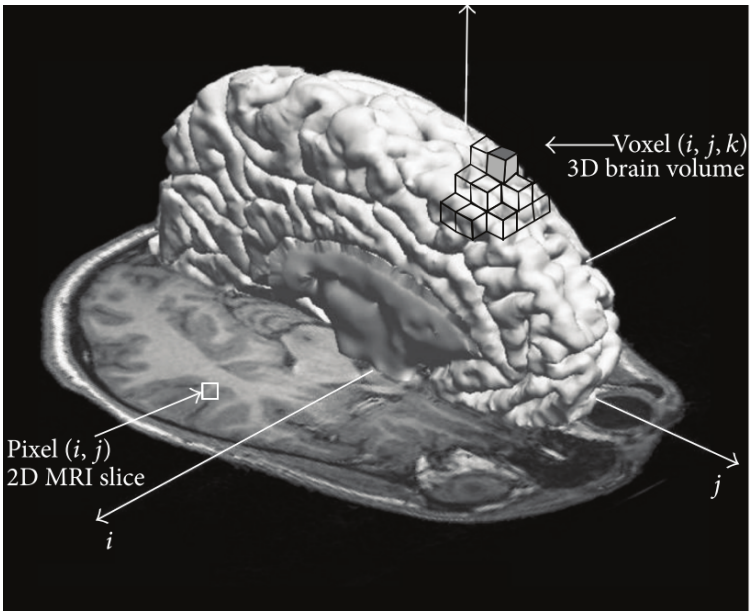
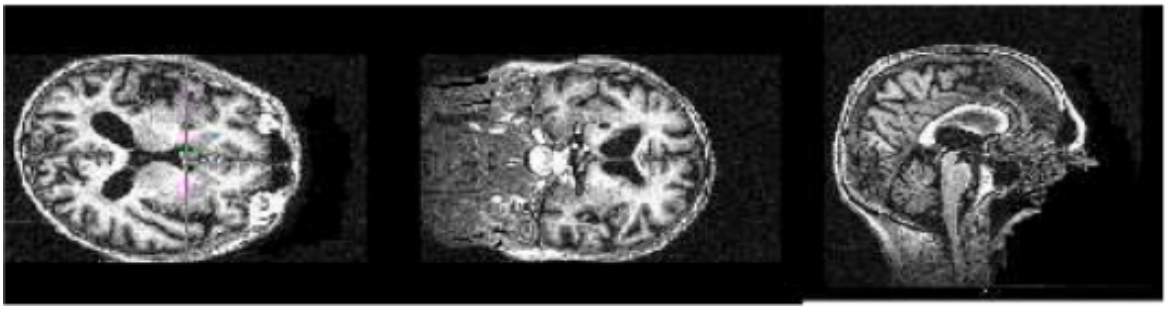


Fig. 2.2 Representation of an image pixel and image voxel in a brain MRI. Source: Despotović et al. [31]

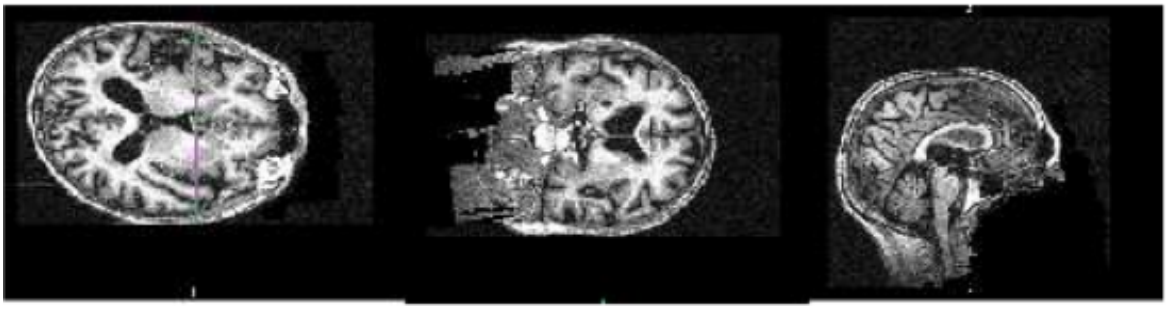
MR image registration. In research as well as in clinical practice, it is common to obtain a set of subject MR images (sessions) across time, even for a short period. In this sense, precise and consistent alignment of scans based on image registration techniques is necessary for motion correction of images taken in the same session (single timepoint), or across time in longitudinal studies where changes in the images can be expected [98]. Figure 2.3 shows an example of 3D image registration applied on three MR images sessions obtained for an individual in a short period.

Bias correction. Bias field signal is a low-frequency and very smooth signal caused by several factors such as magnetic settings and patients' position. These artefacts corrupt MR images degrading further processing based on the image grey level values. For this reason, bias correction is applied before segmentation-based steps or classification. Figure 2.4 shows the resulting image after noise filtering.

Skull stripping. In this process, the non-cerebral tissues, such as the skull and scalp are removed. It is usually achieved by applying image processing techniques, such as tissue threshold [103], border detector, watershed [106] or region growing algorithms, or even a combination of them.

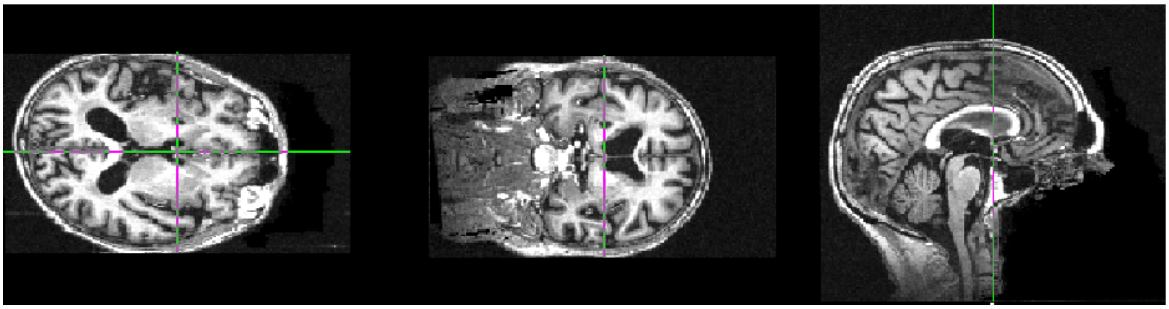


(a)



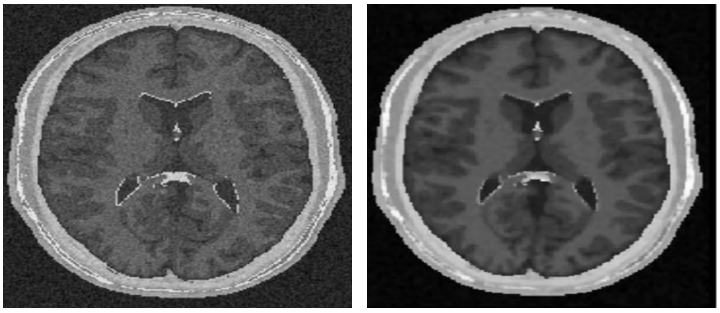
(b)

(c)



(d)

Fig. 2.3 Orthogonal views of three MRI sessions from an individual obtained in short period of time. (a) mpr-1 session. (b) mpr-2 session. (c) mpr-3 session. (d) Image registered from three image sessions.



(a) (b)

Fig. 2.4 Orthogonal views of three MRI sessions obtained in a short period of time. (a) Original MRI slice. (b) Image after noise filtering. Gavidia et al. [50]

Skull stripping is among important initial process in image analysis. Inaccuracies at this step can lead to the introduction of artefacts adversely affecting further analysis; therefore, a robust and accurate automated method for this step is highly desirable [122].

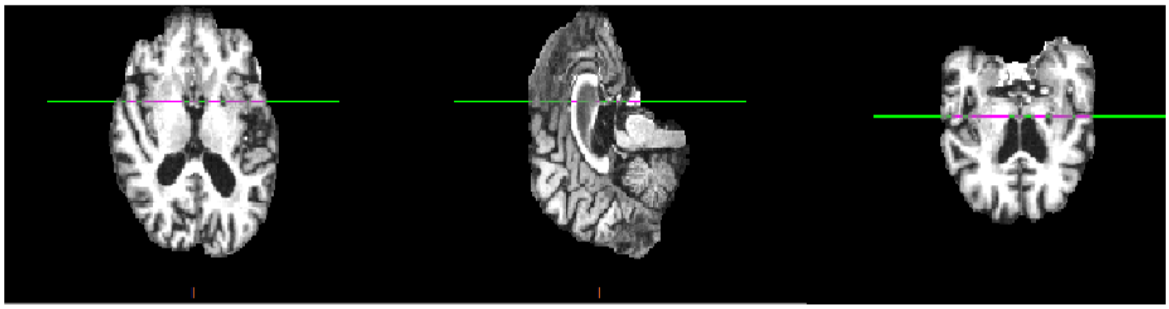


Fig. 2.5 Orthogonal views of a brain image after defacing and skull stripping.

Segmentation

Segmentation is the fully automatic separation of the non-overlapping anatomical structures within the brain including white matter (WM), grey matter (GM), CSF, the cerebral cortex, and more specific ROIs. Pre-processing and image segmentation could be considered as the most difficult tasks due to both the variability of region shapes and image quality. Brain MR images frequently display high-intensity variations throughout regions, but two neighbouring regions can share very similar intensity profiles being more difficult to differentiate them. Also, noise and other image artefacts can cause incorrect structures or boundary discontinuities. For these reasons, despite the more complex algorithms developed so far, the

preprocessing and segmentation of brain images remain highly dependent on the imaging modality and its quality.

There exist an extensive list of MRI segmentation methods applied on the human brain, which could be grouped as follows.

Manual segmentation. The first group is focused on computing measures from brain structures obtained manually. For example, in previous studies, the hippocampus and entorhinal cortex have been outlined manually, and then their volume was computed to quantify their decline [32, 15]. However, manual methods could be inexact and time-consuming activities due to the complexity of brain structures; as a consequence, these methods are applied just on limited regions.

Intensity-based methods. In this group are included well-known methods such as thresholding, region growing, classification, and clustering, where each pixel/voxel is classified based on their intensity [31]. Because the intensity profiles of more specific brain structures overlap, these methods are often applied to distinguish three main classes including WM, GM and CSF, see Figure 2.6. Furthermore, these methods are combined with more sophisticated techniques to bring hybrid segmentation methods, which allow to segment additional brain regions, as explained below.

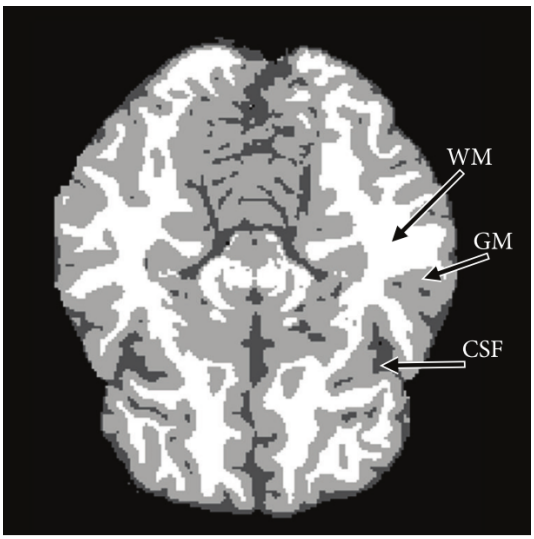
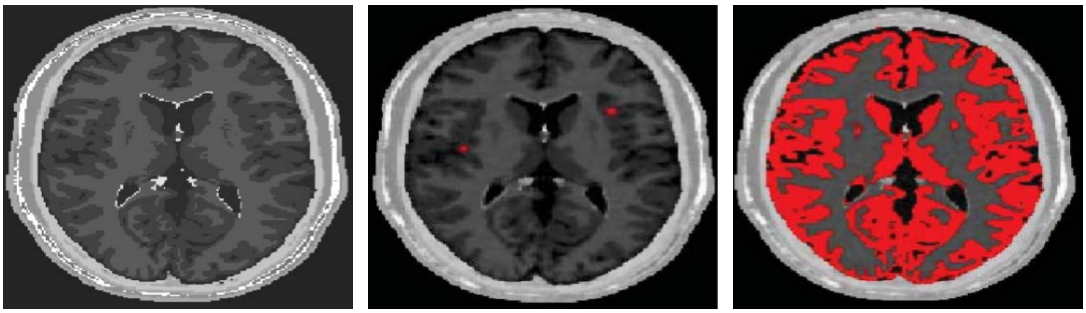


Fig. 2.6 Segmented image with three labels:WM, GM, and CSF. Source: Despotović et al. [31]

Region growing is a technique widely used in medical applications to extract connected body structures and study its pathologies. In its simplest form, region growing starts with one

or more seed points (spherical volume for 3D images) within the region of interest. The seed points can be manually selected or automatically initialised with a seed finding algorithm. The next step analyses the neighbour pixels/voxels to determine if their intensities satisfy a predefined homogeneity criterion. This step is applied until no more elements can be added. Finally, the segmented structure is represented by all the elements accepted during the searching process. Figure 2.7 shows an example of GM segmentation by using this algorithm. Region growing was applied on a brain phantom with dimensions 181x217x181 in directions x, y, z. For visualisation purposes, only the segmentation corresponding to the phantom's 98th slice is presented.



(a) (b) (c)

Fig. 2.7 Gray matter segmentation in a phantom volume. (a) Axial slice number 98 of the original phantom image. (b) Image with two seeds points. (b) Region growing segmentation after selecting the four seed points. (c) Final segmented GM region. Source: Gavidia et al. [50]

Atlas-based methods. In this approach, additional knowledge about the human brain for a specific population of interest is introduced through an atlas image (sometimes also called template). The atlas contains information about the localisation of different brain structures, which have been previously labelled by an expert, and it is used as a reference for segmenting new images. First, the atlas is aligned to the target image by applying a registration method where the similarity between the deformed atlas image and the target image is maximised. Then, all atlas information (e.g. segmentation labels) is transferred to the target image. The main advantage of these methods is the possibility of segmenting any brain structure available in the atlas without any additional cost. However, the accuracy of the segmentation is directly dependent on the quality of the registration method [118, 31]. Several image processing tools included atlas-based methods for brain segmentation, as described in section 2.1.3

Deformable models. In the literature, deformable models are also known by different names, such as snakes or active contours in 2D and active surfaces or balloons in 3D [118, 31]. Whichever the case, deformable models use curves or surfaces within an image domain for delineating region boundaries. Curves and surfaces are deformed under the influence of two forces: (a) internal forces, which are applied to keep the model smooth during deformation; and (2) external forces, which are applied to move the model toward an object boundary or other desired features within an image.

Hybrid segmentation methods. Hybrid approaches combine two or more previous methods to avoid many of the disadvantages of each method alone and improve segmentation accuracy. For example, Nestor et al. [90] combined the region growing procedure with image intensity and shape analysis (using morphological operators) for lateral brain ventricles segmentation on MR images to compare ventricular enlargement after six months in subjects with MCI, AD and normal elderly controls. Segonne et al. [106] proposed a hybrid approach that combines watershed algorithms and deformable models to strip the skull from T1-weighted MRI images.

2.1.2 Approaches of biomarkers extraction

Once brain MR images have been processed, the next step is to obtain quantitative measures from segmented regions to assess changes in brain structures as consequence of AD (e.g. atrophy rates). Many methodologies have been developed to process those images and obtain the most relevant MRI-based biomarkers [38, 69, 28, 58]. Nowadays, there are many methods proposed to quantify MRI-based biomarkers related to AD. Here, by considering the source of where they are computed, we grouped them into the three following categories.

MRI voxel level-based approach In this category, biomarkers are defined at the level of MRI voxel, which can be considered as the most simple, direct and low-level value to be used as features in classification. Several studies have computed these biomarkers as the probability of a given voxel belongs to one of three different tissue classes GM, WM and CSF [121, 69, 95]. Another approach consists in grouping the voxels into anatomical regions using a labelled atlas as it was described in section 2.1.1. For example, Fischl et al. [38] applied this approach to create a brain atlas, where each voxel in a brain MRI volume was assigned to one of 37 subcortical structures, including left and right caudate, putamen, pallidum, thalamus, lateral ventricles, hippocampus, and amygdala, among others regions. Figure 2.8 shows a coronal slice through a T1-weighted volume with regions labelled from his proposed atlas.

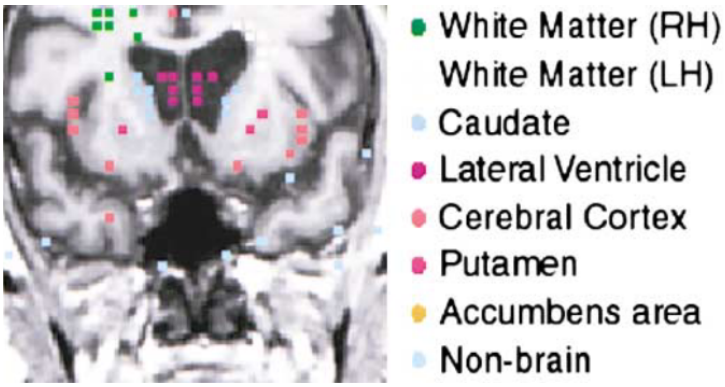


Fig. 2.8 Atlas sample proposed by Fischl et al. [38]

Lao et al. [72] used a labelled atlas of 101 ROIs developed by Noor Kabani at the Montreal Neurological Institute to label MR images and obtaining volumetric measurements of these ROIs, see Figure 2.9. The goal of this study was to detect subtle and spatially complex atrophy by training an SVM classifier.

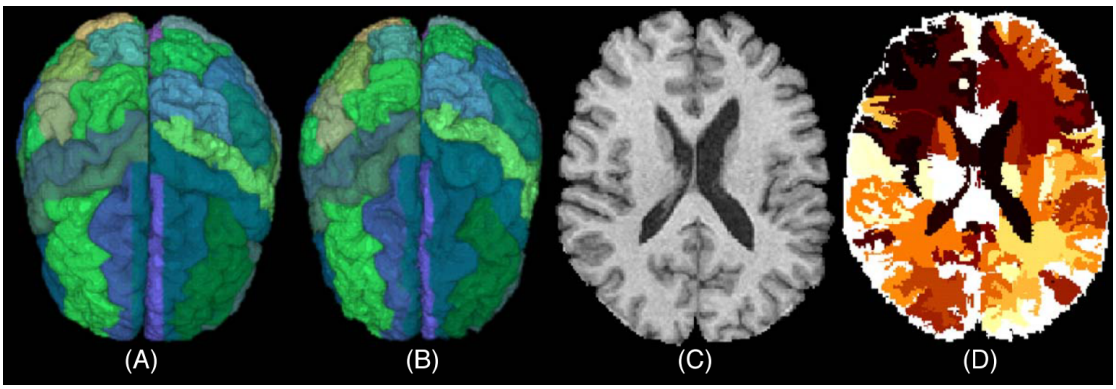


Fig. 2.9 Example of brain atlas used by Lao et al. [72]. (A) Parcellated brain image used as an atlas for the definition of ROIs. (B) Another individual's image parcellated via a high-dimensional elastic warping of the atlas on the left, using the HAMMER method. (C–D) One cross-section of the individual's original MR image and the corresponding labeled (segmented) image. Source: Lao et al. [72]

Vertex level-based biomarkers. In this category, biomarkers are defined at the vertex-level on the cortical surface, i.e. they are computed from the cortex thickness at each vertex of the human cortex. The cerebral cortex has the topology of a 2-D sheet and a highly folded geometry, and it is divided into a large number of different areas [22]. Currently, cortical thickness is considered as a potential candidate to support the early diagnosis of

AD [19, 32, 59]. Several studies have also focused on labelled cortical structures from a reference atlas. Probably, one of the most important studies has been carried out by Desikan et al. [28], who proposed a method to split the human cortical surface within 68 gyral-based ROIs (34 regions per each hemisphere). In this study, data were obtained from MRI scans of 40 subjects that varied widely in age and clinical status (young-adults, middle-aged adults, elderly adults and demented people). Information about regions was then encoded in the form of an atlas, which has been used in many further studies aimed at automatically split the brain into anatomic regions and quantify biomarkers for tracking the evolution of AD-induced changes. This atlas is shown in Figure 2.10. Also, it is import to notice that the atlas does not include the hippocampus as part of the cortical structures.

Individual ROIs-based biomarkers Here, biomarkers are computed from predefined and structured ROIs, either at the voxel or vertex-level. Thanks to the relatively low feature dimensionality and the easy understanding and identification in the whole brain, it is widely used in the literature. For example, one of most studied ROI is the hippocampus region. AD-related atrophy in MTL regions is associated with declining cognitive function, where the hippocampal volumetry on MR images is the most studied to be considered as a potential biomarker [105, 15, 89, 19, 44, 49, 9].

Patch-based biomarkers. In this category, biomarkers are computed from small 3D patches dissected previously from brain areas. In its simplest form, the whole brain is uniformly divided into patches of fixed size without overlapping [74]. However, more sophisticated techniques apply statistical methods to select the most significant voxels, and then 3D patches are extracted from the local neighbourhood of each selected voxel [74, 75].

2.1.3 Software tools.

The development of new tools for analysis, quantification and visualisation of neuroimaging data have made more accessible the quantification of MRI-based biomarkers to researchers and clinicians. These tools incorporate many methods such as the ones describe above and the possibility of analysis another image sources. Following are briefly described five tools used for this purpose, four of them are software packages widely used in the neuroimaging community for structural and functional brain imaging study.

Freesurfer. Freesurfer is a suite of tools for analysing neuroimaging data that provides an array of algorithms to quantify the functional, connectional and structural properties of the human brain [37]. This tool allows for the measurement of neuroanatomic volume, cortical

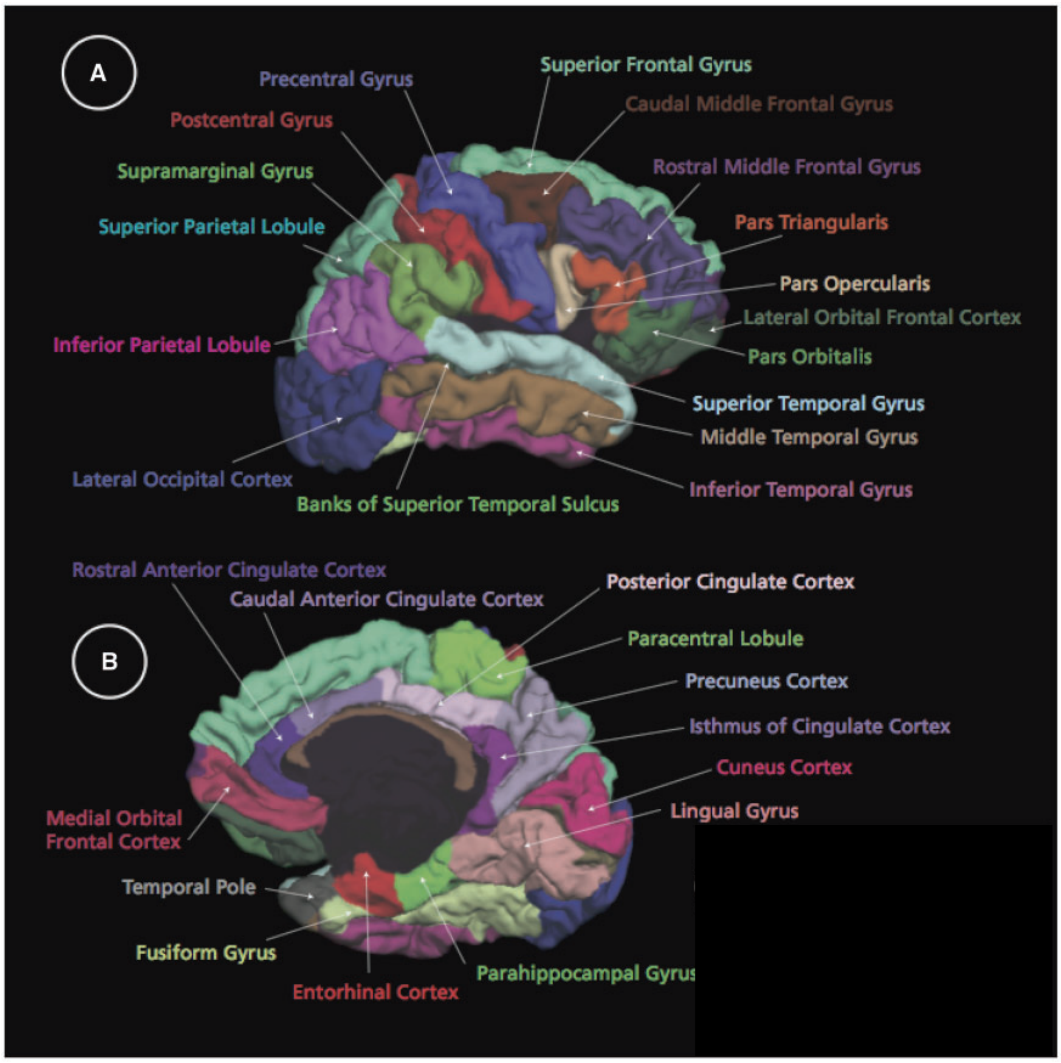


Fig. 2.10 Three-dimensional representations of all gyral-based 34 ROIs from atlas proposed by Desikan et al. [28] (only one hemisphere is shown): (A) lateral and (B) medial views of the grey matter surface: Image Source: Desikan et al. [29]

thickness and surface area of ROIs throughout the whole-brain. It was developed by *Martinos Center for Biomedical Imaging* and is freely available at the "*Freesurfer wiki site*" (http://surfer.nmr.mgh.harvard.edu/fswiki). FreeSurfer has been used in many studies dedicated to automatically obtaining MRI-based measures from the whole brain or specific regions to successfully predict the early conversion from MCI to AD subjects [29, 32, 19, 18, 14, 111, 2]. Technical details of the FreeSurfer processing pipeline have been described in prior publications [38, 106, 28, 37]. Here, the automatic subcortical segmentation of a brain volume is based upon the atlas proposed by [38] (see Section 2.1.2. This atlas contains probabilistic information on the location of structures. Each voxel in the normalized brain volume is assigned one of about 40 labels, including the Cerebral White Matter, Cerebral Cortex, Lateral Ventricle, Inferior Lateral Ventricle, Cerebellum White Matter, Cerebellum Cortex, Thalamus, Caudate, Putamen, Pallidum, Hippocampus, Amygdala, Lesion, Accumbens area, Vessel, Third Ventricle, Fourth Ventricle, Brain Stem and Cerebrospinal Fluid. Regarding the cortical brain regions, one of the Fresurfer methods to segment the human cerebral cortex into gyral based ROIs is based on the Desikan-Killiany atlas [28], which automatically labels 34 cortical ROIs in each hemisphere. This atlas was briefly introduced in Section 2.1.2.

SPM. The Statistical parametric mapping (SPM) is as software package freely available at (http://www.fil.ion.ucl.ac.uk/spm), which was developed by the Wellcome Department of Imaging Neuroscience at University College London [6]. SPM has been designed for the analysis of brain imaging data sequences, such as image normalisation, segmentation, bias correction, among others. The sequences can be a series of images from different cohorts or time-series from the same subject. Currently, SPM12 is the major update to the SPM software, containing substantial theoretical, algorithmic, structural and interface enhancements over previous versions. That release is designed for the analysis of functional MRI (fMRI), PET, Single-photon emission computed tomography (SPECT), Electroencephalography (EEG) and Magnetoencephalography (MEG). Regarding MRI segmentation, SPM was designed to perform segmentation of brain tissues consisting of GM, WM and CSF.

FMRIB Software Library (FSL). The FMRIB Software Library (FSL) (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki) is a comprehensive library of analysis tools for multimodality imaging data (structural MRI, functional MRI and diffusion MRI). It was developed by members of the Oxford Centre for Functional MRI of the Brain (Oxford University). FSL has many different modules for functional and structural MRI data analysis, of which the FMRIB Automated Segmentation Tool (FAST) was developed for segmentation of brain tissues into cortical and subcortical structures.

BrainSuite. BrainSuite (http://brainsuite.org/) is a collection of open source software tools that enable largely automated the processing of MR images of the human brain. It was written by the Laboratory of Neuro Imaging at the University of California and the Biomedical Imaging Research Group at the University of Southern California [107]. As with Freesurfer and FSL, BrainSuite can segment MR images into cortical and subcortical regions.

Pysurfer. PySurfer (http://pysurfer.github.io), a Python library, is a tool for visualising the cortical surface change. PySurfer is mainly intended for use with Freesurfer, but it also can plot data that are drawn from a variety of sources. Besides, PySurfer extends Mayavi's powerful rendering engine with a high-level interface for working with MRI data.

2.2 Statistical Learning Methods

Model building is a crucial stage in AD data analysis. Statistical learning (SL) methods are applied to MRI-based biomarkers to model and understand their change and facilitate the access to them automatically. This section presents an overview of several SL methods applied to cross-sectional and longitudinal clinical data relevant to the study presented in this thesis. Feature selection (FS) method is briefly described in Section 2.2.1. Details of several SL methods applied to cross-sectional and longitudinal data are described in Sections 2.2.2-2.2.5, followed by a brief description of metrics to assess their performance in Section 2.2.6.

2.2.1 Feature selection (FS)

Feature selection (FS), also known as variable or attribute selection, consists in selecting a subset of most relevant features eliminating the non-informative and redundant features from an initial set for being used as input features in model construction. Therefore, an FS step is desirable before model building to reduce possible issues arising from high dimensionality and reach the potential benefits, including [124]:

- Better accuracy of the inference engine,
- Improving scalability,
- Better data visualisation and understanding,
- Reducing measurement and storage requirements,
- Reducing training and inference time.

In a very simple example scenario, given a n x m dataset X consisting of n samples (rows) over m-dimensional feature space R^m representing m features $x_1,...,x_m$ over n samples; and the n x 1 outcome feature vector y; FS can be seen as the combination of an exhaustive search technique for proposing a feature subset $S = x_1,...,x_s, s < m$, which improves the model performance. Many FS algorithms include feature ranking (FR) as a principal or auxiliary selection mechanism because of its simplicity, scalability, and good empirical success [53]. Below it is briefly described the main methods applied in this study for carrying out both approaches.

Feature ranking (FR)

FR applied a score function T to rate each feature $x_1,...,x_m$ according its significance to the outcome y. Methods for defining T could be grouped into two categories: (1) methods applied on individual features, independently of the context of others (e.g. correlation); or (2) methods applied on individual features taking into account its effect in combination with other features, hence together have good predictive power.

Minimum Redundancy and Maximum Relevance (mRMR). The Minimum Redundancy and Maximum Relevance (mRMR) method, introduced by Peng et al. [92], Ding and Peng [33], belong to the second category of FR method. It computes the mutual information of two features based on their probabilistic density function, and it evaluates all of the features by looking at the intrinsic characteristics of the data concerning the outcome classes. Tthe mRMR scores are obtained by the simultaneous optimisation of two criteria, the maximum relevance and the minimum redundancy. Maximum relevance determines how well a feature discriminates between the response classes, while the minimum redundancy measures similarity or correlation between the distribution of attributes and the distribution of labels. For continuous features, a discretisation approach is applied as a preprocessing step, and then the mutual information between features is computed. The, a feature discretisation is applied using $\mu \pm (t * \sigma)$, where μ is the mean value, σ is the standard deviation, and t is a threshold usually ranging from 0.5 to 2. For example, if $t = 1$, it means that each feature is discretized into three states: -1 if it is less than $\mu - \sigma$; 1 if larger than $\mu + \sigma$; and 0 otherwise.

Wrapper methods

Wrapper methods popularized 20 years ago by Ron and George H [101], use a predictive-classification model to score feature subsets iteratively. In its most general formulation, for each subset, these methods train a new model and then test its prediction performance to assess

the relative usefulness of that subset. This train-test process could be very computationally intensive, but usually, provide the best performing feature set for that particular type of model. Wrapper methods incorporate efficient search strategies such as forward selection and backward elimination to reduce computational cost. In forward selection, features are progressively incorporated into growing subsets, whereas in backward elimination the search process starts with the set of all features and progressively eliminates the least promising ones; in both cases by evaluating the prediction model performance.

Many statistical and ML algorithms have been applied for model training, either for FS tasks or for building final prediction models. Several of them are briefly described below.

2.2.2 Early origins of linear models for longitudinal data

The early origins of the statistical analysis of change were based on ANOVA-based approaches including the following methods [39]:

- Mixed-effect ANOVA, known as univariate repeated-measures ANOVA, where a random subject effect was introduced among the repeated measurements on the same subject.
- MANOVA, repeated-measures multivariate analysis of variance,
- Methods focused on computing a derived features by summarising repeated measures by a single value (or set of them). For example, the area under the curve (AUC) is a common measure frequently used to summarise the sequence of repeated measures on any individual.

ANOVA methods resulted useful for very simple studies where data were balanced, with a uniform timing of measurements and discrete covariates. Fitzmaurice and Molenberghs [39] presented a historical perspective and review about the most remarkable developments in statistical methodology for longitudinal data analysis in the past 30 years. Authors referred an extensive review of shortcomings that limit the usefulness of based-ANOVA approaches in longitudinal studies. In Table 2.1 are summarised some of these.

2.2.3 Linear mixed-effect (LME) regression modelling

There are two main reasons for studying multiple subjects over time, the first one is the interest in individual differences and the second one is the interest in what is common to the population. Hence, there are two sources of variability known as within-subject and between-subject variabilities, and both must be taken account for inferences about a population.

Table 2.1 Review of main features and shortcomings of three ANOVA-based approaches. Summarised from Fitzmaurice and Molenberghs [39]

Approaches	References	Features	Shortcomings
Univariate repeated-measures ANOVA	(Airy, 1861), Fisher(1918, 1925), Scheff´e (1956)	It was conceptualized as a model for a single response variable. - Because it includes a single, individual-specific random effect, it induces positive correlation among the repeated measurements. - It can be considered a forerunner of more versatile regression models for longitudinal data by allowing effects to vary randomly from one individual to another.	- To made very restrictive assumptions about the covariance structure for repeated measures on the same individual, by assuming a compound symmetry form for the covariance (that assumption is often unrealistic). - It was originally developed for the analysis of balanced data with discrete covariates.
Repeated-measures multivariate analysis of variance (MANOVA)	(Box, 1950; see also Geisser and Greenhouse, 1958; Greenhouse and Geisser, 1959)	- It is a model for multivariable responses, where repeated measures of the same response feature over time are correlated.	- Require somewhat more advanced computations. - It forces the within-subject covariates to be the same for all individuals, as consequence: it cannot be used when the design is unbalanced over time, it did not allow for general missing-data patterns to arise, and individuals with missing data must be excluded from the analysis.
Derived Variable	Wishart (1938), Box (1950) and Rao (1958)	- After obtaining the single number summary, another ANOVA methods (or non-parametric methods) for the analysis of a univariate response can be applied.	- By summarising repeated measures, it forces to focus on only a single aspect of the repeated measures over time losing information.- Subject including discernibly different response profiles can produce the same summary measure.- The method cannot be applied when covariates are time-varying.- Problems with missing data or irregularly spaced repeated measures.

Models based on fixed effect approach, as linear regression model (LM), take into account just the within-subject variance. LM is a common approach for the analysis of longitudinal data when the response is continuous, however, when applied on repeated measures on the same subjects, this approach is not sufficiently flexible to capture the complex patterns of change in the response, their inherent correlation, as well as their relationships to covariates. Thus, the Type I error rate (i.e., when the null hypothesis H_0 is true but is rejected) could be increased.

Linear Mixed-Effects (LME) provide parsimonious ways to account for group level structure in the data while simultaneous assessing effects within and across groups [62]. These models incorporate both *fixed-effects* and *random-effects* [94]; and describe the relationship between a response and covariates that have been observed along with the response [7]. Like LM models, LME models describe a relationship between a response and one or more covariates. However, in this kind of models at least one of the covariates is categorical and represents the observational unit under study (e.g. humans in medical studies). By focusing on the "effect" of the covariate levels, it arise the two kinds of effects previously mentioned: *fixed-effects* and *random-effects*. In the literature, there are different ways of defining them. Pinheiro and Bates [94] define the *fixed-effects* as the parameters associated with an entire population or with certain repeatable levels of experimental factors, and the *random-effects* as associated with individual experimental units drawn at random from a population. Bates [7] establishes that a covariate is modelled using *fixed-effects* parameters when the set of possible levels of that covariate is fixed and reproducible (e.g. gender). On the other hand, if the levels observed of a covariate represent a random sample from the set of all possible levels, *random-effects* are incorporate in the model (e.g. participant identification). Significantly, it is important to notice that in these definitions, Bates [7] only establishes *fixed-effects* as model parameters, whilst *random-effects* not. For more details about other meanings of *fixed-effects* and *random-effects* see [52].

In the field of clinical longitudinal studies, probably the usefulness of LME models was highlighted in the 1980s by Laird and Ware [70], who applied a unified approach based on the two-stage formulation to fitting both growth models and repeated-measures models on data taken from an epidemiological study of the health effects of air pollution. However, the adoption of these models in medicine has been much slower. LME modelling provides a general and flexible approach to longitudinal data because it allows a wide variety of correlation patterns (or variance-covariance structures) to be explicitly modelled.

LME models offer a flexible framework by which to model the sources of variation and correlation that arise from grouped data. They provide enormous advantages over other approaches, which are listed below.

- Handles clustered data. Inferences can be made by fitting the clustering effect as random.
- Allows to analyse unbalanced and incomplete data, and also, it handles uneven spacing of repeated measurements.
- Works with reduced data.
- Models both linear and nonlinear relationships between response and covariates.
- These models are often more interpretable than other methods applied on repeated measures.

Formulation. The standard form of an LME model is given by Eq (2.1)[94].

$$Y_{ij} = X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij}; i = 1, ..., n \tag{2.1}$$

where y_{ij} is the j^{th} response on the i^{th} subject; X_{ij} is the regressor vector for the j^{th} response on the i^{th} subject; β is the p x 1 vector of fixed effect coefficients; Z_{ij} is the n_i x q model matrix for the random effects for the j^{th} observation in subject i; b_i is the q x 1 vector of random-effect coefficients for group i; ε_{ij} is the n_i x 1 vector of errors for the j^{th} observation in subject i; Ψ is the q x q covariance matrix for the random effects; and $0, \sigma^2\lambda_i$ is the n_i x n_i covariance matrix for the errors in subject i.

The assumptions for the LME model are:

- *Random-effects* vector b and the error vector ε have the following distributions: $b_i \sim N_q(0, \Psi)$, and $\varepsilon_{ij} \sim N_{ni}(0, \sigma^2\lambda_i)$
- b and ε are independent from each other.

2.2.4 Partial least squares regression (PLSR)

The Partial Least Squares Regression (PLSR) is a statistical method introduced by Wold et al. [123], which relates the two data matrices, X and Y, by a linear multivariate model instead of finding hyperplanes of maximum variance between them. In other words, PLSR is used to find the multidimensional direction in the X space that explains the maximum multidimensional variance direction in the Y space. PLSR derives its usefulness from its ability to analyse data with many, noisy, collinear, and even incomplete variables in both X and Y [123]. PLSR has the desirable property that the precision of the model parameters improves with the increasing number of relevant variables and observations.

Formulation. By definition, after observing n data samples from each block of features, PLSR decomposes the $n \times N$ matrix of zero-mean predictors features X and the $n \times M$ matrix of zero-mean responses variables Y into the form shown in Eq (2.2).

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \tag{2.2}$$

where $X \subset R^M$ is an $n \times m$ matrix of predictors and $Y \subset R^P$ is an $n \times p$ matrix of responses. T and U are $n \times l$ matrices that are the l extracted score vectors (projections, components, latent vectors) of X and Y, respectively. The $m \times l$ matrix P and the $p \times l$ matrix Q represent matrices of loadings. The $n \times m$ matrix E and the $n \times p$ matrix F are the matrices of residuals (or error matrices), assumed to be independent and identically distributed random normal variables. The decompositions of X and Y are made to maximise the covariance between T and U.

2.2.5 Support vector machine (SVM) classifier

Support Vector Machine (SVM) is a supervised method of ML developed by Vapnik [120], which have gained wide popularity for solving classification and regression problems. SVM makes use of the well-understood linear classifiers in combination with a projection into a higher-dimensional space, where the original problem can be solved (or at least reasonably well approximated) in a linear manner. A classification problem involves separating data into training and testing sets. Each sample in the training set contains one "observed response value" (i.e. the class labels) and several input features. The goal of SVM is to build a model (based on the training data) which predicts the response values of the test data given only the test data features. Figure 2.11 shows a simple representation of that training and testing processes for a two-class SVM classifier. Note that the original data were split into training and testing sets. The SVM model was built from the training data to predict the response values, and then the model was used for predicting the response values from testing data. Finally, the predicted and observed responses were compared to assess the model performance. Statistical measurements applied for model assessing are described in more detail in Section 2.2.6.

Figure 2.12 illustrates a hypothetical 2-D illustration of how a linear binary classification problem is solved during the training phase. Several different classifiers could correctly separate the two classes (black and white points), but the SVM approaches this problem through the concept of the *margin*, which is defined to be the smallest distance between the

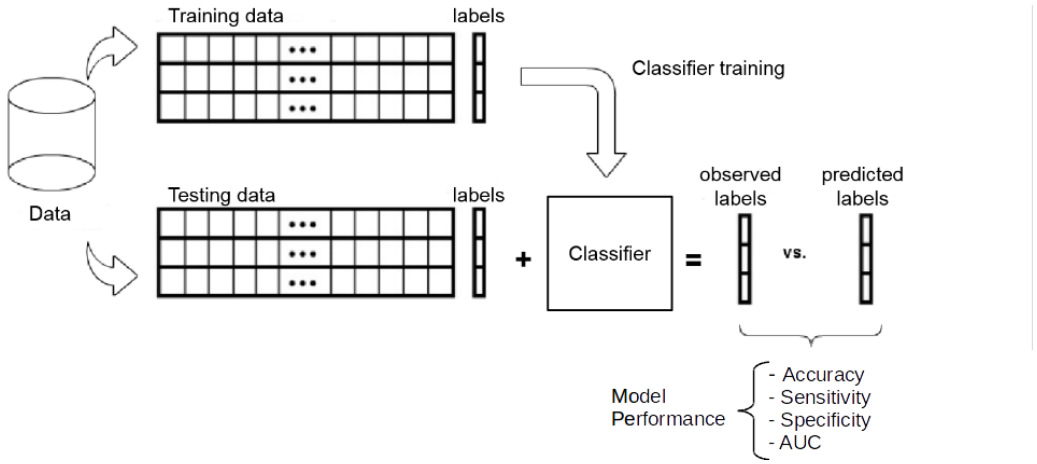


Fig. 2.11 Simplest representation of SVM model building for a binary classification problem.

decision boundary and any of the samples [10]. In this sense, the optimal model, represented by the solid black line, is trained by maximising the margin of separation between the two groups and using those samples lying closest to the separating plane as defining points, which are called as support vectors (blue circle symbols).

Linear SVM formulation. In a binary classification problem, as illustrated in Figure 2.12, SVM is formally described in a easier way as follow.

Given a training set of l points instance-response $(\vec{x}_i, y_i), i = 1, ..., l$; where the binary response $y_i \in [1, -1]^l$ and each $\vec{x}_i$ is a p-dimensional real vector. The decision surface of a linear SVM classifier is described as follow.

$$\vec{w}.\vec{x} - b = 0, \tag{2.3}$$

where the the feature weight vector $\vec{w}$ is normal (or not) to the hyperplane, and the parameter $\frac{b}{\|\vec{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector $\vec{w}$. The support vectors $\vec{x}_i$ are mapped into a higher dimensional space by the function ϕ. SVM finds a linear separating hyperplane with the "maximum-margin hyperplane" in that higher or infinitive dimensional space. As illustrated in Figure 2.12, the support vectors lie on two parallel hyperplanes described by $\vec{w}.\vec{x} - b = 1$ and $\vec{w}.\vec{x} - b = -1$, such that the distance between them is computed by $\frac{2}{\|\vec{w}\|}$.

During the training phase, there are several parameters that need to be optimized. On of them is the SVM-specific parameter $C > 0$, which is a penalty parameter of the error term. Likewise, $K(x_i, x_j) \equiv \phi(x_i^T \phi(x_j)$, which is known as the "kernel function", must be evaluated

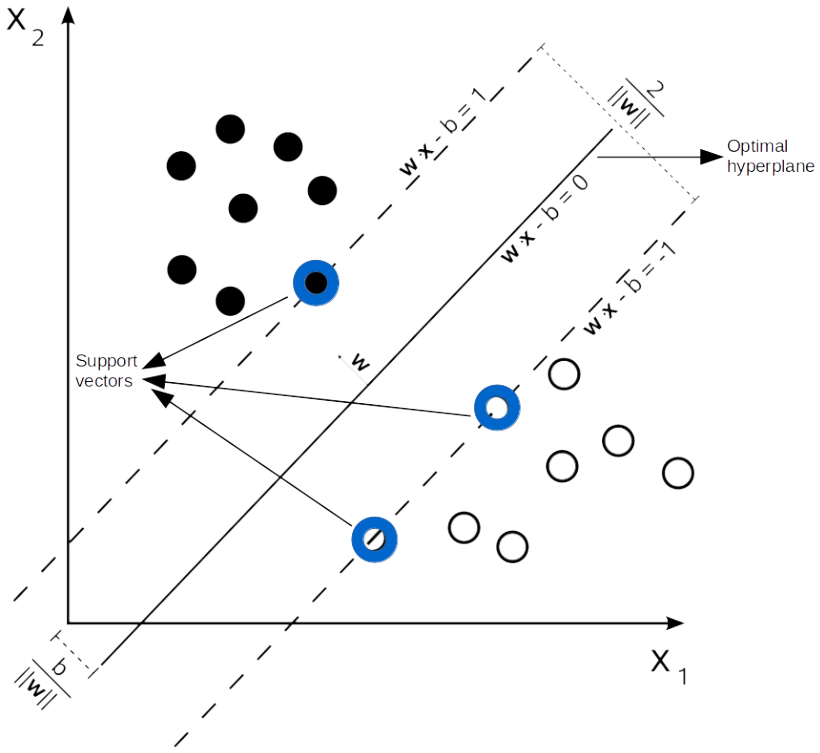


Fig. 2.12 Hypothetical example of a linear classification for a binary problem. The solid black line represents the optimal classifier by using SVM. Here, the support vectors are represented by blue circled symbols

for all possible pairs x_i and x_j. There are numerous forms of kernel functions applied on SVM algorithm, several of them are listed bellow:

- *Linear*:

$$\mathrm{K}(x_i, x_j) = x_i^T (x_j) \tag{2.4}$$

- *Polynomial*:

$$\mathrm{K}(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \tag{2.5}$$

- *Radial Basis Function*:

$$\mathrm{K}(x_i, x_j) = exp(-\gamma \|x_i - xj\|^2), \gamma > 0 \tag{2.6}$$

- *Sigmoid*:

$$\mathrm{K}(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \tag{2.7}$$

where γ, r y d are kernel parameters, which need to be optimised.

2.2.6 Model performance

Additional methods are required to assess the performance of statistical and ML models, and their selection will depend on the kind of model under evaluation. For example, a well-fitting regression model results in predicted values close to the observed response values. These models can be examined using graphical and numerical summaries. A common practice is to plot residuals versus the fitted responses to assess the assumption of constant variance of the error. Also, many numerical summaries of estimated model-specific parameters can be used as measures of "goodness of fit", including the R-squared, overall F-test, Root Mean Square Error (RMSE), Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), among many others.

An important part of this study focuses on assessing classification models built for binary outcomes. There are a variety of metrics for that purpose, several of them are briefly described in the following section.

Assessing classification models

In classification problems, statistical metrics are computed to assess the model performance using testing data, as well as for parameters optimisation during the training stage.

Confusion matrix Confusion matrix, also knows as contingency tables, is used to visualise the performance of classifiers by representing the observed and predicted classifications. A confusion matrix is of size $L \times L$, where L is the number of different classes. In a binary response variable with classes positive (p) and negative (n) , $L = 2$, and thus, the confusion matrix takes the form presented in Table 2.2. In this case, four basic metrics could be computed: True Positive (TP), which represents the number of positive samples correctly labelled as such; True Negative (TN), which represents the number of negative samples correctly labelled as such; False Positive (FP), which represents the number of negative samples incorrectly labelled as positive; and False Negative (FN), which represents the number of positive samples incorrectly labelled as negative.

Table 2.2 Representation of a confusion matrix for a binary classifier

↓ **observed \ predicted** →	**negative** (n)	**positive** (p)
negative (n)	TN	FP
positive (p)	FN	TP

By using the metrics listed above, another evaluation metrics can be derived. Several of them are listed below.

Accuracy(ACC). In the binary classification context, $ACC = (TN + TP)/(TN + FP + FN + TP)$. ACC is the proportion of true results (both TP and TN) among the total number of samples.

Precision (PREC). Also know as Positive Predictive Value (PPV). PREC is calculated as $PREC = TP/(FP + TP)$. It is the proportion of positive cases that were correctly identified.

Negative predictive value (NPV). It is the proportion of negative cases that were correctly identified. It is computed as: $NPV = TN/(TN + FN)$

Sensitivity (SEN). Also knows as True Positive Rate (TPR) or recall, is computed as $TPR = TP/(FN + TP)$. It measures the proportion of positives cases that are correctly identified as such.

Specificity (SPE). Also knows as True Negative Rate (TNR). It is calculated as $TNR = TN/(TN + FP)$. It measures the proportion of negatives cases that are correctly identified as such.

Cross-validation and train-and-test

Cross-validation is a popular strategy for models evaluation and comparison. The basic idea behind CV is to split data into two segments, where the first one is used to learn or train a model, and the other is used to validate the model. The principal interest of cross-validation lies in the universality of the data splitting heuristics, where training and validation samples are independent. Therefore, cross-validation can be applied to (almost) any algorithm in (almost) any framework, such as regression and classification models, among many others [5].

One commonly used method is k-fold cross-validation, in which data are randomly partitioned into k equally (or nearly equally) sized segments, called folds. Then, k iterations of training and validation are performed such that within each iteration a different fold of the data are held-out for validation while the remaining $k-1$ folds are used for learning. Finally, the results are averaged over the folds.

Figure 2.13, obtained from Refaeilzadeh et al. [97], represents an example of 3-fold CV, where the darker section of the data are used for training, while the lighter sections are used for validation.

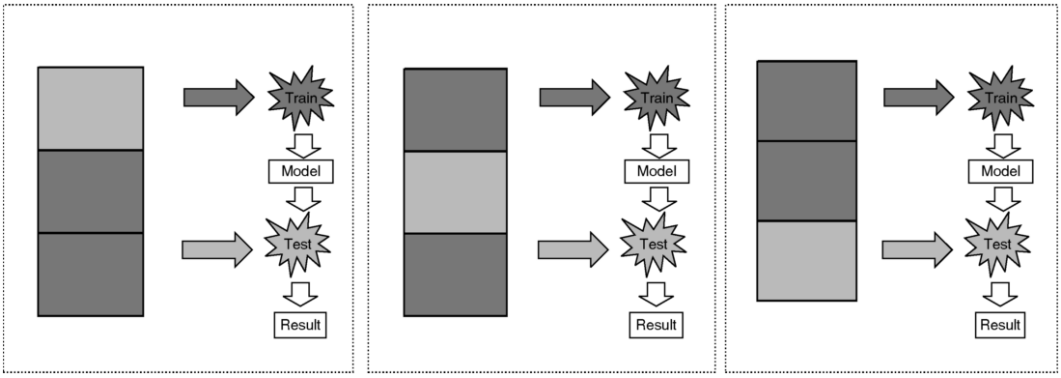


Fig. 2.13 Example of 3-fold CV. Source: Refaeilzadeh et al. [97].

