Computational anatomy strategies for characterization of brain patterns associated with Alzheimer’s disease

Diana Lorena Giraldo Franco

Universidad Nacional de Colombia
Facultad de Ingeniería
Bogotá, Colombia
2022

Universiteit Antwerpen
Faculteit Wetenschappen
Antwerpen, België
2022
Computational anatomy strategies for characterization of brain patterns associated with Alzheimer’s disease

Diana Lorena Giraldo Franco

In fulfillment of the requirements for the degrees of:
PhD in Engineering
Doctor of Science

Advisors:
Eduardo Romero Castro, Ph.D.
Jan Sijbers, Ph.D.
Ben Jeurissen, Ph.D.

Research Groups:
Computer Imaging and Medical Applications Laboratory - CIM@LAB
Imec-Vision Lab

Universidad Nacional de Colombia
Faculty of Engineering
Department of Systems and Industrial Engineering
Bogota, Colombia
2022

University of Antwerp
Faculty of Science
Department of Physics
Antwerp, Belgium
2022
You ask about the effect my work has on others.

If I may speak frivolously, that’s a masculine question.

Men always want to be influential.

Don’t see myself as influential? No, I want to understand.
Acknowledgements

This is the space where I can recognize all the people who made this work possible, the ones who gave me incredible opportunities and support, the ones who gave me the conditions to take those opportunities, and the ones who were around during these years under quite unexpected circumstances.

I want to express my sincere gratitude to Eduardo Romero, who was been a great teacher during the last 10 years. His guidance and advice have played a fundamental role in who I am as a researcher and as a person. I deeply appreciate his constant support, even in the most difficult and uncertain times, and the fact that he always believed in me way more than I did in myself. I need to highlight his patience insistence on teaching me an important lesson for life, that sometimes the best you can do is to shut up and listen.

To Jan Sijbers for giving me the incredible opportunity to work on the Visionlab, opening my eyes to a whole new environment and culture. To Ben Jeurissen for his dedication, time, help, trust, and fruitful conversations that shaped a large part of this thesis. Their kindness and willingness to support me were (and still are) remarkable. I will always appreciate the warmth, joy and consideration with which Jan and Ben received me in Antwerp for the first time during the autumn of 2015, when I was still in a state of confusion about language, weather, food, and all the things related to being in a different continent for the first time in my life.

I was given great opportunities during my career, opportunities that I could take thanks to the hard work of my parents. They encouraged me and gave me the conditions to be admitted to a public university to study Mathematics. Their effort and support on many of the decisions I make have made this road possible.

Es más que necesario escribir estos agradecimientos también en español. Esta tesis representa un punto emblemático de un camino sin final, y este camino no habría sido posible sin el esfuerzo y apoyo de mi papás. A ellos, quiénes trabajaron gran parte de sus vidas para asegurarse de que tuviéramos las condiciones y oportunidades que para ellos fueron escasas, les debo poder estar escribiendo estos agradecimientos.

To the Cim@lab research group for being a great environment to exchange ideas and opinions. Special thanks Germán, Chars, and Josué, people who I can consider my friends. Thanks for your help, support, conversations and for all the unforgettable moments we have shared. The most special thanks go to Josué, my partner, for having my back for the last couple of years. I doubt I would have finished this thesis under these unexpected circumstances without his care and love.

I also want to thank the Visionlab people who welcomed me with kindness, especially to Gabriel and Piet with whom I got the luck to share an office and quickly became my friends.
We had a lot of fun together, and I will never forget Pearl Harbour thanks to the two of you... unless something strange happens with my memory (pun intended).

To all my friends who gave me their unconditional support and were by my side during these years: Mabe, Peri, Laura A., Laura E., Emmis, Rodrigo, Daniel, Beatriz, Cata y Federico. I consider myself a lucky person because there are many people that I can call my friends. I want to extend my gratitude to them, although some of these people still think this thesis is about robotic arms for babies with cancer.

Muchas gracias
Thanks a lot
Dank je wel
Abstract

Computational anatomy strategies for characterization of brain patterns associated with Alzheimer's disease

Alzheimer’s disease (AD) is one of the most complex systematic malfunctions of the nervous system that are known. The clinical symptoms of this neurodegenerative disease are alterations in cognition and behaviour that can lead to the onset of a dementia syndrome. Disease mechanisms that lead to neurodegeneration and cognitive impairment in sporadic AD are not well understood yet, making it difficult to predict the clinical progression of patients at the early stages of the AD continuum. Currently, no single biomarker or exam is sufficient to diagnose AD and existing standard instruments are not sensitive enough to detect subtle changes, predict the clinical course, and recognize heterogeneous forms of AD. This thesis presents two computational anatomy strategies aiming to identify and quantify neurodegeneration patterns associated with different clinical stages along the AD continuum using two different modalities of magnetic resonance imaging. A third contribution consists of a data-driven strategy to develop a set of domain-specific scores that result useful to estimate the risk of and predict the progression from mild cognitive impairment to dementia. Evaluation of these strategies with machine-learning and statistical inference methods demonstrate the potential of the proposed quantitative tools to help patients’ clinical management and monitoring and could be used to improve the evaluation of potential disease-modifying interventions.

Keywords: Alzheimer’s Disease, Neuroimaging, Medical Image Processing, Magnetic Resonance Imaging, Cognitive Impairment.
Resumen

Estrategias de anatomía computacional para la caracterización de patrones cerebrales asociados a la enfermedad de Alzheimer

La enfermedad de Alzheimer (EA) es una de las fallas sistemáticas del sistema nervioso más complejas que se conocen. Los síntomas clínicos de esta enfermedad neurodegenerativa son alteraciones de la cognición y el comportamiento que pueden conducir a la aparición de un síndrome de demencia. Los mecanismos de la enfermedad que conducen a la neurodegeneración y al deterioro cognitivo en la EA aún no se conocen bien, lo que dificulta la predicción de la evolución clínica de los pacientes en las primeras fases de la EA. Actualmente, ningún biomarcador o examen es suficiente para diagnosticar la EA y los instrumentos estándar existentes no son lo suficientemente sensibles para detectar cambios sutiles, predecir el curso clínico o reconocer presentaciones atípicas de EA. Esta tesis presenta dos estrategias de anatomía computacional destinadas a identificar y cuantificar los patrones de neurodegeneración asociados a diferentes etapas clínicas a lo largo del continuo de la EA utilizando dos modalidades diferentes de imágenes de resonancia magnética. Una tercera contribución consiste en una estrategia guiada por datos para desarrollar un conjunto de puntajes específicos por dominio que resultan útiles para estimar el riesgo y predecir la progresión del deterioro cognitivo leve a la demencia. La evaluación de estas estrategias con métodos de aprendizaje automático y de inferencia estadística demuestra el potencial de las herramientas cuantitativas propuestas para ayudar al manejo y el seguimiento clínico de los pacientes y podría utilizarse para mejorar la evaluación de posibles intervenciones que puedan modificar el curso de la enfermedad.

Palabras clave: Enfermedad de Alzheimer, Neuroimágenes, Procesamiento de imágenes médicas, Imágenes de Resonancia Magnética, Deterioro Cognitivo.
Samenvatting

Computationele anatomie strategieën voor karakterisering van hersenpatronen geassocieerd met de ziekte van Alzheimer

De ziekte van Alzheimer (AD) is een van de meest complexe systemische storingen van het zenuwstelsel die bekend zijn. De klinische symptomen van deze neurodegeneratieve ziekte zijn veranderingen in cognitie en gedrag die kunnen leiden tot het ontstaan van een dementiesyndroom. De ziektemechanismen die leiden tot neurodegeneratie en cognitieve stoornissen bij sporadische AD zijn nog niet goed begrepen, waardoor het moeilijk is om de klinische progressie van patiënten in de vroege stadia van het AD continuüm te voorspellen. Momenteel is geen enkele biomarker of onderzoek voldoende om de diagnose AD te stellen en de bestaande standaardinstrumenten zijn niet gevoelig genoeg om subtiele veranderingen te detecteren, het klinische verloop te voorspellen en heterogene vormen van AD te herkennen. Dit proefschrift presenteert twee computationele anatomie strategieën die gericht zijn op het identificeren en kwantificeren van neurodegeneratie patronen geassocieerd met verschillende klinische stadia in het AD continuüm, gebruikmakend van twee verschillende modaliteiten van magnetische resonantie beeldvorming. Een derde bijdrage bestaat uit een data-gestuurde strategie om een reeks van domeinspecifieke scores te ontwikkelen die bruikbaar zijn om het risico in te schatten op en de progressie te voorspellen van milde cognitieve stoornissen naar dementie. Evaluatie van deze strategieën met machine-learning en statistische inferentie methoden tonen het potentieel aan van de voorgestelde kwantitatieve instrumenten om het klinisch management en de monitoring van patiënten te helpen en zouden gebruikt kunnen worden om de evaluatie van potentiële ziekte-modifierende interventies te verbeteren.