In contrast to cross-validation, the train-and-test methodology just employs two completely different data sets as training and testing data (e.g. 70% for training and 30% for test).

One of the main reasons for using k-fold cross-validation instead of train-and-test is when there is not enough data available to partition it into separate training and test sets without losing significant modelling or testing capability.

2.3 The Alzheimer's Disease Neuroimaging Initiative (ADNI)

The Alzheimer's Disease Neuroimaging Initiative (ADNI) (http://www.loni.usc.edu/ADNI) is an extensive, multicenter, longitudinal neuroimaging study, which has addressed the major health problem of AD by including an extensive set of biomarkers, imaging and genetic data, as well as socio-demographic characteristics of hundreds of subjects. It was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organisations, led by Principal Investigator Michael W. Weiner, MD. Its primary goal has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and AD. The complete ADNI project enrolled 819 adult subjects, aged 55–90 years, and recruited from over 50 sites across the United States and Canada.

ADNI has been used by many publications focused on the characterisation of age-related brain changes [43, 44] and the early prediction of conversion to AD [29, 30, 19, 18, 111]. A recent review has been published by Weiner and colleagues in Alzheimer & Dementia journal [122]. Details about the procedures for selection of participants and the full study protocol have been presented in [87, 60].

2.3.1 Sociodemographic and neuropsychological data

ADNI data include demographic features, such as sex, age at baseline, years of education, handedness and APOE-ε4 carrier state, among others, which were documented in the screening visit of ADNI participants. These features have been considered in other dementia studies based on ADNI database [71, 18, 19]. APOE-ε4 carrier state includes three classes: (0) non-carrier; (1) single copy carrier; or (2) two copies carrier.

Regarding neuropsychological features, ADNI includes scores, such as, the MMSE, the CDR and the Global CDR (CDRGLOBAL) scores, which were obtained for each participant visit. CDRGLOBAL indicates the severity of dementia: 0) no dementia; 0.5) very mild dementia; 1) mild dementia; 2) moderate dementia; 3) severe dementia). It is obtained by using an algorithm that weights memory more heavily than the other remaining five categories (orientation, judgment and problem solving, community affairs/involvement, home life and hobbies, and personal care).

Furthermore, participants were clinically diagnosed at baseline and during follow-up visitations. The clinical assessment is described below.

2.3.2 ADNI diagnosis criteria

The general inclusion-exclusion criteria applied by ADNI to a baseline clinical assessment are: 1) Control normal subjects (CN, n=229) had MMSE scores between 24 and 30 (inclusive) and a CDR of 0. They were non-depressed, non MCI, and non-demented; 2) Late MCI (LMCI, n=398) subjects had MMSE scores between 24 and 30 (inclusive), a memory complaint, had objective memory loss measured by education-adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia; 3) AD (n=193) subjects had MMSE scores between 20 and 26 (inclusive), a CDR of 0.5 or 1.0, and met the *NINCDS/ADRDA* criteria for probable AD. It is important to remark that ADNI patient diagnostics are not pathologically confirmed. Thus some uncertainty on the subject's diagnosis may be introduced. In this study, this baseline clinical assessment is called dx_{bl}.

For the longitudinal study, ADNI also provides a clinical assessment of the subjects assigned at each visitation in which MRI images were obtained. Here, subjects are labelled as following: 1) Normal (NL) subjects diagnosed at current visitation; 2) NL to MCI, Subjects diagnosed as MCI at current visit who previously were NL; 3) MCI, subjects diagnosed as stable MCI at current visit who previously were also MCI; 4) NL to Dementia, subjects diagnosed as dementia due to AD at current visit who previously were NL; 5) MCI to Dementia, subjects diagnosed as dementia due to AD at current visit who previously were NL; 6) Dementia, subjects diagnosed as stable dementia due to AD at current visit who previously were MCI.

2.3.3 CSF-based Biomarkers.

ADNI CSF-based data include the $A\beta_{1-42}$, T-τ and P-τ, which were measured during baseline evaluation from 410 subjects(100 mild AD, 196 MCI, and 114 elderly cognitively normal subjects) enrolled at 56 participating centres [108]. Besides, a reduced number of participants have CSF biomarkers measured for months 12 and 24. In order to address the high variability in the assays used to measure these biomarkers, Shaw et al. [109] conducted a seven-centre inter-laboratory standardisation study for these biomarkers as part of the ADNI study. Concentrations of these biomarkers were measured using the xMAP platform (Luminex Corp, Austin, Texas) and INNO-BIA AlzBio3 research-use-only reagents.

2.3.4 MRI-based biomarkers

In the ADNI repository are available structural MR images obtained for each participant, as well as datasets of neurodegeneration features associated with cortical reconstruction and volumetric segmentation performed with the Freesurfer image analysis suite (version 4.4). These MRI-based biomarkers corresponded to cortical measures including the cortical thickness average (TA) provided in units of millimetres (mm), cortical thickness standard deviation (TS) in mm, surface area (SA) in mm^2 and cortical volume (CV) in mm^3, which are obtained from a total 34 cortical ROIs available per each hemisphere; and measures including the subcortical volume (SV) in mm^3 from 25 subcortical brain regions. All these biomarkers were measured along five years follow-up. It is important to remark that ADNI classifies the hippocampus region as a subcortical region, i.e. an SV biomarker, but it is not entirely right. Although the hippocampus is located under the cerebral cortex, it is not truly a subcortical structure. In humans, the hippocampus is located in the medial temporal lobe, which is one of the brain structures making up the limbic system. In fact, the hippocampus is a cortical infolding itself much older and more primitive than the surrounding neocortex.

The ADNI MRI Core, led by Clifford Jack, M.D., is responsible for all aspects of MRI images including determining specific MRI pulse sequences, site qualification, QA and QC of all MRI data, tracking all MRI data acquisition and processing, and performance of all MRI data processing. The ADNI MRI methods are described in more detail in [60]. Also, details of the acquisition of structural MR images of the participants can be found in *ADNI project site* (adni.loni.usc.edu).

Freesurfer was developed by *Martinos Center for Biomedical Imaging* and is freely available at the *"Freesurfer wiki site"* (http://surfer.nmr.mgh.harvard.edu/fswiki). This tool is described with more detail in Section 2.1.3.

2.4 The Open Access Series of Imaging Studies (OASIS)

The Open Access Series of Imaging Studies (OASIS) is a series of neuroimaging datasets that are publicly available for study and analysis. OASIS include data from a cross-sectional study [79] and a longitudinal study [80]. Cross-sectional imaging data includes three or four individual T1-weighted MRI scans obtained in single imaging sessions from 416 adults, ages 18–96. One hundred of the included subjects older than 60 years have been clinically diagnosed with very mild to moderate AD. On the other hand, the longitudinal data consist of a collection 0f 150 subjects aged 60 to 96. Each subject was scanned on two or more visitations, and 3 or 4 individual T1-weighted MRI scans were obtained in single scan sessions. At baseline, 64 of the included subjects were characterised as demented including

51 individuals with mild to moderate AD. Over the study years, 14 subjects, who were characterised as non-demented at baseline, were subsequently characterised as demented.

Both studies include additional subject's sociodemographic information, such as age, gender, handedness, socio-economic status and level of education. Neuropsychological features are also provided, including the CDR and MMSE scores. All participants with $CDR > 0$ were diagnosed with a probable AD. Genetic or CSF-based data are not available in OASIS studies. Like ADNI, a dataset of MRI-based biomarkers including the TA, TS, SA, CV and SV measures, also is provided for the cross-sectional study. These data are openly available at the *"OASIS Website"* (http://www.oasis-brains.org)

2.5 Advances in the early diagnosis of AD

Due to neuropsychological tests commonly used in the clinical practice of AD (see Section 1.2) may lack high reliability [88], there exist a great need for providing complementary information and increasing the accuracy of early and reliable diagnosis of the disease as well as enhancing the ability to predict the progression from early disease stages. Several studies have dedicated a special interest in exploring the influence of socio-demographic characteristics on cognitive impairment and AD development [112, 119, 4, 126, 84, 46] by analyzing data from cross-sectional studies. However, AD-related features including neuroimaging biomarkers, CSF-based biomarkers and APOE genotype (see Section 1.6) have demonstrated to be more robust to improve the diagnostic and prognostic accuracy of AD, mainly when they come from longitudinal studies. These features enable physicians to identify individuals at risk for the disease, even before symptoms appear, and possibly prevent or slow down the progression to dementia. Statistical learning (SL) methods are applied on MRI-based biomarkers to model and understand their change and facilitate the access to them automatically. SL is a recently developed area in statistics and blends with parallel developments in computer science and, in particular, machine learning [63]. It includes, multivariate techniques such as Linear Mixed-Effects (LME) modelling, Orthogonal Partial Least Squares (OPLS), Support Vector Machine (SVM) and Relevance Vector Machine (RVM), among others, which have enabled the correlation and variance of these biomarkers to be evaluated in a more easily interpreted way and with greater statistical power as compared to univariate approaches. The following sections present a review of several studies, which were focused on building SL models using AD-related biomarkers, mainly studies based on MRI-based biomarkers. Several of these models were previously introduced in Section 2.2.

2.5.1 Analysis of biomarkers trajectory

The analysis of biomarkers change over time is key for understanding differences along disease stage and for discriminating them against age-related changes. Mainly, longitudinal studies based on neuroimaging have allowed measuring brain changes within the same individuals over time, independently of such cohort effects [35]. In many studies, MRI biomarkers have been used in combination with other types of biomarkers, such as PET-based biomarkers, CSF-based biomarkers, genetic-based biomarkers and socio-demographic and neuro-psychological features to characterise the AD-related changes and improve early diagnosis performance.

Table 2.3 shows several studies focused mainly on characterising differences in AD-related biomarkers and neuro-psychological features between HC, MCI and AD subjects.

Davatzikos et al. [24] modelled spatial patterns of brain atrophy to distinguish among HC and MCI using a single value, called as the SPARE-AD index (Spatial Pattern of Abnormality for Recognition of Early AD); and whether those patterns are associated with decline cognitive. First, a classifier was trained to recognise spatial patterns of brain atrophy that distinguish AD vs HC. By using that classifier, it was obtained a SPARE-AD index for each subject. More positive SPARE-AD implies a more AD-like pattern of brain atrophy, and more negative SPARE-AD implies a more normal pattern of brain morphology. Then, the classifier was tested on a longitudinal cohort, and the progression of the computed SPARE-AD index for HC and MCI subjects was analysed using LME models. As a result, a significant increase in the rate of SPARE-AD with age was observed in HC and MCI, being the rates of change more significantly greater in MCI than in HC subjects. Furthermore, subjects who had converted to MCI over time showed worse cognitive performance; here the LME model showed a significant association between the SPARE-AD index and MMSE scores.

Risacher et al. [100] analysed different types of structural MRI biomarkers in four subject groups (sAD, cAD, sMCI and HC) and the conversion of MCI to AD using two timepoints (baseline and the final 1-year sample). Neurodegeneration biomarkers were obtained by using multiple image processing methods on MR images, including voxel-based morphometry, ROIs, and automated parcellation. For each subject, annual percent change (APC) estimates of biomarkers were calculated by using mean values from left and right ROIs from two timepoints. Also, they evaluated the influence of APOE genotype on APC rates. As a result, sAD and cAD subjects showed higher atrophy APCs across regions than HC, and APOE genotype was associated with APC in key AD-related structures.

Lo et al. [77] studied longitudinal data from ADNI subjects (HC, MCI, and AD) during up to 36 months of follow-up. Data included 3 categories of biomarkers: (1)CSF-based

Table 2.3 Summary of methods focused on comparing diagnostic group-specific biomarker trajectories.

Study	Method	**AD vs HC**	**MCI vs HC**	**AD vs MCI**	**Data used**
Davatzikos et al. [24]	LME	✓	-	-	MRI (GM,WM,CSF)
Risacher et al. [100]	APC	✓	-	-	MRI (GM, hippocampus, mean lobar densities)
Lo et al. [77]	Repeated measures regression	✓	✓	✓	MRI(hippocampus volume), CSF, FDG-PET, ADAS-cog scores
Bernal-Rusiel et al. [9]*	LME	✓	✓	✓	MRI(total hippocampus volume, entorhinal cortex thickness averaged across hemispheres)
Fjell et al. [43]	APC, One-sample t-tests	✓	-	-	MRI
Fjell et al. [44]	APC, One-sample t-tests	✓	✓	✓	CSF-Aβ, APOE, NM

* This study modelled the biomarkers change for additional subjects groups, which are not listed here. LME, Linear Mixed-Effect Modelling APC, Annual Percentage Change; MRI, Magnetic Resonance Imaging-based features; GM, Grey Matter; WM, White Matter; CSF; Cerebral Spinal Fluid-based biomarkers; NM: Neuro-psychological measures; PET, Positron Emission Tomography-based features; FDG-PET, [18F]fluorodeoxyglucose uptake measured in PET; ADAS-cog, Alzheimer Disease's Assessment Scale cognitive subscale

biomarkers (CSF-$A\beta$, CSF-τ and CSF-Phosphorylatedτ), (2) MRI hippocampal volume, and (3) Fludeoxyglucose F18 (FDG) uptake using positron emission tomography (PET). Population average rates of change of these time-varying biomarkers were evaluated by using repeated measures linear regression [34], which is an extension of generalised linear models [128]. These models were regressed by time and age at baseline. Likewise, to assess the extent to which pathological markers correlated with cognitive decline function over time, the correlation between these biomarkers and the Alzheimer Disease's Assessment Scale (ADAS) cognitive subscale (ADAS-cog) score was also studied. In summary, researchers found that biomarkers observed over time provide a potential approach for early diagnosis of AD and longitudinal patterns of them capture AD pathological states sequentially and that their predictive values for cognitive decline depend on the disease stage.

Bernal-Rusiel et al. [9] analysed well-known pair of AD biomarkers (MRI-derived longitudinal hippocampal volume and entorhinal cortex thickness measurements) from AD, MCI and HC subjects groups observed during 2-year follow up. They provided a methodology to apply LME modelling on longitudinal structural MRI studies in order to characterize the change of both biomarkers between five diagnostic groups: stable HC, stable MCI and stable AD (all these subjects were stable throughout the follow-up period) and converter HC and converter MCI subjects (those who were converted to MCI and AD during follow up). They observed a significant difference between the two biomarkers across HC, stable MCI, and converted MCI subjects, but that difference diminished and became statistically insignificant for the entorhinal cortex thickness measurements when comparing converted MCI subjects and stable AD patients. Their results suggested that LME approach offers superior statistical power in detecting longitudinal group differences in comparison with widely-used alternative methods.

Fjell et al. [43] focused on determining which brain regions show the greatest changes in normal ageing over short time intervals (one or two years), and how those regions change with age. Longitudinal biomarkers were obtained from MRI volumetric ROIs and surface-based ROIs from HC and AD (mild to moderate status) subjects. One-sample t-tests were used to test significant atrophy in each ROI of both diagnostic groups. As an important finding, significant volumetric reductions of the cerebral cortex and subcortical brain structures, as well as an expansion of the ventricles, were seen in HC over periods as short as one year. Several of these age-related changes occur in well-known brain regions vulnerable to AD, while other changes have been observed in areas less characteristic of an early-stage AD. Four years later, in order to differentiate the characteristics of normal aging from presymptomatic AD, this study was extended to compare atrophy rates in HC, MCI and AD subjects using

MRI biomarkers observed to 3-to-4-year follow-up, as well as, CSF-Aβ biomarker, APOE genotype and neuro-psychological features (CDR and MMSE) [44] .

2.5.2 AD classification from neuroimaging data

The development of automatic methods for the accurate classification of subjects into diagnostic groups from neuroimaging data is the aim of multiple studies. Studies are focused on supporting the early diagnosis and prognosis of AD by selecting and combining optimum features from multiple modalities, including MRI, PET, CSF biomarkers, and clinical tests; and building statistical and ML models using those features. Most studies are mainly focused on three differential diagnosis problem: (1) AD vs HC, (2) MCI vs HC, and (3) AD vs MCI. The last one problem is focused on the prodromal AD, and it is mainly directed at discriminating between the stable MCI (sMCI, MCI who had not converted to AD) versus converted to AD (cAD, subjects who had converted to AD).

Table 2.4 compares the prediction accuracy of several methods, which have been evaluated on the ADNI datasets and other data sources and have mainly focused on MRI and its fusion with different data modality. In the table, we only show the most significant results of those studies. Many of those methods applied SVM [121, 18, 129, 65, 73, 113, 114] and Principal Components Analysis (PCA) [47–49] before classification task to improve prediction quality by selecting the most significant brain structures or other feature types. In this way, once selected the most optimal features, the correlation and variance of multivariate models were determined and evaluated in a more easily interpreted way. In summary, for AD vs HC, the methods mentioned achieved high accuracy (ACC) values (up to 96.88% sensitivity and 95.22% specificity); and ACCs from 79.27% to 97.62% for MCI vs HC. However, for the detection of the prodromal AD (AD vs MCI), the ACC was substantially lower (below 76.72%) for all methods.

Table 2.4 also shows that many of classification methods are based on SVM [69, 121, 18, 19, 65, 73, 129, 113]. Klöppel et al. [69] built SVM models to classify the GM voxels of structural MR images from patients with a confirmed diagnosis of AD, probable mild AD and HC subjects. Furthermore, in order to test the ability of the SVM's to differentiate different forms of dementia, researchers also built SVM models to discriminate between confirmed AD and subjects with neuropathologically proven frontotemporal lobar degeneration (FTLD) (this disease is sometimes difficult to distinguish from AD clinically). They found that SVMs can aid the clinical diagnosis of AD and correctly differentiate between different forms of dementia. Plant et al. [95] developed a multi-step data mining framework including FS, clustering and classification to identify the best discriminating regions in brain images, which support the prediction of the conversion from MCI to AD (cAD vs sMCI). For classification

tasks, authors selected three representative approaches: SVM, Bayes statistics, and voting feature intervals (VFI). Using AD and HC as training data and MCI as test data, the best prediction of conversion to AD was achieved with the VFI classifier. Like the previous study, Cui et al. [18] were also focused on supporting the prediction of conversion from MCI to AD by examining multiple features including MRI-based biomarkers, CSF-based biomarkers and neuropsychological and functional (NM) scores. MRI-based biomarkers corresponded to 34 ROIs per hemisphere. Data were downloaded from the ADNI website, where longitudinal structural MR images were processed using Freesurfer software. These biomarkers provide information about the cortical thickness average, the standard deviation of thickness, surface area and cortical volume of each ROI. FS step was carried out to select an optimal feature subset from each modality. Then, SVM models were trained on both HC and AD data at baseline and testing was conducted to discriminate between sMCI and cAD subjects at different times (6 months,12 months, 18 months and 24 months). The SVM model trained with selected MRI, CSF and NM features obtained the best performance, suggesting that their combination may be more useful and practical for clinical diagnosis than single modality of predictors. Cuingnet et al. [19] evaluated the performance of ten multivariate classification methods applied on MR images at baseline. Five methods were based on the voxel-based segmentation, three methods were based on cortical thickness, and the remainders were based on the hippocampal shape or volume. Three classification experiments were performed to compare the different approaches: the classification AD vs HC, cAD vs HC, cAD vs sMCI. Classifiers were built using SVM. High accuracies in distinguishing AD from HC were reported for whole-brain (or the whole cortex) approaches. However, at the prodromal stage, no method was able to predict conversion to AD with higher sensitivity. Zhang and Shen [129] proposed a general learning framework called Multi-Modal Multi-Task (M3T) learning, to jointly predict multiple features from baseline multi-modal data, including MRI, FDG-PET, and CSF data. First, a multi-task FS method was applied to select the relevant features for multiple response features from each data modality, and then a multi-modal SVM (for both regression and classification) was trained which joins the above-selected features from all modalities to predict multiple (regression and classification) variables. Finally, SVM models were validated on two sets of experiments: (1) Estimation of two clinical continuous features (MMSE score and ADAS cognitive subscale) and one categorical diagnostic feature (HC/MCI/AD class label); and (2) Prediction of 2-year changes of MMSE, ADAS cognitive subscale and also the conversion of MCI to AD (cAD vs sMCI).

On the other hand, some methods summarised imaging data into one score, allowing the direct comparison of subjects and their classification into diagnostic groups. Vemuri et al. [121] trained SVM models to classify tissues density values of structural MR images obtained

from HC and AD subjects. Model input data included the previously selected GM, WM and CSF voxels by using SVM-related weights. In order to study the age, gender and genotype effect on brain atrophy, authors added that information to model training. The output of that models was an adjusted structural abnormality index called (aSTAND)-score. This score takes into account the additional subject-specific information for each individual and gives a number that represents the severity of brain atrophy that individual in comparison to the AD and HC cohorts of the same age, gender and APOE genotype. This study found that the aSTAND-scores correlate well with cognitive decline measures, such as MMSE, CDR, and Dementia Rating Scale (DRS). Other studies have proposed an age estimation framework (*BrainAGE*) to estimate individual brain ages based on GM measurements from MR images. Franke et al. [47, 48] trained an RVM model for regression (RVR) from healthy subjects images- Then, Gaser et al. [49] used this model to estimate the age of MCI and AD subjects and recognise faster brain atrophy to predict the conversion from MCI to AD. Differences between the estimated and the true age were used to indicate accelerated (positive values) or decelerated (negative values) brain ageing. Other studies proposed methods to calculate an MRI-based AD severity index from cross-sectional [111] and longitudinal studies [2], where multivariate models based on OPLS regression were built from HC and AD subject data. Then, those models were applied to MCI subjects for early prediction of conversion to AD through subject classification as either HC-like or as AD-like.

Recently, a deep learning (DL) architecture is gaining a considerable attention due to its representational power. When DL is applied to medical images, low to high-level features can be obtained, which allow building more robust classifiers. Suk and Shen [113] proposed a DL-based feature representation for early diagnosis of prodromal AD. They applied a Stacked Auto-Encoder to discover a latent representation from the neuroimaging and biological features. Data included baseline MRI, PET, and CSF features from ADNI. Their method achieved accuracies of 95.9%, 85.0%, and 75.8% for AD, MCI, and MCI-converter diagnosis, respectively. Later, Suk et al. [114] addressed both feature representation and multimodal data fusion for computer-aided AD/MCI diagnosis by combining a patch-based feature representation and Deep Boltzmann Machine. In that way, researchers handled small changes in the brain with more sensitivity. For the differential diagnosis problems, this study obtained the maximal accuracies of 95.35%, 85.67%, and 74.58%, for AD vs HC, MCI vs HC and AD vs MCI, respectively.

Table 2.4 Comparison of method performances focused on subject classification (%).

Study	Method	**AD vs HC**			**MCI vs HC**			**AD vs MCI**			**Data used**
		ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	
Kloeppel et al. [69]	SVM	95.0	95.0	95.0	-	-	-	-	-	-	MRI (GM)
Vemuri et al. [121]	SVM	-	86.0	86.0	-	-	-	-	-	-	MRI (GM,WM,CSF)
	SVM	-	88.0	90.0	-	-	-	-	-	-	MRI (GM,WM,CSF), age, gender
	SVM	-	86.0	92.0	-	-	-	-	-	-	MRI (GM,WM,CSF), age, gender, APOE
Plant et al. [95]	SVM	90.0	96.88	77.78	97.62	95.83	100	75.0	55.56	86.87	MRI (GM,WM)
Franke et al. [47]	RVM	-	-	-	-	-	-	-	-	-	MRI (GM)
Franke et al. [48]	RVM	-	-	-	-	-	-	-	-	-	MRI (GM,WM)
Cui et al. [18]	SVM	-	-	-	-	-	-	67.13	96.43	48.28	NM,CSF,MRI-based biomarkers
		-	-	-	-	-	-	62.24	92.86	42.53	NM, MRI-based biomarkers
		-	-	-	-	-	-	62.24	57.14	65.52	MRI-based biomarkers
Cuingnet et al. [19]	SVM	-	81.0	95.0	-	-	-	-	57.0	78.0	MRI(GM)
Zhang et al. [129]*	multi-modal SVM	93.3	-	-	83.2	-	-	73.9	68.6	73.6	MRI, PET, CSF
Suk et al. [113]	DL, SVM	95.9	-	-	85.0	-	-	75.8	-	-	MRI, PET
Jie et al. [65]	SVM	95.03	94.90	95.00	79.27	85.86	66.64	68.94	64.65	71.79	MRI, FDG-PET
Gaser et al. [49]	RVR	-	-	-	-	-	-	75.00	71.00	84.00	MRI-based age
Spulber et al. [111]	OPLS Regression	88.4	86.1	90.4	-	-	-	67.7	69.6	66.8	MRI-based index
Aguilar et al. [2]	OPLS Regression	-	92.0	75.0	-	-	-	-	92.0	47.0	MRI-based index
Liu et al. [73]	Multi-kernel SVM	94.37	94.71	94.04	78.8	84.85	67.06	67.83	64.88	70.0	MRI, PET
Suk et al. [114]	DL, SVM	92.38	91.54	94.56	84.24	99.58	53.79	72.42	36.70	90.98	MRI
		93.35	94.65	95.22	85.67	95.37	65.87	75.92	48.04	95.23	MRI,PET

* This study applied multi-modal learning (both regression and classification). Here, just classification ACC is reported. MRI, Magnetic Resonance Imaging-based features; CSF; Cerebral Spinal Fluid-based biomarkers; NM: Neuro-psychological measures; PET, Positron Emission Tomography-based features; FDG-PET, [18F]fluorodeoxyglucose uptake measured in PET; MRI-based age, individual estimated age computed from MRI images; MRI-based index, individual severity index computed from MRI images;

2.6 Open issues

As seen in the previous Section, neuroimaging biomarkers are among the most promising areas of research focused on early detection. In fact, structural MRI is considered part of the standard workup for AD to obtain macroscopic visualisation of atrophy caused by neuronal death over time. However, at the moment, there is no agreement on standardised values for brain volume, which would help to determine a significant brain shrinkage for any person at a single point in time. Furthermore, as previously described in Section 1.5, ageing effects on the brain include changes at molecular and morphological levels. Several of these changes are shared by neurodegenerative diseases. Healthy subjects lose brain matter over time, and those changes are not uniform across the whole brain or over subjects. Part of the ageing-related atrophy occurs in areas vulnerable to AD, while other changes are observed in areas less characteristic of the early-stage AD, such as MCI. Regarding molecular changes, cognitively normal elderly have Aβ deposition in the brain with similar levels of the substances observed in subjects with dementia due to AD. Similarly, despite an increase in τ protein having been seen in AD as compared to HC, τ deposition is found in other neurodegenerative diseases. These shared biochemical and morphological characteristics represent opportunities and challenges for researchers and clinicians to discriminate age-related changes in healthy elderly subjects from people affected with the early stage of AD. Also, part of that challenge includes dealing with the current poor understanding of the initial stage of AD. In order to address these issues, a good starting point is to understand and differentiate age-related changes in brain regions in the absence of disease, and then to support early and accurate AD diagnosis.

Multivariate regression and classification algorithms, such as those used in the studies described above, have boosted the major advances made in the early diagnosis of AD in this century. These methods have allowed obtaining multi-source features obtained by identifying and combining different types of AD-related features, such as phenotypes, biomarkers, clinical data, genotype, among others. Multi-source features have more statistical significance than each of them alone. However, most studies commented in Section 2.5 present limitations, which should be considered as starting points for new studies. Although it is well known that ageing-related effects on the brain are widespread; ageing is the major risk factor for sporadic AD—older brains being more susceptible than young or middle-aged ones, and the atrophy rate in elderly adults is higher than in middle-aged adults. most of studies listed on Tables 2.3 and 2.4 have not taken into account age differences in applying the SL methods. On the other hand, even though the most studies have been applied to ADNI data repository, which is highly prized for the complete AD longitudinal study carried out on hundreds of people; very few of them have been focused on analysing the longitudinal data.

Furthermore, the few methods applied to longitudinal data have not been careful enough with likely bias issues due to samples of the same subjects are correlated. In that sense, it is necessary to conduct studies that focus on modelling both between-subject and within-subject changes addressing the challenges inherent in that data type. Finally, no study provided information about the advancement (in years) in the early disease prediction by using their proposed methods. By knowing the number of years of advancement in the early prediction of the disease, researchers would compare methods, identify the most potent biomarkers and others AD-related features; and also clinicians could apply the most reliable treatments according to the disease stage.

Chapter 3

Objectives and thesis contribution

3.1 Main Objective

The main objectives of the thesis are to figure out the ageing-related brain change over time, and to address the differential diagnosis between healthy elderly (HC), mild cognitive impairment (MCI) and demented due to Alzheimer's disease (AD) subjects, as well as the early prediction of conversion to MCI or AD, using statistical learning methods on longitudinal neurodegeneration biomarkers based on MRI.

3.1.1 Specific Objectives

1. Define a set of rules to stratify in a reliable way the HC, MCI and AD cohorts according to their current and future diagnosis.
2. Develop suitable methods to analyse longitudinal MRI-based biomarkers from HC, MCI and AD by taking into account the nature of this kind of data.
3. Figure out the ageing-related changes of the brain in HC, as well as in MCI and AD subjects.
4. Identify the most significant MRI-based neuro-degeneration biomarkers on HC and subjects along AD stages taking into account age and gender.
5. Apply statistical learning algorithms to address the estimation of current diagnosis by carrying out differential diagnostic experiments between HC, MCI and AD subjects.
6. Address the prediction of future diagnosis of HC, MCI and AD subjects by building statistical learning models, which should be capable of predicting advanced stages of the disease from pre-dementia stages.

7. Compute the time of advancement in the prediction of future diagnosis reached by the proposed methods.

8. Validate the proposed methods by using data gathered from a different population.

3.1.2 Expected Contributions

The most important expected contribution is to aid the growing research area focused on the early diagnosis of AD, not only by developing powerful models to subject's diagnosis but also by initially addressing the ageing-specific changes in the brain, in order to differentiate these changes from the changes due to the disease. In particular, by using significant neuro-degeneration biomarkers based on MRI, two methods for HC/MCI/AD subjects discrimination and early prediction of future disease stages, are presented. Also, the ageing-related morphological brain changes are identified and characterised by applying specific methods for modelling longitudinal data, which takes into account the between-subject and within-subject changes over time.

Chapter 4

Early prediction of AD: Method M_{res}

Work in this chapter has, in part, been presented in: Giovana Gavidia-Bovadilla, Samir Kanaan-Izquierdo, María Mataró-Serrat, Alexandre Perera-Lluna, for the Alzheimer's Disease Neuroimaging Initiative. Early Prediction of Alzheimer's Disease Using Null Longitudinal Model-Based Classifiers. PLOS ONE, 12(1):e0168011, jan 2017. ISSN 1932-6203.

4.1 Introduction

In order to understand the specific effects of AD on brain structures, it is important to differentiate their age-related changes in the absence of disease. This chapter presents a sequential method, called M_{res}, which is focused on characterizing the age-related changes in brain structures of healthy elderly to identify the variant (*vr*) and quasi-variant (*qvr*) MRI biomarkers and build ageing-based null models. We hypothesized that, by identifying both *vr* and *qvr* regions, it would be possible to obtain an ageing-based null model, which would characterize their normal atrophy and growth patterns as well as the correlation between them. Null models were built from identified HC subjects (n=46) with a normal CSF-profile. Most importantly, this method address the estimation of current diagnosis (ECD) of HC/MCI/AD subjects and the prediction of future diagnosis (PFD) mainly focused on the early prediction of conversion to AD, by using these null models to estimate the age-related values of *vr* and *qvr* MRI biomarkers for longitudinal data of HC (n=161), MCI (n=209) and AD (n=331) subjects. Residuals were then calculated as deviation scores of observed MRI-based biomarkers from estimated normal MRI-based biomarkers. Support vector machines (SVM) were used to build residual-based classifiers for three experiments: MCI vs HC, AD vs MCI and AD vs HC. The advancement for early disease prediction was calculated as the number of years that the proposed method leads in predicting the last known subject diagnostic. Data

used in this study was obtained from the ADNI study (adni.loni.usc.edu). Details about ADNI database was briefly introduced in Section 2.3.

4.2 Data

4.2.1 Subjects and inclusion criteria.

Participants were selected from original ADNI study if they met the following criteria (at the time of the study, April, 2015): (1) Had all selected longitudinal MRI images correctly processed (2) Had completed demographic and neuropsychological data and were clinically diagnosed at each visitation. In total, longitudinal data of 747 subjects (215 CN, 366 LMCI and 166 AD) were studied. Demographic details of the studied cohort are given in Table 4.1. Summaries are grouped by gender and correspond to baseline stage.

Table 4.1 Statistical descriptors of studied ADNI cohort at baseline.

	Female	Male
	$N = 316$	$N = 431$
Age_{bl}	$_{71.1}\,75.3\,_{79.9}$	$_{71.2}\,75.6\,_{79.9}$
	(74.8 ± 6.8)	(75.4 ± 6.7)
Years of education	$_{0.65}\,0.75\,_{0.85}$	$_{0.70}\,0.80\,_{0.90}$
	(0.75 ± 0.14)	(0.80 ± 0.15)
Ethnicity		
Hisp/Latino	$2.2\%\ \frac{7}{316}$	$1.9\%\ \frac{8}{431}$
Not Hisp/Latino	$96.5\%\ \frac{305}{316}$	$98.1\%\ \frac{423}{431}$
Unknown	$1.3\%\ \frac{4}{316}$	$0.0\%\ \frac{0}{431}$
Race		
Am Indian/Alaskan	$0.32\%\ \frac{1}{316}$	$0.00\%\ \frac{0}{431}$
Asian	$1.58\%\ \frac{5}{316}$	$2.09\%\ \frac{9}{431}$
Black	$6.33\%\ \frac{20}{316}$	$3.25\%\ \frac{14}{431}$
More than one	$0.32\%\ \frac{1}{316}$	$0.23\%\ \frac{1}{431}$
White	$91.46\%\ \frac{289}{316}$	$94.43\%\ \frac{407}{431}$
Marital State		
Divorced	$10.44\%\ \frac{33}{316}$	$3.25\%\ \frac{14}{431}$
Married	$61.08\%\ \frac{193}{316}$	$89.10\%\ \frac{384}{431}$
Never married	$4.43\%\ \frac{14}{316}$	$2.09\%\ \frac{9}{431}$
Unknown	$0.00\%\ \frac{0}{316}$	$0.23\%\ \frac{1}{431}$
Widowed	$24.05\%\ \frac{76}{316}$	$5.34\%\ \frac{23}{431}$
APOE-ε4 *		
0	$52\%\ \frac{164}{316}$	$50\%\ \frac{215}{431}$
1	$37\%\ \frac{116}{316}$	$39\%\ \frac{167}{431}$
2	$11\%\ \frac{36}{316}$	$11\%\ \frac{49}{431}$
MMSE	$_{25.0}\,27.0\,_{29.0}$	$_{25.0}\,27.0\,_{29.0}$
	(26.8 ± 2.7)	(26.8 ± 2.5)
CDRGLOBAL		
0	$32.6\%\ \frac{103}{316}$	$26.0\%\ \frac{112}{431}$
0.5	$54.4\%\ \frac{172}{316}$	$65.7\%\ \frac{283}{431}$
1	$13.0\%\ \frac{41}{316}$	$8.3\%\ \frac{36}{431}$
dx_{bl}		
AD	$25\%\ \frac{80}{316}$	$20\%\ \frac{86}{431}$
CN	$33\%\ \frac{103}{316}$	$26\%\ \frac{112}{431}$
LMCI	$42\%\ \frac{133}{316}$	$54\%\ \frac{233}{431}$
Availability of CSF data **		
Yes	$52\%\ \frac{164}{316}$	$55\%\ \frac{236}{431}$
Not	$48\%\ \frac{152}{316}$	$45\%\ \frac{195}{431}$

$_a\,b\,_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD.N is the number of non–missing values observations. Numbers after percents are frequencies. AD: dementia due to Alzheimer's Disease; CN: Control Normal and LMCI: Late Mild Cognitive Impairment.

* Number of copies of allele.

** Availability of both CSF-Aβ and CSF-$\tau$was data at time of study (April, 2015).

4.2.2 CSF biomarkers.

CSF biomarkers set was conformed by the Aβ peptide extending to position 42 (CSF-Aβ) and the total tau (CSF-τ), in pg/mL, which were briefly described in Section 2.3.3. It should be noted that in Table 4.1 just 164 women and 236 men have available measures of CSF-based biomarkers.

4.2.3 MRI-based biomarkers.

MRI-based biomarkers used in this study were previously described in Section 2.3.4. They correspond to measured of neuro-degeneration available in ADNI at 5 years follow-up. These biomarkers were obtained with the FreeSurfer image analysis suite, version 4.4 (see Section 2.1.3). We only included biomarkers of the cortical volume (CV), cortical thickness average (TA) and the volume estimates of a wide range of sub-cortical structures (SV). Note that in this thesis, we maintained the biomarkers classification established by ADNI. Thus, we kept the hippocampus within the group of SV structures. Biomarkers related to white matter hypointensities, optic chiasm, insula and the unknown regions were excluded from further analysis. Furthermore, biomarkers with missing values for most samples were discarded. Likewise, we only included the ones correctly processed and available for at least two time points. In this sense, we included an unbalanced longitudinal data of 166 longitudinal MRI-based biomarkers, including CV (N=66) and TA (N=66) of 33 cortical regions; and SV (N=34) of 22 subcortical regions. Biomarkers measured at multiple time points: baseline, 6, 12, 18, 24, 36, 48 and 60 months.

4.3 Methods

This study was divided into three main stages: (1) Subject classification, (2) Building the ageing-based null models, and (3) Residuals-based early prediction of conversion to MCI/AD and HC/MCI/AD classification with SVM classifiers. Figure 4.1 illustrates a schematic diagram of the proposed method, M_{res}. All statistical and Machine Learning analyses and graphics were performed in R version 1.0.44 [96].

4.3.1 Classification of subjects.

Since the diagnosis of MCI and AD is progressive, in addition to ADNI clinical assessment established at every visitation, dx_{age} (see Section 2.3.2), we aim to control the last ADNI clinical diagnosis of each subject at time of this study, by building a time-invariant diagnosis

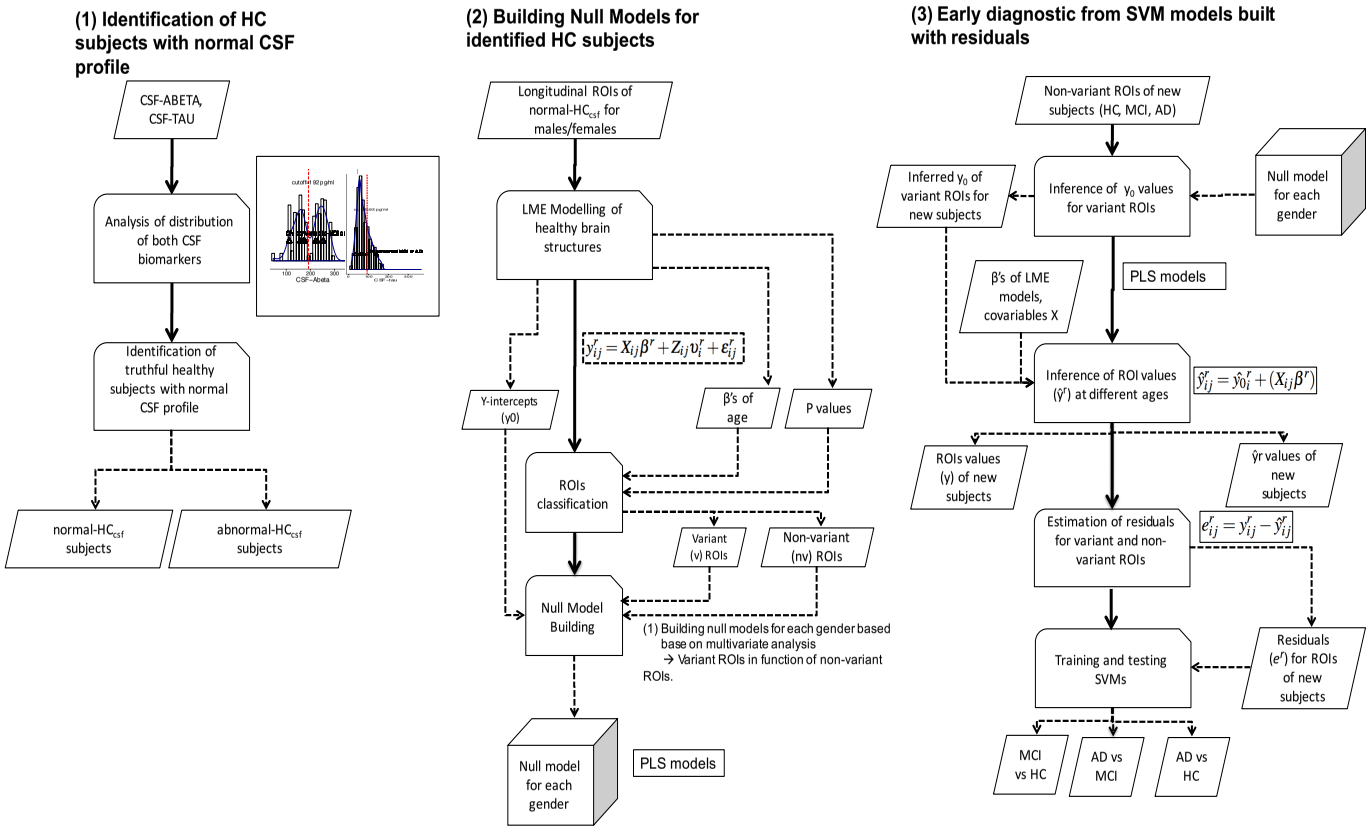


Fig. 4.1 Proposed method: M_{res}.

(1) HC subjects with normal CSF profile are identified from cutoff values calculated from CSF biomarkers distributions. (2) Longitudinal ROIs of these subjects are modelled using LME approach, variant (*vr*) and quasi-variant (*qvr*) ROIs and Y-intercepts (y_0) ROIs values are identified and then null models for both genders are built from these values by applying multivariate modelling. (3) *qvr* ROIs values of new HC, MCI and AD subjects are used within null models to infer the y_0 values of *vr* ROIs. Estimated ROIs values ($\hat{y}$) at different ages are estimated by linear regression by using y_0 and β coefficients of age and education. Residuals are calculated as the difference $y - \hat{y}$; and finally, SVM classifiers are trained for subject classification and addressing the early diagnosis problem: HC vs MCI, MCI vs. AD and HC vs AD. The full workflow of last two stages is applied separately for each gender.

variable (dx_{last}). This variable labels subjects according the following classes: (1) stable HC (sHC), (2) stable MCI (sMCI), (3) converted to MCI (cMCI), (4)stable AD (sAD) and (5) converted to AD (cAD). Because previous studies have found ADNI participants with abnormal concentrations of CSF-based biomarkers even in healthy elderly subjects [117, 108], we decided to study the subjects with available CSF data (see Table 4.1) and to discriminate the ones with normal CSF-profile from the ones with abnormal concentrations. For this propose, we investigated the cut-off values previously established by [108] (CSF-Aβ : $192pg/ml$, CSF-τ : $93pg/ml$) [108] from their study with ADNI-independent autopsy-based samples, on our samples at long-term follow-up times (84 months). As many of these subjects have more than one measurement available for both CSF-based biomarkers, the normal or abnormal profile was evaluated for all observations. In this way, we expected to classify these subjects through the other time-invariant feature (dx_{csf}), which combines their dx_{last} state and their CSF profile. Table 4.2 gives a brief description of all diagnosis variables used in this study. Once the subjects were characterized according to dx_{last} and dx_{csf}, two cohorts were established: (1) The null model cohort, integrated by the stable HC subjects with normal CSF profile (normal-HC_{csf}); and (2) The early prediction cohort, integrated by the remaining subjects who did not meet the previous condition.

4.3.2 Building the ageing-based null models

In this stage, we analyzed data from the null model cohort. Every MRI-based biomarker of the normal-HC_{csf} group was standardized to have zero mean and unit variance. Here, we also refer to an MRI-based biomarker as a ROI when we refer to the modelling process.

Modelling temporal change in ROIs.

To visualize the between-subject and within-group variabilities on normal-HC_{csf}, ROIs were represented by quartiles. The change in cortical and subcortical brain regions over 5 years was calculated by applying the LME approach for every ROI (see Section 2.2.3). The subject age (*age*) at each observation and the years of education (*educ*) were included as covariates in all models. Because we hypothesize that there are important individual-level effects and believe that subjects have similar rates of change over time, we fitted random intercepts LME models. In this type of model, the measured value of ROI r defined as y_{ij} is assumed to have a set of parameters β, fixed across subjects. In addition, for each individual i, a set of random parameters υ_i is assigned that models the deviation from the fixed effect β. For $i = 1, ..., n$, each model reads as follows:

Table 4.2 Diagnosis variables used in this method.

Source	Diagnostic variable	Class	Class description
ADNI	dx_{bl}	CN	Control normal at baseline
		LMCI	Late MCI at baseline
		AD	Early probable AD at baseline
	dx_{age}	NL	Subjects diagnosed as stable normal at current visitation
		NL to MCI	Subjects diagnosed as MCI at current visit who previously were NL
		NL to Dementia	Subjects diagnosed as dementia due to AD at current visit who previously were NL
		MCI	Subjects diagnosed as stable MCI at current visit who previously were also MCI
		MCI to Dementia	Subjects diagnosed as dementia due to AD at current visit who previously were MCI
		Dementia	Subjects diagnosed as stable dementia due to AD at current visit who previously were also MCI
Our study	dx_{last}	sHC	Subjects labelled as HC who remained like HC in all follow-up visits (who did not become MCI or AD)
		sMCI	the MCI subject who did not become AD
		cMCI	Subjects initially labelled as HC who subsequently have converted to MCI
		sAD	Subjects who remained like probable or possible AD all the follow-up visits
		cAD	Subjects labelled as HC or MCI who subsequently have converted to probable or possible AD
	dx_{csf}	normal-HC_{csf}	sHC subjects with normal CSF profile
		abnormal-HC_{csf}	sHC subjects with abnormal CSF profile
		normal-MCI_{csf}	sMCI and cMCI subjects with normal CSF profile
		abnormal-MCI_{csf}	sMCI and cMCI subjects with abnormal CSF profile
		normal-AD_{csf}	sAD and cAD subjects with normal CSF profile
		abnormal-AD_{csf}	sAD and cAD subjects with abnormal CSF profile

$$y_{ij}^r = X_{ij}\beta^r + Z_{ij}\upsilon_i^r + \varepsilon_{ij}^r \tag{4.1}$$

where, y_{ij}^r is the standardized value of ROI r measured for the i^{th} subject in the j^{th} observation; $i = 1,...,n$ subjects, $j = 1,...,n_i$ available observations for subject i and $r = 1,...,n_r$, n_r = 166 ROIs. X_{ij} is a n_i x p design matrix, where p is the number of covariates (*age*, *educ* and the constant term of 1's) on the j^{th} observation of i^{th} subject. β^r is the p x 1 vector of unknown fixed effects or regressor's coefficients, which are: β_1^r (coefficient for constant term or *Intercept*), β_a^r (coefficient for *age*) and β_e^r (coefficient for *educ*). Z_{ij} is a known design matrix of size n_i x q, where q is the number of random effects for the j^{th} observation of subject i. υ_i^r is the $q \times 1$ vector of unknown random effects coefficients $\sim N_q(0,\psi)$ for subject i measured for ROI r. ε_{ij}^r is the $n_i x 1$ residual vector of errors $\sim N_{ni}(0,\sigma^2\lambda_i)$ for the

j^{th} observation in subject i measured for ROI k. ψ is the $q \times q$ covariance matrix for the random effects. $0, \sigma^2\lambda_i$ is the n_i x n_i covariance matrix for the errors in subject i.

LME modelling was performed using the *lme*4 package available for R [8].

Identification of variant and quasi-variant ROIs.

The predictor estimates of LME models were interpreted the same way as the coefficients from a traditional regression. For instance, a one year increase in the regressor *age* corresponded to the effect of *age* (βa) increase or decrease in the outcome. Taking into account the β_a's, we classified as variant (*vr*) ROIs the ones that had both an annual change greater than 1% on ROI standard deviation and a significant change at p-values ≤ 0.05. The ROIs with annual change less than or equal to 1% were considered as quasi-variant (*qvr*), where all of these did not have a significant change.

Inference of variant ROIs y-intercept values from quasi-variant ROIs.

The Y-intercept value (y_0) from each ROI, which represents the subject-specific ROIs measure at basal stage ($age = 0$), was directly obtained from all LME models. We assumed that for healthy elderly people, the *qvr* ROIs values remains basically the same along time, even at basal stage, which is not true in the *vr*. We also assumed there is a strong correlation between the y_0 set of both ROI types. Therefore, we built the HC null model based on the PLSR algorithm [123] (see Section 2.2.4) to infer the set of y_0's for *vr* ROIs in function of the set of y_0's for *qvr* ROIs. PLSR is a linear algorithm particularly useful to analyse data with strongly collinear (correlated), noisy, and numerous predictors variables, and also simultaneously model several response variables [123]. There are several algorithms proposed to implement the PLSR model. In this study, we applied the *kernel* algorithm [25] available in the *pls* package [82] developed in R [96]. To determine the optimal number of components to take into account, it was used leave-one-out (LOO) cross-validation method available in this package. LOO calculates potential models excluding one observation at a time.

4.3.3 Early disease prediction based on residuals.

In this stage, we addressed the estimation of current diagnosis (ECD) and the prediction of future diagnosis (PFD) using data from HC, MCI and AD subjects unused in the previous stage. ROIs of these subjects were standardized according to the mean and standard deviation of the null model cohort.

Inference of age-associated ROIs values for new subjects.

For each subject, the y_0's of *vr* ROIs were inferred from the y_0's of *qvr* ROIs using the same PLSR model described in previous section. Once the y_0 set for both ROIs types was completed, we used these values in combination with the vector of coefficients β, see Eq (4.1) and the observed covariates (*age* and *educ*), to infer the expected values of each ROI ($\hat{y}_{ij}$) according to Eq (4.2). Here, the inferred value represents the value that should measure the ROI at $age \neq 0$ whether the subject is healthy or not.

$$\hat{y}^r_{ij} = \hat{y0}^r_i + (X_{ij}\beta^r) \tag{4.2}$$

where $\hat{y}^r_{ij}$ represents the inferred or predicted value of the r^{th} variant ROI on the j^{th} observation for the i^{th} subject. X_{ij} is the design matrix with observed covariates at $age \neq 0$ without the constant term. β^r is the vector of calculated fixed effects of ROI r obtained from its LME model, according Eq (4.1).

Computing the residuals of variant ROIs.

In order to get a meaningful deviation value of ROIs from an inferred healthy subject-specific trend, the difference between the estimated ($\hat{y}$) and the true ($\hat{y}$) ROIs values, here called residual (e), was computed. The residuals e^r_{ij} for each ROI were calculated as following:

$$e^r_{ij} = y^r_{ij} - \hat{y}^r_{ij} \tag{4.3}$$

where y^r_{ij} and $\hat{y}^r_{ij}$ are the observed and inferred values, respectively, for each ROI r measured on the i^{th} subject in the j^{th} observation. The e's for all ROIs were stored in a matrix **E**.