Sleutelwoorden: Ziekte van Alzheimer, Neurobeeldvorming, Medische Beeldverwerking, Magnetische Resonantie Beeldvorming, Cognitieve Stoornis.
Esta tesis de doctorado se sustentó el 18 de Febrero de 2022 a las 9:00 a.m., y fue evaluada por los siguientes jurados:

Jury members:

Ileana Jelescu, PhD.
Professor
Université de Lausanne

Norberto Malpica, PhD.
Professor
Universidad Rey Juan Carlos

Marleen Verhoye, Phd.
Professor
Universiteit Antwerpen

Sabine Van Doorslaer, PhD.
Professor
Universiteit Antwerpen

Fabio Augusto González Osorio, PhD.
Professor
Universidad Nacional de Colombia

Diego Alexander Garzón Alvarado, PhD.
Professor
Universidad Nacional de Colombia
Contents

Acknowledgements vii
Abstract ix
List of acronyms and abbreviations xvi

1 Introduction 1
  1.1 Alzheimer’s disease dementia . . . . . . . . . . . . . . . . . . . . . . . . . 1
  1.2 Computational anatomy in AD . . . . . . . . . . . . . . . . . . . . . . . . . 10
  1.3 This Thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

2 Comparing region intensity distributions 20
  2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
  2.2 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
     2.2.1 Region description . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
     2.2.2 Automated classification of brain images . . . . . . . . . . . . . . . 23
  2.3 Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
     2.3.1 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
     2.3.2 Cross-validation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
  2.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
     2.4.1 Classification between patients and controls . . . . . . . . . . . . . . 26
     2.4.2 Region importance . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
  2.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
  2.6 Products . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32

3 Investigating tissue-specific abnormalities in AD with DW-MRI 34

4 Improving the quantitative characterization of cognitive impairment 35
  4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
  4.2 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
     4.2.1 Participants data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
     4.2.2 Learning domain composite scores . . . . . . . . . . . . . . . . . . . . 40
     4.2.3 Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  4.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
     4.3.1 Parameters for composite scores calculation . . . . . . . . . . . . . . 44
     4.3.2 Sub-groups of MCI patients . . . . . . . . . . . . . . . . . . . . . . . . 47
     4.3.3 Automated prediction of progression to AD dementia . . . . . . . . . 50
  4.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
List of acronyms and abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>ADAS-Cog</td>
<td>Alzheimer’s Disease Assessment Scale - Cognition</td>
</tr>
<tr>
<td>ADC</td>
<td>Apparent diffusion coefficient</td>
</tr>
<tr>
<td>ADD</td>
<td>Alzheimer’s Disease Dementia</td>
</tr>
<tr>
<td>ADNI</td>
<td>Alzheimer’s Disease Neuroimaging Initiative</td>
</tr>
<tr>
<td>AFD</td>
<td>Apparent fibre density</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>Aβ42</td>
<td>Amyloid-β peptide 42</td>
</tr>
<tr>
<td>CDR</td>
<td>Clinical Dementia Rating</td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory factor analysis</td>
</tr>
<tr>
<td>CI</td>
<td>Cognitive impairment</td>
</tr>
<tr>
<td>CN</td>
<td>Cognitively normal</td>
</tr>
<tr>
<td>CSD</td>
<td>Constrained spherical deconvolution</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CU</td>
<td>Cognitively unimpaired</td>
</tr>
<tr>
<td>DT</td>
<td>Diffusion Tensor</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion Tensor Imaging</td>
</tr>
<tr>
<td>DW-MRI</td>
<td>Diffusion weighted magnetic resonance imaging</td>
</tr>
<tr>
<td>EER</td>
<td>Equal error rate</td>
</tr>
<tr>
<td>EMD</td>
<td>Earth-mover’s distance</td>
</tr>
<tr>
<td>FA</td>
<td>Fractional anisotropy</td>
</tr>
<tr>
<td>FBA</td>
<td>Fixel-based analysis</td>
</tr>
<tr>
<td>FC</td>
<td>Fibre cross-section</td>
</tr>
<tr>
<td>fODF</td>
<td>Fibre orientation distribution function</td>
</tr>
<tr>
<td>FWE</td>
<td>Family-wise error</td>
</tr>
<tr>
<td>GLM</td>
<td>General linear model</td>
</tr>
<tr>
<td>GM</td>
<td>Grey matter</td>
</tr>
<tr>
<td>ICV</td>
<td>Intracranial volume</td>
</tr>
<tr>
<td>ILF</td>
<td>Inferior longitudinal fasciculus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ilr</td>
<td>Isometric log-ratios</td>
</tr>
<tr>
<td>MCI</td>
<td>Mild cognitive impairment</td>
</tr>
<tr>
<td>MD</td>
<td>Mean diffusivity</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini–Mental State Examination</td>
</tr>
<tr>
<td>MoCA</td>
<td>Montreal Cognitive Assessment</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MSMT-CSD</td>
<td>Multi-shell multi-tissue constrained spherical deconvolution</td>
</tr>
<tr>
<td>MTL</td>
<td>Medial temporal lobe</td>
</tr>
<tr>
<td>P-tau</td>
<td>Phosphorylated tau</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PVE</td>
<td>Partial volume effects</td>
</tr>
<tr>
<td>RF</td>
<td>Random forests</td>
</tr>
<tr>
<td>ROC curve</td>
<td>Receiver operating characteristic curve</td>
</tr>
<tr>
<td>SH</td>
<td>Spherical harmonics</td>
</tr>
<tr>
<td>SRB</td>
<td>Standardized regression based</td>
</tr>
<tr>
<td>VBA</td>
<td>Voxel-based analysis</td>
</tr>
<tr>
<td>VBM</td>
<td>Voxel-based morphometry</td>
</tr>
<tr>
<td>WM</td>
<td>White matter</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Alzheimer’s disease dementia