Application of proposed method in a hypothetical example

Figure 4.2 illustrates a hypothetical example of how we have used the LME and PLSR approaches to infer the ROI values at basal stage and over time; and then to infer the residuals. The figure shows an example of LME-based trajectories for hypothetical variant and quasi-variant ROIs fitted on healthy elderly data. In each plot, P_1, P_2 and P_3 represent hypothetical observations of each ROI y for two subjects at three different ages (a_1, a_2 and a_3). The first subject is assumed as HC and the second subject is assumed as AD, and it is

assumed that neither subject was used to build the models. The black lines represent the healthy population regression line calculated for each ROI, where $\hat{y}_0$ represents the vertical y-intercept value of healthy population. The blue and red lines represent the individual regression lines estimated for both subjects by assuming both as healthy; and the points $\hat{P}_1$, $\hat{P}_2$ and $\hat{P}_3$ represent the inferred $\hat{y}$'s for the three ages. Observe that, $\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$ are the subject-specific y-intercepts estimated for HC and AD subjects, respectively. For both cases, $\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$ of *vr* ROI are inferred from the $\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$ of *qvr* ROI using the PLSR model (as described above). The slope β_a is the rate change of the standard deviation of ROI per unit of age; and this slope is the same for both estimated individual regression lines. ε_{HC1}, ε_{HC2}, ε_{HC3}, ε_{AD1}, ε_{AD2} and ε_{AD3} are the residuals of each observation with respect to the estimated individual regression lines, which are computed in general way as $y - \hat{y}$. Here, the figure shows that AD residuals are greater than HC residuals because this subject is possibly affected by further neurodegeneration.

Diagnosis prediction using SVM.

In this stage, we used the matrix of residuals **E** to address the ECD and PFD problems. The first one was focused on the subject classification, where the vector of class labels, used as the outcome, was the diagnostics at the time of the visit, dx_{age} (Table 4.2). The second problem was focused on the early disease prediction, here, we trained a classifier that predicts the future diagnosis of the subjects given their current clinical tests, i.e. what is the expected diagnostic of the subject some years after the current visit. The vector of class labels used to train this classifier was dx_{last} (Table 4.2). In this case, as the feature set used was obtained in previous visits of the subject, the classifier learns to predict the future outcome of a subject, given his present state.

For both problems, we performed three experiments focused on the binary classification problems: (1) HC vs AD, (2) HC vs MCI and (3) MCI vs AD. The MCI vs AD experiment in the early prediction problem was focused on addressing the problem of prodromal stage of AD, by comparing the sMCI (stable MCI over all visitations) with the subject initially diagnosed as MCI who became to AD over time (cAD). Likewise, for each problem, two configurations of the feature set have been tested. Let E_v be the subset of matrix E where only the columns of either the *vr* ROIs or the *qvr* ROIs whose residuals are different from zero were included. The first feature set used in each experiment, F_1, included the residuals from matrix E_v plus the *age* (age of the subject at the time of the visit). The second feature set used, F_2, includes the information in F_1 together with the results from the CDR global score (CDRGLOBAL) and MMSE tests. The goal of testing two training sets was to assess

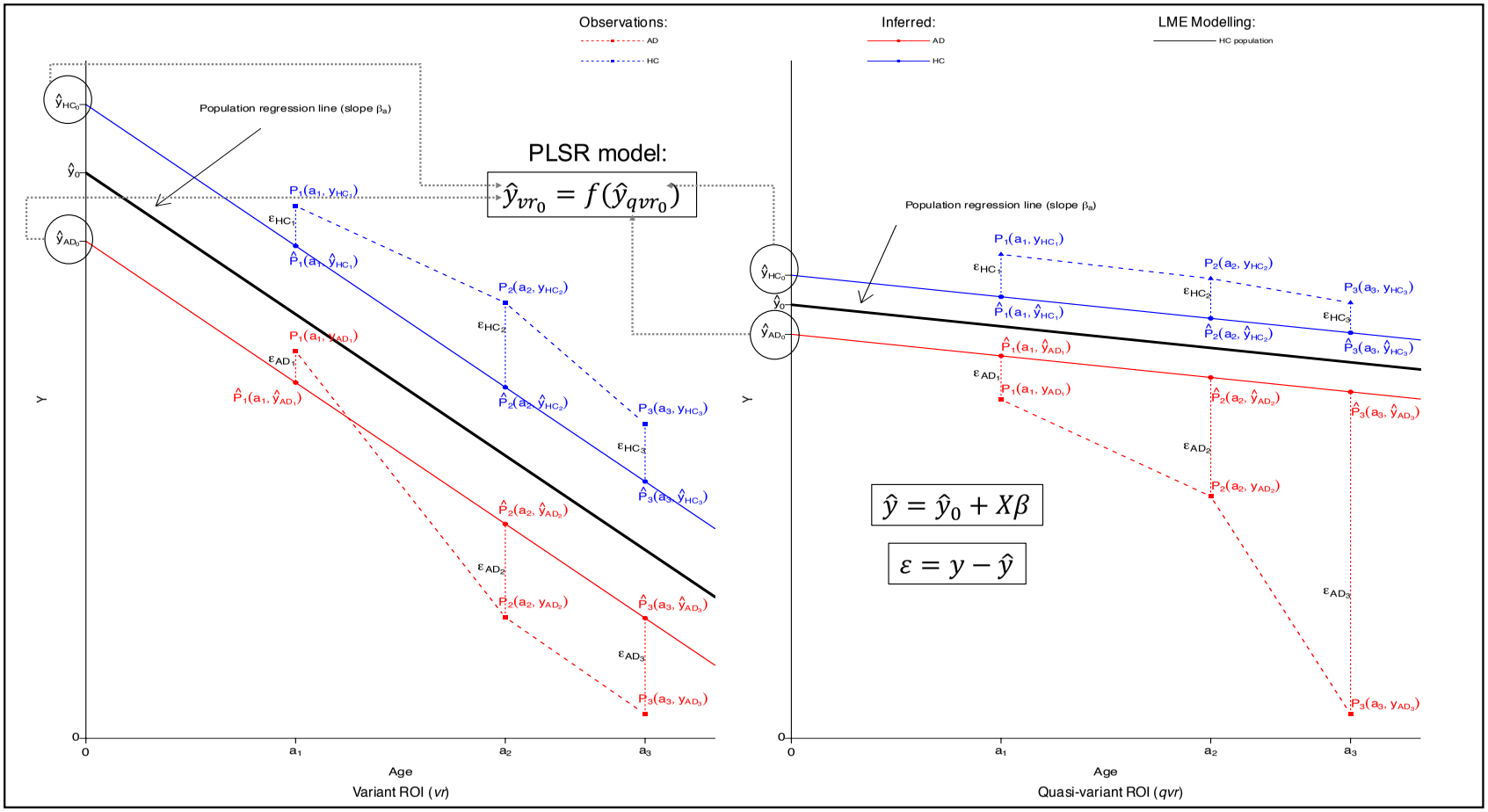


Fig. 4.2 Example of LME modelling for hypothetical variant (*vr*) and quasi-variant(*qvr*) ROIs.
HC and AD are hypothetical subjects. P_1, P_2 and P_3 are observations of each ROI y at three different ages (a_1, a_2 and a_3). Black lines, healthy population regression line calculated from LME. $\hat{y}_0$, vertical y-intercept value of healthy population. Blue and red lines, individual regression lines estimated by assuming both as healthy. Points $\hat{P}_1$, $\hat{P}_2$ and $\hat{P}_3$, inferred $\hat{y}$'s for the three ages.$\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$, the subject-specific y-intercepts estimated for HC and AD subjects, respectively. $\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$ of *vr* ROI are inferred from the $\hat{y}_{HC_0}$ and $\hat{y}_{AD_0}$ of *qvr* ROI through PLSR model. β_a, slope or rate change of the standard deviation of ROI per unit of age. ε_{HC1}, ε_{HC2}, ε_{HC3}, ε_{AD1}, ε_{AD2} and ε_{AD3}, the residuals of each observation with respect to the estimated individual regression lines.

the influence of the neuro-psychological tests in the quality of diagnostic as opposed to the use of ROI residuals with age only.

The classification method used to carry out each experiment is a SVM, configured with a Gaussian radial kernel and a misclassification cost parameter $C = 1$ (see Section 2.2.5). Each experiment has been evaluated using a ten-fold cross validation with a specific constraint: as ADNI is a longitudinal database, there may be multiple samples per subject (multiple visits). The subjects were assigned to either the training or the test subsets, in order to perform a fair evaluation. On each cross validation run, 60% of the subjects were assigned to the training subset and the remaining 40% to the test subset.

Because, it is known that the rate of atrophy increases with age, SVM classifiers were built on age groups to address the EFD problem. Furthermore, since previous studies have suggested that there are gender differences in brain atrophy with aging [17, 127, 68, 64, 104], the full workflow in building the null models and training the SVM models was applied separately for each gender.

4.4 Results

The ADNI data used to evaluate the classification for males and females are summarized in Tables 4.3 and 4.4, respectively. These tables provide demographic information and the number of measurements and diagnoses by each age group.

4.4.1 Healthy elderly subjects with normal CSF profile

Analysis of CSF-Aβ and CSF-τ biomarkers showed us that their distributions were not normal for all diagnosis groups, see figure 4.3. A bimodal distribution was observed in CSF-Aβ levels within each diagnosis group. CSF-τ values in CN group showed a unimodal normal distribution. We confirmed cut-off values determined by [108], classifying as normal the CSF profile of subjects who fulfilled both conditions: CSF-Aβ $>= 192pg/ml$ and CSF-$\tau \leq 93pg/ml$. Several subjects with abnormal CSF profile classified as CN by ADNI at baseline clinical assessment were converted to MCI or AD (triangle dots) along time. Likewise, some subjects classified like 'LMCI' by ADNI were converted to AD later (open circles).

From CSF profiles, we identified 46 normal-HC_{csf}, 33 normal-MCI_{csf}, 11 normal-AD_{csf}, 63 abnormal-HC_{csf}, 75 abnormal-MCI_{csf} and 172 non-AD_{csf} subjects. Figure 4.4 shows CSF-τ concentration versus CSF-Aβ concentration for these six groups, where the blue dots represent the HC subjects with normal CSF profile used to build the null models for

Table 4.3 Statistical descriptors of males used to build the SVM.

	N	55–59	60–64	65–69	70–74	75–79	80–85	85–89	90+
		$N = 23$	$N = 86$	$N = 152$	$N = 459$	$N = 554$	$N = 420$	$N = 209$	$N = 5$
age	1908	57.71 58.67 59.37	62.38 63.47 64.28	66.26 67.90 69.17	71.81 73.03 74.02	76.28 77.41 78.60	80.95 82.18 83.61	85.83 86.78 87.82	90.21 90.35 90.61
		(58.46± 1.19)	(63.19± 1.32)	(67.73± 1.56)	(72.88± 1.35)	(77.44± 1.41)	(82.29± 1.49)	(86.94± 1.33)	(90.40± 0.26)
dx_{csf}	1003								
normal-MCI_{csf}		25.0% $\frac{3}{12}$	30.9% $\frac{13}{42}$	14.3% $\frac{12}{84}$	9.4% $\frac{21}{224}$	4.1% $\frac{13}{315}$	10.2% $\frac{21}{205}$	17.2% $\frac{20}{116}$	0.0% $\frac{0}{5}$
normal-AD_{csf}		0.0% $\frac{0}{12}$	0.0% $\frac{0}{42}$	0.0% $\frac{0}{84}$	4.9% $\frac{11}{224}$	3.2% $\frac{10}{315}$	3.4% $\frac{7}{205}$	5.2% $\frac{6}{116}$	0.0% $\frac{0}{5}$
abnormal-HC_{csf}		0.0% $\frac{0}{12}$	9.5% $\frac{4}{42}$	7.1% $\frac{6}{84}$	21.9% $\frac{49}{224}$	22.9% $\frac{72}{315}$	11.7% $\frac{24}{205}$	14.7% $\frac{17}{116}$	40.0% $\frac{2}{5}$
abnormal-MCI_{csf}		0.0% $\frac{0}{12}$	2.4% $\frac{1}{42}$	23.8% $\frac{20}{84}$	21.0% $\frac{47}{224}$	30.8% $\frac{97}{315}$	24.4% $\frac{50}{205}$	19.0% $\frac{22}{116}$	0.0% $\frac{0}{5}$
abnormal-AD_{csf}		75.0% $\frac{9}{12}$	57.1% $\frac{24}{42}$	54.8% $\frac{46}{84}$	42.9% $\frac{96}{224}$	39.0% $\frac{123}{315}$	50.2% $\frac{103}{205}$	44.0% $\frac{51}{116}$	60.0% $\frac{3}{5}$
dx_{last}	1908								
AD		60.9% $\frac{14}{23}$	43.0% $\frac{37}{86}$	59.2% $\frac{90}{152}$	44.9% $\frac{206}{459}$	39.9% $\frac{221}{554}$	46.2% $\frac{194}{420}$	41.6% $\frac{87}{209}$	60.0% $\frac{3}{5}$
HC		4.3% $\frac{1}{23}$	12.8% $\frac{11}{86}$	9.2% $\frac{14}{152}$	27.2% $\frac{125}{459}$	29.8% $\frac{165}{554}$	16.2% $\frac{68}{420}$	20.1% $\frac{42}{209}$	40.0% $\frac{2}{5}$
MCI		34.8% $\frac{8}{23}$	44.2% $\frac{38}{86}$	31.6% $\frac{48}{152}$	27.9% $\frac{128}{459}$	30.3% $\frac{168}{554}$	37.6% $\frac{158}{420}$	38.3% $\frac{80}{209}$	0.0% $\frac{0}{5}$
dx_{age}	1908								
AD		30.4% $\frac{7}{23}$	33.7% $\frac{29}{86}$	34.9% $\frac{53}{152}$	24.6% $\frac{113}{459}$	26.9% $\frac{149}{554}$	29.1% $\frac{122}{420}$	27.3% $\frac{57}{209}$	60.0% $\frac{3}{5}$
HC		4.3% $\frac{1}{23}$	11.6% $\frac{10}{86}$	9.2% $\frac{14}{152}$	27.2% $\frac{125}{459}$	30.9% $\frac{171}{554}$	19.3% $\frac{81}{420}$	21.5% $\frac{45}{209}$	40.0% $\frac{2}{5}$
MCI		65.2% $\frac{15}{23}$	54.6% $\frac{47}{86}$	55.9% $\frac{85}{152}$	48.1% $\frac{221}{459}$	42.2% $\frac{234}{554}$	51.7% $\frac{217}{420}$	51.2% $\frac{107}{209}$	0.0% $\frac{0}{5}$
APOE-ε4	1908								
0		69.6% $\frac{16}{23}$	54.6% $\frac{47}{86}$	28.3% $\frac{43}{152}$	47.7% $\frac{219}{459}$	48.6% $\frac{269}{554}$	53.1% $\frac{223}{420}$	56.9% $\frac{119}{209}$	40.0% $\frac{2}{5}$
1		13.0% $\frac{3}{23}$	38.4% $\frac{33}{86}$	44.1% $\frac{67}{152}$	37.7% $\frac{173}{459}$	41.2% $\frac{228}{554}$	40.0% $\frac{168}{420}$	38.8% $\frac{81}{209}$	40.0% $\frac{2}{5}$
2		17.4% $\frac{4}{23}$	7.0% $\frac{6}{86}$	27.6% $\frac{42}{152}$	14.6% $\frac{67}{459}$	10.3% $\frac{57}{554}$	6.9% $\frac{29}{420}$	4.3% $\frac{9}{209}$	20.0% $\frac{1}{5}$
MMSE	1906	25.0 26.0 28.5	24.0 27.0 29.0	23.0 26.0 29.0	25.0 27.0 29.0	25.0 28.0 29.0	24.0 27.0 29.0	24.0 27.0 28.0	26.0 27.0 27.0
		(26.7± 2.3)	(26.0± 3.5)	(25.4± 4.2)	(26.4± 3.7)	(26.3± 3.9)	(26.2± 3.6)	(25.8± 3.4)	(25.2± 4.7)
CDRGLOBAL	1899	1.8 2.5 3.2	0.5 1.2 4.0	1.0 2.0 4.0	0.0 1.5 3.0	0.0 1.5 3.5	0.5 1.5 3.5	1.0 2.0 3.5	0.0 4.0 5.5
		(2.7±1.3)	(2.2±2.0)	(2.6±2.2)	(2.1±2.2)	(2.3±2.6)	(2.3±2.3)	(2.7±2.9)	(3.7±3.8)

$_a\,b\,_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. N: number of samples available at each age group, it being possible to have more of one observation per subject.

Table 4.4 Statistical descriptors of females used to build the SVM

	N	55–59	60–64	65–69	70–74	75–79	80–85	85–89	90+
		$N = 37$	$N = 62$	$N = 117$	$N = 338$	$N = 355$	$N = 287$	$N = 101$	$N = 5$
age	1302	56.6 57.5 58.5	61.5 62.6 64.0	66.4 68.0 69.2	71.6 72.8 73.9	76.3 77.5 78.6	81.0 82.2 83.5	85.6 86.5 87.4	90.9 91.4 92.8
		(57.6± 1.3)	(62.6± 1.4)	(67.8± 1.5)	(72.7± 1.4)	(77.5± 1.4)	(82.3± 1.4)	(86.6± 1.2)	(91.6± 1.2)
dx_{csf}	640								
normal-MCI_{csf}		0.0% $\frac{0}{29}$	29.7% $\frac{11}{37}$	16.0% $\frac{12}{75}$	3.0% $\frac{5}{167}$	2.5% $\frac{4}{161}$	12.1% $\frac{16}{132}$	10.5% $\frac{4}{38}$	0.0% $\frac{0}{1}$
normal-AD_{csf}		0.0% $\frac{0}{29}$	0.0% $\frac{0}{37}$	0.0% $\frac{0}{75}$	0.0% $\frac{0}{167}$	2.5% $\frac{4}{161}$	3.8% $\frac{5}{132}$	15.8% $\frac{6}{38}$	100.0% $\frac{1}{1}$
abnormal-HC_{csf}		0.0% $\frac{0}{29}$	0.0% $\frac{0}{37}$	1.3% $\frac{1}{75}$	22.8% $\frac{38}{167}$	33.5% $\frac{54}{161}$	25.0% $\frac{33}{132}$	26.3% $\frac{10}{38}$	0.0% $\frac{0}{1}$
abnormal-MCI_{csf}		17.2% $\frac{5}{29}$	54.0% $\frac{20}{37}$	21.3% $\frac{16}{75}$	21.6% $\frac{36}{167}$	18.6% $\frac{30}{161}$	17.4% $\frac{23}{132}$	0.0% $\frac{0}{38}$	0.0% $\frac{0}{1}$
abnormal-AD_{csf}		82.8% $\frac{24}{29}$	16.2% $\frac{6}{37}$	61.3% $\frac{46}{75}$	52.7% $\frac{88}{167}$	42.9% $\frac{69}{161}$	41.7% $\frac{55}{132}$	47.4% $\frac{18}{38}$	0.0% $\frac{0}{1}$
dx_{last}	1302								
AD		81.1% $\frac{30}{37}$	27.4% $\frac{17}{62}$	64.1% $\frac{75}{117}$	49.1% $\frac{166}{338}$	38.0% $\frac{135}{355}$	45.6% $\frac{131}{287}$	51.5% $\frac{52}{101}$	80.0% $\frac{4}{5}$
HC		0.0% $\frac{0}{37}$	0.0% $\frac{0}{62}$	5.1% $\frac{6}{117}$	30.8% $\frac{104}{338}$	34.1% $\frac{121}{355}$	24.7% $\frac{71}{287}$	28.7% $\frac{29}{101}$	20.0% $\frac{1}{5}$
MCI		18.9% $\frac{7}{37}$	72.6% $\frac{45}{62}$	30.8% $\frac{36}{117}$	20.1% $\frac{68}{338}$	27.9% $\frac{99}{355}$	29.6% $\frac{85}{287}$	19.8% $\frac{20}{101}$	0.0% $\frac{0}{5}$
dx_{age}	1302								
AD		62.16% $\frac{23}{37}$	17.74% $\frac{11}{62}$	35.90% $\frac{42}{117}$	36.39% $\frac{123}{338}$	27.04% $\frac{96}{355}$	29.97% $\frac{86}{287}$	44.55% $\frac{45}{101}$	80.00% $\frac{4}{5}$
HC		0.00% $\frac{0}{37}$	0.00% $\frac{0}{62}$	4.27% $\frac{5}{117}$	33.14% $\frac{112}{338}$	38.87% $\frac{138}{355}$	28.22% $\frac{81}{287}$	29.70% $\frac{30}{101}$	20.00% $\frac{1}{5}$
MCI		37.84% $\frac{14}{37}$	82.26% $\frac{51}{62}$	59.83% $\frac{70}{117}$	30.47% $\frac{103}{338}$	34.08% $\frac{121}{355}$	41.81% $\frac{120}{287}$	24.75% $\frac{25}{101}$	0.00% $\frac{0}{5}$
NL to Dementia		0.00% $\frac{0}{37}$	0.00% $\frac{0}{62}$	0.00% $\frac{0}{117}$	0.00% $\frac{0}{338}$	0.00% $\frac{0}{355}$	0.00% $\frac{0}{287}$	0.99% $\frac{1}{101}$	0.00-% $\frac{0}{5}$
APOE-ε4	1302								
0		59.5% $\frac{22}{37}$	43.5% $\frac{27}{62}$	17.1% $\frac{20}{117}$	35.8% $\frac{121}{338}$	51.0% $\frac{181}{355}$	62.0% $\frac{178}{287}$	81.2% $\frac{82}{101}$	100.0% $\frac{5}{5}$
1		13.5% $\frac{5}{37}$	38.7% $\frac{24}{62}$	35.9% $\frac{42}{117}$	49.7% $\frac{168}{338}$	43.9% $\frac{156}{355}$	34.8% $\frac{100}{287}$	15.8% $\frac{16}{101}$	0.0% $\frac{0}{5}$
2		27.0% $\frac{10}{37}$	17.7% $\frac{11}{62}$	47.0% $\frac{55}{117}$	14.5% $\frac{49}{338}$	5.1% $\frac{18}{355}$	3.1% $\frac{9}{287}$	3.0% $\frac{3}{101}$	0.0% $\frac{0}{5}$
MMSE	1297	20.0 23.0 25.0	25.2 28.0 28.0	24.0 26.0 28.0	24.0 27.0 29.0	24.0 28.0 29.0	23.5 27.0 29.0	21.0 26.0 29.0	24.0 25.5 26.8
		(22.2± 4.6)	(26.5± 3.5)	(25.8± 3.0)	(25.6± 4.8)	(26.2± 4.5)	(26.1± 3.5)	(24.9± 4.3)	(25.2± 3.3)
CDRGLOBAL	1289	3.0 4.0 6.0	1.0 1.5 2.0	1.5 2.0 3.5	0.0 2.0 4.0	0.0 1.0 3.5	0.5 1.5 4.0	0.0 3.0 5.8	2.8 4.2 6.0
		(4.7±2.6)	(1.9±1.9)	(2.6±2.0)	(2.4±2.5)	(2.1±2.7)	(2.4±2.5)	(3.7±3.6)	(4.5±2.5)

$_a\,b\,_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. N: number of samples available at each age group, it being possible to have more of one observation per subject.

men and women. Furthermore, in Figure 4.5 is shown the last available measurements of CSF biomarkers concentrations for the studied ADNI subjects grouped according dx_{csf}. An increase in τ values was observed when comparing normal-HC_{csf} and subjects with abnormal CSF-profile, and when comparing abnormal-HC_{csf} with MCI and AD subjects with abnormal profile. With respect to Aβ levels, we observed a reduction of levels when comparing normal-HC_{csf} with abnormal-HC_{csf} and the remaining groups with abnormal profile.

In total, 226 samples were available for the 46 normal-HC_{csf} subjects (males: 22, females: 24). Table 4.5 provides their demographic information at baseline. Age, education and MMSE scores were similar across gender, but CSF concentrations and CDRGLOBAL were different between men and women. Furthermore, only three subjects (2 males and 1 female) were carriers of APOE-ε4 at allele 1. More detail of demographic and cognitive features measured in ADNI were presented in Section 2.3.

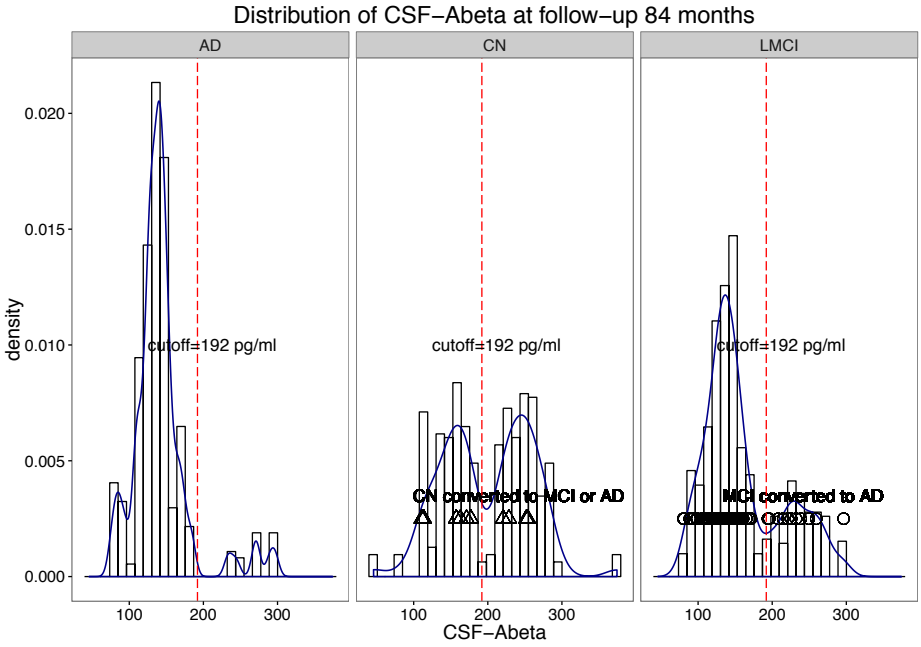
Table 4.5 Baseline statistical descriptors of HC subjects selected for null models building.

	Female	Male
	$N = 24$	$N = 22$
age	71.2 75.1 77.7 (74.7± 5.3)	70.4 71.8 74.0 (72.7± 6.0)
education	0.56 0.62 0.75 (0.67 ±0.12)	0.62 0.78 0.88 (0.77 ±0.20)
APOE-ε4		
0	95.8% $\frac{23}{24}$	90.9% $\frac{20}{22}$
1	4.2% $\frac{1}{24}$	9.1% $\frac{2}{22}$
MMSE	29.0 29.0 30.0 (29.25± 0.67)	28.0 29.0 30.0 (28.54± 1.50)
CDRGLOBAL		
0	100% $\frac{24}{24}$	82% $\frac{18}{22}$
0.5	0% $\frac{0}{24}$	18% $\frac{4}{22}$
CSF-Aβ	234.9 248.5 256.0 (247.23± 20.14)	235.0 257.9 268.2 (253.22± 21.69)
CSF-τ	47.9 55.260.9 (55.41± 15.10)	47.0 59.9 73.8 (59.84± 16.83)

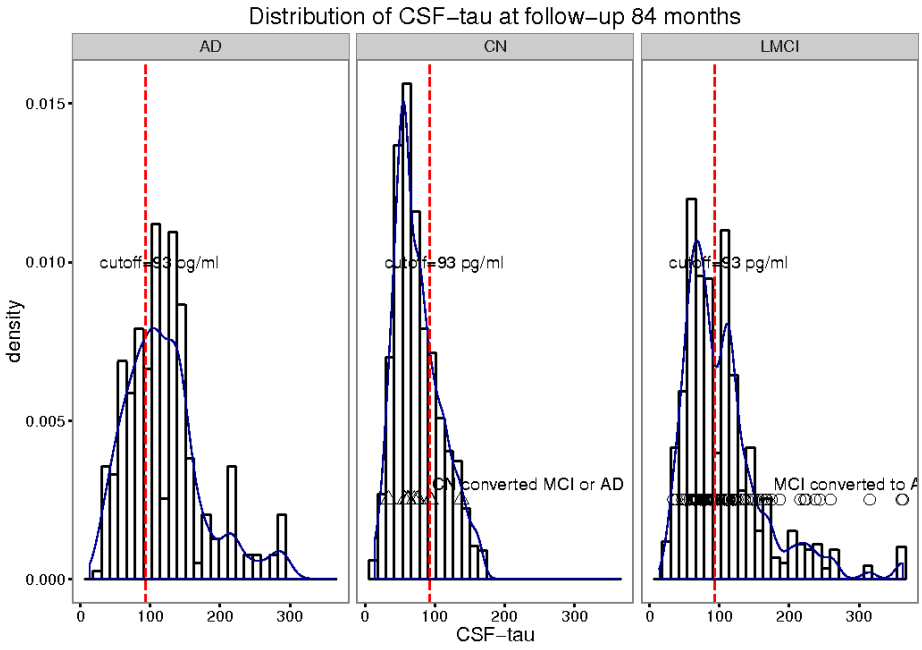
Values of continuous variables are represented by the lower, the median and the upper quartiles; and the *mean*± standard deviation in parentheses.

4.4.2 Ageing-based variant (*vr*) and quasi-variant (*qvr*) ROIs.

When studying the within-subject and between-subject variabilities using the box plots representation, we found that there are significant between-subject variability and strong gender effect for several ROIs. The left hippocampal volume is one of ROIs that showed this pattern, see Figure 4.6. However, for others regions, their change was not very evident.



(a)



(b)

Fig. 4.3 Distribution of CSF-based biomarkers at follow-up 84 months. (a) CSF-Aβ. (b) CSF-τ.
$\triangle$: 8 of 113 CN subjects were converted to MCI and 1 CN subject was converted to AD. $\bigcirc$: 88 of 94 LMCI were converted to AD and 5 were re-converted to HC at follow-up visits. Dotted vertical lines within each diagnosis are the computed cutoff concentrations. CN: Control normal subjects at baseline labelled as such by ADNI. LMCI: Late MCI subjects at baseline labelled as such by ADNI.

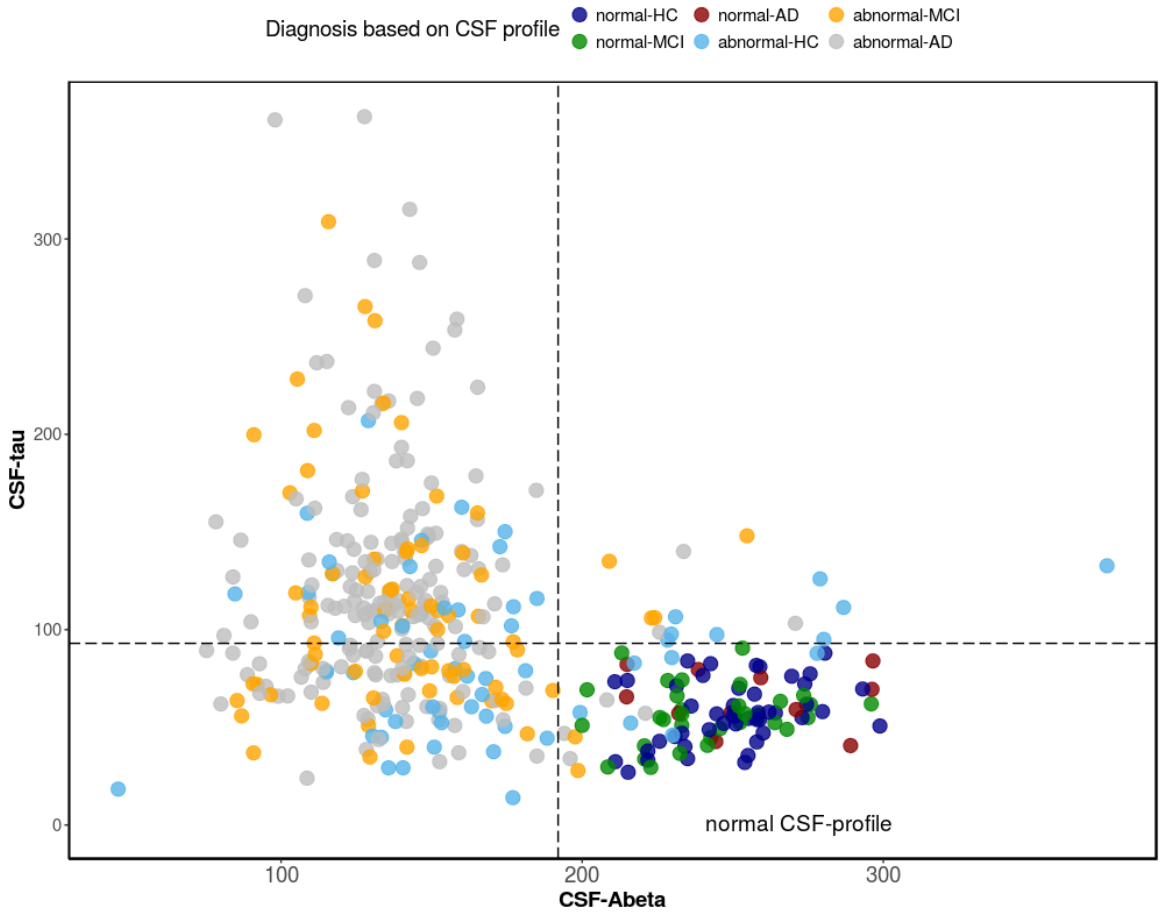
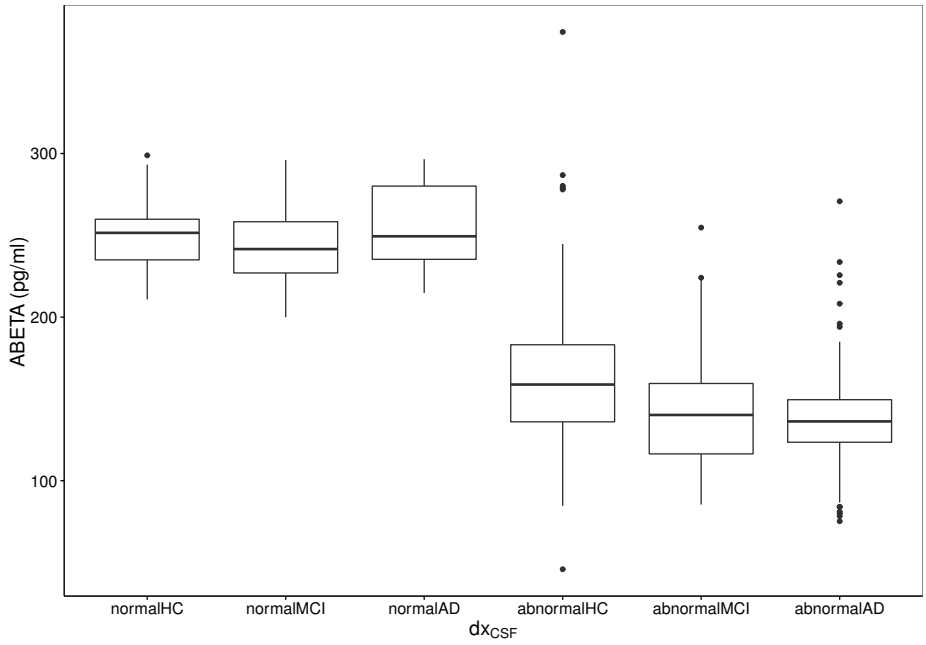
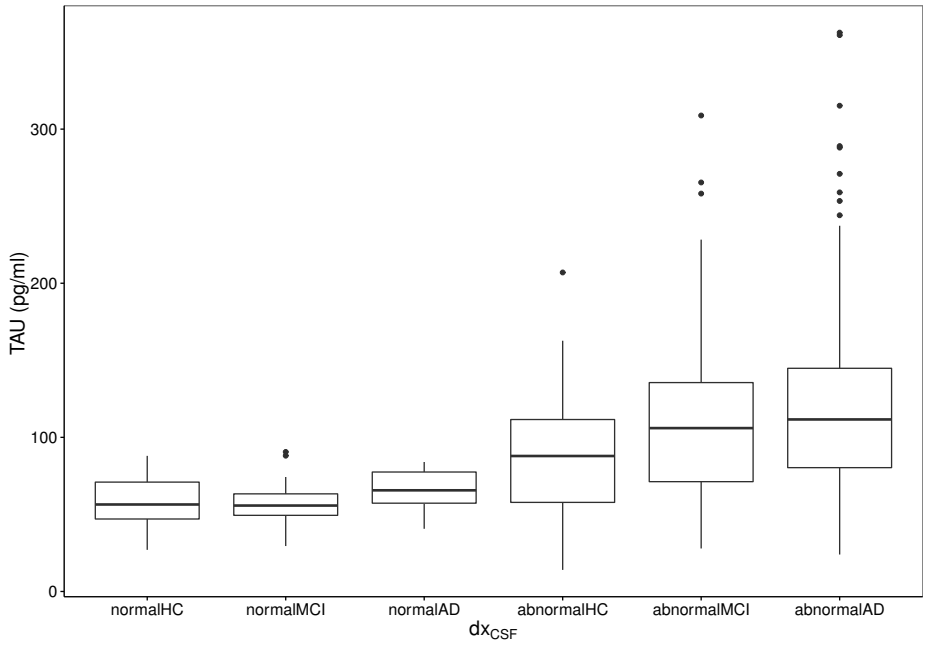


Fig. 4.4 CSF-Aβ vs. CSF-τ concentration available at last subject's observations. Dots represent the last CSF biomarker measured for subjects available at April, 2015. Vertical and horizontal dashed lines split normal CSF-profile from abnormal profile. Null models for characterization of healthy brain structures were built from samples labelled with blue dots.



(a)



(b)

Fig. 4.5 Boxplot of CSF biomarkers concentrations for dx_{csf} diagnostic groups. (a) CSF-$A\beta 1-42$; and (b) CSF-τ.

From LME modelling, we identified 97 *vr* ROIs for males and 109 *vr* ROIs for females. Regarding quasi-variant ROIs, we found 69 *qvr* for males and 57 *qvr* for females. Figure 4.7 shows examples of fitted LME models for both types of ROIs within each gender. In a similar way to these examples, the β coefficients of age (β_a) of the remaining *vr* ROIs were observed as being close to zero, but for *qvr* ROIs, their slopes were not. Figures 4.8 and 4.9 show the summary of ageing-associated differences of biomarkers in control men and women, respectively. Biomarkers are coloured according their type of change: blue for *vr* regions and red for the *qvr* ones. The size of biomarkers with significant *P*-value ≤ 0.05 are bigger than the not significant. Dotted vertical lines separate the increased biomarkers (positive beta values) from the reduced ones (negative beta values). ROIs are represented according to their location into brain hemisphere (lh:left, rh: right or bilateral). The β_a values of TA, CV and SV are presented in measurements units of mm, mm^3 and mm^3 , respectively. These values mean the average population change observed per year in each biomarker. Results suggest important ageing-related reductions in neocortical and subcortical regions and ventricular expansion, with several gender-specific significances. Reductions are observed in the most of cortical volumes and cortical thickness, where men and women showed a similar degree of global thinning. Some of these regions showed prominent atrophy while others showed a more conservative change. For example, the entorhinal volume was observed to be significantly reduced in females but not in males. In relation to the subcortical regions, we also observed few gender-specific differences. In summary, both gender had significant volumetric reductions in both hemispheres of hippocampal volume and significant ventricular expansion; the volume of the optic chiasm region was significantly increased in females, but it was not significant in males; a reduction in amygdala volume was significant in both hemispheres of the male brain but it was not significant in females; and, in general, the volume of bilateral corpus callosum regions are observed as thinned, but without significant differences between the genders. Tables 4.6 and 4.7 summarize the cortical and subcortical *vr* biomarkers for males, respectively; and Tables 4.8 and 4.9 summarize the cortical and subcortical *vr* biomarkers for females, respectively. The last column in all these tables (beta_*AGE_IMG*) shows the β_a's and its measurements units correspond to mm for TA biomarkers, mm^3 for CV biomarkers and mm^3 for SV.

4.4.3 Estimation of current diagnosis based on ROI residuals

Table 4.10 shows the performance of models addressing the ECD problem on each gender. These results include the average prediction accuracy (ACC), sensitivity (SEN) and specificity (SPE) of the test for both the F_1 and F_2 training set configurations. Although the performance

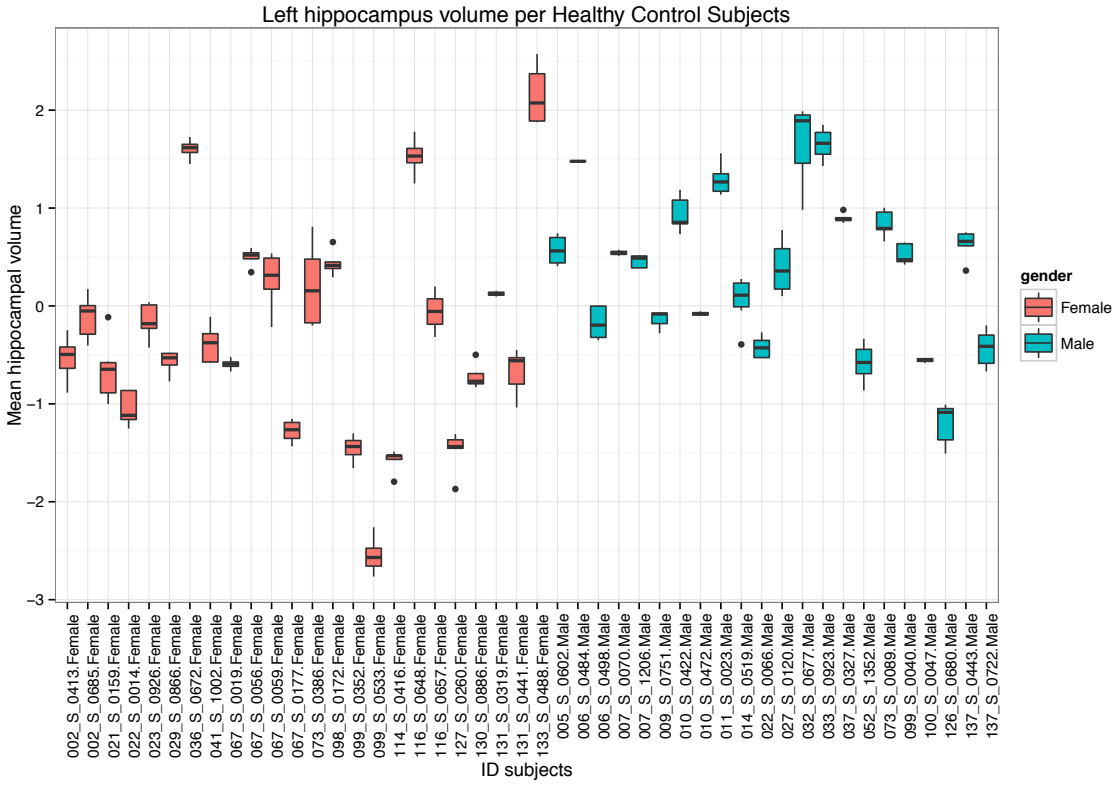
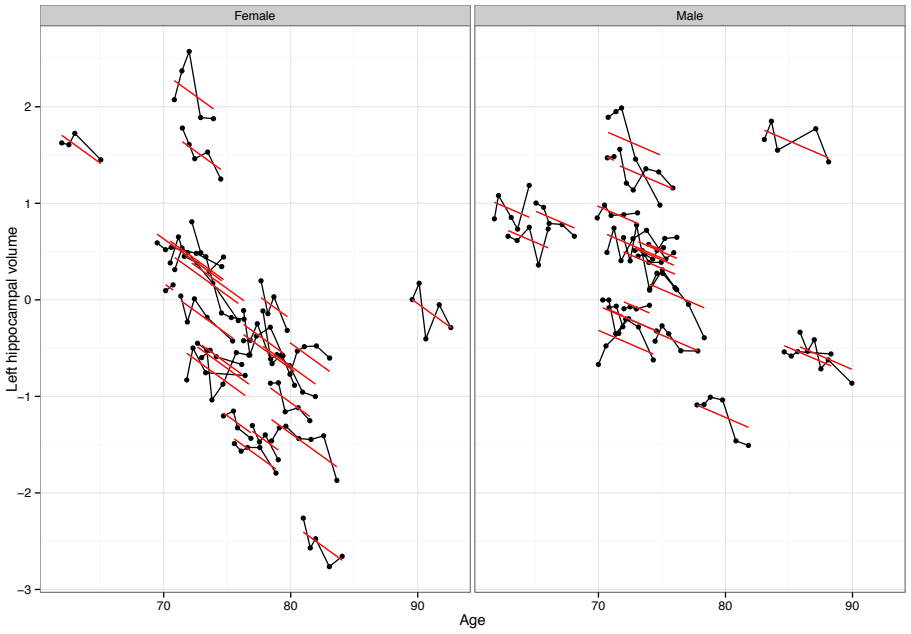
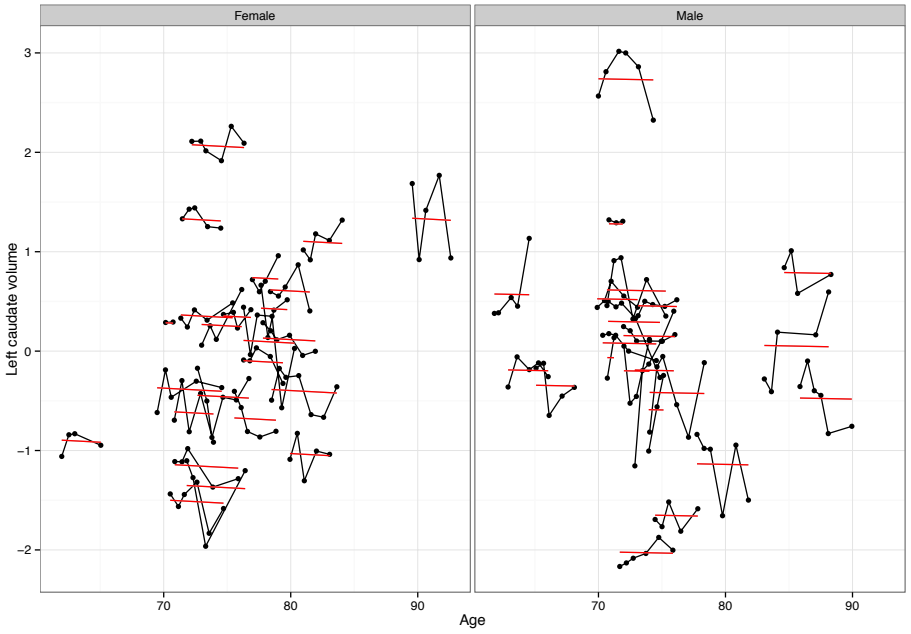


Fig. 4.6 Boxplot of trajectory of left hippocampal volume for normal-HC_{csf} subjects. High between-subject variability is evident, e.g., by comparing subject 099_*S*_0533 with subject 133_*S*_0488. Likewise, there is a strong indication of gender effect over hippocampal volume, female volumes are less than the man ones. Note that we standardized every MRI-based biomarker to have zero mean and unit variance.



(a)



(b)

Fig. 4.7 Examples of variant and quasi-variant ROIs per normal-HC_{csf} subjects stratified by gender. (a) Left hippocampal volume classified as variant (*vr*) ROI. (b) Left caudate volume classified as quasi-variant (*qvr*) ROI.

Note that for the *vr* ROI, the slope of trajectories is not close to zero, in contrast, for the *qvr* ROI, the slope of trajectories is close to zero. Furthermore, both regions, the y-intercept values vary between subjects, but the slope value of each ROI is the same for all subjects.

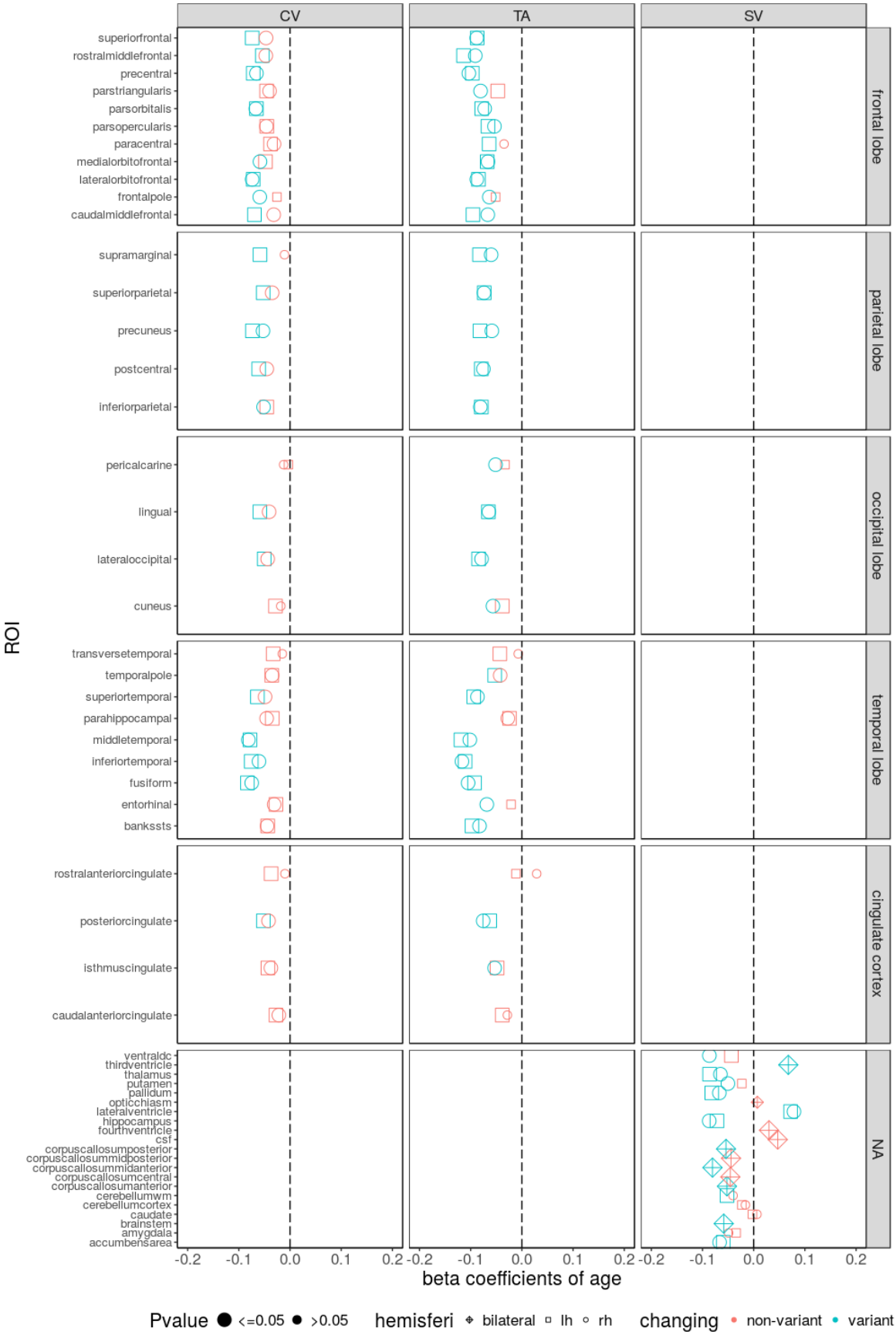


Fig. 4.8 Characterization of ageing-based variant (*vr*) and quasi-variant (*qvr*) ROIs in control males

SV: Subcortical Volume, CV: Cortical volume, SA: Surface Area. lh: left hemisphere. rh: right hemisphere.