Dementia is a syndrome characterized by the progressive deterioration of cognitive function as a result of a brain disease. This syndrome can affect memory, thinking, judgment and behavior up to the point people are unable to perform daily life activities and require constant assistance for the rest of their life. The biggest known risk factor for a person to develop dementia is ageing. A report combining multiple studies estimated the incidence of dementia doubles with every 6.3 year increase in age [127]. The increased life expectancy and aging of the general population have made of dementia a global public health concern. According to global estimates, in 2016 there were 43.8 million individuals living with dementia worldwide [117] and it is expected that by the year 2030 this number will reach 75 million [127]. In addition to the associated mortality, most of the social and monetary impacts of dementia stem from disability, posing an increasing burden on caregivers and healthcare systems.

The most common cause of dementia is Alzheimer’s disease (AD), a neurodegenerative disorder with no effective disease-modifying treatment currently available [54, 108, 138]. This disease is pathologically defined by the presence of Amyloid-β plaques and neurofibrillary tau deposits [143, 78] associated with neuronal and synaptic loss (Figure 1-1). Although these processes might lead to cognitive impairment and dementia, Alzheimer’s pathology can be present in people who did not show symptoms during their lifetime [40].

Excluding the genetic mutations that cause the early-onset hereditary AD and account for less than 5% of AD cases, the etiology of late-onset AD is complex and poorly understood [54]. Experts believe that Alzheimer’s develops as a result of multiple factors such as genetic,
lifestyle and environment. Besides ageing, other risk factors have been identified, including vascular diseases (e.g. hypertension, obesity), genetic susceptibility, and life-style factors such as diet, physical and mental activity, alcohol consumption, and education level [143]. Traditionally, AD has been recognized in terms of its typical clinical manifestation, that is the multi-domain amnestic dementia. This typical expression of AD is characterized by the progressive deterioration of episodic memory and other cognitive domains such as language, executive function, attention, and visuospatial abilities [172]. Examination of the brain in autopsy-confirmed cases of AD has shown a characteristic pattern for the location of AD degeneration: initially, it appears in the entorhinal cortex progressing through the hippocampus and medial temporal structures (Shown in Figure 1-2), to eventually affect association cortices [15, 114]. This neuropathological pathway correlates with the clinical picture of typical AD which starts as an amnestic syndrome of the hippocampal type accompanied by some impairment in executive functions or naming abilities [38].

Figure 1-2: Characteristic AD lesions have been found in the hippocampus and medial temporal lobe. Atrophy of these brain areas has been recognized as a “topographical” biomarker of AD.

Although the typical AD is the most frequent, some atypical presentations of AD have been recognized. These atypical forms vary from the amnestic syndrome presenting a predomi-
1.1 Alzheimer’s disease dementia

nant executive function impairment [167], aphasia and visuospatial dysfunctions [40]. The existence of such atypical forms of AD has been confirmed with neuropathological examinations of brain tissue [114] and analyses of cortical atrophy patterns [36] finding cases with spared hippocampal atrophy but posterior cortical atrophy.