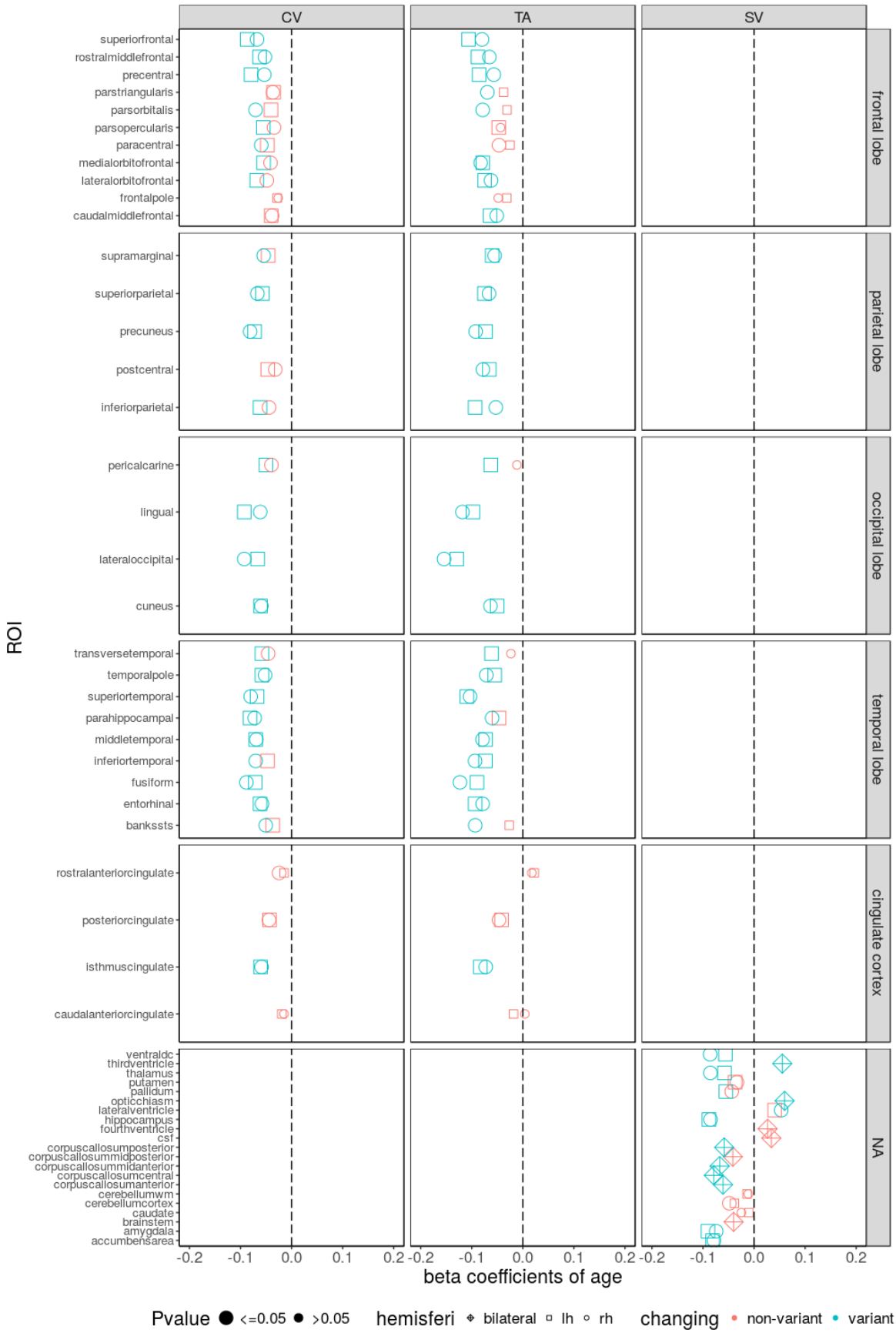


Fig. 4.9 Characterization of ageing-based variant (*vr*) and quasi-variant (*qvr*) ROIs in control females

SV: Subcortical Volume, CV: Cortical volume, SA: Surface Area. lh: left hemisphere. rh: right hemisphere.

Table 4.6 Summary of variant cortical MRI-based biomarkers for males.

	ROI	hemisferi	measuretype	Lobe	β_a
1	lateraloccipital	lh	CV	occipital lobe	-0.050
2	pericalcarine	rh	TA	occipital lobe	-0.051
3	inferiorparietal	rh	CV	parietal lobe	-0.052
4	posteriorcingulate	lh	CV	cingulate cortex	-0.052
5	superiorparietal	lh	CV	parietal lobe	-0.052
6	precuneus	rh	CV	parietal lobe	-0.053
7	temporalpole	lh	TA	temporal lobe	-0.053
8	isthmuscingulate	rh	TA	cingulate cortex	-0.053
9	parsopercularis	rh	TA	frontal lobe	-0.053
10	rostralmiddlefrontal	lh	CV	frontal lobe	-0.054
11	cuneus	rh	TA	occipital lobe	-0.056
12	precuneus	rh	TA	parietal lobe	-0.059
13	supramarginal	lh	CV	parietal lobe	-0.059
14	medialorbitofrontal	rh	CV	frontal lobe	-0.059
15	lingual	lh	CV	occipital lobe	-0.059
16	frontalpole	rh	CV	frontal lobe	-0.059
17	supramarginal	rh	TA	parietal lobe	-0.060
18	inferiortemporal	rh	CV	temporal lobe	-0.061
19	postcentral	lh	CV	parietal lobe	-0.061
20	posteriorcingulate	lh	TA	cingulate cortex	-0.063
21	frontalpole	rh	TA	frontal lobe	-0.064
22	superiortemporal	lh	CV	temporal lobe	-0.064
23	paracentral	lh	TA	frontal lobe	-0.064
24	lingual	rh	TA	occipital lobe	-0.064
25	lingual	lh	TA	occipital lobe	-0.065
26	precentral	rh	CV	frontal lobe	-0.066
27	parsorbitalis	lh	CV	frontal lobe	-0.066
28	medialorbitofrontal	rh	TA	frontal lobe	-0.066
29	parsopercularis	lh	TA	frontal lobe	-0.066
30	caudalmiddlefrontal	rh	TA	frontal lobe	-0.067
31	parsorbitalis	rh	CV	frontal lobe	-0.067
32	medialorbitofrontal	lh	TA	frontal lobe	-0.068
33	entorhinal	rh	TA	temporal lobe	-0.068
34	caudalmiddlefrontal	lh	CV	frontal lobe	-0.070
35	precentral	lh	CV	frontal lobe	-0.072
36	lateralorbitofrontal	lh	CV	frontal lobe	-0.072
37	parsorbitalis	rh	TA	frontal lobe	-0.073
38	precuneus	lh	CV	parietal lobe	-0.073
39	superiorparietal	lh	TA	parietal lobe	-0.074
40	superiorfrontal	lh	CV	frontal lobe	-0.074
41	superiorparietal	rh	TA	parietal lobe	-0.074
42	lateralorbitofrontal	rh	CV	frontal lobe	-0.074
43	fusiform	rh	CV	temporal lobe	-0.075
44	postcentral	rh	TA	parietal lobe	-0.075
45	posteriorcingulate	rh	TA	cingulate cortex	-0.076
46	inferiortemporal	lh	CV	temporal lobe	-0.076
47	middletemporal	lh	CV	temporal lobe	-0.078
48	parsorbitalis	lh	TA	frontal lobe	-0.078
49	lateraloccipital	rh	TA	occipital lobe	-0.079
50	inferiorparietal	lh	TA	parietal lobe	-0.079
51	postcentral	lh	TA	parietal lobe	-0.079
52	parstriangularis	rh	TA	frontal lobe	-0.081
53	inferiorparietal	rh	TA	parietal lobe	-0.081
54	middletemporal	rh	CV	temporal lobe	-0.081
55	precuneus	lh	TA	parietal lobe	-0.082
56	supramarginal	lh	TA	parietal lobe	-0.083
57	bankssts	rh	TA	temporal lobe	-0.083
58	fusiform	lh	CV	temporal lobe	-0.083
59	lateraloccipital	lh	TA	occipital lobe	-0.084
60	lateralorbitofrontal	lh	TA	frontal lobe	-0.085
61	superiortemporal	rh	TA	temporal lobe	-0.087
62	superiorfrontal	lh	TA	frontal lobe	-0.087
63	lateralorbitofrontal	rh	TA	frontal lobe	-0.088
64	superiorfrontal	rh	TA	frontal lobe	-0.088
65	rostralmiddlefrontal	rh	TA	frontal lobe	-0.091
66	fusiform	lh	TA	temporal lobe	-0.092
67	superiortemporal	lh	TA	temporal lobe	-0.094
68	caudalmiddlefrontal	lh	TA	frontal lobe	-0.095
69	precentral	lh	TA	frontal lobe	-0.097
70	bankssts	lh	TA	temporal lobe	-0.097
71	middletemporal	rh	TA	temporal lobe	-0.102
72	precentral	rh	TA	frontal lobe	-0.103
73	fusiform	rh	TA	temporal lobe	-0.105
74	inferiortemporal	lh	TA	temporal lobe	-0.111
75	rostralmiddlefrontal	lh	TA	frontal lobe	-0.114
76	inferiortemporal	rh	TA	temporal lobe	-0.117
77	middletemporal	lh	TA	temporal lobe	-0.119

Table 4.7 Summary of variant subcortical MRI-based biomarkers for males.

	ROI	hemisferi	measuretype	Lobe	β_a
1	lateralventricle	rh	SV		0.079
2	lateralventricle	lh	SV		0.072
3	thirdventricle	bilateral	SV		0.068
4	putamen	rh	SV		-0.050
5	cerebellumwm	lh	SV		-0.052
6	corpuscallosumanterior	bilateral	SV		-0.052
7	corpuscallosumposterior	bilateral	SV		-0.054
8	brainstem	bilateral	SV		-0.058
9	accumbensarea	lh	SV		-0.059
10	thalamus	rh	SV		-0.065
11	accumbensarea	rh	SV		-0.067
12	pallidum	rh	SV		-0.067
13	hippocampus	lh	SV		-0.072
14	corpuscallosummidanterior	bilateral	SV		-0.080
15	pallidum	lh	SV		-0.082
16	thalamus	lh	SV		-0.086
17	ventraldc	rh	SV		-0.087
18	hippocampus	rh	SV		-0.087

of models with the F_2 set was the best for all cases, the experiments performed on the F_1 set had also good results.

4.4.4 Prediction of future diagnosis based on ROI residuals

Table 4.11 shows the performance of models built for addressing the PFD problem. As in previous results, although models built with the F_2 configuration have obtained slightly better accuracies than the ones built on the F_1 configuration, both performances are very satisfactory.

Table 4.12 shows the average number of years this method leads in predicting the subject diagnosis. This advancement in the prediction was only possible in subjects whose dx_{age} (diagnostic at each visitation) was different from their dx_{last} diagnosis (last known diagnostic). Therefore, only these subjects were taken into account and some age groups lack enough data to be shown in this table. The prediction of conversion from HC to AD in females was up to 1.64 years earlier (75–79 age group); and for males, up to 1.73 years earlier (80–84 age group). However, the greatest lead was obtained in the early prediction from MCI to AD in males (80–84 age group), this being 1.85 years earlier. Along with the time advancement in the prediction, Table 4.12 also shows the average prediction accuracy for that differential diagnostic stratified by gender and age group. Accuracies where computed on few samples, namely those that fit the conditions of five-year age group stratification and whose last known diagnostic differ from the diagnostic at the time of the test in the given age group.

4.5 Discussion and conclusions

This chapter has presented the M_{res} method proposed for building ageing-associated null models from longitudinal MRI-based biomarkers, as well as to address both the ECD and PFD problems from residuals calculated through these null models.

Analysis of CSF biomarkers distributions allowed us to confirm the cut-offs values established by [108] and to establish criteria for CSF profiling. We found subjects clinically diagnosed as CN at baseline with abnormal CSF profile, which is consistent with previous studies where the presence of possible AD pathology has been found in ADNI control subjects [108, 116, 27]. These findings may be because ADNI subject diagnosis is made independently of the CSF biomarkers values (see Section 2.3).

We modelled the longitudinal change of an extensive set of MRI-based biomarkers obtained from cortical and subcortical regions. Studies have shown that early diagnosis methods using the whole brain or the whole cortex reached higher specificity (over 90%) than those based on the specific regions like the hippocampus (from 63% to 84%) [15, 19, 23]. LME modelling allowed to classify biomarkers as variant or quasi-variant ROIs, and to build null models for ageing-related changes in men and women. Results suggest important ageing-related reductions in neocortical and subcortical regions and ventricular expansion. In general, few gender-specific differences were observed. The entorhinal volume was observed to be significantly reduced in females but not in males,but in general men and women showed a similar degree of global thinning. As in previous studies [104, 43–45, 64], these results confirm that changing cortical biomarkers in elderly people follow a fronto-temporal pattern and much of this change occur in brain areas related with AD. In relation to the subcortical regions, we also observed few gender-specific differences. For both genders, the hippocampus was the region most reduced in volume, also the most significant expansion was observed in the ventricular system. Similar changes in both the hippocampus and the ventricles in elderly people have been previously reported by [43, 64, 44, 45]

Null models were carried out by making assumptions of correlation between the y-intercepts values (at basal stage) of variant and quasi-variant ROIs, where the first ones were explained in function of last ones using PLSR. By using these null models and LME β coefficients, we calculated residuals, which were established as differences between the observed ROIs and age-related inferred ROIs. These residuals were computed for new cohort of HC, MCI and subjects; and used for training and testing SVM models to address subject classification and early disease prediction.

As of the date of this study, we were unable to find studies where ageing-related null models and residuals-based classifiers were applied to early diagnosis. The performance obtained in all experiments suggests that the proposed method of obtaining the ROI residuals

and their use to train SVM predictors is useful to support the early diagnosis problem, the fundamental challenge in AD research. LME modelling of MRI-based biomarkers was only applied by fitting age and years of education, but future work could assess the impact of different available feature types such as other functional neuroimaging biomarkers, genetic factors, biological markers and other clinical and neuropsychological assessment.

The main contributions of the residual-based classifier presented in this chapter are: (a) From a longitudinal study of 5 years follow-up, the ability to predict the future diagnostic of the subjects up to 1.85 years earlier than the standard clinical procedure. (b) Use of relatively common clinical tests such as MRI and neuropsychological tests, as opposed to methods that rely on more expensive or invasive tests such as PET-based, CSF-based and Genotype-based biomarkers. (c) In the ECD for AD vs HC: highest sensitivity among the state of the art methods; a classification accuracy of 94%, higher than all MRI-only methods except for [69]. (d) In the ECD for AD vs MCI: highest classification accuracy among the state of the art methods. (e) In most experiments, the sensitivity (the ability of a predictor to correctly classify a subject as 'diseased') was higher than the specificity (the ability of predictor to correctly classify a subject as 'disease-free'). This may be due to the fact that ADNI clinical diagnosis is based on neuro-psychological tests, but neurodegeneration occurs many years before the onset of clinical symptoms. Possibly, residual-based SVM predictors may determine that subjects are into early stages of disease (MCI and prodromal AD) but this finding is not consistent with clinical diagnosis because the subject does not yet present clinical symptoms. Abnormalities of CSF profile observed on several subjects diagnosed as HC and MCI (see Figure 4.4) may support this hypothesis.

The use of MMSE and CDRGLOBAL tests (F_2 method) yields significantly better results than using the residuals alone (F_1 method). These are two of the most common neuropsychological tests routinely applied to patients in the primary clinical practice. However this should not be seen as a mere contribution of MMSE and CDRGLOBAL, as these tests on their own have several limitations such as relatively low specificity and reliability. However, they complement and enhance the present method here without adding a significant cost or invasive clinical tests.

Table 4.8 Summary of variant cortical MRI-based biomarkers for females.

	ROI	hemisferi	measuretype	Lobe	β_a
1	cuneus	lh	TA	occipital lobe	-0.050
2	pericalcarine	lh	CV	occipital lobe	-0.050
3	bankssts	rh	CV	temporal lobe	-0.051
4	caudalmiddlefrontal	rh	TA	frontal lobe	-0.051
5	temporalpole	rh	CV	temporal lobe	-0.052
6	rostralmiddlefrontal	rh	CV	frontal lobe	-0.052
7	inferiorparietal	rh	TA	parietal lobe	-0.053
8	precentral	rh	CV	frontal lobe	-0.053
9	supramarginal	rh	CV	parietal lobe	-0.055
10	medialorbitofrontal	lh	CV	frontal lobe	-0.055
11	temporalpole	lh	TA	temporal lobe	-0.055
12	supramarginal	rh	TA	parietal lobe	-0.055
13	parsopercularis	lh	CV	frontal lobe	-0.055
14	precentral	rh	TA	frontal lobe	-0.057
15	superiorparietal	lh	CV	parietal lobe	-0.057
16	entorhinal	rh	CV	temporal lobe	-0.058
17	temporalpole	lh	CV	temporal lobe	-0.058
18	transversetemporal	lh	CV	temporal lobe	-0.058
19	isthmuscingulate	rh	CV	cingulate cortex	-0.059
20	cuneus	rh	CV	occipital lobe	-0.059
21	paracentral	rh	CV	frontal lobe	-0.059
22	supramarginal	lh	TA	parietal lobe	-0.060
23	parahippocampal	rh	TA	temporal lobe	-0.060
24	isthmuscingulate	lh	CV	cingulate cortex	-0.061
25	cuneus	lh	CV	occipital lobe	-0.061
26	transversetemporal	lh	TA	temporal lobe	-0.061
27	lingual	rh	CV	occipital lobe	-0.061
28	entorhinal	lh	CV	temporal lobe	-0.062
29	inferiorparietal	lh	CV	parietal lobe	-0.062
30	lateralorbitofrontal	rh	TA	frontal lobe	-0.062
31	pericalcarine	lh	TA	occipital lobe	-0.063
32	rostralmiddlefrontal	lh	CV	frontal lobe	-0.063
33	cuneus	rh	TA	occipital lobe	-0.063
34	caudalmiddlefrontal	lh	TA	frontal lobe	-0.064
35	rostralmiddlefrontal	rh	TA	frontal lobe	-0.065
36	postcentral	lh	TA	parietal lobe	-0.066
37	superiorparietal	rh	TA	parietal lobe	-0.066
38	lateraloccipital	lh	CV	occipital lobe	-0.066
39	superiorparietal	rh	CV	parietal lobe	-0.067
40	superiortemporal	lh	CV	temporal lobe	-0.067
41	superiorfrontal	rh	CV	frontal lobe	-0.068
42	lateralorbitofrontal	lh	CV	frontal lobe	-0.068
43	middletemporal	rh	CV	temporal lobe	-0.069
44	parstriangularis	rh	TA	frontal lobe	-0.069
45	inferiortemporal	rh	CV	temporal lobe	-0.070
46	middletemporal	lh	CV	temporal lobe	-0.070
47	parsorbitalis	rh	CV	frontal lobe	-0.070
48	fusiform	lh	CV	temporal lobe	-0.071
49	temporalpole	rh	TA	temporal lobe	-0.071
50	parahippocampal	rh	CV	temporal lobe	-0.072
51	precuneus	lh	CV	parietal lobe	-0.072
52	isthmuscingulate	rh	TA	cingulate cortex	-0.073
53	middletemporal	lh	TA	temporal lobe	-0.073
54	precuneus	lh	TA	parietal lobe	-0.073
55	inferiortemporal	lh	TA	temporal lobe	-0.073
56	lateralorbitofrontal	lh	TA	frontal lobe	-0.075
57	superiorparietal	lh	TA	parietal lobe	-0.075
58	postcentral	rh	TA	parietal lobe	-0.078
59	medialorbitofrontal	lh	TA	frontal lobe	-0.078
60	parsorbitalis	rh	TA	frontal lobe	-0.079
61	entorhinal	rh	TA	temporal lobe	-0.079
62	middletemporal	rh	TA	temporal lobe	-0.079
63	precentral	lh	CV	frontal lobe	-0.079
64	superiorfrontal	rh	TA	frontal lobe	-0.080
65	superiortemporal	rh	CV	temporal lobe	-0.080
66	precuneus	rh	CV	parietal lobe	-0.081
67	parahippocampal	lh	CV	temporal lobe	-0.081
68	medialorbitofrontal	rh	TA	frontal lobe	-0.083
69	isthmuscingulate	lh	TA	cingulate cortex	-0.083
70	precentral	lh	TA	frontal lobe	-0.086
71	superiorfrontal	lh	CV	frontal lobe	-0.087
72	rostralmiddlefrontal	lh	TA	frontal lobe	-0.088
73	fusiform	rh	CV	temporal lobe	-0.089
74	fusiform	lh	TA	temporal lobe	-0.090
75	precuneus	rh	TA	parietal lobe	-0.092
76	lateraloccipital	rh	CV	occipital lobe	-0.093
77	lingual	lh	CV	occipital lobe	-0.093
78	bankssts	rh	TA	temporal lobe	-0.093
79	entorhinal	lh	TA	temporal lobe	-0.093
80	inferiorparietal	lh	TA	parietal lobe	-0.093
81	inferiortemporal	rh	TA	temporal lobe	-0.094
82	lingual	lh	TA	occipital lobe	-0.097
83	superiortemporal	rh	TA	temporal lobe	-0.103
84	superiorfrontal	lh	TA	frontal lobe	-0.106
85	superiortemporal	lh	TA	temporal lobe	-0.110
86	lingual	rh	TA	occipital lobe	-0.118
87	fusiform	rh	TA	temporal lobe	-0.123
88	lateraloccipital	lh	TA	occipital lobe	-0.129
89	lateraloccipital	rh	TA	occipital lobe	-0.154

Table 4.9 Summary of variant subcortical MRI-based biomarkers for females.

	ROI	hemisferi	measuretype	Lobe	β_a
1	opticchiasm	bilateral	SV		0.060
2	thirdventricle	bilateral	SV		0.055
3	lateralventricle	rh	SV		0.053
4	pallidum	lh	SV		-0.055
5	ventraldc	lh	SV		-0.056
6	thalamus	lh	SV		-0.058
7	corpuscallosumposterior	bilateral	SV		-0.058
8	corpuscallosumanterior	bilateral	SV		-0.061
9	corpuscallosummidanterior	bilateral	SV		-0.067
10	amygdala	rh	SV		-0.074
11	accumbensarea	rh	SV		-0.079
12	corpuscallosumcentral	bilateral	SV		-0.079
13	accumbensarea	lh	SV		-0.081
14	hippocampus	rh	SV		-0.085
15	thalamus	rh	SV		-0.085
16	ventraldc	rh	SV		-0.086
17	hippocampus	lh	SV		-0.088
18	amygdala	lh	SV		-0.090

Table 4.10 Performances of classification for current diagnosis.

	Experiment	Features	ACC	SEN	SPE
Females	AD vs HC	F_1	91.7	92.8	90.0
		F_2	94.1	95.2	92.5
	MCI vs HC	F_1	79.5	92.9	52.9
		F_2	85.7	93.6	70.0
	AD vs MCI	F_1	66.8	57.3	74.3
		F_2	81.6	80.9	82.1
Males	AD vs HC	F_1	87.1	89.0	84.2
		F_2	94.1	97.7	88.4
	MCI vs HC	F_1	75.5	97.3	10.0
		F_2	82.5	93.5	49.7
	AD vs MCI	F_1	72.3	68.4	75.7
		F_2	73.8	64.7	81.3

F_1 is the features set integrated with the MRI-based biomarkers and age; F_2 is integrated with the MRI-based biomarkers, age, MMSE (Mini-Mental Examination Score) and CDR-GLOBAL (Clinical dementia rating global scale).

Table 4.11 Performances of classification models built to address the prediction of future diagnosis.

	Experiment	Features	ACC	SEN	SPE
Females	AD vs HC	F_1	86.6	90.7	77.7
		F_2	91.7	92.9	89.2
	MCI vs HC	F_1	70.0	70.0	70.0
		F_2	77.1	75.3	79.2
	AD vs MCI	F_1	69.8	86.5	45.0
		F_2	72.5	87.4	50.4
Males	AD vs HC	F_1	85.6	94.8	64.4
		F_2	90.0	96.8	74.4
	MCI vs HC	F_1	62.9	83.5	29.4
		F_2	77.6	85.0	65.6
	AD vs MCI	F_1	65.4	85.0	54.0
		F_2	69.0	85.1	52.9

F_1 is the features set integrated with the MRI-based biomarkers and age; F_2 is integrated with the MRI-based biomarkers, age, MMSE (Mini-Mental Examination Score) and CDR-GLOBAL (Clinical dementia rating global scale).

Table 4.12 Last known diagnostic prediction advancement.

		Advancement for early prediction in years		
	Experiment	70-74 yrs.	75-79 yrs.	80-84 yrs.
Females	AD vs HC	1.32(80.92)	1.64(83.92)	1.28(75.93)
	AD vs MCI	1.35(76.52)	1.75(71.24)	1.34(77.21)
Males	AD vs HC	1.51(84.72)	1.22(82.70)	1.73(89.30)
	AD vs MCI	1.68(72.91)	1.41(81.50)	1.85(79.19)

(ACC) represents the prediction accuracy of F_2 method in %.

Chapter 5

Early prediction of AD: Method M_{raw}

5.1 Introduction

This chapter presents a second method, called M_{raw}, which also addresses both ageing-related brain change understanding and the early prediction and diagnostic prediction advancement of MCI and AD by combining univariate and multivariate statistical learning methods. In contrast to previous proposed method (M_{res}), this method is focused on directly analysing the raw MRI-based biomarkers values (without derived residuals), stratified on five-year age group-specific atrophy. Furthermore, instead of identifying variant biomarkers, M_{raw} includes a differential diagnosis-specific feature selection (FS) method, which is applied before classification. One hypothesis of the present work is that the age of the subjects is a factor that qualitatively influences the diagnosis of a subject. Consequently, noticeable differences are expected to exist in the optimal subset of features related to the diagnosis in a specific age group. In the beginning of the method, subjects with a stable last diagnostic (dx_{last}), i.e., who remained as HC (sHC), MCI (sMCI) or AD (sAD) for the duration of the study were identified and labelled as such. For each dx_{last} group, data were stratified in five-year age groups, including subjects aged 55–90 years. To study the age effect on the change of cortical brain regions, we fitted random intercepts LME models by using as covariate the subject's age at each visitation (*age*). After descriptive analysis, we focused on a addressing the estimation of current diagnosis and the prediction of future diagnosis for each age group, where three main experiments were established: sHC vs sAD, sHC vs sMCI and sMCI vs sAD. First, an FS approach was applied on training data in order to identify the MRI-based biomarkers that best discriminate comparisons in the three experiments. The FS stage was performed with two main steps: firstly, features were ranked and sorted according the minimum redundancy and maximum relevance (mRMR) method and, secondly, an SVM wrapper was built to determine the inclusion or exclusion of each biomarker within

the final set of significant features. Finally, age-associated HC/MCI/AD prediction and diagnostic prediction advancement was performed using SVM models with radial kernels. These models were built and evaluated on testing data including the remaining subjects with stable diagnosis that had never been used on training and also those subjects who did not have a stable diagnostic.

Details about ADNI database whose participants are used in this study was briefly introduced in Section 2.3. Furthermore, ADNI MRI-based biomarkers used in this work are previously described in Section 4.2. ADNI has been used by many number of publications focused on the characterization of age-related brain changes [43, 44, 51] and the early prediction of conversion to AD [29, 30, 19, 18, 111, 51]. A recent review has been published by Weiner and colleagues in *Alzheimer's & Dementia: Journal of the Alzheimer's Association* [122].

5.2 Data

Participants were selected from the original ADNI study if they met the following criteria (at the time of this study , December, 2016): 1) Had all selected longitudinal MRI images correctly processed; 2) Had completed demographic and neuropsychological data and were clinically diagnosed at each visitation; and 3) Had available CSF-based biomarkers for at least one visitation. ADNI features were described in Section 2.3). In total, the longitudinal data of 456 subjects (128 HC, 135 MCI and 193 AD) was studied. The initial features set was integrated by 174 variables including demographic (N=3), neuropsychological (N=2) and clinical diagnostic features (N=1), as well as 166 cortical and subcortical MRI-based biomarkers.

5.2.1 Clinical Feature Collection.

The clinical features set was integrated by demographic and neuropsychological characteristics, as well as the ADNI clinical diagnostic at each visitation (dx_{age}), see Section 2.3.2. The demographic features set was integrated by *gender*, age at each visitation (*age*) and normalized years of education (*educ*). For the neuropsychological features set, we included the MMSE scores [46] and the Clinical Dementia Rating-Sum of Boxes (*CDRSB*)[86].

5.2.2 Morphological Features Collection.

MRI-based biomarkers were directly obtained from ADNI processed data repository available at 5 years follow-up. These biomarkers were previously described in Section 2.3.4.

Biomarkers related to the unknown regions were excluded from further analysis. Furthermore, biomarkers with missing values for most samples were discarded, we only included the ones correctly processed and available for at least two time points. In total, 166 longitudinal MRI-based biomarkers, including 66 CV's and 66 TA's of 33 cortical regions, and 34 SV's of 22 subcortical regions; were used in this method.

5.3 Methods

We hypothesize that AD-related brain changes are age-dependent and as a consequence, that certain MRI-based biomarkers are more sensitive and relevant for some age groups as opposed to others. Furthermore, we argue these changes are also differentiated within each disease condition. For these reasons, this study was focused on studying the brain structures changes within stable HC, MCI and AD subjects. Then, an early estimation and prediction of their diagnosis was also extended to those subjects with unstable diagnosis. The proposed M_{raw} consists of two main stages. The first stage was focused on a descriptive analysis of the age-specific effect on each MRI biomarkers by following two main steps 1) Data stratification, where subjects were first classified according their visitation-specific diagnosis and last known diagnosis, then assigned within five-year age groups and, finally, assigned to training or testing datasets, followed by 2) Analysis of age-related effect of cortical-related biomarkers. The second stage was focused on early differential diagnosis problems. For this purpose, three main steps were carried out: 1) Feature ranking to sort MRI biomarkers according their significance within each experiment; 2) Feature selection, where the most relevant MRI-based biomarkers were selected; 3) SVM classifiers building to subjects diagnosis and diagnostic prediction advancement.

5.3.1 Data stratification.

In this study we used both dx_{age} and dx_{last} diagnostic variables, which were previously introduced in Table 4.2. In order to facilitate the analysis, classes of variable dx_{age} were grouped into three main classes such as HC, MCI and AD. Table 5.1 gives a brief description of both diagnosis variables.

Once the subjects were characterized according to dx_{last}, five-year age groups were established as follows: [55–60) (55 to 60 years), [60-65) (60 to 65 years), [65–70) (65 to 70 years), [70–75) (70 to 75 years), [75–80) (75 to 80 years), [80–85) (80 to 85 years), [85–90) (85 to 90 years) and 90+ (greater than or equal to 90 years)—the given intervals are right-open.

Table 5.1 Diagnosis variables used in this study.

Source	Diagnostic variable	Class		Class description
ADNI	dx_{age}	HC*	NL	Subjects diagnosed as stable normal at current visitation
		MCI*	NL to MCI	Subjects diagnosed as MCI at current visit who previously were NL
			MCI	Subjects diagnosed as stable MCI at current visit who previously were also MCI
		AD*	NL to Dementia	Subjects diagnosed as dementia due to AD at current visit who previously were NL
			MCI to Dementia	MCI to Dementia &Subjects diagnosed as dementia due to AD at current visit who previously were MCI
			Dementia	Subjects diagnosed as stable dementia due to AD at current visit who previously were also MCI
Our study	dx_{last}	sHC		Subjects labelled as HC who remained like HC in all follow-up visits (who did not become MCI or AD)
		sMCI		MCI subjects who did not become AD
		cMCI		Subjects initially labelled as HC who subsequently have converted to MCI
		sAD		Subjects who remained like probable or possible AD all the follow-up visits
		cAD		Subjects labelled as HC or MCI who subsequently have converted to probable or possible AD

*For simplification, dx_{age} values were grouped into three main classes.

For each 5-year group, the data from the sHC, sMCI and sAD groups were randomly divided into training (60%) and testing (40%) datasets. Each individual only was assigned once to training or testing dataset. The remaining data from cMCI and cAD subjects were also included within the testing dataset.

When profiling data, the subject characteristics were compared according to age groups. Continuous variables were described by their percentiles, while categorical variables were described by frequencies and percentages. The years of education variable was normalized to [0, 1].

5.3.2 Analysis of age-related effect on brain structures change

In order to visualize the brain regions showing a level of atrophy within each stable diagnosis group, the mean change of biomarkers related to cortical brain regions was quantified by applying the LME approach (see Section 2.2.3). Because we hypothesize that there are important individual-level effects and believe that subjects have similar rates of change over time, we established random intercepts for subjects; the subject's age at each visitation (*age*) was specified as a single fixed effect. Such approach have been also applied in the method M_{res} and recently published in [51]. The observed MRI biomarker r defined as y_{ij} is assumed to have a set of parameters β, fixed across subjects. In addition, for each individual i, a set of random parameters υ_i is assigned to model the deviation from the fixed effect β. For $i = 1,...,n$, each model reads as follows:

$$y_{ij}^r = X_{ij}\beta^r + Z_{ij}\upsilon_i^r + \varepsilon_{ij}^r \tag{5.1}$$

where, y_{ij}^r is the standardized value of biomarker r measured for the i^{th} subject in the j^{th} observation; $i = 1,...,n$ subjects, $j = 1,...,n_i$ available observations for subject i and $r = 1,...,n_r$, n_r = 166 MRI biomarkers. X_{ij} is a n_i x p design matrix, where p is the number of covariates (*age* and the constant term of 1's) on the j^{th} observation of i^{th} subject. β^r is the p x 1 vector of unknown fixed effects or regressor's coefficients, which are: β_1^r (coefficient for constant term or $Intercept$) and β_a^r (coefficient for *age*). Z_{ij} is a known design matrix of size n_i x q, where q is the number of random effects for the j^{th} observation of subject i. υ_i^r is the $q \times 1$ vector of unknown random effects coefficients $\sim N_q(0,\psi)$ for subject i measured for r. ε_{ij}^r is the $n_i x 1$ residual vector of errors $\sim N_{ni}(0,\sigma^2\lambda_i)$ for the j^{th} observation in subject i measured for r. ψ is the $q \times q$ covariance matrix for the random effects. $0,\sigma^2\lambda_i$ is the n_i x n_i covariance matrix for the errors in subject i.

LME modelling was performed using the *lme*4 package available for R. [8]

5.3.3 Feature ranking (FR)

Three differential diagnosis problems were established from dx_{last} classes: sAD vs sHC, sMCI vs sHC and sAD vs sMCI. By using training data, feature ranking was applied separately for each five-year age group to obtain lists of MRI-based biomarkers or features (F) sorted according their significance for each experiment. Ranking used the minimum redundancy and maximum relevance (mRMR) method described previously in Section 2.2.1. Since MRI biomarkers are continuous values, a discretization approach was applied as

a preprocessing step before the mutual information between r's was computed. Feature discretisation was applied using a threshold $t = 1$, which means that each continuous feature was discretised into three classes [-1, 1, 0]. The Mutual Information Difference (MID) method was selected to combine the relevance and redundancy values defined from the mutual information.

5.3.4 Feature selection (FS)

Based on previous studies, where researchers have found that only specific brain regions are significantly affected by ageing, we included a FS step to obtain the set of most significant features for each differential diagnosis problem. FS methods are usually applied before classification because eliminating the non-informative features tends to improve the classification performance. Therefore, a selection of significant r's is desirable before performing classification. Likewise, FS should help to reduce possible issues arising from high dimensionality of data (i.e. number of features vs number of training samples).

Once the MRI biomarkers were ranked and sorted according their mRMR scores, a forward FS scheme (based on an SVM-wrapper approach) was applied to obtain the relevant feature set (F') for each five-year age group. The SVM Wrapper method was performed for each combination of differential diagnosis problem and age group and included the following main steps:

Null SVM model. By including only *age*, *gender* and *education* as initial classifiers, a null SVM model (SVM_{null}) was built on training data using the RBF (Radial Basis Function) kernel. By using this kernel, there were two parameters used to fit the models: 1) the cost parameter (C) of SVM models that allows some flexibility in separating the classes, controlling the trade off between them allowing training errors and forcing rigid margins; and 2) the Gaussian kernel width (γ). As it is not known beforehand which C and γ values are the best for classification problem; it was necessary to apply a method to find the optimal parameters values. For this, we applied the "grid-search" algorithm using 10-fold cross-validation (CV). This algorithm tried various pairs of (C ,γ) by using exponentially growing sequences of these ($\log_2 C$ from $[-4,4]$ and $\log_2 \gamma$ from $[-4,0]$). The combination of parameters, which provided the best accuracy (ACC) of CV was selected. Then, the SVM_{null} model was trained again by including these parameters and its ACC was computed.

Computing performance of features. The SVM parameter searching, model re-training and ACC computation processes described above were repeated by incrementally adding each r from vector F (sorted from highest to lowest mRMR scores). At each iteration, the

training data were linearly normalized to the range [-1,1]. The feature inclusion-exclusion criterion was based on the ACC of an SVM model computed by adding that feature. Thus, if the ACC was increased by including r, then that feature was included into the list of optimal features for the specific age group F', otherwise it was excluded.

The SVM wrapper was implemented in the R package [96]. Tasks included in SVM building were performed under the R interface to LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm)[12] available in package e1071 [83].

5.3.5 Differential diagnosis experiments

The motivation of the experiments described in this section is to use the age-specific features selected at the previous stage, as described in Section 5.3.4, to generate an array of classifiers whose output is the diagnosis of the subjects under study.

One hypothesis of the present work is that the age of the subjects is a factor that qualitatively influences the diagnosis of a subject. Consequently, noticeable differences are expected to exist in the optimal subset of features F' related to the diagnosis in a specific age group. The experiments described in this section also follow this premise, clustering the diagnosis of the subjects in the study by age groups. For this reason, a different classifier is trained for each combination of age group and differential diagnosis.

In order to perform reliable experiments and produce trustworthy results, the subjects involved in the experiments described in this section have not been used on the previous experiments, i.e. the subjects used here (referred as S_E) are a subset of the subjects in the original dataset.

The input data given to the classifiers is a set of aggregated features known as F^{add}. These feature sets were obtained by grouping the age-specific features from F' with the features selected for all the previous age groups (in chronological order). When required, categorical features were converted to real numbers.

The motivation for using F^{add} instead of F' is twofold. Firstly, from a conceptual point of view, if a feature has had a relevant role with respect to the diagnosis at a given age, this implies that it is somehow related to the neurological condition. In that case, it is reasonable to think that it will be related to the condition subsequently, even if its relative importance may decline in favour of other features. Secondly, as the experimental results show (see Section 5.4), the estimation of the diagnosis is better if F^{add} is used rather than F'.

Two experiments are next described: the estimation of the current diagnosis (referred to as ECD) and the prediction of a future diagnosis (referred to as PFD).

Estimation of current diagnosis (ECD).

The goal of this experiment is to generate an array of classifiers that correctly estimate the current diagnosis of a subject (i.e. their dx_{age}), given their F^{add} feature set. As explained above, a different classifier is trained for each combination of age group and differential diagnosis. Each classifier was trained given F^{add} as input variables and dx_{age} as classification target.

The classifiers used in the experiments are SVM with Gaussian radial basis kernel. In order to tune their hyperparameters (γ of the Gaussian function and the misclassification cost C), the set of subjects used in this experiment (S_E) is further divided into three subsets: training, validation and test. A grid search on the hyperparameter space is performed, training on the training subset and measuring the ACC on the validation subset. After the hyperparameters are set, the classifier is evaluated on the test subset. This process is repeated 10 times with different partitioning of the data (10-fold CV).

The performance of each classifier is measured by the classification ACC, sensitivity (SEN) and specificity (SPE).

Because the remaining age groups do contain enough data to perform the experiments as described, only the samples of subjects aged 65–90 years have been used.

Prediction of future diagnosis (PFD).

The objective of this experiment was to develop an array of classifiers that predict the future diagnosis of a subject (dx_{last}), given their current features F^{add}. The main difference with respect to the ECD experiment is the use of the last known diagnosis dx_{last} as classification target instead of the diagnosis at the age of the sample, dx_{age}. The overall methodology of this experiment was basically the same as that in ECD. There are, however, some important differences:

- Only the subjects whose diagnosis changes in course of the ADNI study are considered, i.e. subjects whose dx_{last} is either cMCI or cAD, as per to the description given in Table 5.1;
- Therefore, only a subset $S_{E\prime} \subseteq S_E$ of subjects is used in this experiment;
- The PFD is focused on two differential diagnosis: HC to AD and MCI to AD. There are not enough subjects to run the experiment on the HC to MCI diagnosis;
- The data available in $S_{E\prime}$ only allows us to run this experiment on age groups 70 to 85.

Besides the classification ACC, SPE and SEN, another relevant finding from this experiment was the average number of years by which the PFD-predicted diagnosis precedes the clinical diagnosis as stated in the dataset. This value reflects the reduced diagnostic time lag that the method presented here may offer.

5.4 Results

In total, 214 subjects were included within training cohort with 75 sHC, 75 sMCI and 64 sAD subjects, all diagnosed as such based on the stability of their diagnosis during the study period. The testing cohort comprised 242 subjects including 40% of subjects with stable diagnosis (53 sHC, 51 sMCI and 39 sAD) and subjects with unstable diagnosis (9 cMCI and 90 cAD). A summary of the number of samples, by five-year age groups, is shown in Table 5.2. The table shows there are several age groups with few or no samples for certain diagnosis groups. We took this issue into account when model building, discarding models for such age groups.

Table 5.3 shows the clinical characteristics of the training dataset. Tables 5.4 and 5.5 show the clinical characteristics of the testing datasets comprising subjects with stable and unstable diagnosis, respectively. Both cohorts were referred as S_E in the diagnosis estimation and prediction stage. Note that dx_{last} in Table 5.3 has no samples for cMCI and cAD subjects because this dataset was used in the feature selection stage, which was defined for selecting features from subjects with a stable diagnosis over the duration of the study. sHC subjects had the greatest MMSE values as compared to other groups and their CDRSB index was always equal to zero; however some of them were APOE-ε4 carriers. With respect to CSF-based biomarkers, we observed that sHC subjects tend to have higher CSF-ABETA values than do sMCI and sAD; they also showed lower CSF-TAU values than sMCI and sAD. In relation to subjects with unstable diagnosis, we observed that CSF-ABETA values in cAD tend to be lower than cMCI; their CDF-TAU values alsotend to be higher.

Table 5.2 Number of studied samples stratified by dx_{last} and age groups.

dx_{last}	55–60 N=29		60–65 N=47		65–70 N=121		70–75 N=335		75–80 N=273		80–85 N=273		85–90 N=117		+90 N=6	
	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M
sAD	8	4	5	7	6	11	15	22	22	32	12	27	13	8	–	–
sHC	–	–	1	1	3	4	44	60	62	58	35	22	7	8	1	2
sMCI	1	1	15	3	16	20	22	36	19	61	20	36	2	29	–	–
cMCI	–	–	–	4	–	1	4	5	5	4	8	10	–	4	–	–
cAD	10	5	1	10	34	26	55	72	36	86	33	70	8	38	–	3

F: Female, M: Male

Table 5.3 Clinical characteristics of subjects with stable diagnostic included in training groups

	stable HC						
	60–65	65–70	70–75	75–80	80–85	85–90	+90
	$N = 2$	$N = 5$	$N = 29$	$N = 39$	$N = 20$	$N = 6$	$N = 2$
AGE_IMG	62.2 62.4 62.6	65.4 65.9 69.5	70.8 72.0 73.1	75.5 76.0 77.0	80.5 80.8 82.7	85.3 86.3 87.1	90.1 90.2 90.2
gender							
Female	50.0% $\frac{1}{2}$	40.0% $\frac{2}{5}$	51.7% $\frac{15}{29}$	53.9% $\frac{21}{39}$	45.0% $\frac{9}{20}$	33.3% $\frac{2}{6}$	50.0% $\frac{1}{2}$
Male	50.0% $\frac{1}{2}$	60.0% $\frac{3}{5}$	48.3% $\frac{14}{29}$	46.1% $\frac{18}{39}$	55.0% $\frac{11}{20}$	66.7% $\frac{4}{6}$	50.0% $\frac{1}{2}$
educ	0.594 0.688 0.781	0.562 0.750 0.875	0.625 0.750 0.875	0.625 0.750 0.875	0.609 0.750 0.875	0.656 0.812 0.969	0.812 0.875 0.938
MMSE	28.5 29.0 29.5	28.0 29.0 30.0	29.0 29.0 30.0	29.0 29.0 30.0	29.0 30.0 30.0	28.0 29.0 30.0	29.0 29.0 29.0
CDRSB							
0	50.00% $\frac{1}{2}$	100.00% $\frac{5}{5}$	100.00% $\frac{29}{29}$	92.31% $\frac{36}{39}$	80.00% $\frac{16}{20}$	100.00% $\frac{6}{6}$	100.00% $\frac{2}{2}$
0.5	50.00% $\frac{1}{2}$	0.00% $\frac{0}{5}$	0.00% $\frac{0}{29}$	5.13% $\frac{2}{39}$	20.00% $\frac{4}{20}$	0.00% $\frac{0}{6}$	0.00% $\frac{0}{2}$
2	0.00% $\frac{0}{2}$	0.00% $\frac{0}{5}$	0.00% $\frac{0}{29}$	2.56% $\frac{1}{39}$	0.00% $\frac{0}{20}$	0.00% $\frac{0}{6}$	0.00% $\frac{0}{2}$
APOE4							
0	50.0% $\frac{1}{2}$	60.0% $\frac{3}{5}$	79.3% $\frac{23}{29}$	84.6% $\frac{33}{39}$	70.0% $\frac{14}{20}$	66.7% $\frac{4}{6}$	50.0% $\frac{1}{2}$
1	50.0% $\frac{1}{2}$	40.0% $\frac{2}{5}$	20.7% $\frac{6}{29}$	15.4% $\frac{6}{39}$	30.0% $\frac{6}{20}$	33.3% $\frac{2}{6}$	50.0% $\frac{1}{2}$
ABETA	187 203 220	168 171 222	162 234 255	139 215 252	147 200 237	150 196 252	189 220 250
TAU	43.4 49.2 55.1	37.5 37.9 55.7	54.6 72.2 82.9	52.1 70.0 98.5	46.4 59.7 98.0	62.5 94.9 106.5	93.5 99.0 104.6
	stable MCI						
	55–60	60–65	65–70	70–75	75–80	80–85	85-90
	$N = 2$	$N = 6$	$N = 12$	$N = 21$	$N = 26$	$N = 18$	$N = 10$
AGE_IMG	55.3 55.5 55.8	61.1 62.5 62.8	65.7 66.5 68.5	71.1 72.8 73.9	75.4 76.3 78.3	80.2 81.2 83.0	85.3 85.5 86.9
gender							
Female	50.0% $\frac{1}{2}$	66.7% $\frac{4}{6}$	58.3% $\frac{7}{12}$	42.9% $\frac{9}{21}$	23.1% $\frac{6}{26}$	22.2% $\frac{4}{18}$	20.0% $\frac{2}{10}$
Male	50.0% $\frac{1}{2}$	33.3% $\frac{2}{6}$	41.7% $\frac{5}{12}$	57.1% $\frac{12}{21}$	76.9% $\frac{20}{26}$	77.8% $\frac{14}{18}$	80.0% $\frac{8}{10}$
educ	0.781 0.812 0.844	0.750 0.781 0.906	0.500 0.625 0.828	0.500 0.750 0.875	0.578 0.750 0.875	0.625 0.750 0.875	0.656 0.875 0.875
MMSE	27.5 28.0 28.5	27.2 28.0 28.8	25.8 27.5 29.0	25.0 27.0 28.0	27.2 28.0 28.0	25.0 26.0 28.0	26.0 28.5 29.0
CDRSB	2.000 2.000 2.000	1.000 1.250 1.500	0.500 1.250 2.125	0.500 1.000 2.000	0.625 1.500 2.000	1.000 1.000 1.500	1.000 1.500 1.500
APOE4							
0	50.0% $\frac{1}{2}$	50.0% $\frac{3}{6}$	33.3% $\frac{4}{12}$	47.6% $\frac{10}{21}$	46.1% $\frac{12}{26}$	55.6% $\frac{10}{18}$	90.0% $\frac{9}{10}$
1	50.0% $\frac{1}{2}$	50.0% $\frac{3}{6}$	50.0% $\frac{6}{12}$	52.4% $\frac{11}{21}$	42.3% $\frac{11}{26}$	44.4% $\frac{8}{18}$	10.0% $\frac{1}{10}$
2	0.0% $\frac{0}{2}$	0.0% $\frac{0}{6}$	16.7% $\frac{2}{12}$	0.0% $\frac{0}{21}$	11.5% $\frac{3}{26}$	0.0% $\frac{0}{18}$	0.0% $\frac{0}{10}$
ABETA	185 209 234	152 233 240	122 142 225	127 142 224	130 150 178	131 146 226	217 237 257
TAU	76.6 97.5 118.4	46.3 57.2 74.9	47.5 95.4 140.3	63.7 99.2 170.0	65.4 90.1 114.9	55.3 81.9 109.7	55.2 57.6 81.5
	stable AD						
	55–60	60–65	65–70	70–75	75–80	80–85	85–90
	$N = 4$	$N = 5$	$N = 7$	$N = 15$	$N = 20$	$N = 17$	$N = 8$
AGE_IMG	58.0 58.6 59.1	62.4 62.8 64.1	65.8 67.6 68.7	70.8 71.5 72.9	75.8 77.5 78.7	80.2 80.8 83.2	85.2 85.4 87.6
gender							
Female	100.0% $\frac{4}{4}$	0.0% $\frac{0}{5}$	42.9% $\frac{3}{7}$	53.3% $\frac{8}{15}$	40.0% $\frac{8}{20}$	52.9% $\frac{9}{17}$	37.5% $\frac{3}{8}$
Male	0.0% $\frac{0}{4}$	100.0% $\frac{5}{5}$	57.1% $\frac{4}{7}$	46.7% $\frac{7}{15}$	60.0% $\frac{12}{20}$	47.1% $\frac{8}{17}$	62.5% $\frac{5}{8}$
educ	0.594 0.688 0.781	0.750 0.875 0.875	0.594 0.750 0.812	0.500 0.562 0.750	0.562 0.750 0.812	0.625 0.750 0.875	0.469 0.625 0.906
MMSE	22.2 23.0 23.5	21.0 24.0 24.0	23.5 25.0 26.0	21.0 24.0 24.5	22.0 24.0 25.0	23.0 24.0 26.0	20.0 23.0 26.0
CDRSB	4.38 4.75 6.00	4.00 4.00 6.00	2.75 4.00 4.25	3.75 4.50 5.25	3.00 4.50 5.12	3.00 3.50 5.00	3.38 5.25 7.75
APOE4							
0	25.00% $\frac{1}{4}$	40.00% $\frac{2}{5}$	28.57% $\frac{2}{7}$	20.00% $\frac{3}{15}$	20.00% $\frac{4}{20}$	52.94% $\frac{9}{17}$	50.00% $\frac{4}{8}$
1	0.00% $\frac{0}{4}$	60.00% $\frac{3}{5}$	14.29% $\frac{1}{7}$	46.67% $\frac{7}{15}$	55.00% $\frac{11}{20}$	41.18% $\frac{7}{17}$	50.00% $\frac{4}{8}$
2	75.00% $\frac{3}{4}$	0.00% $\frac{0}{5}$	57.14% $\frac{4}{7}$	33.33% $\frac{5}{15}$	25.00% $\frac{5}{20}$	5.88% $\frac{1}{17}$	0.00% $\frac{0}{8}$
ABETA	109 126 141	118 128 136	118 130 144	106 127 147	128 140 147	131 142 171	130 139 179
TAU	122.9 161.5 207.6	56.1 76.9 146.2	84.6 109.5 130.0	68.5 111.2 138.0	76.0 109.0 133.6	113.2 141.2 193.4	60.2 107.3 150.1

Values of continuous variables are represented by the lower, the median and the upper quartiles. Categorical variables are represented by percentages and frequencies. *N* is the number of non-missing values.