All the known clinical signs and symptoms of AD are related to disturbances in cognition and behavior. Depending on the presentation and stage of the disease, one or multiple cognitive domains can be affected. Such symptoms are evaluated with the neuropsychological examination consisting of tests to assess the overall level of Cognitive Impairment (CI), specific tests to detect alterations in particular domains like memory, and interviews with the patient and relatives to grade the severity of dementia-like symptoms. The clinical onset of AD is determined by the clinical diagnosis of dementia (or major neurocognitive disorders), requiring the cognitive deficits to interfere with the ability to perform everyday activities [6]. However, cognitive impairment can be detected before it compromises the daily functioning, i.e. in a pre-dementia stage of the disease.

**Mild cognitive impairment**

When individuals show some cognitive decline in one or more cognitive domains, but this decline does not interfere with their activities or behavior, they are diagnosed with mild cognitive impairment (MCI). This broad category includes patients with similar clinical features but a variety of different causes including neurodegenerative disease such as AD, psychiatric conditions like depression, or even side effects of certain medications can be responsible for the perceived cognitive impairment.

Traditionally, MCI has been classified into two sub-types: amnestic and non-amnestic [123], depending on whether there is memory impairment or not. In non-amnestic MCI, cognitive domains like language, visuospatial skills, or executive function show some impairment that drives the distinction from normal ageing. Initially, amnestic MCI was thought to represent a prodromal form of AD [123] but nowadays it is known that multiple pathologies can cause amnestic MCI and not all AD cases show memory dysfunction at pre-clinical stages.

Although the group of subjects with amnestic MCI has a homogeneous clinical phenotype, it is highly heterogeneous in terms of the underlying biology, compromise of other cognitive domains, and specific domain decline trajectories [116] [42] [22]. This heterogeneity makes it difficult to predict clinical progression for MCI patients because they can progress to dementia, remain stable or even revert to cognitively normal. That is also the reason why identifying which individuals with MCI are more likely to develop AD dementia is a research priority [149] and an active field of study [1].

**Clinical diagnosis of AD**

Clinical diagnosis of Alzheimer has been guided by the NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association) criteria published in 1984 [107], it stratifies the confidence of the diagnosis in probable and definite depending on the level of certainty that dementia syndrome is caused by Alzheimer’s pathology:
- **Probable AD**: based on the diagnosis of dementia with progressive impairment of memory and other cognitive functions with no presence of other diseases that can cause cognitive deficits.

- **Definite AD**: based on a diagnosis of probable AD while the patient was alive and evidence of AD pathology from tissue examination post-mortem [75].

According to these criteria, the amnestic syndrome was a core feature for the clinical diagnosis of probable AD, recognizing only the typical manifestation of AD. Update proposals for diagnostic criteria acknowledge the atypical forms of AD by including other clinical phenotypes different from memory impairment [40].

Diagnosis criteria give the general guidelines to diagnose AD in clinical settings but the way the guidelines are implemented is highly variable across medical facilities. The diagnosis of dementia mainly relies on clinical examination and neuropsychological testing. However, the diagnostic guidelines do not specify which tests or population standards should be applied, therefore each medical center chooses the tests and cut-offs used to ascertain whether the patient’s results are normal or abnormal. For example, a systematic analysis of the global burden of dementia found 230 different diagnostic procedures across 237 studies in the dementia literature [117] and the main source of this heterogeneity is the use of different tests and cut-off scores during cognitive screening.

Given the research advances to prove Alzheimer’s pathology with *in vivo* biomarkers, there have been various proposals to include the use of biomarkers for AD diagnosis in dementia [39, 40, 78] and pre-dementia stages of the disease [5, 41]. These recommendations are specially useful in research settings where early detection of Alzheimer’s pathology can be part of the inclusion criteria for clinical trials.

The use of biomarkers for the “early” detection of AD pathology without symptoms of dementia in clinical practice is a matter of current debate [144, 63, 76]. Considering that there are not disease-modifying treatments available and it is not certain if the pathological signs do inevitably lead to dementia syndrome, an early diagnosis of Alzheimer’s could create a psychological burden in patients that may never develop dementia [144].

## Biomarkers for AD

The definite diagnosis of AD is done by the assessment of characteristic structural lesions through the pathological examination of brain tissue. These characteristic lesions are formed by abnormal accumulation of proteins, specifically: extracellular deposit of amyloid-β and neurofibrillary tangles of the protein tau. The two proteins define two groups of *in vivo* pathophysiological markers of Alzheimer’s pathology [38, 78]:

- **Markers of amyloidosis**: low levels of amyloid-β peptide 42 (Aβ42) or Aβ42/Aβ02 ratio in cerebrospinal fluid (CSF), and high cortical binding values for positron emission tomography (PET) with Pittsburgh compound B (PiB).

- **Markers of tauopathy**: high CSF total tau (t-tau) and phosphorylated tau (P-tau), and PET with tau ligands.
These CSF and PET biomarkers have demonstrated to be correlated with the pathological marks of AD: amyloid plaques and neurofibrillary tau deposits, however, they are largely static and give little information about disease stage or progression. Other biomarkers assess subsequent pathological brain changes related with AD progression such as synaptic and neuronal loss. These include Fluorodeoxyglucose (FDG) - PET, which measures glucose uptake and it is sensitive to neuronal dysfunction, and structural Magnetic Resonance Imaging (MRI) to detect atrophy in certain areas such as the medial temporal lobe (MTL) and hippocampus. Although these “topographical” biomarkers are less specific for AD, they do correlate with disease severity, and can improve disease characterization and prediction of cognitive decline in MCI patients. Biomarkers give evidence of Alzheimer’s pathology and AD-related pathological changes in any stage of the disease and MCI subjects with a combination of positive biomarkers are more likely to progress to dementia than those with negative biomarkers. However, none of them alone is sufficient enough to diagnose AD and the consistent finding across studies is that the combination of different biomarkers significantly improves the diagnostic accuracy and prediction of future cognitive decline.

Although the use of biomarkers is widely accepted and implemented in research settings, there are important concerns that prevent their extended use in general clinical practice. First, almost all of them are subject to methodologic variations, in particular, CSF biomarkers are highly variable across laboratories and techniques, which makes it difficult to standardize cut-off points for abnormality. Secondly, validation of the clinical usefulness of biomarkers is still incomplete because most of the studies have been conducted in selected samples that might not be representative of real-world populations. Lastly, CSF and PET biomarkers are invasive and expensive, therefore the access to them is limited in different settings.

Biomarkers have opened up the possibility of detecting Alzheimer’s pathology before any cognitive symptom, i.e. at the preclinical stage. It should be noted that recent studies have shown that most individuals with positive biomarkers for AD are not symptomatic and might remain cognitively healthy during their lifetime. Under biomarker-based criteria these subjects would be diagnosed with the disease without certainty they will actually develop the dementia syndrome. Although this group of individuals, the ones with “preclinical AD” or “at risk of AD”, might be useful in clinical trials to test possible early interventions, a clinical diagnosis based only on biomarker positivity would pose an unnecessary burden on them and their relatives.

Physiopathology: the underlying chain of events

Despite all the scientific efforts and advances during the last decades, there is little clarity about the mechanisms that lead from the disease-defining proteinopathies to neurodegeneration and cognitive impairment. The underlying biological processes in AD are not well understood and they are, probably, one of the most complex systematic malfunctions of the nervous system that are known. A large body of research, including disease-modifying trials, has been based on the amyloid cascade hypothesis proposed almost 30 years ago. According to this hypothesis, the
Introduction

Abnormal aggregation of the amyloid-β peptide is the initial cause of a linear sequence of pathological changes in AD: formation of macroscopic plaques and tau deposits, neurodegeneration, and cognitive impairment. Cross-sectional studies suggest that the entire process from amyloid-β accumulation to the onset of dementia can take up to 20 years [161]. Aligned with the amyloid cascade hypothesis, a temporal model outlining the change of biomarkers was postulated by Jack and colleagues [79]. According to this model, the first biomarker to change is the CSF Aβ42 followed by amyloid PET and CSF tau, then FDG-PET and structural MRI become abnormal indicating neuronal injury and atrophy, and finally, the cognitive symptoms appear. Although this biomarker model recognizes that the two proteinopathies might be initiated independently, it does incorporate the idea that amyloid-β changes can accelerate antecedent tauopathy [79]. The basic structure of this model agrees with the amyloid cascade hypothesis in the assumption that pathological changes occur in a linear sequence initiated by amyloid-β accumulation.

The amyloid cascade hypothesis is supported by the observation that the genetic mutations associated with the hereditary form of AD are also known to over-express amyloid-β [88]. Although this hypothesis might explain the physiopathology in hereditary AD, it is not sufficient to explain the development of sporadic AD, which is the most prevalent form of the disease (> 95% of cases). Growing evidence from animal models, human studies, and failed clinical trials for disease-modifying therapies suggests that the relation between amyloid-β accumulation and Alzheimer’s dementia is not as direct as stated in the amyloid cascade hypothesis [73, 88]. Studies based on post-mortem tissue examination and in vivo biomarker analysis have shown that a considerable amount of cognitively intact elder people have amyloid aggregation in their brains that could be considered pathological [73, 80]. Additionally, the amount of amyloid-β plaques does not correlate with neurodegeneration and cognitive decline [88]. With the available evidence, there is no definite explanation yet for how the amyloid-β deposition could lead to neurodegeneration and cognitive impairment [167].