Table 5.4 Clinical characteristics of subjects with stable diagnostic included in testing groups

stable HC							
	N	65–70	70–75	75–80	80–85	85–90	+90
		$N = 2$	$N = 75$	$N = 81$	$N = 37$	$N = 9$	$N = 1$
AGE_IMG	205	66.3 67.5 68.7	71.8 72.8 73.9	76.3 77.6 78.7	80.8 81.5 82.4	86.3 87.6 88.5	90.6 90.6 90.6
gender	205						
Female		50.0% $\frac{1}{2}$	38.7% $\frac{29}{75}$	50.6% $\frac{41}{81}$	70.3% $\frac{26}{37}$	55.6% $\frac{5}{9}$	0.0% $\frac{0}{1}$
Male		50.0% $\frac{1}{2}$	61.3% $\frac{46}{75}$	49.4% $\frac{40}{81}$	29.7% $\frac{11}{37}$	44.4% $\frac{4}{9}$	100.0% $\frac{1}{1}$
educ	205	0.562 0.625 0.688	0.656 0.812 0.875	0.750 0.750 0.875	0.562 0.750 0.875	0.688 0.688 0.750	0.688 0.688 0.688
MMSE	204	29.2 29.5 29.8	29.0 30.0 30.0	29.0 30.0 30.0	28.8 29.0 30.0	28.0 28.0 29.0	27.0 27.0 27.0
CDRSB	203	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
APOE4	205						
0		100.00% $\frac{2}{2}$	100.00% $\frac{75}{75}$	64.20% $\frac{52}{81}$	86.49% $\frac{32}{37}$	100.00% $\frac{9}{9}$	100.00% $\frac{1}{1}$
1		0.00% $\frac{0}{2}$	0.00% $\frac{0}{75}$	32.10% $\frac{26}{81}$	8.11% $\frac{3}{37}$	0.00% $\frac{0}{9}$	0.00% $\frac{0}{1}$
2		0.00% $\frac{0}{2}$	0.00% $\frac{0}{75}$	3.70% $\frac{3}{81}$	5.41% $\frac{2}{37}$	0.00% $\frac{0}{9}$	0.00% $\frac{0}{1}$
ABETA	205	235 236 236	173 235 257	144 161 234	153 181 216	177 230 245	245 245 245
TAU	205	66.7 72.5 78.2	51.0 60.5 106.6	45.6 57.6 82.7	52.4 78.0 97.8	97.5 97.5 111.8	97.5 97.5 97.5

stable MCI							
	N	60–65	65–70	70–75	75–80	80–85	85–90
		$N = 12$	$N = 24$	$N = 37$	$N = 54$	$N = 38$	$N = 21$
AGE_IMG	186	61.8 62.4 63.5	67.0 68.2 69.2	71.8 73.4 74.4	76.1 77.7 79.0	81.2 82.7 83.6	85.7 86.3 87.1
gender	186						
Female		91.67% $\frac{11}{12}$	37.50% $\frac{9}{24}$	35.14% $\frac{13}{37}$	24.07% $\frac{13}{54}$	42.11% $\frac{16}{38}$	0.00% $\frac{0}{21}$
Male		8.33% $\frac{1}{12}$	62.50% $\frac{15}{24}$	64.86% $\frac{24}{37}$	75.93% $\frac{41}{54}$	57.89% $\frac{22}{38}$	100.00% $\frac{21}{21}$
educ	186	0.500 0.500 0.750	0.719 0.812 0.875	0.625 0.812 0.938	0.625 0.750 0.859	0.750 0.750 0.875	0.750 0.812 0.875
MMSE	186	25.8 28.0 29.2	26.8 28.0 29.0	25.0 26.0 29.0	26.0 28.0 29.0	27.2 29.0 29.0	25.0 27.0 28.0
CDRSB	183	1.000 1.500 2.000	0.875 1.000 1.500	1.000 1.500 2.000	1.000 1.500 2.500	0.500 1.000 1.500	1.000 1.500 2.500
APOE4	186						
0		8.33% $\frac{1}{12}$	37.50% $\frac{9}{24}$	27.03% $\frac{10}{37}$	59.26% $\frac{32}{54}$	89.47% $\frac{34}{38}$	71.43% $\frac{15}{21}$
1		83.33% $\frac{10}{12}$	25.00% $\frac{6}{24}$	54.05% $\frac{20}{37}$	33.33% $\frac{18}{54}$	5.26% $\frac{2}{38}$	28.57% $\frac{6}{21}$
2		8.33% $\frac{1}{12}$	37.50% $\frac{9}{24}$	18.92% $\frac{7}{37}$	7.41% $\frac{4}{54}$	5.26% $\frac{2}{38}$	0.00% $\frac{0}{21}$
ABETA	186	142 142 190	110 156 202	139 147 152	92 131 165	160 221 261	134 134 140
TAU	186	68.9 129.3 139.9	69.2 72.3 79.0	80.1 100.0 128.0	45.1 72.0 111.2	59.0 66.3 109.8	102.1 110.9 127.0

stable AD							
	55–60	60–65	65–70	70–75	75–80	80–85	85–90
	$N = 8$	$N = 7$	$N = 10$	$N = 22$	$N = 34$	$N = 22$	$N = 13$
AGE_IMG	56.8 57.2 57.8	61.3 62.6 63.6	66.3 67.6 69.2	71.9 72.9 73.8	76.0 77.3 78.7	81.3 82.3 83.5	86.5 87.1 88.3
gender							
Female	50.0% $\frac{4}{8}$	71.4% $\frac{5}{7}$	30.0% $\frac{3}{10}$	31.8% $\frac{7}{22}$	41.2% $\frac{14}{34}$	13.6% $\frac{3}{22}$	76.9% $\frac{10}{13}$
Male	50.0% $\frac{4}{8}$	28.6% $\frac{2}{7}$	70.0% $\frac{7}{10}$	68.2% $\frac{15}{22}$	58.8% $\frac{20}{34}$	86.4% $\frac{19}{22}$	23.1% $\frac{3}{13}$
educ	0.750 0.875 1.000	0.656 0.688 0.875	0.422 0.562 0.875	0.750 0.875 0.875	0.500 0.656 0.750	0.500 0.750 0.875	0.562 0.625 0.875
MMSE	20.5 23.0 23.8	17.0 21.0 23.0	20.0 21.5 24.5	18.0 23.5 24.8	20.0 23.0 25.0	19.0 22.5 24.0	20.0 21.0 23.0
CDRSB	4.12 5.25 6.88	3.75 4.50 6.00	4.50 5.00 5.00	3.00 3.75 6.75	4.00 5.00 7.75	3.12 4.75 7.00	6.00 7.00 9.00
APOE4							
0	50.00% $\frac{4}{8}$	71.43% $\frac{5}{7}$	20.00% $\frac{2}{10}$	4.55% $\frac{1}{22}$	17.65% $\frac{6}{34}$	18.18% $\frac{4}{22}$	46.15% $\frac{6}{13}$
1	0.00% $\frac{0}{8}$	28.57% $\frac{2}{7}$	50.00% $\frac{5}{10}$	50.00% $\frac{11}{22}$	64.71% $\frac{22}{34}$	81.82% $\frac{18}{22}$	53.85% $\frac{7}{13}$
2	50.00% $\frac{4}{8}$	0.00% $\frac{0}{7}$	30.00% $\frac{3}{10}$	45.45% $\frac{10}{22}$	17.65% $\frac{6}{34}$	0.00% $\frac{0}{22}$	0.00% $\frac{0}{13}$
ABETA	116 137 158	110 140 141	127 127 128	114 142 149	111 151 153	137 141 163	137 137 234
TAU	101.6 107.0 112.5	82.3 100.4 120.1	134.8 148.1 161.4	92.5 118.6 148.0	65.4 90.8 113.2	76.5 97.0 109.4	112.0 140.1 144.2

Values of continuous variables are represented by the lower, the median and the upper quartiles. Categorical variables are represented by percentages and frequencies. *N* is the number of non-missing values.

Table 5.5 Clinical characteristics of subjects with unstable diagnostic included in testing groups

	Converted to MCI					
	60–65	65–70	70–75	75–80	80–85	85–90
	$N = 4$	$N = 1$	$N = 9$	$N = 9$	$N = 18$	$N = 4$
AGE_IMG	63.0 63.4 63.9	65.7 65.7 65.7	71.5 72.7 73.5	78.0 78.5 79.0	81.1 82.1 83.2	85.4 85.9 86.6
gender						
Female	0.0% $\frac{0}{4}$	0.0% $\frac{0}{1}$	44.4% $\frac{4}{9}$	55.6% $\frac{5}{9}$	44.4% $\frac{8}{18}$	0.0% $\frac{0}{4}$
Male	100.0% $\frac{4}{4}$	100.0% $\frac{1}{1}$	55.6% $\frac{5}{9}$	44.4% $\frac{4}{9}$	55.6% $\frac{10}{18}$	100.0% $\frac{4}{4}$
educ	0.500 0.500 0.500	0.500 0.500 0.500	0.562 0.750 0.750	0.562 0.938 1.000	0.516 0.625 0.891	0.688 0.719 0.750
MMSE	25.0 25.0 25.8	26.0 26.0 26.0	28.0 29.0 29.0	28.0 29.0 29.0	28.0 29.0 29.8	27.8 28.5 29.2
CDRSB						
0	0.0% $\frac{0}{4}$	0.0% $\frac{0}{1}$	55.6% $\frac{5}{9}$	77.8% $\frac{7}{9}$	50.0% $\frac{9}{18}$	33.3% $\frac{1}{3}$
0.5	0.0% $\frac{0}{4}$	0.0% $\frac{0}{1}$	11.1% $\frac{1}{9}$	22.2% $\frac{2}{9}$	16.7% $\frac{3}{18}$	33.3% $\frac{1}{3}$
1	50.0% $\frac{2}{4}$	0.0% $\frac{0}{1}$	22.2% $\frac{2}{9}$	0.0% $\frac{0}{9}$	16.7% $\frac{3}{18}$	33.3% $\frac{1}{3}$
1.5	50.0% $\frac{2}{4}$	100.0% $\frac{1}{1}$	11.1% $\frac{1}{9}$	0.0% $\frac{0}{9}$	16.7% $\frac{3}{18}$	0.0% $\frac{0}{3}$
APOE4						
0	100.0% $\frac{4}{4}$	100.0% $\frac{1}{1}$	100.0% $\frac{9}{9}$	55.6% $\frac{5}{9}$	61.1% $\frac{11}{18}$	100.0% $\frac{4}{4}$
1	0.0% $\frac{0}{4}$	0.0% $\frac{0}{1}$	0.0% $\frac{0}{9}$	44.4% $\frac{4}{9}$	38.9% $\frac{7}{18}$	0.0% $\frac{0}{4}$
ABETA	233 233 233	233 233 233	177 221 221	111 171 171	114 157 171	252 253 254
TAU	56.7 56.7 56.7	56.7 56.7 56.7	33.8 33.8 93.8	70.5 93.1 93.1	62.3 66.4 76.2	57.0 59.0 61.0

	Converted to AD							
	55–60	60–65	65–70	70–75	75–80	80–85	85–90	+90
	$N = 15$	$N = 11$	$N = 60$	$N = 127$	$N = 122$	$N = 103$	$N = 46$	$N = 3$
AGE_IMG	56.6 57.6 58.2	61.7 62.8 63.8	66.2 67.5 68.8	71.8 73.0 74.0	76.0 77.2 78.6	80.9 81.9 83.2	86.2 87.3 88.5	90.2 90.4 90.5
gender								
Female	66.67% $\frac{10}{15}$	9.09% $\frac{1}{11}$	56.67% $\frac{34}{60}$	43.31% $\frac{55}{127}$	29.51% $\frac{36}{122}$	32.04% $\frac{33}{103}$	17.39% $\frac{8}{46}$	0.00% $\frac{0}{3}$
Male	33.33% $\frac{5}{15}$	90.91% $\frac{10}{11}$	43.33% $\frac{26}{60}$	56.69% $\frac{72}{127}$	70.49% $\frac{86}{122}$	67.96% $\frac{70}{103}$	82.61% $\frac{38}{46}$	100.00% $\frac{3}{3}$
educ	0.625 0.688 0.875	0.750 0.875 0.875	0.625 0.750 0.875	0.625 0.750 0.875	0.625 0.750 0.875	0.625 0.750 0.875	0.625 0.750 0.875	0.625 0.750 0.750
MMSE	22.5 24.0 25.0	26.0 27.0 28.5	23.0 25.0 27.0	23.0 25.0 27.5	21.0 24.0 26.0	23.0 26.0 28.0	23.0 25.0 26.0	21.5 26.0 26.5
CDRSB	2.50 3.00 3.75	2.25 3.50 4.25	2.00 2.50 4.12	2.00 2.50 4.00	2.00 3.25 4.50	1.50 3.00 4.50	2.00 3.50 4.50	4.75 5.50 7.25
APOE4								
0	100.00% $\frac{15}{15}$	72.73% $\frac{8}{11}$	8.33% $\frac{5}{60}$	25.20% $\frac{32}{127}$	31.15% $\frac{38}{122}$	59.22% $\frac{61}{103}$	63.04% $\frac{29}{46}$	33.33% $\frac{1}{3}$
1	0.00% $\frac{0}{15}$	18.18% $\frac{2}{11}$	58.33% $\frac{35}{60}$	47.24% $\frac{60}{127}$	59.84% $\frac{73}{122}$	28.16% $\frac{29}{103}$	23.91% $\frac{11}{46}$	33.33% $\frac{1}{3}$
2	0.00% $\frac{0}{15}$	9.09% $\frac{1}{11}$	33.33% $\frac{20}{60}$	27.56% $\frac{35}{127}$	9.02% $\frac{11}{122}$	12.62% $\frac{13}{103}$	13.04% $\frac{6}{46}$	33.33% $\frac{1}{3}$
ABETA	123 148 154	128 128 171	131 145 149	108 131 150	128 142 162	120 134 160	129 138 153	110 138 182
TAU	79.3 90.3 146.9	68.0 89.0 100.8	112.6 122.2 159.3	76.1 107.5 136.1	76.9 104.0 143.1	67.4 81.0 110.0	97.1 118.0 119.1	97.9 98.7 142.6

Values of continuous variables are represented by the lower, the median and the upper quartiles. Categorical variables are represented by percentages and frequencies. N is the number of non-missing values.

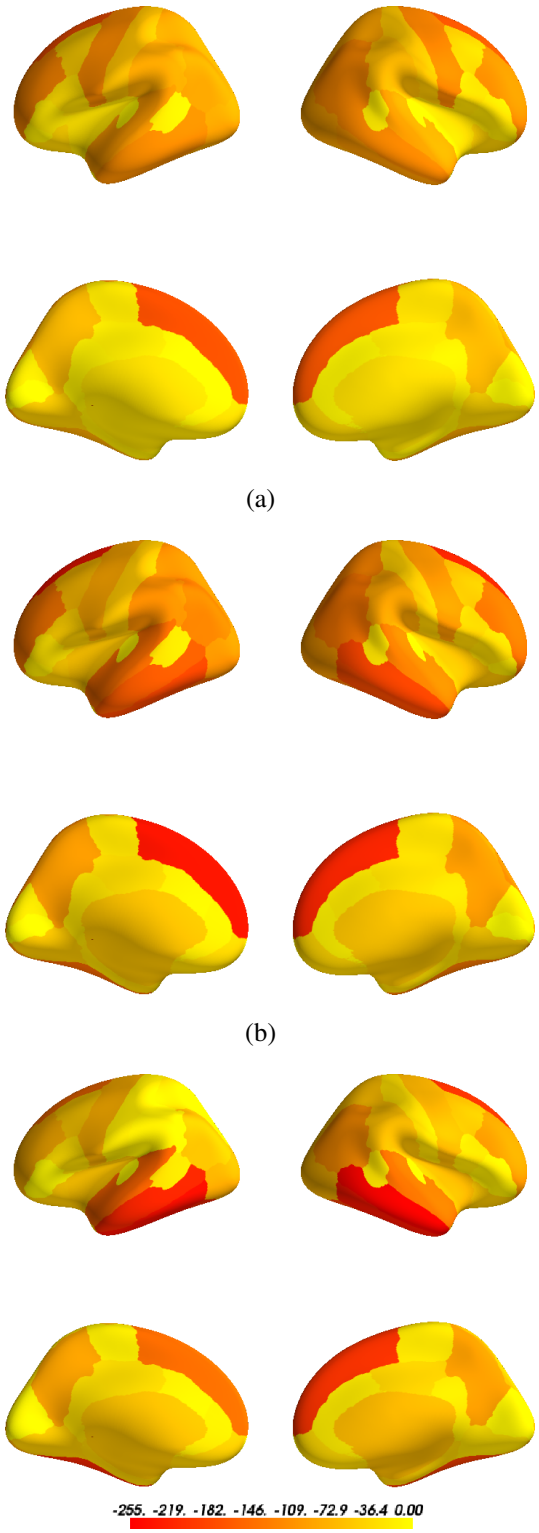
5.4.1 Annual percentage of cortical brain change

The age-related changes in 66 CV biomarkers for sHC, sMCI and sAD are illustrated in Figure 5.1. Similarly, Figure 5.2 shows the age-related changes in 66 TA biomarkers for each of these subject's groups. The figures show lateral and medial views of both hemispheres (from left to right). Brain regions are coloured according the β coefficient values of *age* (β_a) computed for MRI biomarkers in LME models (see Equation 5.1). These values are presented in measurements units of mm for TA biomarkers and in mm^3 for CV biomarkers

In Figure 5.1, red colours represent regions of the cortex with greater one-year reduction as compared to the other regions. In contrast, yellow colours represent less reduced areas. For the three kind of stable diagnostic subjects we found cortical volume reductions across the whole brain surface. In sHC subjects, the frontal lobe (including the superior frontal gyrus), the rostral middle frontal gyrus, and the precentral gyrus showed the highest degree of change. Substantial changes were also observed in the temporal lobe (including the middle temporal, the inferior temporal and the superior temporal), and the parietal lobe (inferior and superior parietal). For sMCI subjects, the frontal lobe (superior frontal gyrus) and the temporal lobe (inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus and the fusiform gyrus), showed the greatest reductions. Substantial reductions were also observed in several regions of the parietal lobe (inferior parietal, superior parietal, supramarginal and precuneus). For sAD, the temporal lobe, including the inferior temporal gyrus, the middle temporal gyrus and the superior temporal gyrus constituted the greatest change. Important reductions were also shown in the fusiform region and the superior frontal gyrus. For both sHC and sMCI, age-related atrophy was significant ($P \leq 0.01$) in most regions of both hemispheres with the exception of the pericalcarine region. In contrast, for sAD there were a limited number of regions which did not show significant longitudinal change and the cuneus was the common not significant region for both hemispheres. For all diagnostic groups, the frontal pole, the pericalcarine, and the transversal temporal gyrus showed the smallest volume reduction.

A well as cortical volume, accelerated or reduced estimates of decline in cortical thickness average with increasing age were found for all brain regions. In Figure 5.2, red-orange colours represent regions of greater cortical thinning than in the blue-cyan coloured ones. For sHC, the most accelerated decline shows up in three regions: 1) the medial temporal lobe including the entorhinal cortex, the temporal pole, the parahippocampal cortex (mainly in its left hemisphere) and the bankssts (right hemisphere); 2) the lateral lobe (inferior temporal, middle temporal and superior temporal); 3) the frontal lobe (precentral gyrus). For sMCI, the temporal lobe also showed the most accelerated thinning in both hemispheres, mainly including regions such as the enthorinal, the temporal pole, the parahippocampal, the inferior temporal, the middle temporal, the fusiform and the superior temporal. As happened

with sMCI group, sAD subject also showed the greatest thinning in the temporal lobe; and important declines were also observed in the insular cortex, the rostral anterior cingulate cortex and the medial orbitofrontal cortex. In comparison with sHC, the thinning of the temporal lobe in the sMCI and sAD groups was greater.



(a)

(b)

(c)

Fig. 5.1 Annual change in cortical volumes measured longitudinally in a sample of: (a) sHC (N=103, 61–93 years). (b) sMCI (N=98, 55–90 years); and (c) sAD (N=93, 56–90 years). Results are displayed on the inflated cortical surface of the FreeSurfer "fsaverage" average brain template. Left side: left hemisphere. Right side: right hemisphere. Top: regions on lateral view. Bottom: regions on medial views. Regions are coloured according the beta coefficient values computed for age.

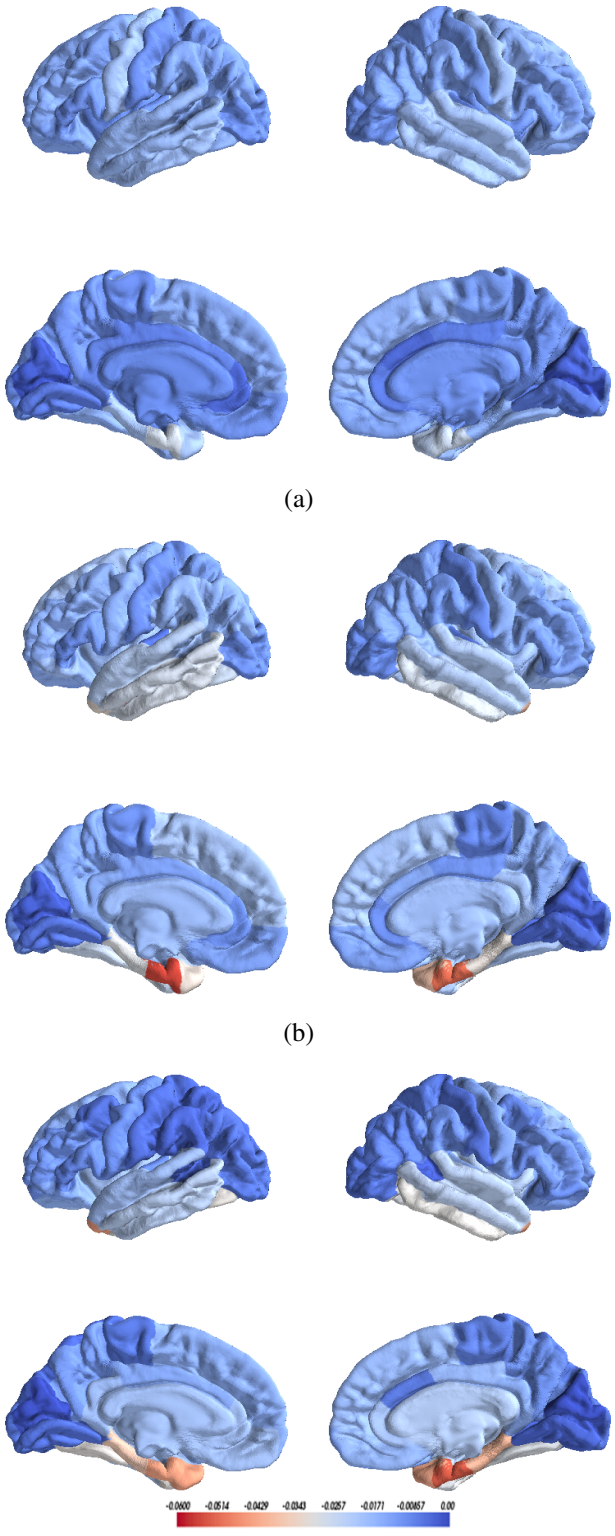
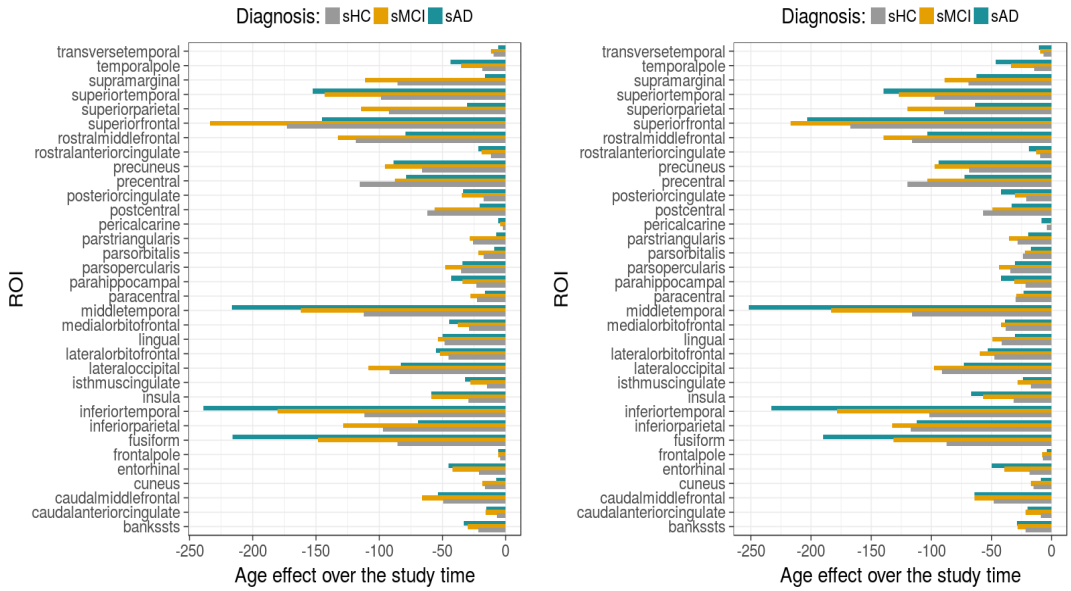


Fig. 5.2 Annual change in cortical thickness measured longitudinally in a sample of: (a) sHC (N=103, 61–93 years); (b) sMCI (N=98, 55–90 years); and (c) sAD (N=93, 56–90 years). Results are displayed on the inflated cortical surface of the FreeSurfer "fsaverage" average brain template. Left side: left hemisphere. Right side: right hemisphere. Top: regions on lateral view. Bottom: regions on medial views. Regions are coloured according the beta coefficient values computed for age.

Similarly, in order to observe more clearly differences of the age effect (β_a) for all CV and TA biomarkers, these values are summarized in Figures 5.3 and 5.4, respectively.



(a) (b)

Fig. 5.3 Summary of age effect on CV biomarkers for (a) left hemisphere and (b) right hemisphere. CV: Cortical Volume.

5.4.2 MRI features ranking

The mRMR ranking computed for each feature in the sHC vs sMCI, sHC vs sAD and sMCI vs sAD experiments for the frontal lobe, the temporal lobe and the subcortical regions are shown in Figures 5.5, 5.6 and 5.7, respectively. The mRMR order was coloured in a range from yellow to red. Red colours indicate that these features were lowest in the ranking; in contrast, blue colours were highest. The Pearson correlation was used as the distance metric. Note that certain features which were placed in the lower half of the ranking for the early age groups, were promoted to the highest positions for the older groups. However, other features remained in the upper half of the ranking for all age groups. For example, for the sHC vs sAD experiment (Figure 5.7(a)), the hippocampal volume of the left hemisphere ($LeftHippocampus.ST29SV$) occupied the first two positions in subjects aged 65 to 80 years and it remained on the upper half for the other age groups. Finally, some features were always placed in the lowest positions.

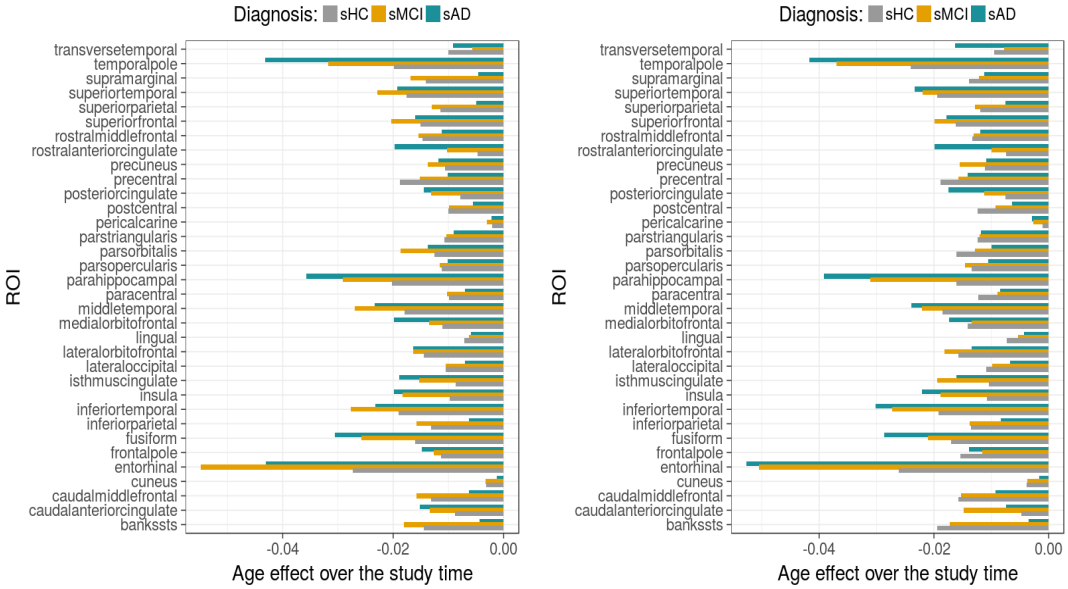


Fig. 5.4 Summary of age effect on TA biomarkers for (a) left hemisphere and (b) right hemisphere. TA: Thickness Average

5.4.3 Selected features

The SVM wrapper method allowed us to obtain the most significant features sets F' for each five-year age group and differential diagnosis problem, see Table 5.6. In total, 33 MRI biomarkers were selected and the remaining 133 MRI biomarkers were discarded. There were biomarkers selected for more than one age group and most of the selected ones correspond to the temporal and frontal lobes. Regarding subcortical regions, only the volumes of left hippocampus, right amygdala, right inferior lateral ventricle, theft choroid plexus, and the corpus callosum central were selected. Also, the table shows that *age*, *gender* and *educ* were always selected because they were included a priori. As mentioned in Section 5.3.5, for the SVM models building for early diagnosis estimation and prediction, the aggregated set of features F^{add} (instead of F') was established as input. The subset F^{add} was obtained by grouping the age-specific features from previous F' (see Table 5.7). Observe that, within all the SVM models, *gender*, *age* and *educ* were invariably included as explanatory features, and MMSE and CDRSB were included only in several models.

5.4.4 Classification of current diagnosis

Tables 5.8 and 5.9 show the results of models focused on the ECD problem when including or not neuro-psychological test as input features. These models were trained on the diagnosis at the time of the clinical tests, structured in 5-year blocks and three differential diagnosis experiments. The last column of both tables also shows the weighted average across all the age groups. The classification quality indicators are accuracy (ACC), sensitivity (SEN), and specificity (SPE).

The models including the MMSE and the CDRSB features as input show very high values for accuracy, sensitivity and specificity for all age groups, mainly for the AD vs HC results, where all the indicators achieved values well above 86%, and values of 100% for some age groups. For the AD vs MCI experiments, which are intrinsically harder than AD vs HC, show an average accuracy above 84%, the performance across all age groups being quite consistent. Finally, the MCI vs HC models show an average accuracy above 82%, with a classification performance that apparently declines with age.

In contrast, the models that do not include the neuro-psychological features as input show lower average accuracies than previous models, obtaining 86.23% for AD vs HC, 68.73% for AD vs MCI and 72.75% for MCI vs AD.

By comparing these results with comparable results from the literature previously presented in Section 2.5 (see Table 7.2), our method shows the highest indicators in the AD vs HC classification, the highest specificity in MCI vs HC, as well as the highest accuracy in AD vs MCI.

5.4.5 Diagnosis prediction advancement

Tables 5.10 and 5.11 show the detailed results of the experiments carried out to address the PFD problem, where classifiers were trained on the last known diagnosis of each subject in the dataset, including or not the MMSE and the CDRSB features as model input. The results are structured both by age group and by differential diagnosis. The absence of enough suitable subjects necessarily limited the age groups and differential diagnoses we studied. The diagnosis advancement rows indicate the average number of years improvement in predicting the future diagnosis of the subjects. The possible advancement depends on the specific clinical history of the subjects in the dataset, i.e. the difference in years from their clinical tests to their last known diagnosis.

When including the neuro-psychological tests as input for the AD vs HC experiment, all classification indicators are well above 95%, with a highest accuracy of 98.75% on the 70–75 age group. The diagnosis advancement ranges from 1.48 to 1.85 years. The results of

the AD vs MCI experiments, though lower than in AD vs HC due to their intrinsic greater difficulty, also show high performance indicators, especially for classification accuracy and sensitivity. On the other hand, when the MMSE and the CDRSB features were excluded from the models, accuracies for AD vs HC were lower than the accuracies obtained from the previous model, but also remained above 90%. Furthermore, for AD vs MCI, the classifiers achieved accuracies up to 90%, slightly lower than in the previous case.

5.5 Conclusions

To address the early diagnosis and prognosis of AD, the proposed method was focused on a univariate and multivariate analysis of the brain structures change through data stratification by diagnosis and age groups. The current results confirm predominant cortical fronto-temporal patterns of atrophy in sHC subjects across the adult age-span (Figures 5.1(a) and 5.2(a)). Moreover, most selected MRI-based biomarkers corresponded to the temporal and frontal lobes, which also support findings from the previous univariate analysis stage. This pattern has been previously described in other cross-sectional [104] and longitudinal studies [43, 64, 44] and could explain the ageing-related decline in specific cognitive abilities. Also, a fronto-temporal pattern was also observed in sMCI (Figures 5.1(b) and 5.2(b)) and in sAD (figures 5.1(c) and 5.2(c)) subjects, where the change rate observed in sAD is slightly different from the change observed in MCI and is more differentiated than sHC. All these results related with the brain change also coincide to a large extent with the results obtained from M_{res} (see Section 4.5). As described previously by Fjell et al. [45], all these findings could confirm that the fronto-temporal change corresponds to an inevitable process related to normal ageing and is not necessarily an AD-specific one. Furthermore, it is possible that the vulnerability of these areas to normal ageing-related decline contributes to their vulnerability to AD-related atrophy.

Regarding differential diagnosis applied on the estimation of the current diagnosis, classifiers including neuro-psychological features produce better results than comparable methods in the literature for most classification quality indicators, specially on AD vs HC experiment where all the indicators rank in the first place. Furthermore, when comparing to the models performance obtained through our first method (Table 7.2), we confirmed both the inclusion of an FS and the approach of aggregating selected features as age groups get older, prove to be a robust approach for improving classification performance of ECD models (over 98% ACC for AD vs HC, over 84% ACC for AD vs MCI and over 82% ACC for MCI vs HC). In fact, the FS task not only helped us in improving model accuracy, but was also useful for visualization and interpretation of data.

Likewise, the method$_{raw}$ also improves the prediction of future diagnosis given the current clinical tests, both in prediction quality indicators as well as in the amount of time by which the diagnosis is advanced. Results of models focused on the early diagnosis of disease progression outperform results obtained by the previous method (M$_{res}$), while also increasing slightly the span of the age advancement in the diagnosis.

In summary, these results demonstrate that M$_{raw}$ focused on combining diagnosis-specific feature selection and age stratification is an efficient approach to address the main problems established in this thesis.

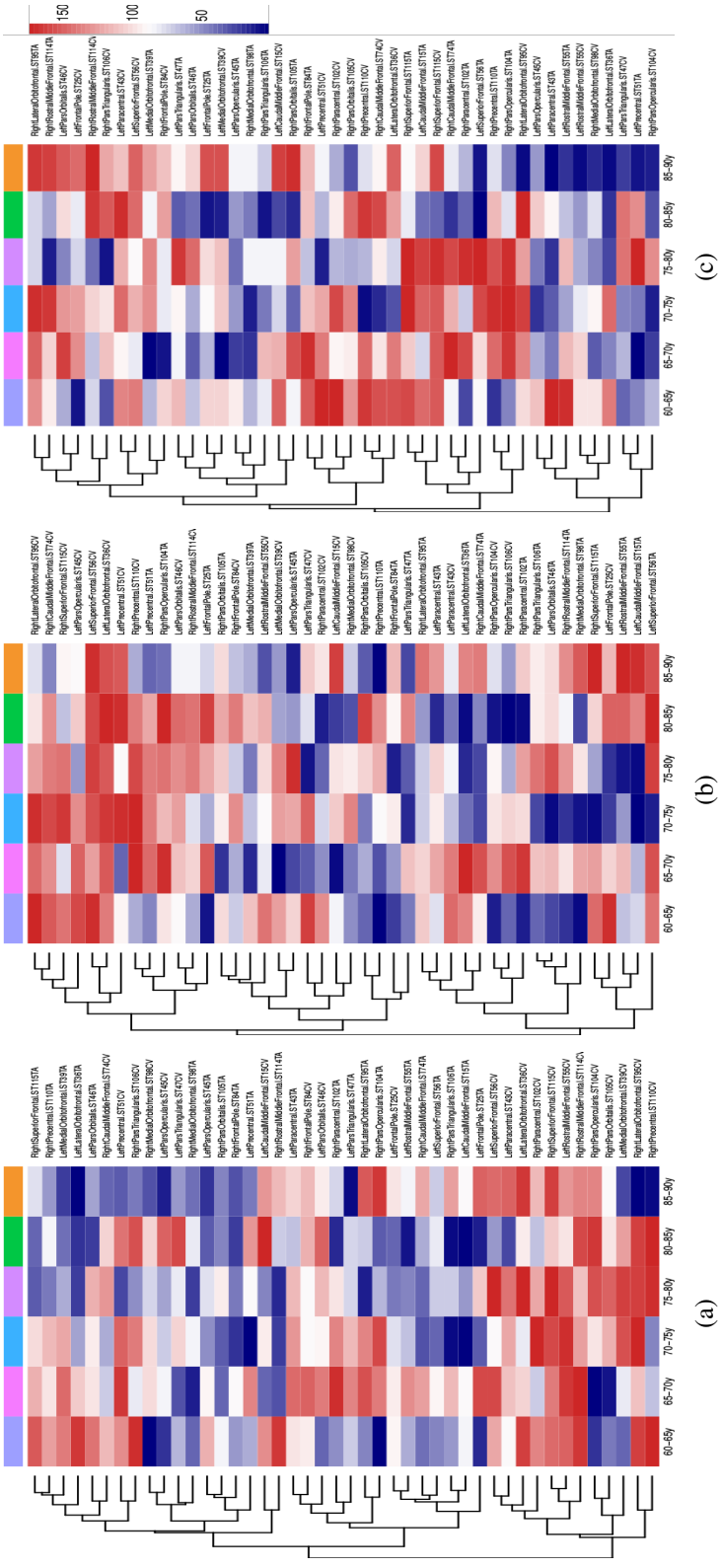


Fig. 5.5 mRMR ranking achieved by the MRI biomarkers of the frontal lobe. (a) AD vs HC. (b) MCI vs HC. (c) AD vs MCI Red colours mean that features were in the lowest ranking positions; blue colours that the features ranked in the first positions.

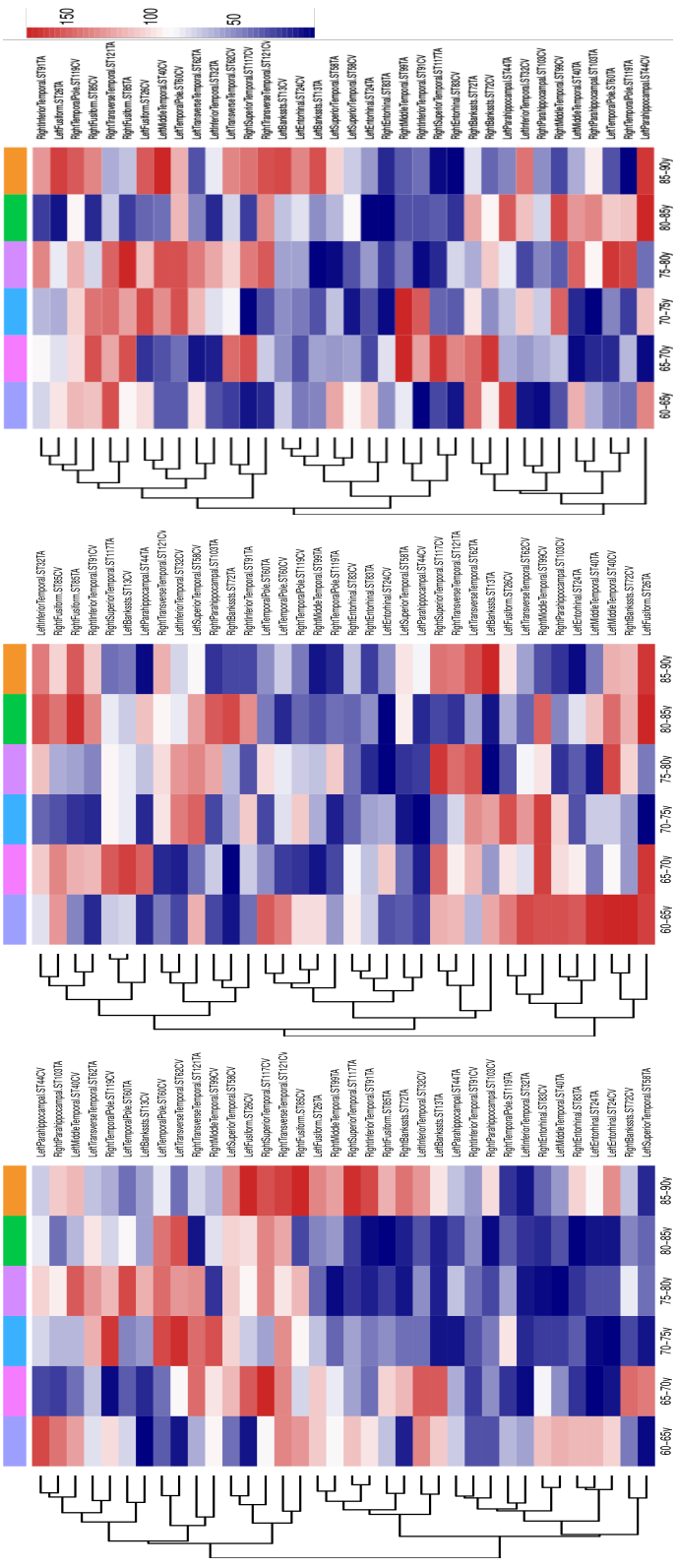


Fig. 5.6 mRMR ranking achieved by the MRI biomarkers of the temporal lobe. (a) AD vs HC. (b) MCI vs HC. (c) AD vs MCI Red colours mean that features were in the lowest ranking positions; blue colours that the features ranked in the first positions.

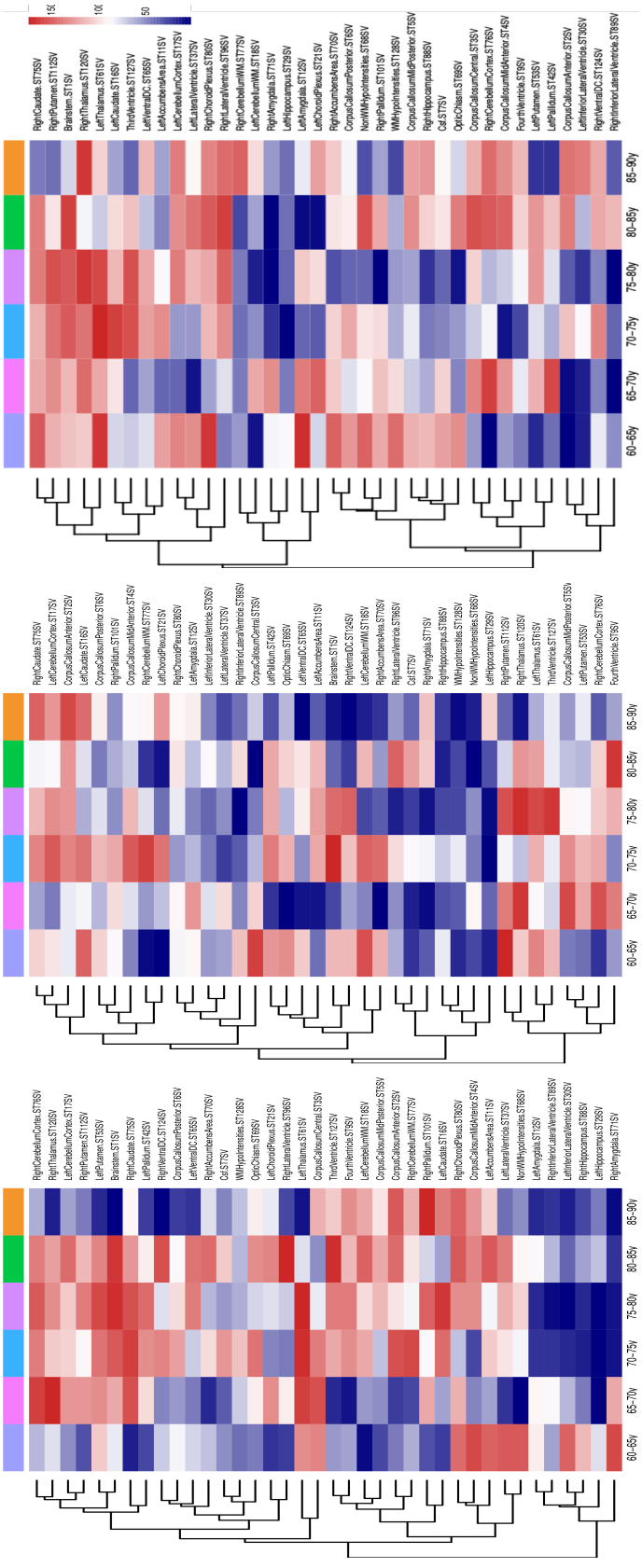


Fig. 5.7 mRMR ranking achieved by the MRI biomarkers of the subcortical region. (a) AD vs HC. (b) MCI vs HC. (c) AD vs MCI
Red colours mean that features were in the lowest ranking positions; blue colours that the features ranked in the first positions.

Table 5.6 Features selected for each experiment/age group by using the SVM wrapper

			sHC vs sAD						sHC vs sMCI						sMCI vs sAD						
	Main Region	Selected Features	60–65	65–70	70–75	75–80	80–85	85–90	60–65	65–70	70–75	75–80	80–85	85–90	55–60	60–65	65–70	70–75	75–80	80–85	85–90
1		age	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	—	educ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	—	gender	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	cingulate cortex	LeftIsthmusCingulate					✓														
5	cingulate cortex	RightRostralAnteriorCingulate										✓									
6	frontal lobe	LeftCaudalMiddleFrontal					✓				✓										
7	frontal lobe	RightLateralOrbitofrontal						✓													
8	frontal lobe	LeftMedialOrbitofrontal								✓											
9	frontal lobe	RightParacentral											✓								
10	frontal lobe	LeftParsOrbitalis									✓										
11	frontal lobe	RightParsTriangularis					✓														
12	frontal lobe	RightPrecentral							✓												
13	frontal lobe	LeftPrecentral			✓																
14	frontal lobe	RightSuperiorFrontal									✓										
15	frontal lobe	LeftSuperiorFrontal																		✓	
16	occipital lobe	RightCuneus	✓																		
17	occipital lobe	RightLateralOccipital							✓												
18	occipital lobe	LeftLateralOccipital							✓												
19	parietal lobe	LeftPrecuneus														✓					
20	subcortical brain	RightAmygdala																	✓	✓	
21	subcortical brain	LeftChoroidPlexus							✓											✓	
22	subcortical brain	CorpusCallosumCentral											✓								
23	subcortical brain	LeftHippocampus			✓						✓										
24	subcortical brain	RightInferiorLateralVentricle															✓				
25	temporal lobe	LeftBankssts										✓									
26	temporal lobe	RightBankssts								✓											
27	temporal lobe	LeftEntorhinal		✓																✓	
28	temporal lobe	RightEntorhinal.ST83TA					✓													✓	
29	temporal lobe	LeftEntorhinal										✓	✓								
30	temporal lobe	LeftFusiform									✓										
31	temporal lobe	RightFusiform					✓														
32	temporal lobe	RightInferiorTemporal					✓														
33	temporal lobe	LeftMiddleTemporal				✓															
34	temporal lobe	LeftParahippocampal									✓										
35	temporal lobe	RightSuperiorTemporal																✓			
36	temporal lobe	RightTemporalPole																			✓

Table 5.7 Features sets used for each five-year age prediction and diagnosis models

				sHC vs sAD						sHC vs sMCI						sMCI vs sAD						
	Lobe	Feature	Type	60–65	65–70	70–75	75–80	80–85	85–90	60–65	65–70	70–75	75–80	80–85	85–90	55–60	60–65	65–70	70–75	75–80	80–85	85–90
1	—	age_img		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	—	educ		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	—	gender		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	—	*MMSE		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	—	*CDRSB		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6	cingulate cortex	leftisthmuscingulate	TA					✓	✓													
7	cingulate cortex	rightrostralanteriorcingulate	CV										✓	✓	✓							
8	frontal lobe	leftcaudalmiddlefrontal	TA					✓	✓			✓	✓	✓	✓							
9	frontal lobe	rightlateralorbitofrontal	CV						✓													
10	frontal lobe	leftmedialorbitofrontal	CV								✓	✓	✓	✓	✓							
11	frontal lobe	rightparacentral	TA											✓	✓							
12	frontal lobe	leftparsorbitalis	TA									✓	✓	✓	✓							
13	frontal lobe	rightparstriangularis	TA					✓	✓													
14	frontal lobe	rightprecentral	TA							✓	✓	✓	✓	✓	✓							
15	frontal lobe	leftprecentral	TA			✓	✓	✓	✓													
16	frontal lobe	rightsuperiorfrontal	TA									✓	✓	✓	✓							
17	frontal lobe	leftsuperiorfrontal	TA																		✓	✓
18	occipital lobe	rightcuneus	CV	✓	✓	✓	✓	✓	✓													
19	occipital lobe	rightlateraloccipital	TA							✓	✓	✓	✓	✓	✓							
20	occipital lobe	leftlateraloccipital	TA							✓	✓	✓	✓	✓	✓							
21	parietal lobe	leftprecuneus	TA														✓	✓	✓	✓	✓	✓
22	subcortical brain	rightamygdala	SV																	✓	✓	✓
23	subcortical brain	leftchoroidplexus	SV							✓	✓	✓	✓	✓	✓						✓	✓
24	subcortical brain	corpuscallosumcentral	SV											✓	✓							
25	subcortical brain	lefthippocampus	SV			✓	✓	✓	✓			✓	✓	✓	✓							
26	subcortical brain	rightinferiorlateralventricle	SV															✓	✓	✓	✓	✓
27	temporal lobe	leftbankssts	TA										✓	✓	✓							
28	temporal lobe	rightbankssts	TA								✓	✓	✓	✓	✓							
29	temporal lobe	leftentorhinal	TA		✓	✓	✓	✓	✓												✓	✓
30	temporal lobe	rightentorhinal	TA					✓	✓												✓	✓
31	temporal lobe	leftentorhinal	CV										✓	✓	✓							
32	temporal lobe	leftfusiform	TA									✓	✓	✓	✓							
33	temporal lobe	rightfusiform	TA					✓	✓													
34	temporal lobe	rightinferiortemporal	TA					✓	✓													
35	temporal lobe	leftmiddletemporal	TA				✓	✓	✓													
36	temporal lobe	leftparahippocampal	CV									✓	✓	✓	✓							
37	temporal lobe	rightsuperiortemporal	CV																✓	✓	✓	✓
38	temporal lobe	righttemporalpole	TA																			✓

* These features were excluded when we focused on addressing the classification problems excluding neuro-psychological tests.