Nowadays it is recognized that amyloid aggregation is not sufficient to cause AD and that a complex interaction between multiple factors could be a better explanation than a simple linear sequence model [73, 78]. The long list of evidence-based alternatives to the amyloid cascade hypothesis favor other initial causes for AD (e.g., failure of autophagy, mitochondrial function, cell cycle control, Ca²⁺ homeostasis) and consider multiple factors that may be responsible for neuronal damage and cognitive impairment such as inflammation, glucose metabolism and DNA damage [73, 167, 94]. More research is needed to disentangle and understand the biological mechanisms and pathological processes involved in sporadic AD physiopathology [11], closing those gaps in knowledge are important to develop successful disease-modifying treatments [54] and make more accurate predictions about progression at the early stages of the disease.

Neuroanatomical changes

The common characteristic among most of the possible disease models is the final pathway in which neurodegeneration is the pathological feature most proximate to cognitive decline [78]. Recent studies support the observation that neurodegeneration is not the result of
1.1 Alzheimer’s disease dementia

a linear cascade of events but the result of the interaction between multiple mechanisms involving positive and negative feedback loops [167]. Accepted neurodegeneration biomarkers capture different scales of this process: FDG-PET and CSF total tau are indicators of neuronal metabolism and damage [78, 38] while MRI biomarkers give information about macro-structural atrophy of brain tissue. These MRI biomarkers are the ones that become abnormal in the closest temporal proximity to the cognitive impairment, however, macro-structural changes could be detected with structural MRI up to ten years before the onset of clinical symptoms [163].

Early research with T1-weighted MRI established macro-structural landmarks of the disease such as hippocampal and MTL atrophy [77, 142, 35], although these alterations are not specific for AD [156, 48], they are nowadays accepted as topographical biomarkers for disease staging and risk assessment [11, 78]. Longitudinal analyses of grey matter loss [160, 146] have resulted in defined sequential patterns of cortical atrophy starting in temporal and limbic cortices, particularly the entorhinal cortex, progressing with time to frontal and occipital brain regions matching the trajectory of brain lesions observed post-mortem, this pattern of atrophy is observed first in the left hemisphere and occurs faster than in its right counterpart [160, 169]. In the advanced stages of the disease, there is noticeable shrinkage in most neocortical areas accompanied by significant expansion of the ventricles.

Atrophy of the hippocampus and MTL are characteristic of typical AD that manifests predominantly with memory impairment. Studies in relatively large samples of patients have shown there is heterogeneity in cortical and subcortical grey matter atrophy patterns and this heterogeneity is related to atypical manifestations or differences in compromised cognitive domains [167, 171, 36, 43]. Research with structural MRI [175, 133, 158] has confirmed the existence of atrophy patterns that are consistent with the three subtypes that were defined pathologically [114]: typical AD, hippocampal-sparing AD, and limbic-predominant AD. The observed heterogeneity supports the idea that there might be different pathways that lead to neurodegeneration [167], thus restricting the anatomical markers to independent volumetric measures of very specific regions might be an oversimplification that hampers the identification of patient subgroups.

Alzheimer’s has been considered a grey matter disease because the defining brain lesions, intra-neuronal neurofibrillary tangles and extracellular senile plaques, occur mainly in the grey matter. However, there is evidence that pathological changes also occur in the white matter [19], including abnormal levels of Aβ42 observed in post-mortem tissue examination [23], regional atrophy [139], presence of lesions [111], reduction of microstructural integrity [2, 7, 180, 61, 148, 110], and connectivity failures [85, 155]. White matter micro and macrostructure can be examined in vivo using different modalities of MRI. Studies with Diffusion Tensor Imaging (DTI) have revealed diffusion abnormalities in white matter regions such as the splenium of the corpus callosum, superior, middle and inferior longitudinal fasciculi, corticospinal tracts, and limbic system tracts including the fornix, cingulum bundle, and parahippocampal gyrus [2, 37]. Although some works considered the white matter changes a consequence of neuronal degeneration in the grey matter explained by Wallerian degeneration, there is growing evidence suggesting that abnormalities in the white matter might occur independently and could be detected before grey matter changes [23, 85]. These findings support the idea that other pathological mechanisms like neuroinflammation and prion-like
propagation might play an important role in disease physiopathology [19, 85, 94]. Some MRI-based analyses have found that neuroanatomical changes in AD are related to cognitive alterations. For instance, MTL atrophy is correlated with impairment in memory and language [142, 68, 34, 158], and thickness of the parieto-occipital cortex is associated with visuomotor speed [34], visuospatial and executive functioning [158]. Although such correlations exist, and it is known that abrupt damage in certain brain areas can affect specific cognitive functions, it would be inaccurate to attribute certain cognitive skills to a single brain region. Nowadays it is recognized that complex brain functions involve a variety of brain regions functionally connected, and that brain structures are involved in a wide variety of cognitive and functional processes. In this context, the initial neurodegeneration in particular areas might not be enough to cause the characteristic AD decline in specific cognitive functions, and that failures in the functional networks could be the ones responsible for clinical symptoms [85].