Table 5.8 Performance of diagnostic classifier including neuropsychological features

5–yr	65–70	70–75	75–80	80–85	85–90	Weighted Average
AD vs HC						
No. test subjects	12	97	115	57	22	
Accuracy	100.00%	100.00%	99.13%	94.74%	100.00%	98.68%
Specificity	100.00%	100.00%	98.77%	100.00%	100.00%	99.53%
Sensitivity	100.00%	100.00%	100.00%	86.36%	100.00%	97.43%
AD vs MCI						
No. test subjects	34	59	87	58	34	
Accuracy	91.18%	86.44%	82.76%	75.86%	94.12%	84.56%
Specificity	95.83%	86.49%	84.91%	80.55%	90.47%	86.38%
Sensitivity	80.00%	86.36%	79.41%	68.18%	100.00%	81.17%
MCI vs HC						
No. test subjects	26	112	134	71	30	
Accuracy	88.46%	84.82%	79.10%	87.32%	70.00%	82.31%
Specificity	00.00%	85.33%	81.48%	97.14%	0.00%	77.38%
Sensitivity	95.83%	83.78%	75.47%	77.78%	100.00%	81.80%

Table 5.9 Performance of diagnostic classifier excluding neuro-psychological features

5–yr	65–70	70–75	75–80	80–85	85–90	Weighted Average
AD vs HC						
No. test subjects	12	97	115	57	22	
Accuracy	58.33%	84.53%	86.08%	89.83%	100.00%	86.23%
Specificity	0%	89.33%	82.71%	100.00%	100.00%	86.16%
Sensitivity	70.00%	68.18%	94.11%	72.72%	100.00%	81.21%
AD vs MCI						
No. test subjects	34	59	87	58	34	
Accuracy	73.52%	72.88%	64.77%	65.00%	73.52%	68.73%
Specificity	95.83%	94.59%	92.96%	65.78%	90.47%	77.83%
Sensitivity	20.00%	36.36%	67.65%	63.64%	46.15%	51.51%
MCI vs HC						
No. test subjects	26	112	134	71	30	
Accuracy	92.30%	76.78%	62.96%	78.66%	70.00%	72.75%
Specificity	00.00%	81.33%	79.01%	81.08%	00.00%	68.40%
Sensitivity	100.00%	67.57%	38.88%	76.32%	100.00%	63.87%

Table 5.10 Prediction of the future diagnosis including neuropsychological features

5–yr	70–75	75–80	80–85	Weighted Average
AD vs HC				
No. test subjects	40	40	27	
Accuracy	98.75%	97.25%	95.56%	97.38%
Sensitivity	99.60%	98.33%	95.00%	97.96%
Specificity	97.33%	95.62%	97.14%	96.64%
Time advancement (years)	1.55	1.62	1.87	
AD vs MCI				
No. test subjects	32	34	27	
Accuracy	88.12%	85.30%	93.33%	88.60%
Sensitivity	97.60%	92.92%	99.00%	96.30%
Specificity	54.29%	67.00%	77.14%	65.57%
Time advancement (years)	1.50	1.66	1.64	

Table 5.11 Prediction of the future diagnosis excluding neuropsychological features

5–yr	70–75	75–80	80–85	Weighted Average
AD vs HC				
No. test subjects	40	40	27	
Accuracy	92.00%	90.00%	92.96%	91.49%
Sensitivity	96.80%	96.25%	96.00%	96.39%
Specificity	84.00%	80.63%	84.28%	82.81%
Time advancement (years)	1.52	1.56	1.91	
AD vs MCI				
No. test subjects	32	34	27	
Accuracy	88.44%	82.06%	90.00%	85.18%
Sensitivity	98.80%	87.92%	96.00%	94.00%
Specificity	51.43%	68.00%	72.86%	63.71%
Time advancement (years)	1.54	1.32	1.56	

Chapter 6

Validation of methods.

6.1 Introduction

The reliability of both methods M_{res} and the M_{raw} to correctly identify AD vs HC subjects was evaluated and compared by testing these methods on a new dataset obtained from the publicly accessible database of OASIS cross-sectional study (Section 2.4). Because in that dataset, the clinical diagnostic assigned to participants was directly obtained from CDRGLOBAL scores, methods comparison tasks were focused on MRI-based biomarkers, excluding information about neuro-psychological features (CDR-based and MMSE scores). Furthermore, few features were customised or excluded to facilitate such comparison. Models derived from both methods were firstly re-trained from ADNI training data and then tested separately on remaining ADNI data, as well as, on OASIS data.

6.2 Data

Data was integrated with MRI-based biomarkers and sociodemographic variables (years of education, gender and age) from ADNI and OASIS databases. ADNI data included 313 subjects clinically diagnosis as stable HC (N=151) and stable AD (N=162) during five years follow up. Details about ADNI database was briefly introduced in Section 2.3. The criteria for including MRI biomarkers and participants took into account in M_{res} and the M_{raw} have been previously described in Sections 4.2 and 5.2, respectively. On the other hand, OASIS data included 134 subjects (93 AD and 41 HC) observed at baseline. In this dataset, two biomarkers associated with the inferior lateral ventricle were missed. Thus the initial set of MRI based biomarkers (N=166) utilised in both previous methods was reduced to 164, including measures for CV (N=66), TA (N=66) and SV (N=32).

6.3 Methods

Preliminary data processing was necessary to match OASIS and ADNI data. Because education variable observed in OASIS subjects correspond to the following levels of education 1: less than high school grade, 2: high school grade, 3: some college, 4: college grade and 5: beyond college; it was necessary to convert those levels into years of education in order to match them with ADNI subject's education values. Thus, education levels 1, 2, 3, 4 and 5 were replaced by 8, 12, 15, 18 and 20 education years, respectively. Then, the feature *educ* was normalised to [0,1].

In several cases, models derived from both M_{res} and M_{raw} methods were re-built for adjusting them to new data characteristics. After that, ADNI data were used to train SVM classifiers and then the performance of those models was evaluated by testing them separately on new data from ADNI and OASIS.

For the M_{res}, 107 variant ROIs (*vr*) and 57 quasi-variant ROIs (*qvr*) were utilised to re-built the female ageing-based null model from 24 women diagnosed as stable HC. Likewise, 95 variant ROIs (*vr*) and 69 quasi-variant ROIs (*qvr*) were included into the male null model, which was built from 22 stable HC men. In Section 4.3.2 was described with more detail the ageing-based null model approach and in Table 4.5 were described the baseline characteristics of those 66 stable HC subjects included on building those models. After adjustment of the null model, residuals were computed for ADNI HC and AD remainders, as well as, for OASIS HC and AD subjects. As it was described in Section 4.3.3, residuals represent the deviation values of observed MRI-based biomarkers from an inferred healthy subject-specific trend computed through null models. Then, SVM models for the HC vs AD classification experiment were trained separately for males and females residuals taking into account the same configuration described in Section 4.3.3. For that purpose, only residuals of *vr* biomarkers and age (age of the subject at the time of the visit) were included as input features.

Likewise, for the M_{raw} method, which is focused on age group-depended brain changes, it was not necessary to re-train the SVM wrapper for feature selection tasks because the OASIS dataset included the 33 MRI-based biomarkers selected by that wrapper, see Section 5.6. Because the OASIS diagnosis label is generated directly from CDR-related scores, the inclusion of that feature could bias and compromise the classifier. Thus, the SVM models built for each five-year age group for the HC vs AD experiment were re-trained excluding neuropsychological features such as the CDR-related and MMSE scores. Details about implementation of the SVM models for the estimation of the current diagnosis (ECD) were described in Section 5.3.5. ADNI remainders not used in feature selection tasks and all OASIS data were used to assess the classifier performance. Because testing age groups of

both datasets do not contain enough data to perform the experiments as described, only the samples of subjects aged 65–90 years were analysed. As with the previous method, possible differences between five-year age groups in ADNI and OASIS populations were examined through quartiles representation of MRI-based biomarkers stratified by age group, diagnosis and data source.

To evaluate possible differences between ADNI and OASIS populations, it was carried out a normality analysis followed by statistical tests to determine whether the means of two groups are equal to each other.

Furthermore, regardless of the method, all SVM classifiers were built through the repeated random sampling approach described in Section 2.2.6. In summary, models were built from 80% of ADNI subjects randomly selected for training, and the ADNI remaining 20%, as well as, the 100% of OASIS subjects were separately used for testing. This approach was repeated ten times, then the estimates of classifiers performance were obtained by averaging all ten accuracies, sensitivities and specificities. For splitting ADNI data, we were careful that the subjects selected for training were not selected for testing.

6.4 Results

Tables 6.1 and 6.2 show the statistical descriptives of ADNI and OASIS data used to re-train and test classifiers. There were not available information about APOE genotype in OASIS subjects.

Table 6.1 Statistical descriptors of ADNI data at baseline grouped by diagnostic group

	AD	HC
	$N = 162$	$N = 151$
Gender		
Female	48% $\frac{78}{162}$	46% $\frac{69}{151}$
Male	52% $\frac{84}{162}$	54% $\frac{82}{151}$
Age	$_{70.7}$ 75.3 $_{80.4}$ (75.0 ± 7.5)	$_{72.7}$ 76.0 $_{79.0}$ (76.3 ± 5.0)
Educ	$_{0.50}$ 0.69 $_{0.75}$ (0.67 ±0.20)	$_{0.66}$ 0.75 $_{0.88}$ (0.77 ±0.18)
APOE4		
0	33.3% $\frac{54}{162}$	70.9% $\frac{107}{151}$
1	45.7% $\frac{74}{162}$	26.5% $\frac{40}{151}$
2	21.0% $\frac{34}{162}$	2.6% $\frac{4}{151}$
MMSE	$_{22.0}$ 23.0 $_{25.0}$ (23.4 ± 1.9)	$_{29.0}$ 29.0 $_{30.0}$ (29.1 ± 1.0)
CDGLOBAL		
0	0% $\frac{0}{162}$	100% $\frac{151}{151}$
0.5	53% $\frac{86}{162}$	0% $\frac{0}{151}$
1	47% $\frac{76}{162}$	0% $\frac{0}{151}$

$_a\,b\,_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD.

Table 6.2 Statistical descriptors of OASIS data grouped by diagnostic group

	AD	HC
	$N = 93$	$N = 41$
Gender		
Female	58% $\frac{54}{93}$	76% $\frac{31}{41}$
Male	42% $\frac{39}{93}$	24% $\frac{10}{41}$
Age	$_{72.0}$ 77.0 $_{81.0}$ (76.7 ± 7.3)	$_{67.0}$ 71.0 $_{80.0}$ (73.6 ± 9.2)
Educ	$_{0.50}$ 0.50 $_{0.88}$ (0.63 ±0.24)	$_{0.69}$ 0.88 $_{1.00}$ (0.79 ±0.22)
MMSE	$_{22.0}$ 26.0 $_{28.0}$ (24.5 ± 4.1)	$_{30.0}$ 30.0 $_{30.0}$ (30.0 ± 0.0)
CDGLOBAL		
0	0.0% $\frac{0}{93}$	100.0% $\frac{41}{41}$
0.5	73.1% $\frac{68}{93}$	0.0% $\frac{0}{41}$
1	24.7% $\frac{23}{93}$	0.0% $\frac{0}{41}$
2	2.1% $\frac{2}{93}$	0.0% $\frac{0}{41}$

$_a\,b\,_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD.

6.4.1 Analysis of residual's distribution in ADNI and OASIS data

Distributions of residuals for each *vr* biomarker in the HC group of both datasets were checked for normality using the Shapiro-Wilks test. For many features, the null hypothesis was rejected (at 0.05 significance level). Thus, the Wilcoxon-Matt-Whitney test was used to test the hypothesis that distribution of each *vr* residual in the HC OASIS population is the same than in the HC ADNI population without assuming them to follow a normal distribution. At a 0.05 significance level, it was observed that the distribution of few residuals is identical for both populations. Tables 6.3 and 6.4 show the list of biomarkers showing identical distributions (p-value$\geq$ 0.05) for HC females (30 out of 107 *vr*) and HC males (48 out of 107 *vr*), respectively.

Table 6.3 Wilcoxon-Matt-Whitney test for biomarkers showing similar distribution in HC females.

	MRI biomarker	Hemisphere	Measure type	Lobe	p-value
1	rh_isthmuscingulate_volume	rh	CV	cingulate cortex	0.061
2	lh_superiortemporal_volume	lh	CV	temporal lobe	0.076
3	leftamygdala	lh	SV		0.079
4	rh_rostralmiddlefrontal_thickness	rh	TA	frontal lobe	0.113
5	rh_bankssts_thickness	rh	TA	temporal lobe	0.123
6	lh_superiorparietal_thickness	lh	TA	parietal lobe	0.155
7	rh_entorhinal_volume	rh	CV	temporal lobe	0.171
8	thirdventricle	bilateral	SV		0.177
9	leftpallidum	lh	SV		0.192
10	righthippocampus	rh	SV		0.202
11	rh_supramarginal_volume	rh	CV	parietal lobe	0.229
12	rh_superiortemporal_thickness	rh	TA	temporal lobe	0.235
13	lh_cuneus_volume	lh	CV	occipital lobe	0.280
14	rh_temporalpole_volume	rh	CV	temporal lobe	0.298
15	rh_middletemporal_volume	rh	CV	temporal lobe	0.336
16	lh_middletemporal_volume	lh	CV	temporal lobe	0.338
17	lh_lateralorbitofrontal_volume	lh	CV	frontal lobe	0.344
18	lh_temporalpole_thickness	lh	TA	temporal lobe	0.361
19	rh_inferiortemporal_volume	rh	CV	temporal lobe	0.399
20	rh_parahippocampal_volume	rh	CV	temporal lobe	0.434
21	rh_bankssts_volume	rh	CV	temporal lobe	0.514
22	lh_lateraloccipital_thickness	lh	TA	occipital lobe	0.569
23	lh_rostralmiddlefrontal_thickness	lh	TA	frontal lobe	0.639
24	rh_inferiorparietal_thickness	rh	TA	parietal lobe	0.641
25	lh_superiorparietal_volume	lh	CV	parietal lobe	0.669
26	lefthippocampus	lh	SV		0.745
27	corpuscallosummidanterior	bilateral	SV		0.778
28	corpuscallosumcentral	bilateral	SV		0.818
29	lh_isthmuscingulate_thickness	lh	TA	cingulate cortex	0.824
30	rh_inferiortemporal_thickness	rh	TA	temporal lobe	0.961

CV: Cortical volume. SV: Subcortical volume. TA: Thickness Average. rh: Right hemisphere. lh: Left hemisphere.

Table 6.4 Wilcoxon-Matt-Whitney test for biomarkers showing similar distribution in HC males.

	MRI biomarker	Hemisphere	Measure type	Lobe	p-value
1	lh_lingual_thickness	lh	TA	occipital lobe	0.059
2	rh_medialorbitofrontal_volume	rh	CV	frontal lobe	0.069
3	rh_frontalpole_thickness	rh	TA	frontal lobe	0.071
4	rh_middletemporal_volume	rh	CV	temporal lobe	0.074
5	lh_superiorparietal_volume	lh	CV	parietal lobe	0.088
6	lh_middletemporal_thickness	lh	TA	temporal lobe	0.094
7	lh_caudalmiddlefrontal_thickness	lh	TA	frontal lobe	0.097
8	rightaccumbensarea	rh	SV		0.138
9	rh_lateralorbitofrontal_thickness	rh	TA	frontal lobe	0.141
10	rh_medialorbitofrontal_thickness	rh	TA	frontal lobe	0.141
11	rh_parstriangularis_thickness	rh	TA	frontal lobe	0.151
12	rh_lateralorbitofrontal_volume	rh	CV	frontal lobe	0.182
13	lh_inferiortemporal_thickness	lh	TA	temporal lobe	0.190
14	lh_superiortemporal_thickness	lh	TA	temporal lobe	0.207
15	lh_rostralmiddlefrontal_thickness	lh	TA	frontal lobe	0.210
16	rh_bankssts_thickness	rh	TA	temporal lobe	0.224
17	lh_posteriorcingulate_thickness	lh	TA	cingulate cortex	0.228
18	lh_parsorbitalis_thickness	lh	TA	frontal lobe	0.233
19	lh_superiorfrontal_thickness	lh	TA	frontal lobe	0.240
20	lh_superiortemporal_volume	lh	CV	temporal lobe	0.265
21	rh_fusiform_thickness	rh	TA	temporal lobe	0.277
22	rh_parsopercularis_thickness	rh	TA	frontal lobe	0.302
23	rh_parsorbitalis_thickness	rh	TA	frontal lobe	0.319
24	lh_superiorparietal_thickness	lh	TA	parietal lobe	0.347
25	lh_supramarginal_thickness	lh	TA	parietal lobe	0.347
26	lh_inferiorparietal_thickness	lh	TA	parietal lobe	0.367
27	leftthalamus	lh	SV		0.373
28	lh_temporalpole_thickness	lh	TA	temporal lobe	0.388
29	lh_lateralorbitofrontal_thickness	lh	TA	frontal lobe	0.410
30	rh_postcentral_thickness	rh	TA	parietal lobe	0.419
31	rh_superiorparietal_thickness	rh	TA	parietal lobe	0.473
32	lh_supramarginal_volume	lh	CV	parietal lobe	0.499
33	lh_middletemporal_volume	lh	CV	temporal lobe	0.539
34	rh_rostralmiddlefrontal_thickness	rh	TA	frontal lobe	0.550
35	corpuscallosumanterior	bilateral	SV		0.580
36	rh_cuneus_thickness	rh	TA	occipital lobe	0.585
37	brainstem	bilateral	SV		0.597
38	lh_fusiform_thickness	lh	TA	temporal lobe	0.718
39	rh_precuneus_volume	rh	CV	parietal lobe	0.718
40	rh_inferiortemporal_volume	rh	CV	temporal lobe	0.720
41	rh_superiorfrontal_thickness	rh	TA	frontal lobe	0.790
42	rightlateralventricle	rh	SV		0.865
43	rh_supramarginal_thickness	rh	TA	parietal lobe	0.876
44	rightpallidum	rh	SV		0.911
45	leftlateralventricle	lh	SV		0.924
46	rh_inferiorparietal_thickness	rh	TA	parietal lobe	0.926
47	lh_precuneus_thickness	lh	TA	parietal lobe	0.937
48	rh_inferiorparietal_volume	rh	CV	parietal lobe	0.999

CV: Cortical volume. SV: Subcortical volume. TA: Thickness Average. rh: Right hemisphere. lh: Left hemisphere.

Furthermore, in order to visualise more specific differences between groups, quartiles of residuals were generated for all SV, CV and TA MRI-based biomarkers stratified by gender, diagnosis groups and five-year age groups. In many cases, it was observed that boxplots computed for biomarkers from HC OASIS subjects are quite different from their counterpart in ADNI, despite belonging to the same age group. For example, Figures 6.1 and 6.2 show quartiles comparison of females and males residuals, respectively, for SV MRI-based biomarkers measured in the left hemisphere. These figures show clear differences between ADNI and OASIS cohorts, not only regarding the gender and diagnostic group but also within the same age group. Likewise, we observed that there are age groups with no data for both cohorts. The remaining plots are shown in Section A1.1.

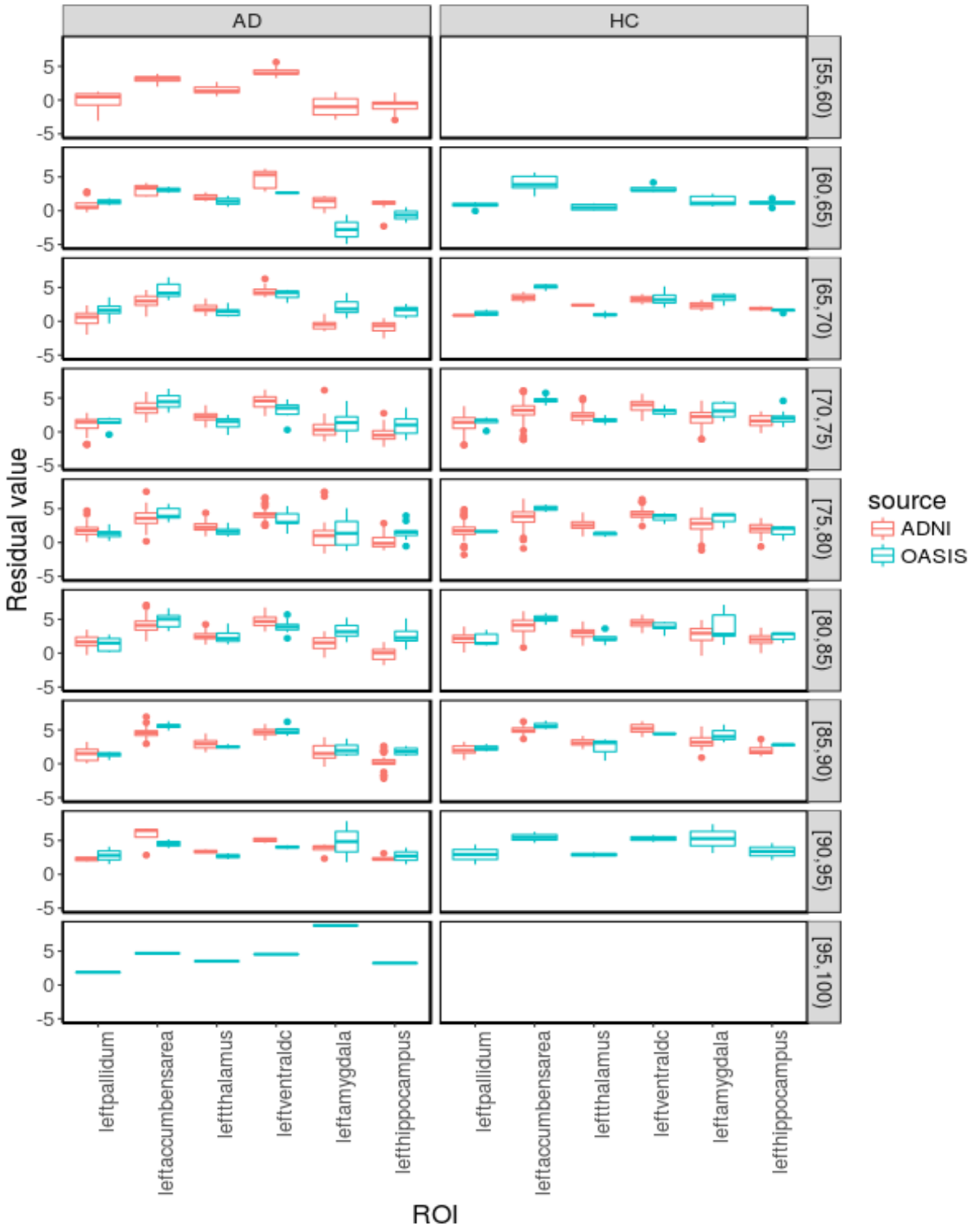


Fig. 6.1 Left hemisphere: Quantiles comparison of SV biomarker residuals for females of ADNI and OASIS data.

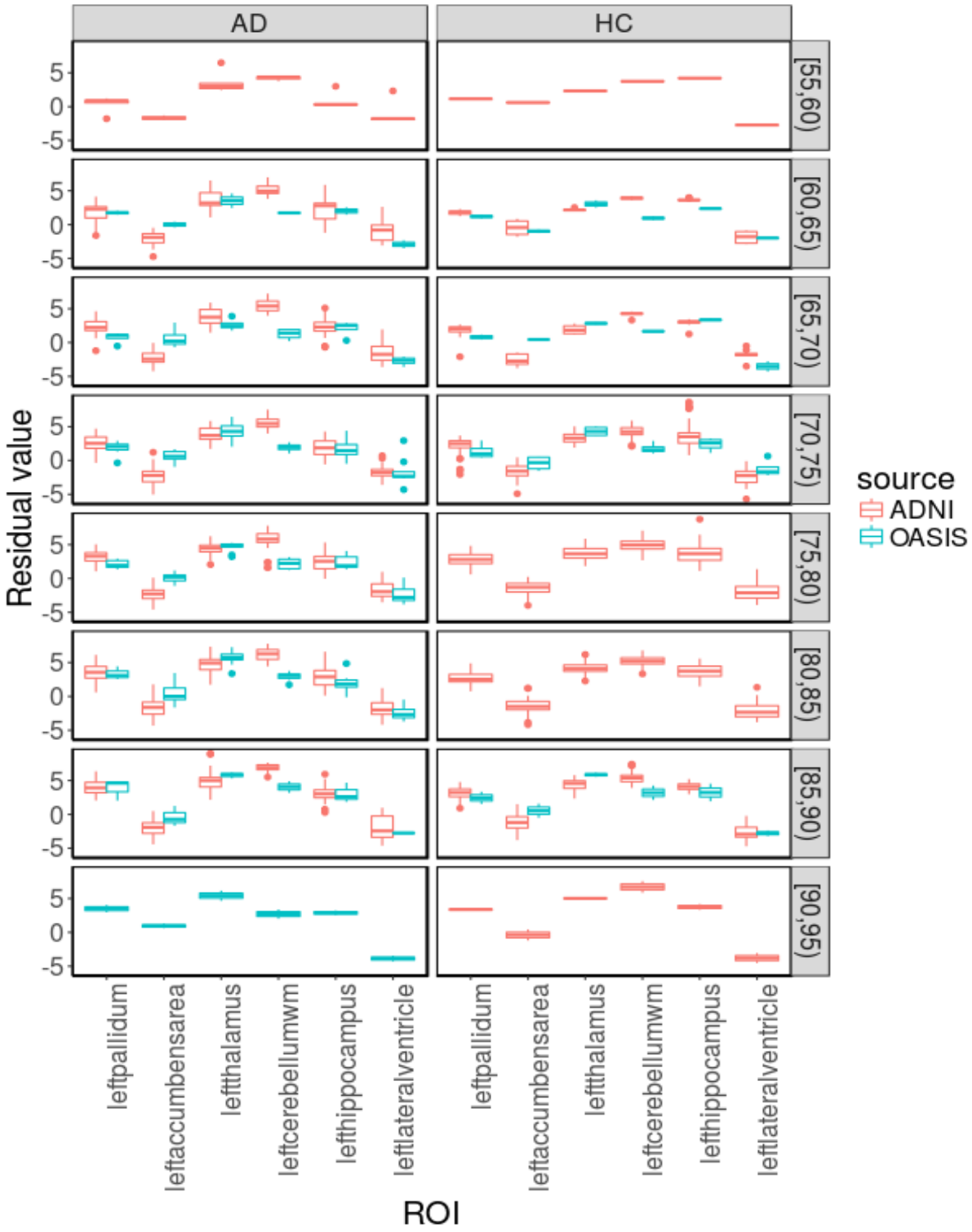


Fig. 6.2 Left hemisphere: Quantiles comparison of SV biomarker residuals for males of ADNI and OASIS data.

6.4.2 AD vs HC classification experiments.

The weighted average of ACC, SEN and SPE metrics for AD vs HC models from the method M_{res} is shown in Table 6.5. Although the initial SVM models built from ADNI data have obtained slightly better accuracies (Table 4.10) than these re-trained models including fewer biomarkers (107 *vr* MRI biomarkers for females and 95 *vr* MRI biomarkers for males), the performances of both models are very satisfactory (weighted average ACC $> 87.34\%$). When evaluating models on OASIS test data, accuracies for females and males models were 62.94% and 65.71% respectively, obtaining better weighted average values for sensitivities ($SEN > 84\%$) than specificities ($SPE \leq 20\%$).

Table 6.5 Performances of classification for AD vs HC experiment from M_{res} method.

Gender	Female	Male	Weighted Average
ADNI: AD vs HC			
No. test subjects	AD: 78, HC: 69	AD: 84, HC: 82	
Accuracy	88.93%	85.94%	87.34%
Specificity	92.31%	89.38%	90.76%
Sensitivity	86.00%	82.50%	84.14%
OASIS: AD vs HC			
No. test subjects	AD: 54, HC: 31	AD: 39, HC: 10	
Accuracy	62.94%	65.71%	63.95%
Specificity	19.03%	21.00%	19.75%
Sensitivity	88.15%	77.18%	84.14%

Regarding the method M_{raw}, Table 6.6 summarises the classification quality indicators for AD vs HC classifiers structured in 5-year age groups. The last column also shows the weighted average across all these groups. For the ADNI test data, we observed high values for ACC, SEN and SPE for most groups, with indicators well above 68%, and values equal to 100% for some of them. For the OASIS subjects, model's performance varied across the five age groups. Accuracies values were above 76% for three of five groups, but the weighted average specificity and sensitivity were above 81.0%.

6.5 Discussion

Differences observed between residuals of many MRI biomarkers may be due to several facts related to the diagnosis criteria, the follow-up of participants and the data source, among others. ADNI criteria used to classify individuals are applied at each visitation by combining various neuropsychological tests, such as the MMSE, CDR and the NINCDS/ADRDA criteria. In contrast, classification of OASIS subjects is only based on CDR scores. Furthermore,

Table 6.6 Performance of AD vs HC classifier from the M_{raw} method.

5–yr	65–70	70–75	75–80	80–85	85–90	Weighted Average
ADNI: AD vs HC						
No. test subjects	AD:10, HC:2	AD:22, HC:75	AD:34, HC:81	AD:22, HC:37	AD:13, HC:9	
Accuracy	58.33%	84.54%	86.08%	89.83%	100.00%	86.22%
Specificity	0.00%	89.33%	82.71%	100.00%	100.00%	86.15%
Sensitivity	70.00%	68.18%	94.11%	72.72%	100.00%	81.20%
OASIS: AD vs HC						
No. test subjects	AD:11, HC:6	AD:24, HC:12	AD:20, HC:3	AD:22, HC:5	AD:7, HC:5	
Accuracy	76.47%	50.00%	60.87%	77.77%	91.67%	66.96%
Specificity	33.33%	75.00%	100.00%	40.00%	80.00%	86.38%
Sensitivity	100.00%	37.50%	55.00%	86.36%	100.00%	81.17%

ACC: Accuracy. SEN: Sensitivity. SPE: Specificity.

ADNI subjects included in this study have been identified as diagnostic stable (stable HC or stable AD) during five years of follow-up; however, OASIS subjects only were observed at baseline; thus it is possible that some of them were affected with an incipient AD. Besides, the Freesurfer versions and analysis approaches used in ADNI and OASIS studies were different. The OASIS study used the version 4.3, and the pipeline to process images was specific to cross-sectional studies. However, the ADNI study used the Freesurfer v.4.4 with a processing pipeline focused on longitudinal studies.

Because OASIS diagnostic is performed directly from CDR scores, we did not include neuro-psychological features as input within the classifiers. Thus, only those models who met this condition were compared. In all cases, the average accuracies of models tested on ADNI data were higher than the results obtained on OASIS test data. However, when comparing model results from ADNI data with results presented in Tables 4.10 and 5.9, specifically, in those classifiers built by excluding MMSE and CDR-related scores, all accuracies are lower than results achieved by the models that included those scores. Furthermore, in most cases, low accuracies while testing on OASIS data were due to very low values in specificity. This fact may be because the low reliability of using only CDR score to subject diagnosis. Possibly, several subjects from OASIS diagnosed as healthy controls have early brain changes, but those subjects still do not have AD symptoms.

Chapter 7

Discussion

7.1 Summary of the results

Method $\mathbf{M}_{res}$. Analysis of CSF biomarkers values for subjects clinically diagnostic as HC, MCI and AD, allowed to identify cut-off values to differentiate the subjects with normal and abnormal CSF profile. Results confirmed the cut-off values previously determined by Shaw et al. [108], classifying as normal the CSF profile of subjects who fulfilled both conditions: CSF-Aβ $>= 192pg/ml$ and CSF-$\tau \leq 93pg/ml$. Consequently, only 46 of 119 subjects diagnosed as HC by ADNI have shown a normal CSF profile (normal-HC_{csf}) over time. Results of the comparison of the CSF profile between the diagnostic groups presented in Figure 4.5, showed an increase in τ values when comparing normal-HC_{csf} and subjects with abnormal CSF-profile, and when comparing abnormal-HC_{csf} with MCI and AD subjects with abnormal profile. Concerning Aβ levels, we observed a reduction of levels when comparing normal-HC_{csf} with abnormal-HC_{csf} and the remaining groups with an abnormal profile.

In order to identify the *vr* and *qvr* biomarkers due to ageing effects, each of the 166 longitudinal MRI-based biomarkers (66 CV, 34 TA and 66 SV) from the normal-AD_{csf} subjects was modelled by applying an LME approach based on random intercepts. Here, we used the subject age (*age*) at each observation and the years of education (*educ*) as covariates. In this type of LME model, the measured value of ROI is assumed to have a set of parameters β, fixed across subjects, but the y-intercept value varies by each subject. We identified 69 *qvr* biomarkers for males and 57 *qvr* biomarkers for females. The main characteristic of these two types is that the age's coefficient β of the *qvr* ROIs is close to zero, but for *vr* ROIs, their slopes are not. By looking in the set of *vr* biomarkers, we found reductions for most thickness and volume biomarkers of cortical regions in both genders following a frontotemporal pattern (Tables 4.6 and 4.8). Also, we observed atrophy in subcortical regions. In men, the greatest atrophy was found in the hippocampus, followed by the ventral diencephalon (ventraldc),

the thalamus, the pallidum and several regions of the corpus callosum (Table 4.7). For females, the greatest reduction was found in the amygdala, followed by the hippocampus, the ventraldc and the thalamus (Table 4.9). On the other hand, we also observed the expansion of regions of the ventricular system (lateral ventricle and third ventricle). Moreover, the optic chiasm region showed the greatest expansion in women.

During classifiers building stage, the SVM models were built using inferred residuals for HC, MCI and AD subjects to address two problems: (1) the estimation of current diagnosis (ECD), i.e. the subject diagnostic at each visitation, and (2) the prediction of future diagnosis (PFD), i.e. the early prediction of progression to AD. For the first problem, in average, all these classifiers achieved high accuracies for the AD vs HC experiment (F_1: above 91%, F_2: over 94%) and the MCI vs HC experiment (F_1: above 79%, F_2: above 85%). Even for the AD vs MCI experiment, which addresses the early diagnosis of prodromal AD, the models have shown great performance, achieving accuracies values up to 81.6%, see Table 7.2. Regarding the PFD problem, all classifiers achieved high accuracies for the AD vs HC experiment (up to 91.7% in females). However, for the remaining experiments, the accuracies values were up to 77.6%, being penalised by low specificities in comparison to the high sensitivities achieved in all cases. Furthermore, in both problems were observed that models built with the $F2$ configuration obtained better accuracies than the ones built on the $F1$ configuration. Finally, regarding the advancement in the prediction of subject's conversion, the greatest lead was obtained in the early prediction from MCI to AD in males, this being 1.85 years earlier. Likewise, for the conversion from HC to AD, these models allowed us to advance up to 1.64 years earlier for females (75–79 age group); and up to 1.73 years earlier for males (80–84 age group), see Table 4.12.

Method $\mathbf{M}_{raw}$. We modelled the annual percentage of cortical change by applying the LME approach on longitudinal each CV (N=66) and TA (N=66) MRI-based biomarkers from sHC, sMCI and sAD. These subjects are characterised by having the same diagnosis over time. Figures 5.1 and 5.2 show the age-related changes for CV and TA biomarkers for each diagnosis group, respectively. Results confirm predominant frontotemporal patterns of atrophy in sHC subjects across the adult age-span (Figures 5.1(a) and 5.2(a)). Furthermore, we also observed a frontotemporal pattern in sMCI and sAD subjects, where the change rate observed in sAD is slightly different from the change observed in sMCI and is more differentiated than sHC. Moreover, most selected MRI-based biomarkers corresponded to the temporal and frontal lobes, which also support findings from the LME models.

As a first step of the multivariate analysis, it was applied a SVM wrapper to select the most powerful features of each five-year age group. Only 33 MRI biomarkers were selected,

and the remaining 133 features were discarded, see Table 5.6. Some MRI-based biomarkers were selected for more than one age group, and most of the selected ones correspond to the temporal and frontal lobes. After that, the final set of MRI-biomarkers selected for each age group was established as the union of all biomarkers previously selected for younger groups.

The final stage of this method was focused on both ECD and PFD problems. For the AD vs HC experiment in the ECD problem, models including neuro-psychological features obtained better results than models built without include these features. In the first case, the SVM model performance shows very high values for accuracy, sensitivity and specificity for all age groups, with all weighted average indicators above 97%, and values of 100% for many age groups. For AD vs MCI, which address the diagnosis of prodromal AD and is intrinsically harder than HC vs AD, results show a weighted average accuracy above 84%, the performance across all age groups being quite consistent. For MCI vs HC, models achieved a weighted average accuracy above 82%, with a classification performance that apparently declines with age. In fact, by comparing these results with comparable results from the literature, see Section 2.5 (Table 7.2), our method M_{raw} shows the highest indicators in the AD vs HC classification, the highest specificity in MCI vs HC, as well as the highest accuracy in AD vs MCI.

For the PFD problem, models including neuro-psychological features also obtained better results than models built without these features. From the first approach, for the AD vs HC experiment, all accuracies are well above 88%, with an accuracy of 98.75% for the 70–74 age group, see Table 5.10. The advancement of early prediction ranges from 1.55 to 1.87 years for AD vs HC and from 1.50 to 1.66 years for AD vs MCI . The results of the AD vs MCI experiment, though lower than in AD vs HC due to their intrinsic greater difficulty, also show high-performance indicators, especially for classification accuracy and sensitivity. When comparing these models with those in the literature, M_{raw} outperforms the only comparable method on all the indicators, while also increasing the span of the age advancement in the diagnosis slightly, see Table 7.3.

Validation of methods. We found some feature-related differences between ADNI and OASIS datasets. In OASIS data, information related to education is a categorical feature that corresponds to levels. In contrast, ADNI provides the total years of education. Thus, to match OASIS data with ADNI education values, the OASIS data were converted to equivalent years of education. On the other hand, two MRI biomarkers related to the inferior lateral ventricle, which were selected as variant ROIs by M_{res}, are not available in OASIS dataset. For that reason, the pool of *vr* biomarkers was reduced from 166 to 164 biomarkers.

Tables 6.3 and 6.4 show the list of biomarkers from ADNI and OASIS datasets showing identical distribution for HC females and HC males, respectively. These tables show that when comparing both datasets, less than half of the initial number of biomarkers have identical distribution for HC and AD subjects.

Because of the specific features of OASIS cohort, it was only addressed the ECD problem, specifically for the AD vs HC experiment. Classification results by using M_{res} are summarized in Table 6.5. The model achieved a weighted average accuracy above 87% on ADNI test data. Accuracies for females and males models on OASIS test data were 62.94% and 65.71% respectively, obtaining better results in the weighted average values for sensitivity ($SEN > 84\%$) than specificity ($SEN < 20\%$). Regarding M_{raw}, Table 6.6 summarizes the performance of models built to address the AD vs HC experiment. For the ADNI test data, the weighted average values for ACC, SEN and SPE were above 81.00%, and the metrics improved as the age progressed. For the OASIS subjects, model's performance varied across all groups. Accuracies values were high for only three groups ($ACC > 76\%$), but the weighted average indicators of ACC, SEN and SPE were above 66.9% for all of them.

Summary of models performance. Results of the ECD problem previously mentioned above are summarised in Table 7.1. Observe that classifiers which included neuro-psychological features obtained the best performance on each method. Also, the method M_{raw} always obtained the best results when models where tested.

Table 7.1 Comparison of models built to address the ECD problem(%).

Method	**Type**	**Dataset**	**AD vs HC** (%)			**MCI vs HC** (%)			**AD vs MCI** (%)			**Data used**
			ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	
M_{res}	initial	ADNI^{*-}	89.22	90.73	86.85	77.08	95.53	26.98	70.30	64.26	75.14	MRI-based residuals, age, gender, educ
M_{res}	initial	ADNI^{**-}	**94.11**	**96.54**	**90.28**	**83.77**	**93.55**	**57.55**	**76.72**	**70.77**	**81.62**	MRI-based residuals, age, gender, educ, MMSE, CDR-GLOBAL
M_{res}	comparison	ADNI^{*-}	87.34	84.14	90.76	-	-	-	-	-	-	MRI-based residuals, age, gender, educ
M_{res}	comparison	OASIS^{*-}	63.95	84.14	19.75	-	-	-	-	-	-	MRI-based residuals, age, gender, educ
M_{raw}	initial	ADNI^{*--}	86.23	81.21	86.16	70.00	63.87	68.40	68.73	51.51	77.83	MRI-based biomarkers, age
M_{raw}	initial	ADNI^{**--}	**98.68**	**97.43**	**99.53**	**82.31**	**81.80**	**77.38**	**84.56**	**81.17**	**86.38**	MRI-based biomarkers, age, MMSE, CDRSB
M_{raw}	comparison	ADNI^{*--}	86.22	81.20	86.15	-	-	-	-	-	-	MRI-based biomarkers, age
M_{raw}	comparison	OASIS^{*--}	66.96	81.17	86.38	-	-	-	-	-	-	MRI-based biomarkers, age

* This method does not include neuropsychological features. ** This method does include neuropsychological features. -Results of this method correspond to the average of the performance recorded for men and women. - -Results of this method correspond to the weighted average of the performance recorded for all five-years age group. MRI, Magnetic Resonance Imaging-based features; MMSE, Mini-Mental Clinical Dementia Rating; CDRSB, Clinical Dementia Rating–Sum of Boxes; CDRGLOBAL, Clinical Dementia Rating global score.

Both M_{res} and M_{raw} achieved better results than comparable methods in the literature for most classification quality indicators. Table 7.2 compares the prediction accuracy of the proposed methods (by averaging the accuracy for both genders and age groups) with existing methods, which have been mainly evaluated on the ADNI dataset. Observe that we only showed the most significant results of those studies, mainly focusing on MRI and its combination with other types of features.

Table 7.2 Comparison of methods focused on the estimation of current diagnosis (%).

Method	**AD vs HC**			**MCI vs HC**			**cAD vs sMCI**			**Data used**
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	
Kloeppel et al. [69]	95.0	95.0	95.0	-	-	-	-	-	-	MRI
Vemuri et al. [121]	-	86.0	86.0	-	-	-	-	-	-	MRI
	-	88.0	90.0	-	-	-	-	-	-	MRI, age, gender
	-	86.0	92.0	-	-	-	-	-	-	MRI, age, gender, APOE
Cui et al. [18]	-	-	-	-	-	-	67.13	96.43	48.28	NM, CSF, MRI
	-	-	-	-	-	-	62.24	92.86	42.53	NM, MRI
	-	-	-	-	-	-	62.24	57.14	65.52	MRI
Cuingnet et al. [19]	-	81.0	95.0	-	-	-	-	57.0	78.0	MRI
Zhang et al. [129]*	93.3	-	-	83.2	-	-	73.9	68.6	73.6	MRI, PET, CSF
Suk et al. [113]	95.9	-	-	85.0	-	-	75.8	-	-	MRI,PET
Jie et al. [65]	95.03	94.90	95.00	79.27	85.86	66.64	68.94	64.65	71.79	MRI, FDG-PET
Gaser et al. [49]	-	-	-	-	-	-	75.00	71.00	84.00	MRI-based age
Spulber et al. [111]	88.4	86.1	90.4	-	-	-	67.7	69.6	66.8	MRI-based index
Aguilar et al. [2]	-	92.0	75.0	-	-	-	-	92.0	47.0	MRI-based index
Liu et al. [73]	94.37	94.71	94.04	78.8	84.85	67.06	67.83	64.88	70.0	MRI, PET
Suk et al. [114]	92.38	91.54	94.56	84.24	99.58	53.79	72.42	36.70	90.98	MRI
	93.35	94.65	95.22	85.67	95.37	65.87	75.92	48.04	95.23	MRI, PET
*$\mathbf{M}_{res}$	**89.22**	**90.73**	**86.85**	**77.08**	**95.53**	**26.98**	**70.30**	**64.26**	**75.14**	MRI-based residuals, age
*$\mathbf{M}_{res}$	**94.11**	**96.54**	**90.28**	**83.77**	**93.55**	**57.55**	**76.72**	**70.77**	**81.62**	MRI-based residuals, age, MMSE, CDR-GLOBAL
*$\mathbf{M}_{raw}$	**86.23**	**81.21**	**86.16**	**70.00**	**63.87**	**68.40**	**68.73**	**51.51**	**77.83**	MRI-based biomarkers, age
*$\mathbf{M}_{raw}$	**98.68**	**97.43**	**99.53**	**82.31**	**81.80**	**77.38**	**84.56**	**81.17**	**86.38**	MRI-based biomarkers, age, MMSE,CDRSB

* Results of this method correspond to the average of performance recorded for men and women or age groups. MRI, Magnetic Resonance Imaging-based features. CSF: Cerebral Spinal Fluid-based biomarkers. NM: Neuro-psychological measures. PET: Positron Emission Tomography-based features. FDG-PET: [18F]fluorodeoxyglucose uptake measured in PET. MRI-based age: individual estimated age computed from MRI images. MRI-based index: individual severity index computed from MRI images. MMSE: Mini-Mental Clinical Dementia Rating. CDRGLOBAL: CDR Global Score.

Regarding the PFD problem, Table 7.3 summarises the performance of models trained to address that problem. Table shows M_{raw} outperforms M_{res} on all indicators, while also increasing the span of advancement in the diagnosis slightly up to an average of 1.68 year early for AD vs HC.

Table 7.3 Comparison of models built to address the PFD problem (%).

Method	**Type**	**Dataset**	**AD vs HC (%)**			**AD vs MCI** (%)			**Data used**
			ACC	SEN	SPE	ACC	SEN	SPE	
M_{res}	initial	ADNI^{*-}	86.10	90.75	71.05	67.60	85.75	49.50	MRI-based residuals, age, gender, educ
M_{res}	initial	ADNI^{**-}	**90.85**	**94.85**	**81.10**	**70.75**	**86.25**	**28.15**	MRI-based residuals, age, gender, educ, MMSE, CDR-GLOBAL
M_{raw}	initial	ADNI^{*--}	91.49	96.39	82.81	85.18	94.00	63.71	MRI-based biomarkers, age
M_{raw}	initial	ADNI^{**--}	**97.38**	**97.96**	**96.64**	**88.60**	**96.30**	**65.57**	MRI-based biomarkers, age, MMSE, CDRSB

* This method does not include neuropsychological features. ** This method does include neuropsychological features. -Results of this method correspond to the average of the performance recorded for men and women. - -Results of this method correspond to the weighted average of the performance recorded for all five-years age group. MRI, Magnetic Resonance Imaging-based features; MMSE, Mini-Mental Clinical Dementia Rating; CDRSB, Clinical Dementia Rating–Sum of Boxes; CDRGLOBAL, Clinical Dementia Rating global score.

7.2 Discussion

This thesis addresses two main problems in the AD research field, i.e. the ageing-related brain change understanding and the subject diagnosis classification of the current state, as well as the progression to AD. The results found throughout the two proposed methods confirm the complexity of studying this disease. Because AD affect brain regions, which

also change due to ageing, even in its early stage, we aim to look by reliable techniques to differentiate these changes taking into account factors such as age, CSF profile and gender, and two methods were proposed. Despite the differences between techniques or the type of data used in each method, both methods converge on firstly to outline each diagnostic group to understand their similarities and differences. After the understanding phase, we applied different statistical learning methods to select the most informative features and to train SVM classifiers to address both the estimation of the current diagnosis (ECD) and the prediction of the future diagnosis (PFD). In both cases, all classifiers were focused on three main experiments: AD vs HC, MCI vs HC and AD vs MCI.

More in detail, as a starting point in this study, we focused on identifying and understanding the brain changes over time, either due to ageing or illness. For M_{res}, we used cut-off values for CSF-Aβ and CSF-τ biomarkers to outline each diagnostic group according to a normal or abnormal CSF profile. We found that several subjects clinically diagnosed as healthy control in the ADNI study have abnormal CSF profiling. This finding is consistent with previous studies [108, 116, 27], where the presence of possible AD pathology has been observed in ADNI control subjects. It may be because ADNI subject diagnosis is made independently of the CSF biomarkers values. Once identified the HC subjects with normal CSF profile, we were able to build LME models to study the change over five years follow-up of an extensive set of MRI-based biomarkers obtained from cortical and subcortical regions for these HC subjects. LME modelling allowed to classify biomarkers as variant or quasi-variant ROIs and to build null models for ageing-related changes in men and women. As in previous studies [104, 43, 64], regions identified in older adults as changing over time confirm that part of these changes occurs in brain areas related with AD. Reductions found in this study for most thickness and volume biomarkers of cortical regions coincide with the frontotemporal change observed in previous studies [104, 43, 64]. Furthermore, the decline in subcortical regions including the hippocampus, the amygdala, the thalamus, the ventraldc and regions of the corpus callosum; as well as expansion in the ventricular system (lateral ventricle and third ventricle), also have been reported by [43]. Regarding the expansion observed in the optic chiasm from healthy females, it may be a consequence of vision loss due to ageing [1]. On the other hand, in M_{raw}, the annual percentage of cortical change was modelled by applying the LME approach on longitudinal MRI-based biomarkers from subjects characterised by having the same diagnosis over time (i.e. sHC, SMCI and sAD). As in the previous method, results confirm predominant frontotemporal patterns of atrophy in sHC subjects across the adult age-span. This pattern has been previously described in other longitudinal studies [44] and could explain the ageing-related decline in specific cognitive abilities. Furthermore, we also observed a frontotemporal pattern in sMCI and sAD subjects,

where the change rate in sAD is slightly different from the change observed in sMCI, and it is more differentiated than sHC. As described previously by Fjell et al. [45], the frontotemporal pattern observed from both methods could confirm that the observed brain structures changes belong to an inevitable process related to normal ageing and is not necessarily an AD-specific one. Furthermore, it is possible that the vulnerability of these areas to normal ageing-related decline contributes to their vulnerability to AD-related atrophy.

The knowledge about how the brain regions change in each diagnosis group was used to approach the other important part of this study, i.e. the ECD and PFD problems. Firstly, we aimed to select the most significant MRI-based biomarkers to differentiate the subject diagnosis groups and also reduce the data dimensionality. In M_{res}, residuals, established as differences between the observed ROIs and ageing-related inferred ROIs, were valuable to differentiate the brain change in MCI and AD subjects from ageing-related change. The residuals from *vr* biomarkers were directly used as input features to build the SVM classifiers to subject classification and early disease prediction. In contrast, M_{raw} was focused on directly analysing the raw MRI-based biomarkers values, and a feature selection method was applied on each combination of the five-year age group and the differential diagnosis to select the most powerful features. Several MRI-based biomarkers were selected for more than one age group, mainly as the ages increase. Then, the final feature set for each of these combinations was obtained by grouping the age group-specific features with the features selected for all the previous age groups (in chronological order). We applied this approach of grouping features because we hypothesised that whether a feature has had a relevant role concerning the diagnosis at a given age, that feature will continue being relevant to the condition subsequently, even if its relative importance declines in favour of other features. Also, indistinctly to the age group or the differential diagnosis, the most selected features also correspond to the frontal and temporal lobes.

Regarding classification performance, both M_{res} and M_{raw} achieved better results than comparable methods in the literature for most classification quality indicators. In summary, both methods provide the following main contributions:

1. By using relatively common clinical tests such as MRI and neuropsychological tests, as opposed to methods that rely on more expensive or invasive tests such as PET-based, CSF-based and Genotype-based biomarkers, Both M_{res} and M_{raw} improve the estimation of current diagnosis:

    - For AD vs HC: By using M_{raw}, classifiers reached higher accuracy values (above 98%) than models from other state-of-the-art methods.

- MCI vs HC: Classification accuracies up to 83.77% obtained by M_{res}, where models also achieved high sensitivities (up to 95.53%).
- AD vs MCI: By using M_{raw}, the model obtained the highest classification accuracy (84.56%) among other models from state of the art.

2. In most experiments, the sensitivity (the ability of a predictor to correctly classify a subject as 'diseased') was higher than the specificity (the ability of predictor to correctly classify a subject as 'disease-free'). This fact may be because ADNI clinical diagnosis is based on neuropsychological tests, but neurodegeneration occurs many years before the onset of clinical symptoms. Besides, ADNI patient diagnostics are not pathologically confirmed. Possibly, our SVM predictors may determine that subjects are into early stages of disease (MCI and prodromal AD) but this finding is not consistent with the clinical diagnosis because the subject does not yet present clinical symptoms. Abnormalities of CSF profile observed in several subjects diagnosed as HC and MCI (see Figure 4.4) may support this hypothesis.

3. When comparing models performance of both methods, we confirmed that the approach applied by the method M_{raw} have proved to be more powerful and robust to improve classification performance than the method M_{res}. This fact may be because AD-related brain changes are more affected by age than gender. Also, this method includes the training of classifiers throughout aggregating pre-selected features as age groups get older, which seems to be the more efficient way to address this problem.

4. Both M_{res} and M_{raw} improve the prediction of future diagnosis given the current clinical tests, both in prediction quality indicators as well as in the amount of time by which the diagnosis is advanced. As of the date of this study, M_{res} and M_{raw} are the first methods applied on a longitudinal study of 5 years follow-up, which provide information about the advancement (in years) in the early disease prediction. The greatest lead was obtained by M_{raw} that predicted in subjects aged 80-84 the conversion from HC to AD up to 1.87 years earlier.

5. As of the date of this study, we were unable to find studies where ageing-related null models and residuals-based classifiers were applied to early diagnosis or current diagnosis. The performance obtained in all experiments suggests that M_{res}, which is focused on computing ROI residuals is powerful to support ECD and PFD problems, the fundamental challenges in AD research.

6. The reliability of both methods was evaluated by applying them to a new dataset from the cross-sectional OASIS study. Because OASIS data only includes subjects diagnosed

as HC or AD, and only there is one observation per subject; methods comparison only was applied to evaluate the classifiers for the ECD problem, specifically for AD vs HC. We found that less than half of the initial number of *vr* biomarkers have identical distribution for HC and AD subjects in both datasets. These differences may be due to several factors related to diagnosis criteria, the lack of follow-up of participants and the data source, among others. ADNI criteria used to classify individuals is based on combining various neuropsychological tests, such as the MMSE, CDR and the NINCDS/ADRDA criteria. In contrast, classification of OASIS subjects is only based on the CDR scores. Furthermore, ADNI subjects included in this comparison have been identified as stable HC or stable AD during five years of follow-up; however, such classification was not viable for OASIS subjects due to having only one observation at baseline, being possible that some of them were affected with an incipient AD. Also, the Freesurfer versions and analysis approaches used in ADNI and OASIS studies were different. The OASIS study used the version 4.3, and the pipeline to process images was specific to cross-sectional studies. However, the ADNI study used the Freesurfer v.4.4 with a processing pipeline focused on longitudinal studies.

7. Finally, none of the studies reviewed of the state-of-the-art, which trained models from ADNI data, has assessed their model's performance using data from OASIS. Moreover, few of them have used a different source of data for training, testing and validation tasks of their proposed methods. This fact constitutes one of the strengths of this study and increases the added value of the methods proposed in this thesis.

Chapter 8

Conclusions

8.1 Main Conclusions

The research presented in this thesis contributes to two fundamental challenges in AD research, i.e. ageing-related brain change understanding, and the diagnostic classification of the current subject state, as well as the progression to MCI or AD. This study has investigated statistical learning algorithms to propose two methods, M_{res} and M_{raw}, which were applied on longitudinal MRI-based biomarkers observed in HC, MCI and AD subjects at five years follow-up. As a starting point, both methods aimed to understand the ageing-related effects in the brain and to distinguish brain changes differences between healthy elderly, MCI and AD subjects, to subsequently to apply that knowledge to approach the diagnosis estimation. Longitudinal MRI-based biomarkers provide robust information about morphological brain changes over time, not only related to the disease, but also with ageing. Furthermore, statistical learning methods carefully applied on these biomarkers provide valuable mechanisms to support the accurate diagnosis as well as the early diagnosis.

The main contribution of this thesis are listed below:

1. This study found significant changes in brain structures over five years in healthy elderly people that are similar to several changes occur in brain areas related with AD. Reductions found for most thickness and volume biomarkers of cortical structures follow a frontotemporal pattern. Furthermore, this study also found a frontotemporal pattern in stable MCI and stable AD subjects, where the change rate in the stable AD subjects is slightly different from the change observed in the stable MCI and is more differentiated than stable HC. These findings indicate that the frontotemporal pattern of change corresponds to an inevitable process related to normal ageing, which could be driving that those regions changing due to normal ageing-related decline are more

vulnerable to be reduced when the disease is present. We also observed significant changes in the subcortical region. The hippocampus, the amygdala, the thalamus, the ventraldc and regions of the corpus callosum showed a predominant decline in HC subjects over time. Likewise, We observed a significant expansion of the ventricular system (lateral ventricle and third ventricle).

2. Both methods M_{res} and M_{raw} have the potential for addressing the estimation of current diagnosis and the estimation of future diagnosis with similar o better performance than state-of-the-art methods. Likewise, both methods show results promising even for the prodromal AD (progression from MCI to AD). Indeed, through the method M_{res} it has been introduced the concept of MRI-based Residuals, which represent the differences between the observed MRI biomarkers and the ageing-related inferred ones.

3. This study demonstrated that applying feature selection to keep only informative features at each age group and to reduce the data dimensionality is meaningful to improve classifiers performance. Although studies focused on analysing neurodegeneration biomarkers from the whole-brain have obtained better results than studies aimed at specific brain regions, it is also a fact that not all brain regions are informative to discriminate between subject diagnostics. Furthermore, that performance obtained by applying an approach of aggregating selected features as age groups get older and then using that set of aggregated features as classifier input could confirm that the effect of the change of a region remains active over time even if there are new brain regions showing greater significance.

4. From the longitudinal study of five years follow-up, this research attained an advancement in the early diagnosis of incipient AD (progression from HC to AD) up to 1.87 years earlier. As of the date of this study, no other study focused on early diagnosis has reported such result. In general, no study has provided the obtained diagnosis advancement.

5. Low specificities penalised the most cases where classifiers obtained low accuracies. Low specificities mean that our models classified as 'diseased', subjects labelled as healthy or in a stage previous to the disease. However, it is possible that these subjects incorrectly classified are truly diseased, even in an early stage, but do not yet present clinical symptoms. There exist several facts that may support this hypothesis: 1) This study found abnormalities of CSF profile on several subjects diagnosed as HC and MCI (see Figure 4.4); 2) The ADNI and OASIS clinical diagnosis only takes into account neuropsychological tests; and 3) ADNI and OASIS patients diagnosis are not

pathologically confirmed, thus some uncertainty on the subject's diagnosis may be introduced.

6. There exist many brain morphological differences between ADNI and cross-sectional OASIS cohorts. Biomarkers from HC OASIS subjects are quite different from their counterpart from ADNI, despite belonging to the same age group and gender. In fact, very few biomarkers show an identical distribution in their respective counterparts. These differences may be due to several factors related to diagnosis criteria, the lack of follow-up of participants and the data source, among others.

7. In Machine Learning, a classifier not only learns the underlying patterns in the training set to make predictions, it also could learn the "peculiarities" of that data. We addressed that issue and attempt to assess our models by using another cohort for testing and methods comparison. In fact, this study is one of the few studies which have addressed the early diagnosis of AD by training and testing classifiers by using different cohorts (i.e. ADNI and OASIS datasets).

8.2 Limitations and future work

Despite promising results, there are several limitations to our study. Data used here correspond to research participants, who meet the inclusion and exclusion criteria established by ADNI and thus are not from a general population.

The available observations of CSF biomarkers do not correspond in number or time points with the available MRI-based observations. In most cases, we had just CSF values at the baseline stage, so it is impossible to track the reliability of CSF profile at the final stages.

This study has only assessed neurodegeneration biomarkers based on MRI, and certain clinical features, such as age, the gender and years of education. In fact, longitudinal modelling of these biomarkers was only applied by fitting age and years of education. Future work could assess the impact of other feature types such neuroimaging biomarkers from PET, genotype data, psychometric scores, CSF markers, and might include biological markers and other clinical and neuropsychological assessments. Besides, other feature ranking and selection methods could be considered.

Finally, this research has only considered high-level features such as cortical volume and thickness, subcortical volumes. Because MRI biomarkers are measures obtained from structurally or functionally predefined brain regions, they could omit more specific information due to small or subtle changes involved in the brain diseases. Recently, the Deep Learning architecture is gaining a great attention due to its representational power. When it is applied

to medical images, low to high-level features can be obtained, which allow building more robust classifiers.

Chapter 9

Publications

9.1 Journal Papers

Submitted papers:

- Josep Lupón, Alexandre Perera-Lluna, Jorge López-Ayer, Marta de Antonio, Giovana Gavidia; Mar Domingo, Elena Ferrer,Elisabet Zamora, Pedro Moliner, Javier Santesmases, Antoni Bayes-Genis. Left Ventricular Ejection Fraction Trajectories in Heart Failure with Reduced Systolic Function: A 15 year Longitudinal Study. The New England Journal o f Medicine.

- Giovana Gavidia-Bovadilla, Samir Kanaan-Izquierdo, María Mataró, and Alexandre Perera-Lluna, for the Alzheimer's Disease Neuroimaging Initiative. Early Age-specific Prediction and Prognostic Advancement of Alzheimer's Disease. IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.

Published papers:

- Laura Martin-Fernandez, Giovana Gavidia-Bovadilla, Irene Corrales, Helena Brunel, Lorena Ramírez, Sonia López, Juan Carlos Souto, Francisco Vidal, José Manuel Soria. Next generation sequencing to dissect the genetic architecture of KNG1 and F11 loci using factor XI levels as an intermediate phenotype of thrombosis. PLOS ONE, 12 (4), Apr 2017. ISSN 19326203.

- Giovana Gavidia-Bovadilla, Samir Kanaan-Izquierdo, María Mataró-Serrat, Alexandre Perera-Lluna, for the Alzheimer's Disease Neuroimaging Initiative. Early Prediction of

Alzheimer's Disease Using Null Longitudinal Model-Based Classifiers. PLOS ONE, 12(1):e0168011, jan 2017. ISSN 1932-6203.

- Gavidia, G., Soudah, E., Martín-Landrove, M., and Cerrolaza, M. (2011). Generación de modelos discretos de tejidos del ser humano a través del preprocesamiento y segmentación de imágenes médicas. Rev. Int. Métodos Numéricos para Cálculo y Diseño en Ing., 27(3):200-226.

- Cerrolaza, M., Gavidia, G., Soudah, E., and Martín-Landrove, M. (2014). MODELING HUMAN TISSUES: AN EFFICIENT INTEGRATED METHODOLOGY. Biomed. Eng. Appl. Basis Commun., 26(01):1450012.

9.2 Conference Papers

- Laura Martin-Fernandez, Giovana Gavidia, Irene Corrales, Helena Brunel, Lorena Ramírez, Juan Carlos Souto, Francisco Vidal, José Manuel Soria. NEXT GENERATION SEQUENCING TO UNRAVEL THE GENETIC ARCHITECTURE OF KNG1 AND F11 LOCI. European Human Genetics Conference (ESHG) 2016, Barcelona, Spain. 21-24/05/2016.

- G. Gavidia; A. Perera. Title: Longitudinal Analysis of Alzheimer's Neuroimage Data. 7TH ANNUAL CONFERENCE CIBER-BBN. Torremolinos, Spain. 21-22/11/2013

- Gavidia GG, Ziyatdinov A, Fernández-Albert, Soria JM and Perera-Lluna. Study of the Random Effects in the Longitudinal Analysis of Diastolic and Systolic Blood Pressure. The Genetic Analysis Workshop 19. City: Vienna, Austria. 23-27/08/2014

- Gavidia, G., López, R., and Soudah, E. (2012). Computer-aided diagnosis of dementia using medical imaging processing and artificial neural networks. In Comput. Vis. Med. Image Process. Proc. VipIMAGE 2011 - 3rd ECCOMAS Themat. Conf. Comput. Vis. Med. Image Process., pages 51-55.

Bibliography

[1] Victoria M. Addis, Heather K. DeVore, and Michael E. Summerfield. Acute Visual Changes in the Elderly, 2013. ISSN 07490690.

[2] Carlos Aguilar, J-Sebastian Muehlboeck, Patrizia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kloszewska, Hilkka Soininen, Simon Lovestone, Lars-Olof Wahlund, Andrew Simmons, and Eric Westman. Application of a MRI based index to longitudinal atrophy change in Alzheimer disease, mild cognitive impairment and healthy older individuals in the AddNeuroMed cohort. *Frontiers in aging neuroscience*, 6:145, jan 2014. ISSN 1663-4365. doi: 10.3389/fnagi.2014.00145. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4094911{&}tool=pmcentrez{&}rendertype=abstract.

[3] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, Peter J Snyder, Maria C Carrillo, Bill Thies, and Creighton H Phelps. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*, 7(3):270–279, may 2011. doi: 10.1016/j.jalz.2011.03.008. URL http://dx.doi.org/10.1016/j.jalz.2011.03.008.

[4] Tracy M Anderson, Perminder S Sachdev, Henry Brodaty, Julian N Trollor, and Gavin Andrews. Effects of sociodemographic and health variables on Mini-Mental State Exam scores in older Australians. *Am J Geriatr Psychiatry*, 15(6):467–476, jun 2007. doi: 10.1097/JGP.0b013e3180547053. URL http://dx.doi.org/10.1097/JGP.0b013e3180547053.

[5] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4(0):40–79, 2010. ISSN 1935-7516. doi: 10.1214/09-SS054. URL http://projecteuclid.org/euclid.ssu/1268143839.

[6] John Ashburner. SPM: A history. *NeuroImage*, 62(2):791–800, aug 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.10.025. URL http://linkinghub.elsevier.com/retrieve/pii/S1053811911011888.

[7] Douglas Bates. *lme4: mixed-effects modeling with R*. Springer, 2010. URL http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf.

[8] Douglas Bates, Martin Maechler, Benjamin M Bolker, and Steven Walker. {lme4}: Linear mixed-effects models using Eigen and S4, 2014. URL http://arxiv.org/abs/1406.5823.

[9] Jorge L Bernal-Rusiel, Douglas N Greve, Martin Reuter, Bruce Fischl, Mert R Sabuncu, and Alzheimer's Disease Neuroimaging Initiative. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage*, 66:249–60, feb 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.10.065. URL http://www.ncbi.nlm.nih.gov/pubmed/23123680http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3586747.

[10] Christopher M Bishop. *Pattern Recognition and Machine Learning*. 2006.

[11] Ron Brookmeyer, Elizabeth Johnson, Kathryn Ziegler-Graham, and H. Michael Arrighi. Forecasting the global burden of Alzheimer's disease. *Alzheimer's and Dementia*, 3(3):186–191, 2007. ISSN 15525260. doi: 10.1016/j.jalz.2007.04.381.

[12] Chih-Chung Chang and Chih-Jen Lin. {LIBSVM}: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1—-27:27, 2011.

[13] G Chételat, B Landeau, F Eustache, F Mézenge, F Viader, V de la Sayette, B Desgranges, and J.-C. Baron. Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *NeuroImage*, 27(4):934–946, 2005. ISSN 1053-8119. URL http://www.sciencedirect.com/science/article/pii/S1053811905003277.

[14] Youngsang Cho, Joon-Kyung Seong, Yong Jeong, Sung Yong Shin, and Alzheimer's Disease Neuroimaging Initiative. Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *Neuroimage*, 59(3):2217–2230, feb 2012. doi: 10.1016/j.neuroimage.2011.09.085. URL http://dx.doi.org/10.1016/j.neuroimage.2011.09.085.

[15] Marie Chupin, Emilie Gérardin, Rémi Cuingnet, Claire Boutet, Louis Lemieux, Stéphane Lehéricy, Habib Benali, Line Garnero, Olivier Colliot, and Alzheimer's Disease Neuroimaging Initiative. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus*, 19(6):579–587, jun 2009. doi: 10.1002/hipo.20626. URL http://dx.doi.org/10.1002/hipo.20626.

[16] C M Clark, C Davatzikos, A Borthakur, A Newberg, S Leight, V M-Y Lee, and J Q Trojanowski. Biomarkers for early detection of Alzheimer pathology. *Neurosignals*, 16 (1):11–18, 2008. doi: 10.1159/000109754. URL http://dx.doi.org/10.1159/000109754.

[17] C E Coffey, J F Lucke, J A Saxton, G Ratcliff, L J Unitas, B Billig, and R N Bryan. Sex differences in brain aging: a quantitative magnetic resonance imaging study. *Archives of neurology*, 55(2):169–79, 1998. ISSN 0003-9942. doi: 10.1001/archneur.55.2.169. URL http://www.ncbi.nlm.nih.gov/pubmed/9482358.

[18] Yue Cui, Bing Liu, Suhuai Luo, Xiantong Zhen, Ming Fan, Tao Liu, Wanlin Zhu, Mira Park, Tianzi Jiang, and Jesse Jin. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One*, 6(7): e21896, 2011.

[19] Remi Cuingnet, Emilie Gerardin, Jérôme Tessieras, Guillaume Auzias, Stéphane Lehéricy, Marie-Odile Habert, Marie Chupin, Habib Benali, and Olivier Colliot. Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, 2011. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.06.013. URL http://www.sciencedirect.com/science/article/pii/S1053811910008578.

[20] Jeffrey Cummings. Alzheimer's disease: clinical trials and the amyloid hypothesis. *Annals of the Academy of Medicine, Singapore*, 40(7):304–6, jul 2011. ISSN 0304-4602. URL http://www.ncbi.nlm.nih.gov/pubmed/21870020.

[21] Jeffrey L Cummings. Alzheimer's disease. *N Engl J Med*, 351(1):56–67, jul 2004. doi: 10.1056/NEJMra040223. URL http://dx.doi.org/10.1056/NEJMra040223.

[22] Anders M. Dale, Bruce Fischl, and Martin I. Sereno. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2):179–194, 1999. ISSN 10538119. doi: 10.1006/nimg.1998.0395.

[23] Sandhitsu R Das, Brian B Avants, John Pluta, Hongzhi Wang, Jung W Suh, Michael W Weiner, Susanne G Mueller, and Paul A Yushkevich. Measuring longitudinal change in the hippocampal formation from in vivo high-resolution T2-weighted MRI. *Neuroimage*, 60(2):1266–1279, apr 2012. doi: 10.1016/j.neuroimage.2012.01.098. URL http://dx.doi.org/10.1016/j.neuroimage.2012.01.098.

[24] Christos Davatzikos, Feng Xu, Yang An, Yong Fan, and Susan M Resnick. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain : a journal of neurology*, 132(Pt 8):2026–35, aug 2009. ISSN 1460-2156. doi: 10.1093/brain/awp091. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2714059{&}tool=pmcentrez{&}rendertype=abstract.

[25] Bhupinder S Dayal and John F. MacGregor. Improved PLS Algorithms. *Journal of Chemometrics*, 11(January 1996):73–85, 1997. ISSN 0886-9383. doi: 10.1002/(SICI)1099-128X(199701)11:1<73::AID-CEM435>3.0.CO;2-#.

[26] Renée F A G de Bruijn and M Arfan Ikram. Cardiovascular risk factors and future risk of Alzheimer's disease. *BMC medicine*, 12:130, nov 2014. ISSN 1741-7015. doi: 10.1186/s12916-014-0130-5. URL http://www.ncbi.nlm.nih.gov/pubmed/25385322http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4226863.

[27] Geert De Meyer, Fred Shapiro, Hugo Vanderstichele, Eugeen Vanmechelen, Sebastiaan Engelborghs, Peter Paul De Deyn, Els Coart, Oskar Hansson, Lennart Minthon, Henrik Zetterberg, Kaj Blennow, Leslie Shaw, and John Q Trojanowski. Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people. *Archives of neurology*, 67(8):949–56, 2010. ISSN 1538-3687. doi: 10.1001/archneurol.2010.179. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2963067{&}tool=pmcentrez{&}rendertype=abstract.

[28] Rahul Desikan, Florent Sigonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, Marilyn S Albert, and Ronald J Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006. ISSN 1053-8119. URL http://www.sciencedirect.com/science/article/pii/S1053811906000437.

[29] Rahul Desikan, Howard Cabral, Christopher Hess, William Dillon, Christine Glastonbury, Michael Weiner, Nicholas Schmansky, Douglas Greve1, David Salat, Randy

Buckner, and Bruce Fischl. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain*, 132(8):2048–2057, 2009.

[30] Rahul S Desikan, Howard J Cabral, Fabio Settecase, Christopher P Hess, William P Dillon, Christine M Glastonbury, Michael W Weiner, Nicholas J Schmansky, David H Salat, Bruce Fischl, and Alzheimer's Disease Neuroimaging Initiative. Automated MRI measures predict progression to Alzheimer's disease. *Neurobiol Aging*, 31(8): 1364–1374, aug 2010.

[31] Ivana Despotović, Bart Goossens, Wilfried Philips, and Wilfried Philips. MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Computational and Mathematical Methods in Medicine*, 2015:1–23, mar 2015. ISSN 1748-670X. doi: 10.1155/2015/450341. URL http://www.hindawi.com/journals/cmmm/2015/450341/.

[32] Bradford C Dickerson, Akram Bakkour, David H Salat, Eric Feczko, Jenni Pacheco, Douglas N Greve, Fran Grodstein, Christopher I Wright, Deborah Blacker, H Diana Rosas, Reisa A Sperling, Alireza Atri, John H Growdon, Bradley T Hyman, John C Morris, Bruce Fischl, and Randy L Buckner. 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. *Cereb Cortex*, 19(3):497–510, mar 2009. doi: 10.1093/cercor/bhn113. URL http://dx.doi.org/10.1093/cercor/bhn113.

[33] Chris Ding and Hanchuan Peng. Minimum Redundancy Feature Selection from Microarray Gene Expression Data. *Journal of Bioinformatics and Computational Biology*, 3(2):185–2005, 2005.

[34] Allan Donner and FS Wood. Linear regression analysis with repeated measurements. *Journal of Chronic Diseases*, 37(6):441–448, jan 1984. ISSN 00219681. doi: 10.1016/0021-9681(84)90027-4. URL http://linkinghub.elsevier.com/retrieve/pii/0021968184900274.

[35] Michael Ewers, Giovanni B Frisoni, Stefan J Teipel, Lea T Grinberg, Edson Amaro, Helmut Heinsen, Paul M Thompson, and Harald Hampel. Staging Alzheimer's disease progression with multimodality neuroimaging. *Prog Neurobiol*, 95(4):535–546, dec 2011. doi: 10.1016/j.pneurobio.2011.06.004. URL http://dx.doi.org/10.1016/j.pneurobio.2011.06.004.

[36] Michael Ewers, Reisa A Sperling, William E Klunk, Michael W Weiner, and Harald Hampel. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's

disease dementia. *Trends Neurosci*, 34(8):430–442, aug 2011. doi: 10.1016/j.tins.2011.05.005. URL http://dx.doi.org/10.1016/j.tins.2011.05.005.

[37] Bruce Fischl. FreeSurfer. *Neuroimage*, jan 2012. doi: 10.1016/j.neuroimage.2012.01.021. URL http://dx.doi.org/10.1016/j.neuroimage.2012.01.021.

[38] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M Dale. Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron*, 33 (3):341–355, 2002. ISSN 0896-6273.

[39] Garrett Fitzmaurice and Geert Molenberghs. Advances in longitudinal data analysis: An historical perspective. In Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs, editors, *Longitudinal Data Analysis*, pages 3–27. Crc press edition, 2008. ISBN 978-1-58488-658-7. URL https://doi.org/10.1201/9781420011579.

[40] Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*. 2004. ISBN 9780470380277 (hardback)\r0470380276 (hardback). doi: 10.1198/jasa.2005.s24.

[41] Garrett M. Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. *Longitudinal data analysis*. CRC Press, 2009. ISBN 9781584886587.

[42] Anders M Fjell and Kristine B Walhovd. New tools for the study of Alzheimer's disease: what are biomarkers and morphometric markers teaching us? *Neuroscientist*, 17(5):592–605, oct 2011. doi: 10.1177/1073858410392586. URL http://dx.doi.org/10.1177/1073858410392586.

[43] Anders M Fjell, Kristine B Walhovd, Christine Fennema-Notestine, Linda K McEvoy, Donald J Hagler, Dominic Holland, James B Brewer, and Anders M Dale. One-year brain atrophy evident in healthy aging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(48):15223–31, dec 2009. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.3252-09.2009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2827793{&}tool=pmcentrez{&}rendertype=abstract.

[44] Anders M Fjell, Linda McEvoy, Dominic Holland, Anders M Dale, and Kristine B Walhovd. Brain changes in older adults at very low risk for Alzheimer's disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(19):

8237–42, 2013. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.5506-12.2013. URL http://www.ncbi.nlm.nih.gov/pubmed/23658162.

[45] Anders M. Fjell, Linda McEvoy, Dominic Holland, Anders M. Dale, and Kristine B. Walhovd. What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, 117, 2014. ISSN 18735118. doi: 10.1016/j.pneurobio.2014.02.004.

[46] M F Folstein, S E Folstein, and P R McHugh. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

[47] Katja Franke, Gabriel Ziegler, Stefan Klöppel, and Christian Gaser. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *NeuroImage*, 50(3):883–92, apr 2010. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2010.01.005. URL http://www.sciencedirect.com/science/article/pii/S1053811910000108.

[48] Katja Franke, Eileen Luders, Arne May, Marko Wilke, and Christian Gaser. Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. *NeuroImage*, 63(3):1305–12, nov 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.08.001. URL http://www.ncbi.nlm.nih.gov/pubmed/22902922.

[49] Christian Gaser, Katja Franke, Stefan Klöppel, Nikolaos Koutsouleris, and Heinrich Sauer. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. *PLoS ONE*, 8(6), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0067346.

[50] G. Gavidia, E. Soudah, M. Martín-Landrove, and M. Cerrolaza. Generación de modelos discretos de tejidos del ser humano a través del preprocesamiento y segmentación de imágenes médicas. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, 27(3):200–226, 2011. ISSN 02131315. doi: 10.1016/j.rimni.2011.07.002. URL http://linkinghub.elsevier.com/retrieve/pii/S0213131511000034.

[51] Giovana Gavidia-Bovadilla, Samir Kanaan-Izquierdo, María Mataró-Serrat, Alexandre Perera-Lluna, JW Suh, and MW Weiner. Early Prediction of Alzheimer's Disease Using Null Longitudinal Model-Based Classifiers. *PLOS ONE*, 12(1):e0168011, jan 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0168011. URL http://dx.plos.org/10.1371/journal.pone.0168011.

[52] Andrew Gelman. Analysis of variance — why it is more important than ever. *The Annals of Statistics*, 33(1):1–53, 2005. ISSN 0090-5364. doi: 10.1214/009053604000001048. URL http://projecteuclid.org/Dienst/getRecord?id=euclid.aos/1112967698/.

[53] I Guyon and A Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[54] Harald Hampel, Richard Frank, Karl Broich, Stefan J Teipel, Russell G Katz, John Hardy, Karl Herholz, Arun L W Bokde, Frank Jessen, Yvonne C Hoessler, Wendy R Sanhai, Henrik Zetterberg, Janet Woodcock, and Kaj Blennow. Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nat Rev Drug Discov*, 9(7):560–574, jul 2010. doi: 10.1038/nrd3115. URL http://dx.doi.org/10.1038/nrd3115.

[55] Harald Hampel, Gordon Wilcock, Sandrine Andrieu, Paul Aisen, Kaj Blennow, K Broich, Maria Carrillo, Nick C Fox, Giovanni B Frisoni, Maria Isaac, Simon Lovestone, Agneta Nordberg, David Prvulovic, Christina Sampaio, Philip Scheltens, Michael Weiner, Bengt Winblad, Nicola Coley, Bruno Vellas, and Oxford Task Force Group. Biomarkers for Alzheimer's disease therapeutic trials. *Prog Neurobiol*, 95(4):579–593, dec 2011. doi: 10.1016/j.pneurobio.2010.11.005. URL http://dx.doi.org/10.1016/j.pneurobio.2010.11.005.

[56] Harald Hampel, Simone Lista, and Zaven S. Khachaturian. Development of biomarkers to chart all Alzheimer's disease stages: The royal road to cutting the therapeutic Gordian Knot, 2012. ISSN 15525260.

[57] J. Hardy and G. Higgins. Alzheimer's disease: the amyloid cascade hypothesis. *Science*, 256(5054):184–185, 1992. ISSN 0036-8075. doi: 10.1126/science.1566067. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1566067.

[58] Rolf A Heckemann, Shiva Keihaninejad, Paul Aljabar, Katherine R Gray, Casper Nielsen, Daniel Rueckert, Joseph V Hajnal, and Alexander Hammers. Automatic morphometry in Alzheimer's disease and mild cognitive impairment. *NeuroImage*, 56 (4):2024–2037, 2011. ISSN 1053-8119. URL http://www.sciencedirect.com/science/article/pii/S1053811911002734.

[59] Xue Hua, Alex D Leow, Suh Lee, Andrea D Klunder, Arthur W Toga, Natasha Lepore, Yi-Yu Chou, Caroline Brun, Ming-Chang Chiang, Marina Barysheva, Clifford R

Jack, Matt A Bernstein, Paula J Britson, Chadwick P Ward, Jennifer L Whitwell, Bret Borowski, Adam S Fleisher, Nick C Fox, Richard G Boyes, Josephine Barnes, Danielle Harvey, John Kornak, Norbert Schuff, Lauren Boreta, Gene E Alexander, Michael W Weiner, Paul M Thompson, and Alzheimer's Disease Neuroimaging Initiative. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. *Neuroimage*, 41(1):19–34, may 2008. doi: 10.1016/j.neuroimage.2008.02.010. URL http://dx.doi.org/10.1016/j.neuroimage.2008.02.010.

[60] C R Jack, M A Bernstein, and N C Fox. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*, 27(4):685–691, 2008.

[61] Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, jan 2010. ISSN 14744422. doi: 10.1016/S1474-4422(09)70299-6. URL http://dx.doi.org/10.1016/S1474-4422(09)70299-6.

[62] T. Florian Jaeger, Peter Graff, William Croft, and Daniel Pontillo. Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology*, 15(2):281–319, 2011. ISSN 14300532. doi: 10.1515/LITY.2011.021.

[63] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to Statistical Learning*, volume 7. 2000. ISBN 978-1-4614-7137-0. doi: 10.1007/978-1-4614-7138-7.

[64] Jiyang Jiang, Perminder Sachdev, Darren M Lipnicki, Haobo Zhang, Tao Liu, Wanlin Zhu, Chao Suo, Lin Zhuang, John Crawford, Simone Reppermund, Julian Trollor, Henry Brodaty, and Wei Wen. A longitudinal study of brain atrophy over two years in community-dwelling older individuals. *NeuroImage*, 86:203–11, feb 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.08.022. URL http://www.ncbi.nlm.nih.gov/pubmed/23959201.

[65] Biao Jie, Daoqiang Zhang, Bo Cheng, and Dinggang Shen. Manifold regularized multi-task feature selection for multi-modality classification in Alzheimer's disease. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16 (Pt 1):275–83, jan 2013. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4109068{&}tool=pmcentrez{&}rendertype=abstract.

[66] G B Karas, P Scheltens, S.A.R.B. Rombouts, P J Visser, R A van Schijndel, N C Fox, and F Barkhof. Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 23(2):708–716, 2004.

[67] R J Killiany, T Gomez-Isla, M Moss, R Kikinis, T Sandor, F Jolesz, R Tanzi, K Jones, B T Hyman, and M S Albert. Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Ann Neurol*, 47(4):430–439, apr 2000.

[68] András Király, Nikoletta Szabó, Eszter Tóth, Gergő Csete, Péter Faragó, Krisztián Kocsis, Anita Must, László Vécsei, and Zsigmond Tamás Kincses. Male brain ages faster: the age and gender dependence of subcortical volumes. *Brain Imaging and Behavior*, 10(3):901–910, 2016. ISSN 19317565. doi: 10.1007/s11682-015-9468-3.

[69] Stefan Klöppel, Cynthia M Stonnington, Carlton Chu, Bogdan Draganski, Rachael I Scahill, Jonathan D Rohrer, Nick C Fox, Clifford R Jack, John Ashburner, and Richard S J Frackowiak. Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131(Pt 3):681–689, mar 2008. doi: 10.1093/brain/awm319. URL http://dx.doi.org/10.1093/brain/awm319.

[70] Nan M. Laird and James H. Ware. Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4):963, 1982. ISSN 0006341X. doi: 10.2307/2529876. URL http://www.jstor.org/stable/2529876?origin=crossref.

[71] S M Landau, D Harvey, C M Madison, E M Reiman, N L Foster, P S Aisen, R C Petersen, L M Shaw, J Q Trojanowski, C R Jack, M W Weiner, W J Jagust, and Alzheimer's Disease Neuroimaging Initiative. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*, 75(3):230–238, jul 2010.

[72] Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M Resnick, and Christos Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, jan 2004. ISSN 1053-8119. URL http://www.ncbi.nlm.nih.gov/pubmed/14741641.

[73] Feng Liu, Chong-Yaw Wee, Huafu Chen, and Dinggang Shen. Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. *NeuroImage*, 84:466–75, jan 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.09.015. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3849328{&}tool=pmcentrez{&}rendertype=abstract.

[74] Manhua Liu, Daoqiang Zhang, Dinggang Shen, and the Alzheimer's Disease Neuroimaging Alzheimer's Disease Neuroimaging Initiative. Ensemble sparse classification of Alzheimer's disease. *NeuroImage*, 60(2):1106–16, apr 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2012.01.055. URL http://www.ncbi.nlm.nih.gov/pubmed/22270352http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3303950.

[75] Manhua Liu, Daoqiang Zhang, Dinggang Shen, and the Alzheimer's Disease Neuroimaging Alzheimer's Disease Neuroimaging Initiative. Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. *Human brain mapping*, 35(4):1305–19, apr 2014. ISSN 1097-0193. doi: 10.1002/hbm.22254. URL http://www.ncbi.nlm.nih.gov/pubmed/23417832http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4132886.

[76] Yawu Liu, Valtteri Julkunen, Teemu Paajanen, Eric Westman, Lars-Olof Wahlund, Andrew Aitken, Tomasz Sobow, Patrizia Mecocci, Magda Tsolaki, Bruno Vellas, Sebastian Muehlboeck, Christian Spenger, Simon Lovestone, Andrew Simmons, and Hilkka Soininen. Education increases reserve against Alzheimer's disease–evidence from structural MRI analysis. *Neuroradiology*, 54(9):929–38, sep 2012. ISSN 1432-1920. doi: 10.1007/s00234-012-1005-0. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3435513{&}tool=pmcentrez{&}rendertype=abstract.

[77] Raymond Y Lo, Alan E Hubbard, Leslie M Shaw, John Q Trojanowski, Ronald C Petersen, Paul S Aisen, Michael W Weiner, William J Jagust, and Alzheimer's Disease Neuroimaging Initiative. Longitudinal change of biomarkers in cognitive decline. *Arch Neurol*, 68(10):1257–1266, oct 2011. doi: 10.1001/archneurol.2011.123. URL http://dx.doi.org/10.1001/archneurol.2011.123.

[78] a Lobo and Lj Launer. Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurology*, 54(June 2000):S4–S9, 2000. ISSN 0028-3878.

[79] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=17714011.

[80] Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *Journal of cognitive neuroscience*, 22(12):2677–84, dec 2010. ISSN 1530-8898. doi: 10.1162/jocn.2009.21407. URL http://www.ncbi.nlm.nih.gov/pubmed/19929323http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2895005.

[81] Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, Richard C Mohs, John C Morris, Martin N Rossor, Philip Scheltens, Maria C Carrillo, Bill Thies, Sandra Weintraub, and Creighton H Phelps. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*, 7(3):263–269, may 2011. doi: 10.1016/j.jalz.2011.03.005. URL http://dx.doi.org/10.1016/j.jalz.2011.03.005.

[82] Bjørn-Helge Mevik and Ron Wehrens. The pls Package: Principle Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18(2):1–24, 2007. ISSN 15487660. doi: 10.1002/wics.10. URL http://www.jstatsoft.org/v18/i02/paper.

[83] David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. Misc functions of the Department of Statistics (e1071), TU Wien, 2014. ISSN 1613-9011. URL http://hosho.ees.hokudai.ac.jp/{~}kubo/Rdoc/library/e1071/html/00Index.html.

[84] Camila Moraes, Jony Pinto, Marcos Lopes, Julio Litvoc, and Cassio Bottino. Impact of sociodemographic and health variables on mini-mental state examination in a community-based sample of older people. *European Archives of Psychiatry and Clinical Neuroscience*, 260(7):535–542, 2010. ISSN 0940-1334. URL http://dx.doi.org/10.1007/s00406-010-0104-3.

[85] Gary P Morris, Ian A Clark, and Bryce Vissel. Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta neuropathologica communications*, 2:135, jan 2014. ISSN 2051-5960. doi: 10.1186/s40478-014-0135-5. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4207354{&}tool=pmcentrez{&}rendertype=abstract.

[86] J C Morris. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, 43:2412–2414, 1993.

[87] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am*, 15(4):869—-77, xi–xii, nov 2005. doi: 10.1016/j.nic.2005.09.008. URL http://dx.doi.org/10.1016/j.nic.2005.09.008.

[88] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford R Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement*, 1(1):55–66, jul 2005. doi: 10.1016/j.jalz.2005.06.003. URL http://dx.doi.org/10.1016/j.jalz.2005.06.003.

[89] Susanne G Mueller, Norbert Schuff, Kristine Yaffe, Catherine Madison, Bruce Miller, and Michael W Weiner. Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease. *Hum Brain Mapp*, 31(9):1339–1347, sep 2010. doi: 10.1002/hbm.20934. URL http://dx.doi.org/10.1002/hbm.20934.

[90] Sean M Nestor, Raul Rupsingh, Michael Borrie, Matthew Smith, Vittorio Accomazzi, Jennie L Wells, Jennifer Fogarty, Robert Bartha, and Alzheimer's Disease Neuroimaging Initiative. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain*, 131(Pt 9):2443–2454, sep 2008.

[91] H Niu, I Álvarez-Álvarez, F Guillén-Grima, and I Aguinaga-Ontoso. Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis. *Neurologia (Barcelona, Spain)*, 2016. ISSN 1578-1968. doi: 10.1016/j.nrl.2016.02.016. URL http://www.elsevier.es/es-revista-neurologia-295-avance-resumen-prevalencia-e-incidencia-enfermedad-alzheimer-S021348

[92] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.159.

[93] R Peters. Ageing and the brain. *Postgraduate medical journal*, 82 (964):84–8, feb 2006. ISSN 1469-0756. doi: 10.1136/pgmj.2005.036665.

URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2596698{&}tool=pmcentrez{&}rendertype=abstract.

[94] José C. Pinheiro and Douglas M. Bates. *Mixed-Effects Models in S and S-Plus. - Springer Science & Business Media*. 2000. ISBN 0387989579. doi: 10.1198/jasa.2001.s411.

[95] Claudia Plant, Stefan J Teipel, Annahita Oswald, Christian Böhm, Thomas Meindl, Janaina Mourao-Miranda, Arun W Bokde, Harald Hampel, and Michael Ewers. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage*, 50(1):162–174, mar 2010. doi: 10.1016/j.neuroimage.2009.11.046. URL http://dx.doi.org/10.1016/j.neuroimage.2009.11.046.

[96] R Core Team. R: A Language and Environment for Statistical Computing, 2014. URL http://www.r-project.org/.

[97] Payam Refaeilzadeh, Lei Tang, and Huan Lui. Cross-Validation. In LING LIU and M. TAMER OZSU, editors, *Encyclopedia of Database Systems*, pages 532–538. Springer US, 2009. ISBN 9783805596992. doi: 10.1007/978-0-387-39940-9_565. URL http://link.springer.com/referenceworkentry/10.1007{%}2F978-0-387-39940-9{_}565.

[98] Martin Reuter, Herminia Diana Rosas, and Bruce Fischl. Highly Accurate Inverse Consistent Registration: A Robust Approach. *NeuroImage*, 53(4):1181–1196, 2010. doi: 10.1016/j.neuroimage.2010.07.020. URL http://dx.doi.org/10.1016/j.neuroimage.2010.07.020.

[99] Shannon L Risacher and Andrew J Saykin. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annual review of clinical psychology*, 9:621–48, 2013. ISSN 1548-5951. doi: 10.1146/annurev-clinpsy-050212-185535. URL http://www.ncbi.nlm.nih.gov/pubmed/23297785http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3955298.

[100] Shannon L Risacher, Li Shen, John D West, Sungeun Kim, Brenna C McDonald, Laurel A Beckett, Danielle J Harvey, Clifford R Jack Jr., Michael W Weiner, and Andrew J Saykin. Longitudinal MRI atrophy biomarkers: Relationship to conversion in the ADNI cohort. *Neurobiology of Aging*, 31(8):1401–1418, 2010.

[101] Kohavi Ron and John George H. Wrappers for Feature Subset Selection. *ARTIFICIAL INTELLIGENCE*, 97(1):273–324, 1997.

[102] W. G. Rosen, R. C. Mohs, and K. L. Davis. A new rating scale for Alzheimer's disease, 1984. ISSN 0002953X.

[103] Shaswati Roy and Pradipta Maji. A simple skull stripping algorithm for brain MRI. In *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, pages 1–6. IEEE, jan 2015. ISBN 978-1-4799-7458-0. doi: 10.1109/ICAPR.2015.7050671. URL http://ieeexplore.ieee.org/document/7050671/.

[104] David Salat, R L Buckner, A Z Snyder, Douglas N Greve, R S Desikan, Evelina Busa, J C Morris, Anders Dale, and Bruce Fischl. Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14:721–730, 2004.

[105] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack, M. W. Weiner, and Alzheimer's Disease Neuroimaging Initiative. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. *Brain*, 132(4):1067–1077, may 2008. ISSN 0006-8950. doi: 10.1093/brain/awp007. URL http://www.ncbi.nlm.nih.gov/pubmed/19251758http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2668943https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awp007.

[106] F Segonne, A M Dale, E Busa, M Glessner, D Salat, H K Hahn, and B Fischl. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22 (3):1060–1075, 2004. ISSN 1053-8119. doi: DOI:10.1016/j.neuroimage.2004.03.032. URL http://www.sciencedirect.com/science/article/B6WNP-4CF5CNY-1/2/33cc73136f06f019b2c11023e7a95341.

[107] David W Shattuck and Richard M Leahy. BrainSuite: an automated cortical surface identification tool. *Medical image analysis*, 6(2):129–42, jun 2002. ISSN 1361-8415. URL http://www.ncbi.nlm.nih.gov/pubmed/12045000.

[108] Leslie M Shaw, Hugo Vanderstichele, Malgorzata Knapik-Czajka, Christopher M Clark, Paul S Aisen, Ronald C Petersen, Kaj Blennow, Holly Soares, Adam Simon, Piotr Lewczuk, Robert Dean, Eric Siemers, William Potter, Virginia M-Y. Lee, John Q Trojanowski, and Alzheimer's Disease Neuroimaging Initiative. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*, 65(4):403–413, apr 2009.

[109] Leslie M Shaw, Hugo Vanderstichele, Malgorzata Knapik-Czajka, Michal Figurski, Els Coart, Kaj Blennow, Holly Soares, Adam J Simon, Piotr Lewczuk, Robert A

Dean, Eric Siemers, William Potter, Virginia M-Y Lee, John Q Trojanowski, and Alzheimer's Disease Neuroimaging Alzheimer's Disease Neuroimaging Initiative. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta neuropathologica*, 121(5):597–609, may 2011. ISSN 1432-0533. doi: 10.1007/s00401-011-0808-0. URL http://www.ncbi.nlm.nih.gov/pubmed/21311900http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3175107.

[110] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack, Jeffrey Kaye, Thomas J Montine, Denise C Park, Eric M Reiman, Christopher C Rowe, Eric Siemers, Yaakov Stern, Kristine Yaffe, Maria C Carrillo, Bill Thies, Marcelle Morrison-Bogorad, Molly V Wagster, and Creighton H Phelps. 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. *Alzheimers Dement*, 7(3):280–292, may 2011. doi: 10.1016/j.jalz.2011.03.003. URL http://dx.doi.org/10.1016/j.jalz.2011.03.003.

[111] G Spulber, A Simmons, J-S Muehlboeck, P Mecocci, B Vellas, M Tsolaki, I Kłoszewska, H Soininen, C Spenger, S Lovestone, L-O Wahlund, and E Westman. An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment. *Journal of internal medicine*, 273(4):396–409, apr 2013. ISSN 1365-2796. doi: 10.1111/joim.12028. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3605230{&}tool=pmcentrez{&}rendertype=abstract.

[112] Y Stern, B Gurland, T K Tatemichi, M X Tang, D Wilder, and R Mayeux. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*, 271(13):1004–1010, apr 1994.

[113] Heung-Il Suk and Dinggang Shen. Deep learning-based feature representation for AD/MCI classification. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 2):583–90, jan 2013. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4029347{&}tool=pmcentrez{&}rendertype=abstract.

[114] Heung-Il Suk, Seong-Whan Lee, and Dinggang Shen. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *NeuroImage*, 101:569–82, nov 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2014.06.077.

URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4165842{&}tool=pmcentrez{&}rendertype=abstract.

[115] Courtney L Sutphen, Anne M Fagan, and David M Holtzman. Progress update: fluid and imaging biomarkers in Alzheimer's disease. *Biological psychiatry*, 75 (7):520–6, apr 2014. ISSN 1873-2402. doi: 10.1016/j.biopsych.2013.07.031. URL http://www.ncbi.nlm.nih.gov/pubmed/24012326http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3947397.

[116] Jon B. Toledo, Sharon X. Xie, John Q. Trojanowski, and Leslie M. Shaw. Longitudinal change in CSF Tau and ABeta biomarkers for up to 48 months in ADNI. *Acta Neuropathologica*, 126(5):659–670, 2013. ISSN 00016322. doi: 10.1007/s00401-013-1151-4.

[117] Duygu Tosun, Norbert Schuff, Diana Truran-Sacrey, Leslie M Shaw, John Q Trojanowski, Paul Aisen, Ronald Peterson, Michael W Weiner, and Alzheimer's Disease Neuroimaging Initiative. Relations between brain tissue loss, CSF biomarkers, and the ApoE genetic profile: a longitudinal MRI study. *Neurobiol Aging*, 31(8): 1340–1354, aug 2010.

[118] Fedde van der Lijn. *Automated Atlas-Based Segmentation of Brain Structures in MR Images: Application to a Population-Based Imaging Study*. PhD thesis, Erasmus University Rotterdam, 2010.

[119] S A H van Hooren, A M Valentijn, H Bosma, R W H M Ponds, M P J van Boxtel, and J Jolles. Cognitive functioning in healthy older adults aged 64-81: a cohort study into the effects of age, sex, and education. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, 14(1):40–54, jan 2007. doi: 10.1080/138255890969483. URL http://dx.doi.org/10.1080/138255890969483.

[120] Vladimir N Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[121] Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Jennifer L Whitwell, Kejal Kantarci, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Clifford R Jack. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage*, 39(3):1186–1197, feb 2008. doi: 10.1016/j.neuroimage.2007.09.073. URL http://dx.doi.org/10.1016/j.neuroimage.2007.09.073.

[122] M W Weiner, D P Veitch, P S Aisen, L A Beckett, N J Cairns, J Cedarbaum, R C Green, D Harvey, C R Jack, W Jagust, J Luthman, J C Morris, R C Petersen, A J Saykin, L Shaw, L Shen, A Schwarz, A W Toga,

and J Q Trojanowski. 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's and Dementia*, 11(6):e1–e120, 2015. ISSN 1552-5260 1552-5279. doi: 10.1016/j.jalz.2014.11.001. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84931096931{&}partnerID=40{&}md5=f2f920e6431f111aa98f53765f5586ee.

[123] Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2):109–130, oct 2001. ISSN 01697439. doi: 10.1016/S0169-7439(01)00155-1. URL http://www.sciencedirect.com/science/article/pii/S0169743901001551.

[124] L. Wolf and A. Shashua. Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weighted-based approach. *Proceedings Ninth IEEE International Conference on Computer Vision*, 6:1855–1887, 2003. ISSN 1532-4435. doi: 10.1109/ICCV.2003.1238369.

[125] Michael S Wolfe. The role of tau in neurodegenerative diseases and its potential as a therapeutic target. *Scientifica*, 2012:796024, 2012. ISSN 2090-908X. doi: 10.6064/2012/796024. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3820460{&}tool=pmcentrez{&}rendertype=abstract.

[126] Ming-Shiang Wu, Tsuo-Hung Lan, Chun-Min Chen, Herng-Chia Chiu, and Tzuo-Yun Lan. Socio-demographic and health-related factors associated with cognitive impairment in the elderly in Taiwan. *BMC Public Health*, 11:22, 2011. doi: 10.1186/1471-2458-11-22. URL http://dx.doi.org/10.1186/1471-2458-11-22.

[127] J Xu, S Kobayashi, S Yamaguchi, K Iijima, K Okada, and K Yamashita. Gender effects on age-related changes in brain structure. *AJNR. American journal of neuroradiology*, 21(1):112–118, 2000. ISSN 0195-6108.

[128] Scott L. Zeger, Kung-Yee Liang, and Paul S. Albert. Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, 44(4):1049, 1988. ISSN 0006341X. doi: 10.2307/2531734. URL http://www.jstor.org/stable/2531734?origin=crossref.

[129] Daoqiang Zhang and Dinggang Shen. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*, 59(2):895–907, jan 2012. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.09.069. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3230721{&}tool=pmcentrez{&}rendertype=abstract.

Appendix A1

Supporting information: Validation of methods.

A1.1 Boxplot of residuals for ADNI and OASIS datasets.

A1.1.1 SV biomarkers

Females

Figures A1.1-A1.3 show quartiles comparison of female's residuals for SV MRI-based biomarkers.

Males

Figures A1.4-A1.6 show quartiles comparison of male's residuals for SV MRI-based biomarkers.

A1.1.2 CV biomarkers

Females

Quartiles comparison of female's residuals for CV MRI-based biomarkers are presented in Figures A1.7 and A1.8.

Males

Quartiles comparison of male's residuals for CV MRI-based biomarkers are presented in Figures A1.9 and A1.10.

A1.1.3 TA biomarkers

Females

Quartiles comparison of female's residuals for TA MRI-based biomarkers are presented in Figures A1.11 and A1.12.

Males

Finally, quartiles comparison of male's residuals for TA MRI-based biomarkers are presented in Figures A1.13 and A1.14.

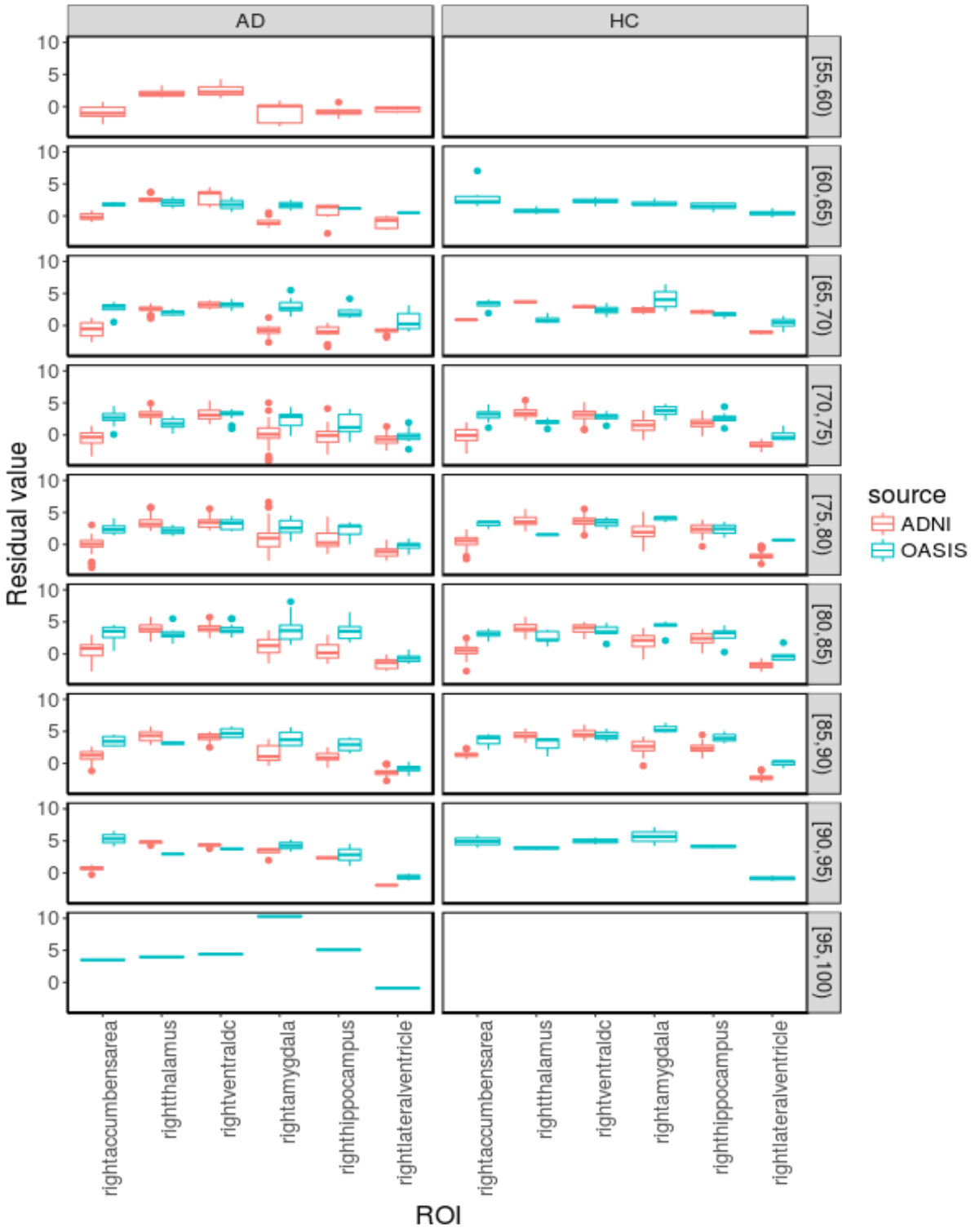


Fig. A1.1 Right hemisphere: Quantiles comparison of the SV biomarker residuals for females of ADNI and OASIS data.

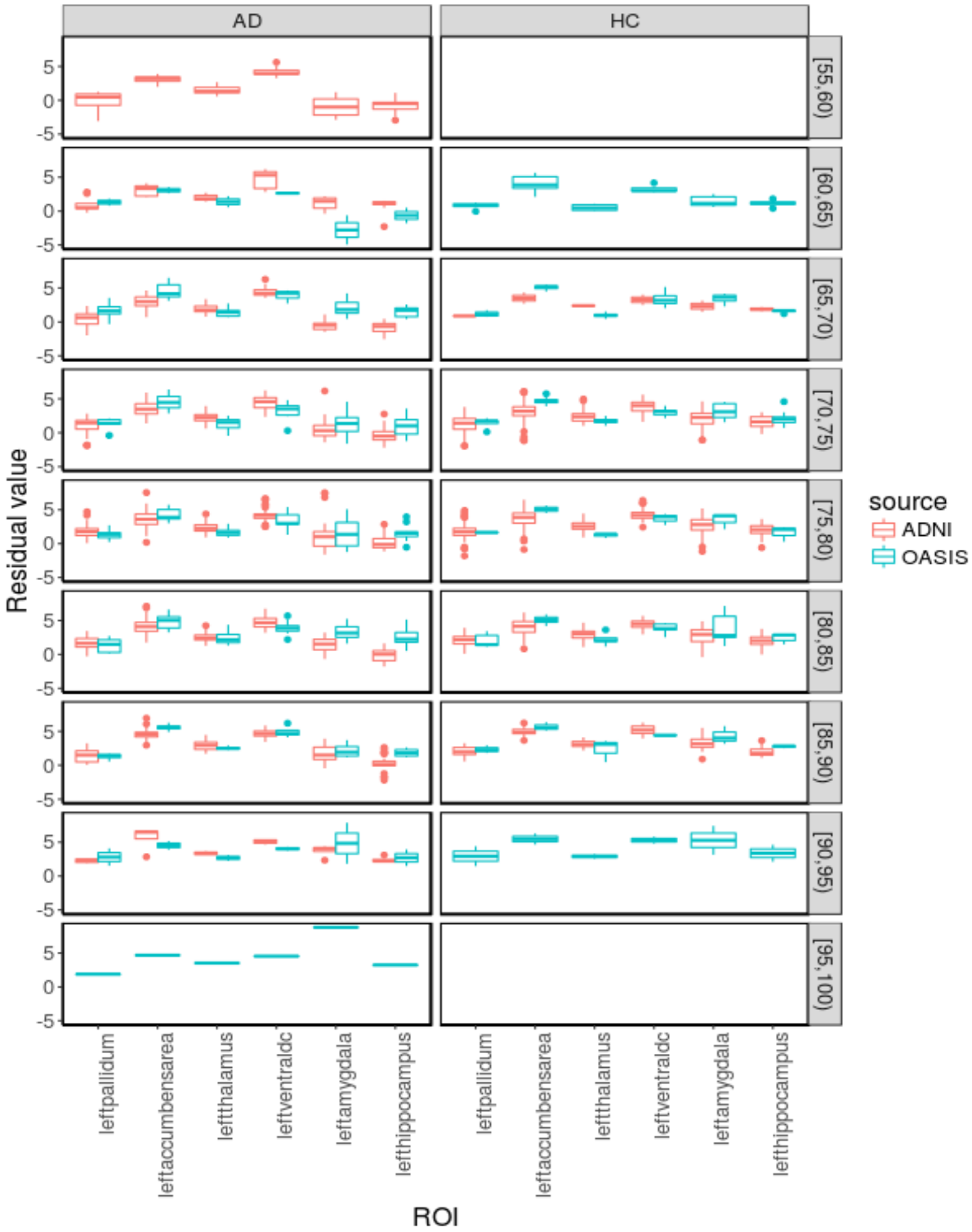


Fig. A1.2 Left hemisphere: Quantiles comparison of SV biomarker residuals for females of ADNI and OASIS data.

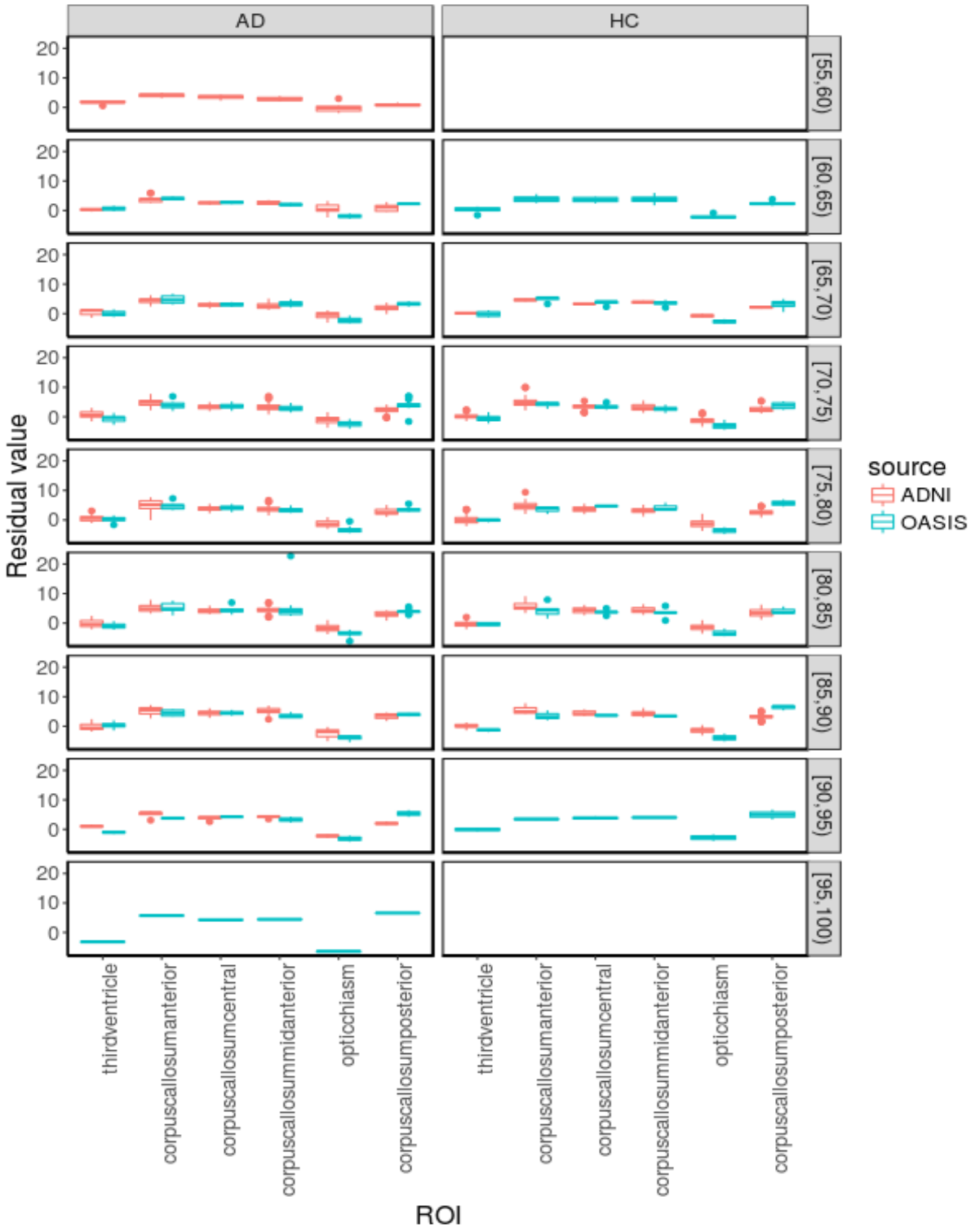


Fig. A1.3 Bilateral regions: Quantiles comparison of SV biomarker residuals for females of ADNI and OASIS data.

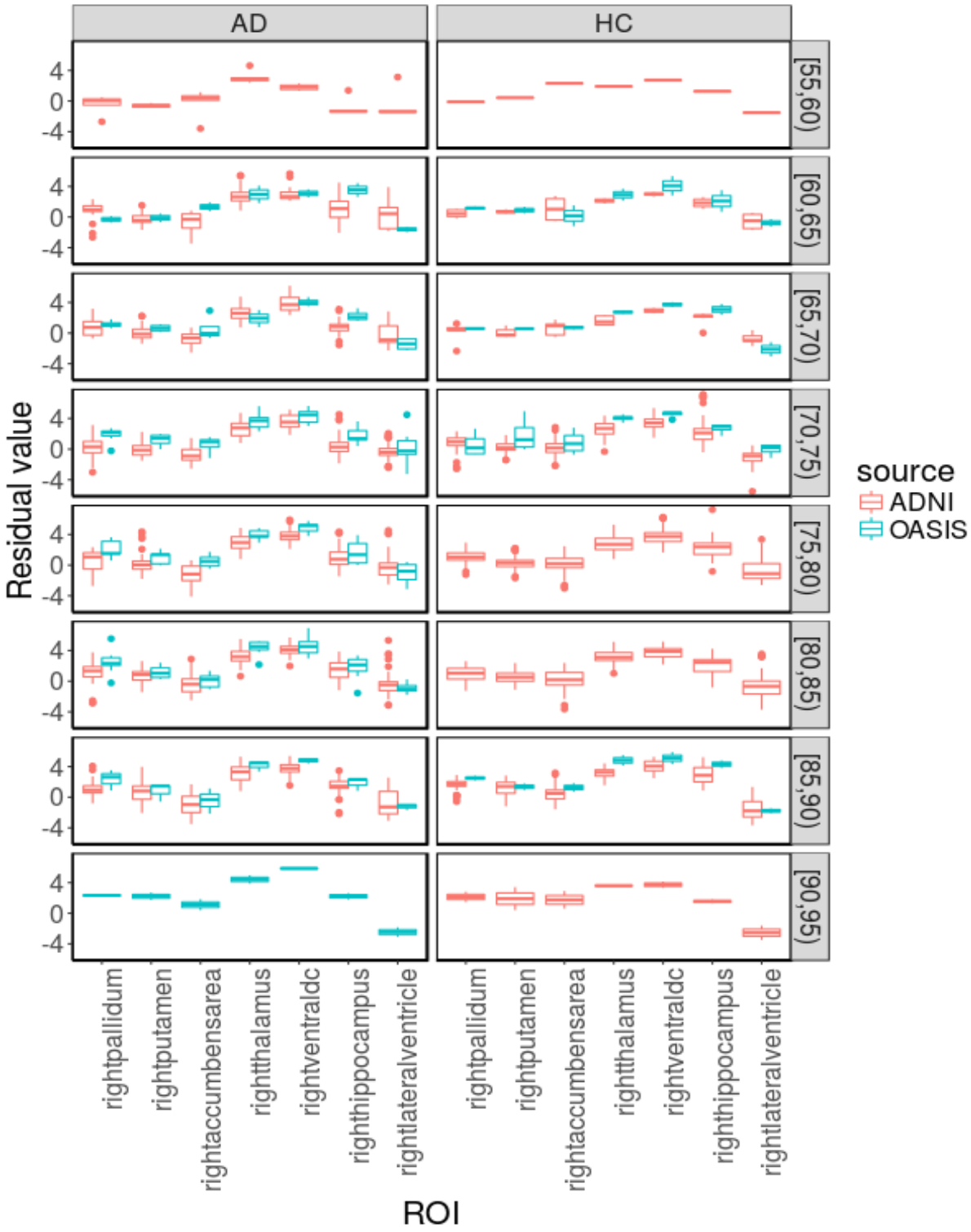


Fig. A1.4 Right hemisphere: Quantiles comparison of the SV biomarker residuals for males of ADNI and OASIS data.

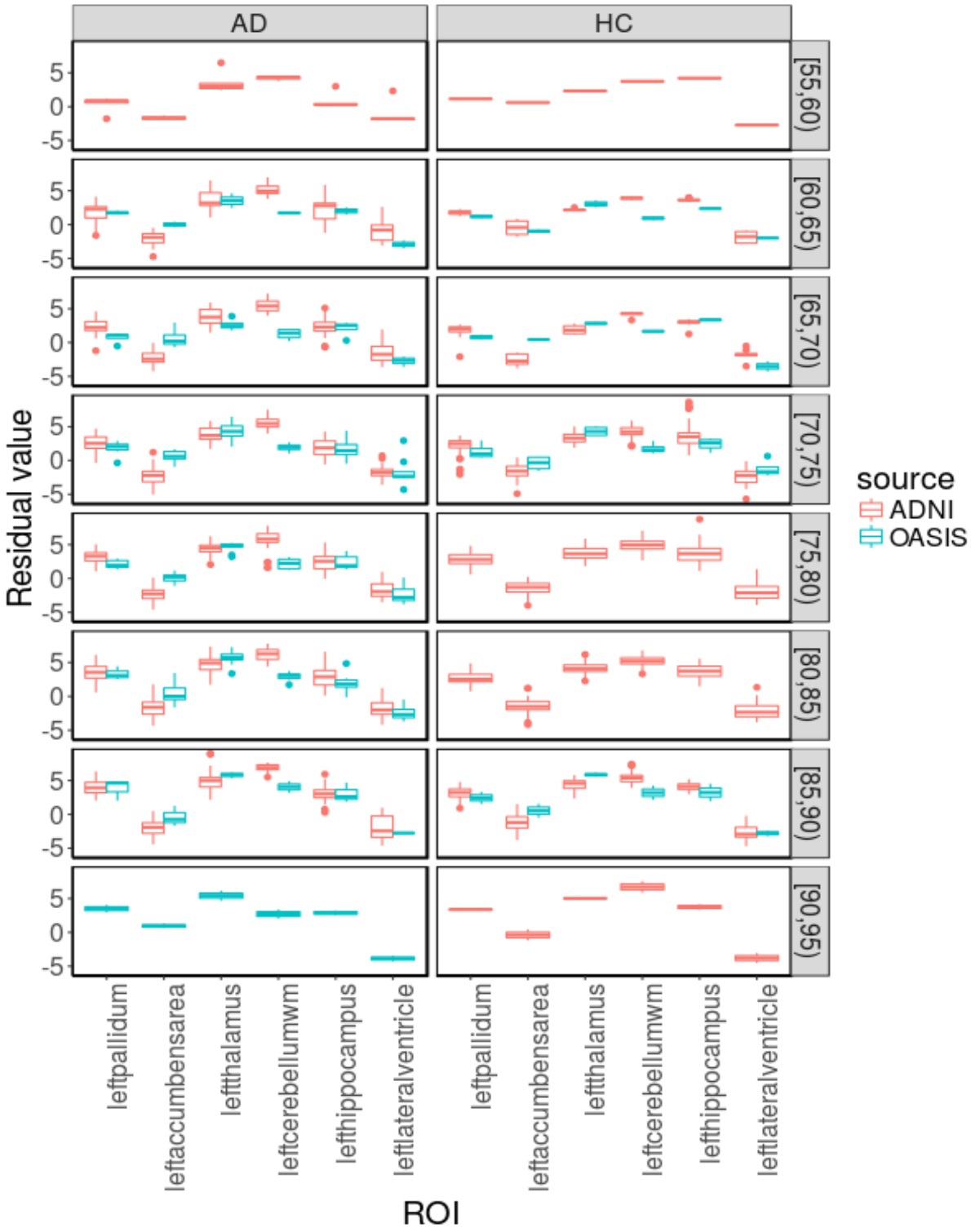


Fig. A1.5 Left hemisphere: Quantiles comparison of SV biomarker residuals for males of ADNI and OASIS data.

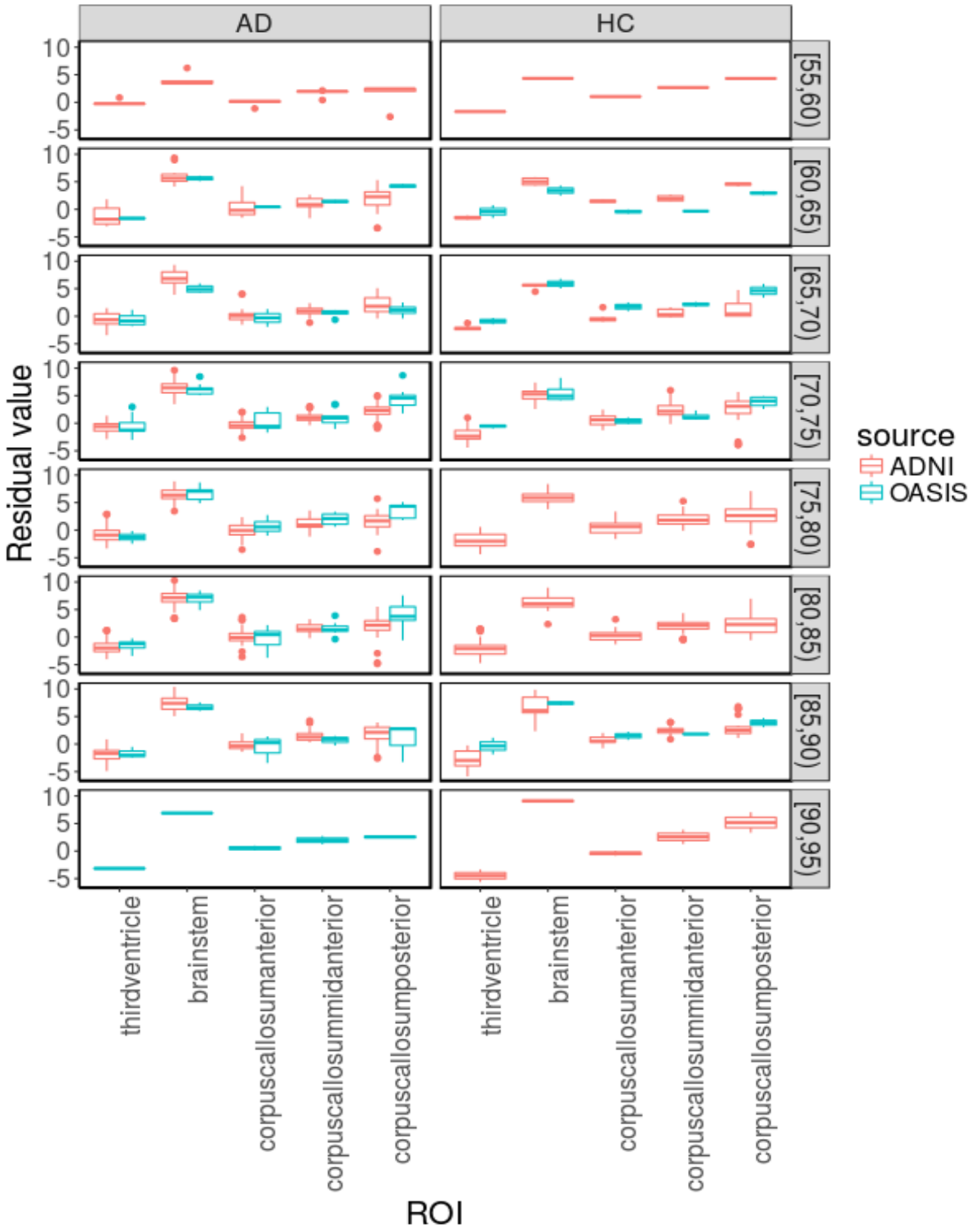


Fig. A1.6 Bilateral regions: Quantiles comparison of SV biomarker residuals for males of ADNI and OASIS data.

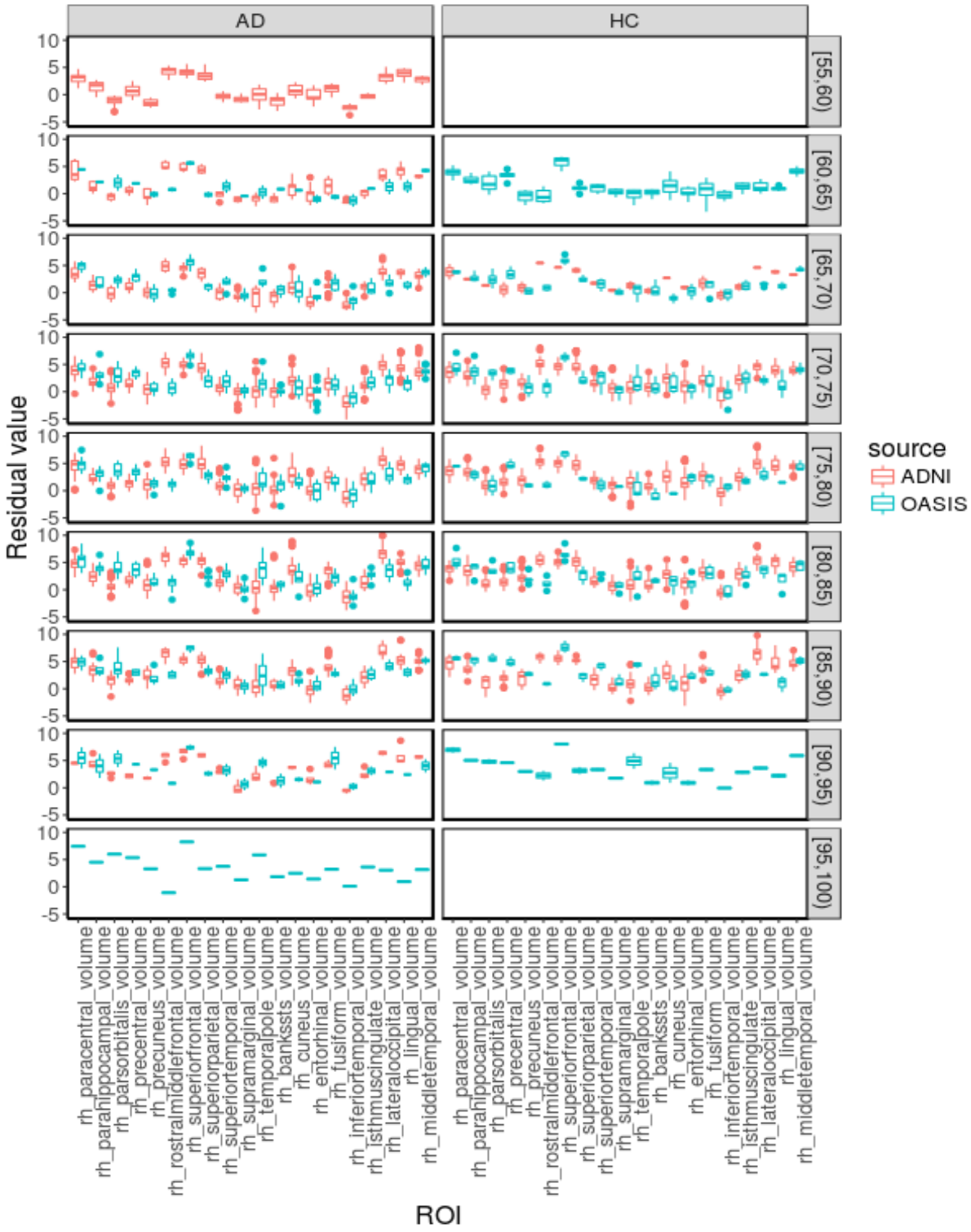


Fig. A1.7 Right hemisphere: Quantiles comparison of the CV biomarker residuals for females of ADNI and OASIS data.

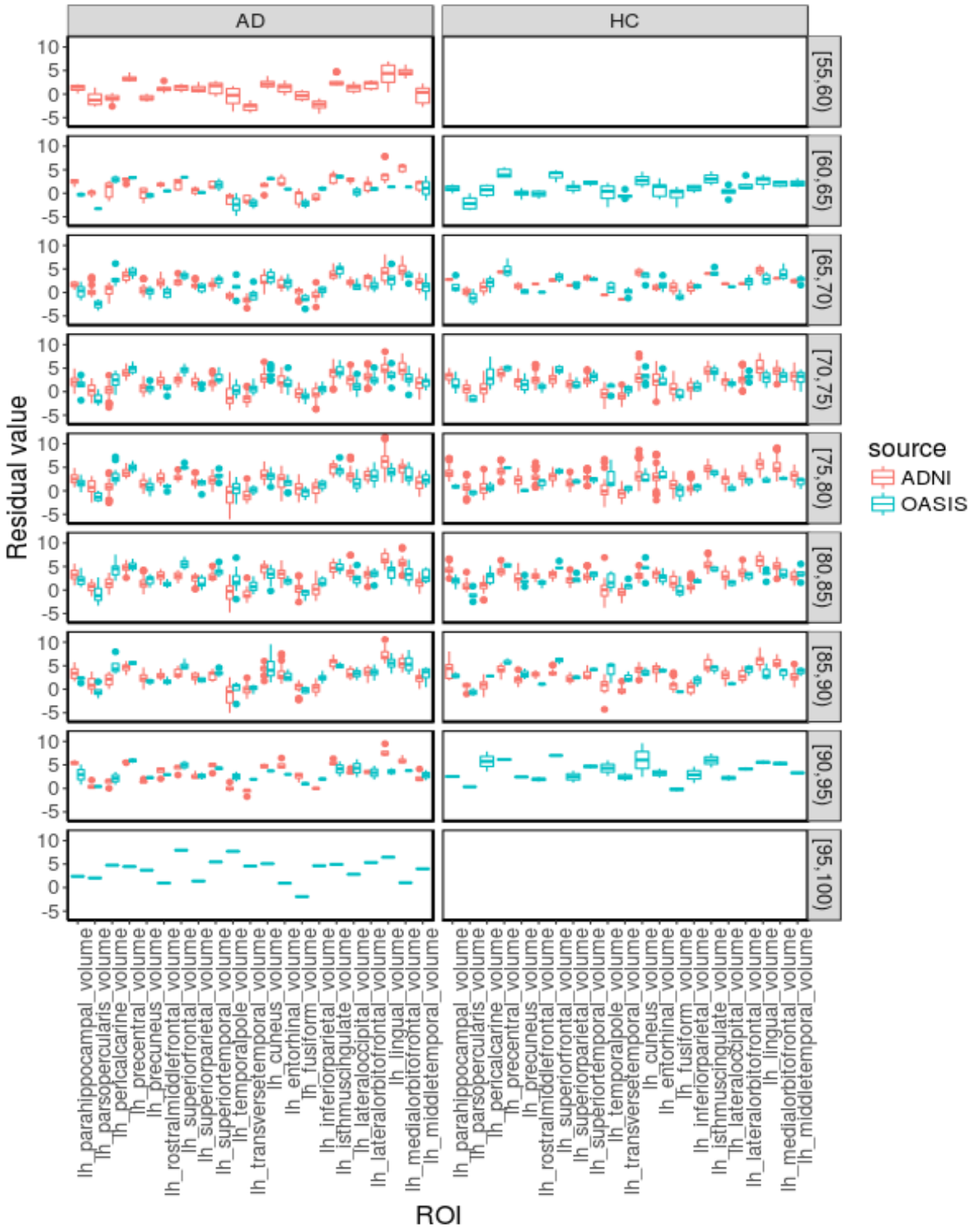


Fig. A1.8 Left hemisphere: Quantiles comparison of CV biomarker residuals for females of ADNI and OASIS data.

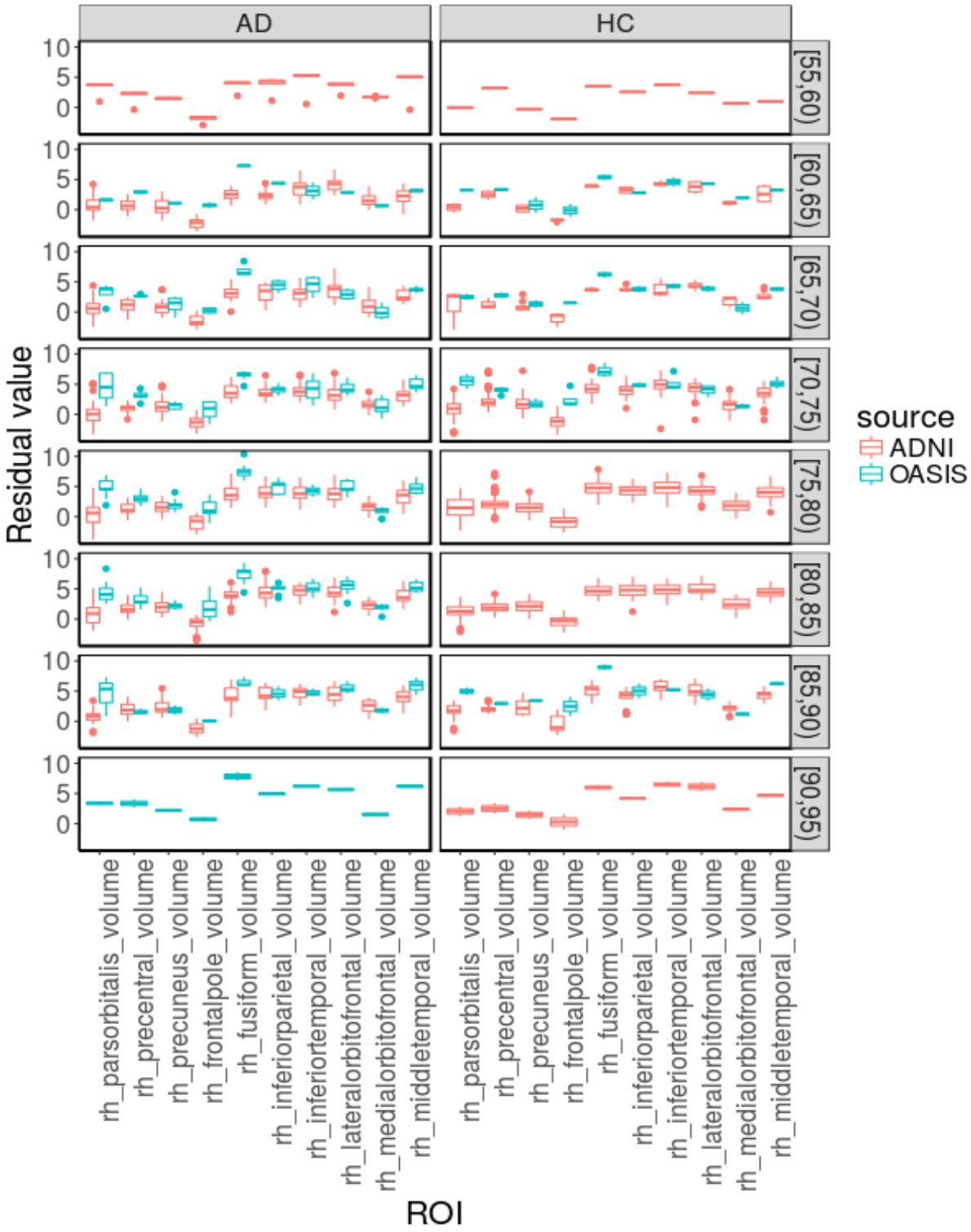


Fig. A1.9 Right hemisphere: Quantiles comparison of the CV biomarker residuals for males of ADNI and OASIS data.

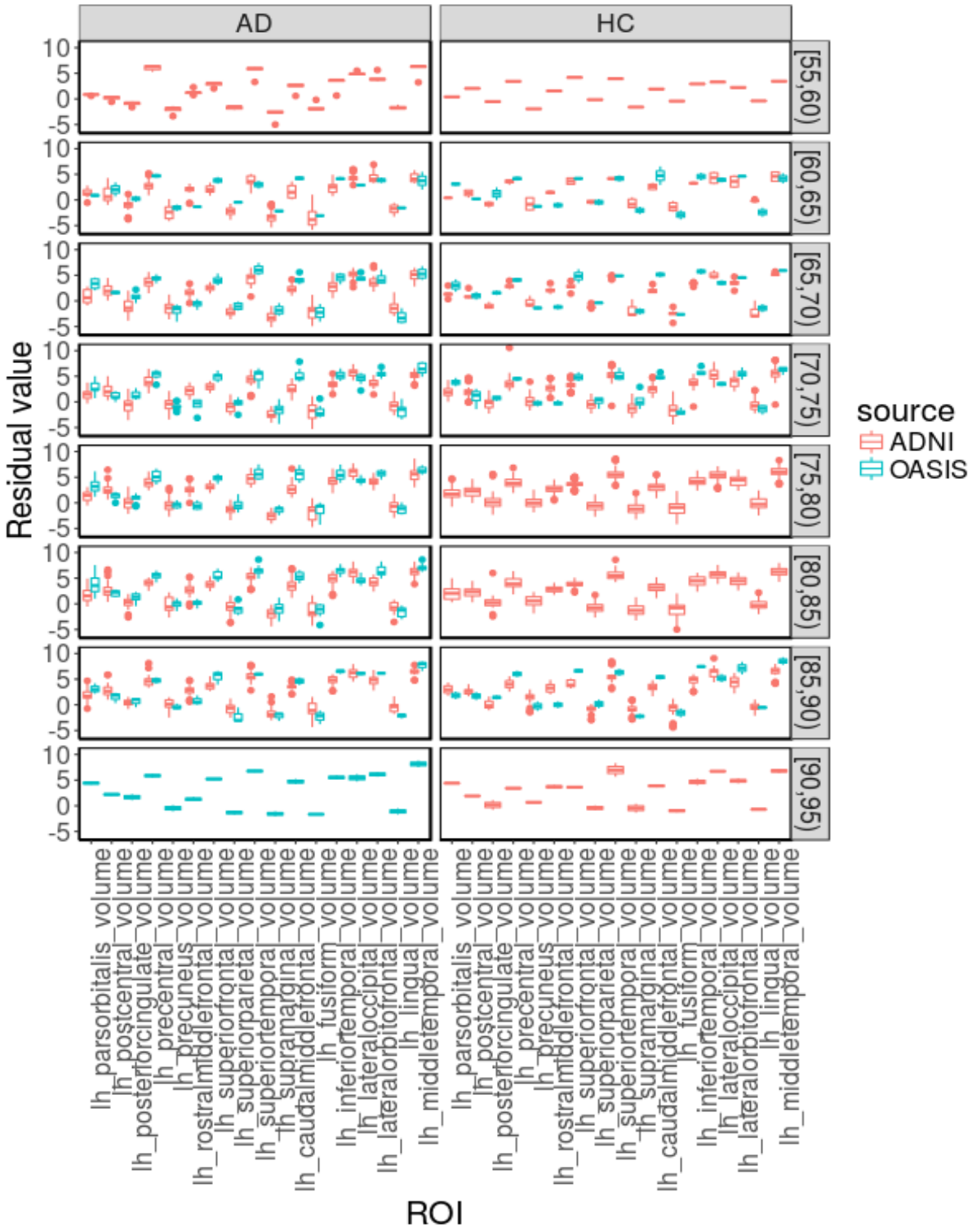


Fig. A1.10 Left hemisphere: Quantiles comparison of CV biomarker residuals for males of ADNI and OASIS data.

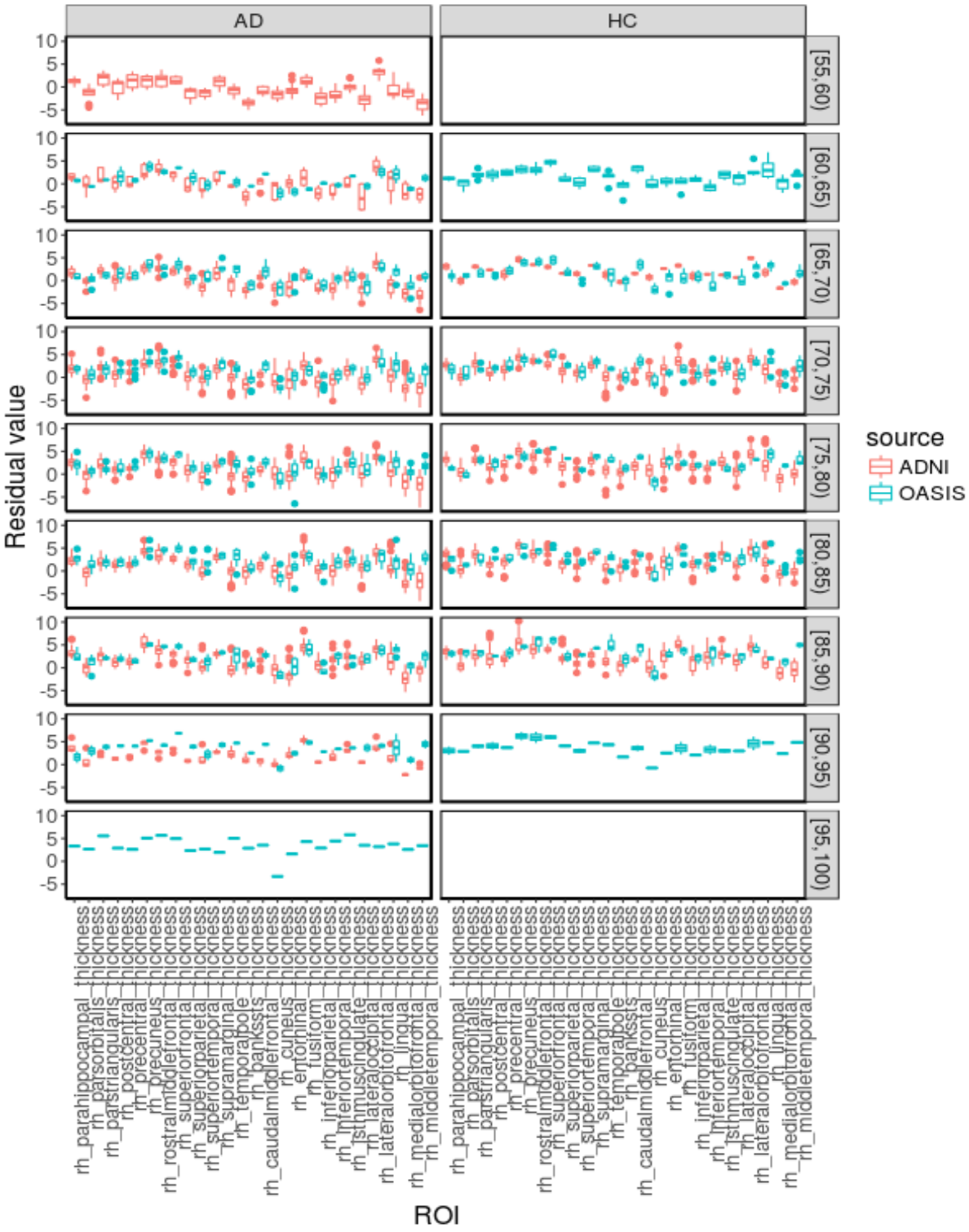


Fig. A1.11 Right hemisphere: Quantiles comparison of the TA biomarker residuals for females of ADNI and OASIS data.

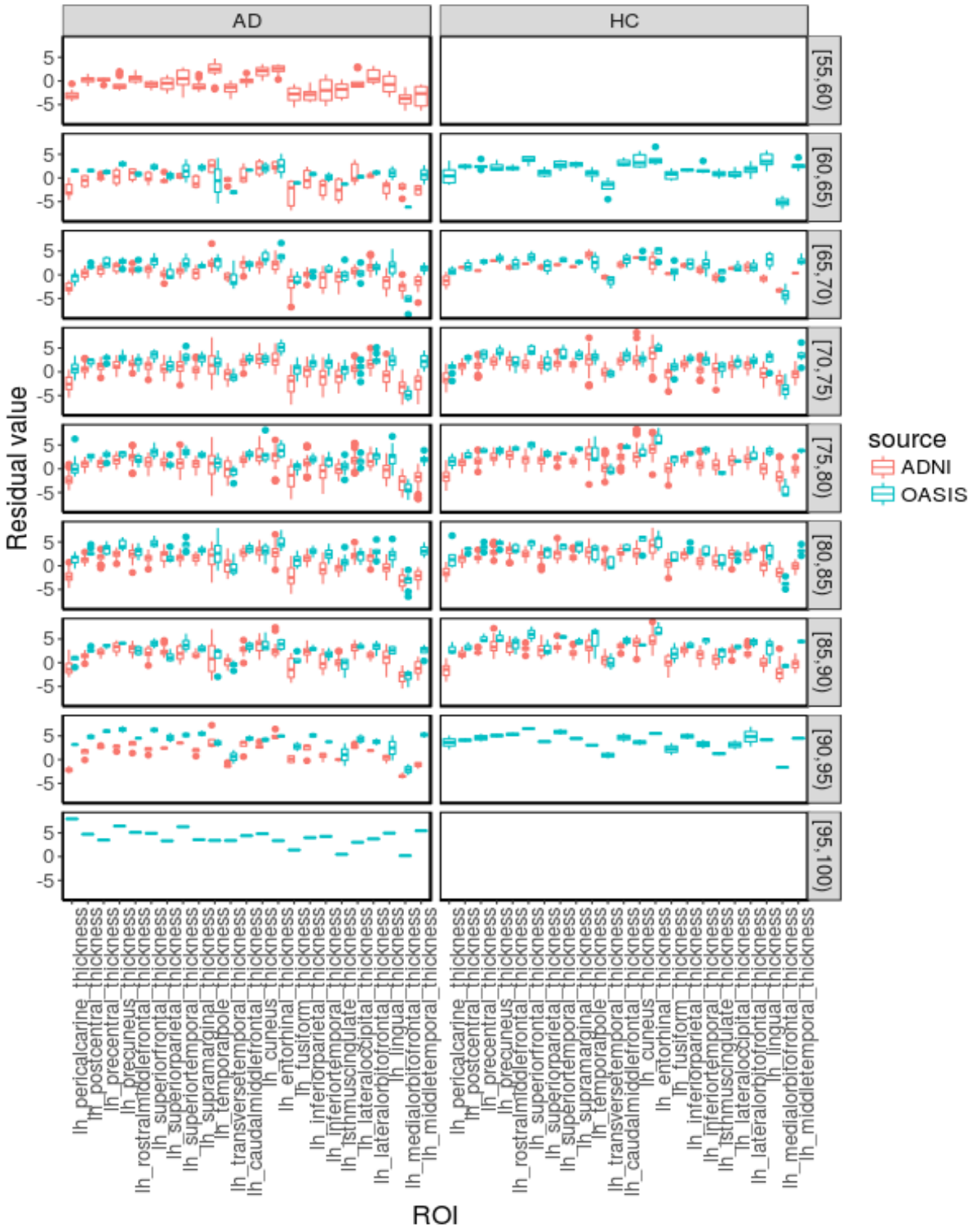


Fig. A1.12 Left hemisphere: Quantiles comparison of TA biomarker residuals for females of ADNI and OASIS data.

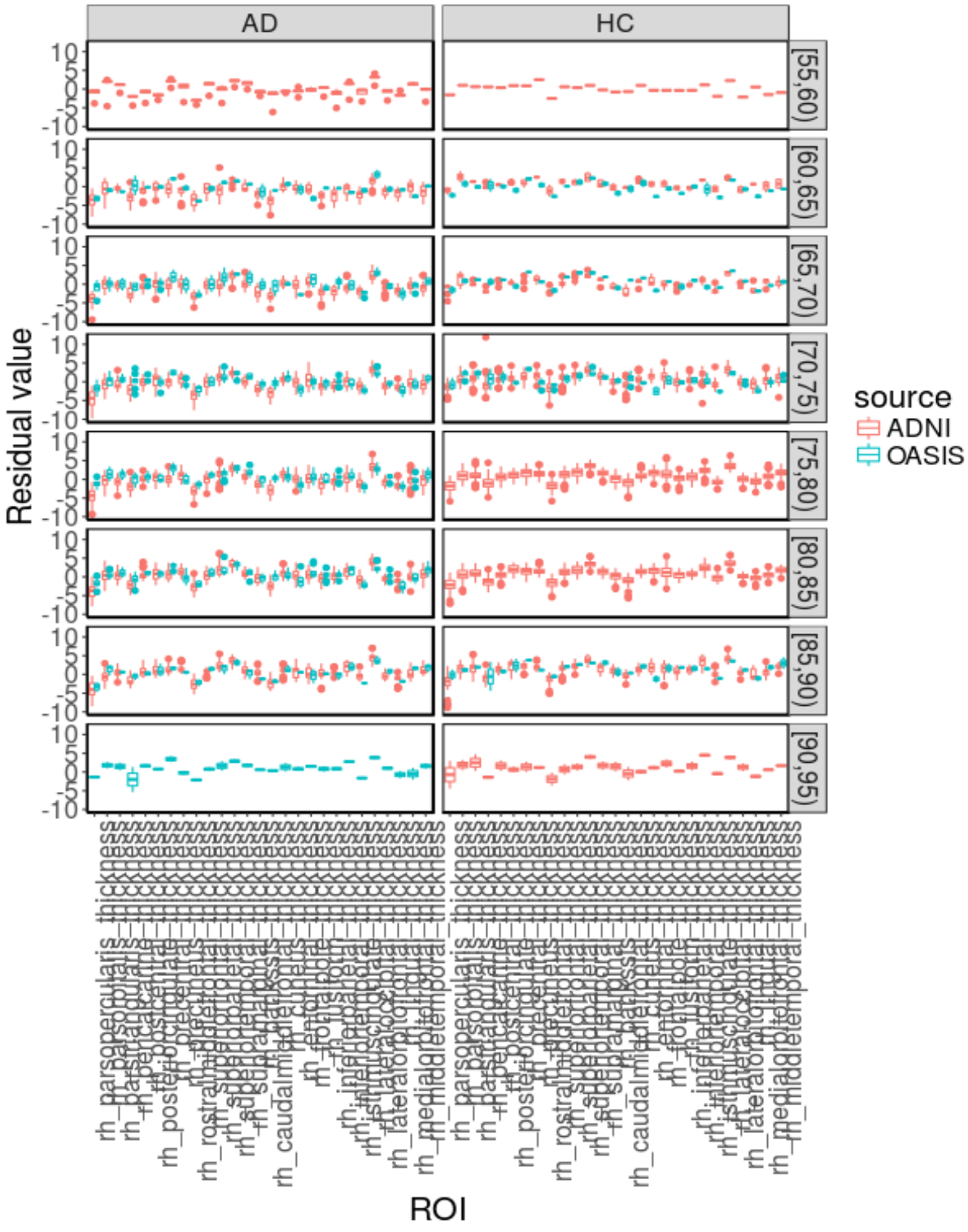


Fig. A1.13 Right hemisphere: Quantiles comparison of the TA biomarker residuals for males of ADNI and OASIS data.

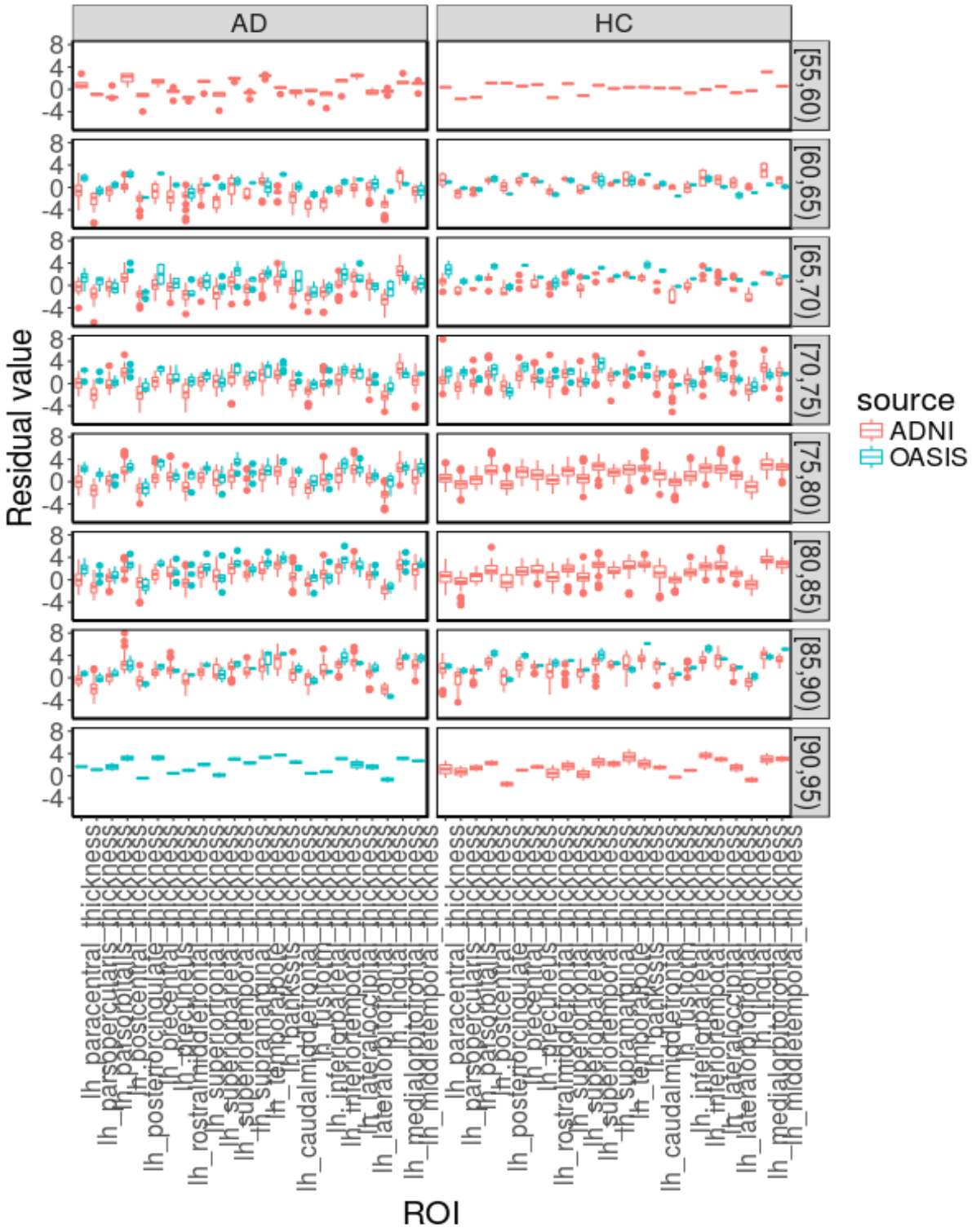


Fig. A1.14 Left hemisphere: Quantiles comparison of TA biomarker residuals for males of ADNI and OASIS data.