Measuring disease outcomes

Clinical stages of the disease are defined by the severity of the symptoms, i.e., the level of cognitive and functional impairment. Therefore, evolution of AD is assessed with neuropsychological tests which evaluate the cognitive abilities and behavior of the patient. Some of the most used scales to determine the severity of the disease are the Clinical Dementia Rating (CDR) [112], and the Alzheimer’s Disease Assessment Scale - Cognition (ADAS-Cog) [134]. The neuropsychological test battery used for diagnosis and monitoring also includes screening tests to measure the overall cognition such as the Mini–Mental State Examination (MMSE) [47] and the Montreal Cognitive Assessment (MoCA) [115]; tests to assess the compromise of specific cognitive domains like the Rey auditory verbal learning test (AVLT), the Logical Memory test [170], the Clock Drawing test [64], the Category Fluency test [113], and the Trail Making test [131]; and tests to evaluate the ability to perform everyday activities such as the Functional Assessment Questionnaire (FAQ) [125]. It is important to point out that the neuropsychological tests are the first clinical tests a patient is subject to when there are suspicions of cognitive decline or self-reported memory concern. Therefore these tests are the entry point to other assessments such as evaluation of neuroimages, biomarkers, risk factors, and longitudinal monitoring of cognition and behavior.

Level of cognitive decline determines the disease severity and marks two of the disease milestones: diagnosis of MCI and dementia onset. Monitoring the small cognitive changes consistent with disease progression is a challenging task due to the existence of a long clinically silent phase in AD [63], combined with a large variety of temporary factors that could also alter cognitive performance in individuals with or without the disease. This issue is particularly relevant in the design of clinical trial for disease-modifying interventions where outcome measures, or end points, are defined to assess the effectiveness of the intervention. Indeed, one of the possible reasons for the long list of failed disease-modifying trials could be the poor performance of outcome measures to detect cognitive changes due to low sensitivity and high measurement variance [108, 138]. Although there are no standard measures for clinical outcomes, and different assessments from the neuropsychological test battery have been used across clinical trials [54], one of the
1.1 Alzheimer’s disease dementia

most widely used measures is the ADAS-Cog \[138\]. This test evaluates multiple cognitive areas and combines the results to give a single number that should indicate the level of overall cognitive impairment. However, some studies have showed that the ADAS-Cog has: low reliability for measuring cognitive change \[67\], high variance due to measurement errors, ceiling effects of its sub-scores, and it is insensitive for patients in mild stages \[108, 90, 138\]. Improved measures of disease outcomes should be:

- Robust and sensitive enough to detect cognitive changes at early stages \[4\].
- Include parts that would be sensitive for heterogeneous forms of AD \[51\].
- Useful to make predictions about patients’ progression and evaluate the risk of developing dementia.

Problem

A lot of unknowns

In sporadic AD, the causes of the disease are largely unknown, the biological mechanisms that lead from proteinopathies to cognitive impairment are unclear \[167, 144\], the complexity and heterogeneity of those mechanisms are not well understood yet \[54\], and the transition between what is considered healthy, or normal, ageing and AD is not well defined \[4\]. There is also no certainty about whether someone with the AD pathological signs will develop some cognitive impairment during their lifetime \[144\], neither about whether a patient with MCI could go back to cognitively normal, will remain as MCI or progress to dementia. Although there are accepted biomarkers for diagnosis and monitoring, not a single one is sufficient to diagnose AD or predict disease progression \[40, 78, 51\]. Moreover, it remains unclear what are thresholds, anatomical distributions, or combinations of abnormal biomarkers that better predict the emergence and evolution of clinical symptoms \[4\].

What is needed

Existing instruments for patient evaluation and monitoring are not sensitive enough to detect subtle changes, predict progression, and recognize heterogeneous forms of AD. Although there have been several advances in this field, more research is needed to develop and validate markers that help to identify disease patterns or profiles that could predict the clinical course of the disease \[149, 4\]. Better markers and strategies to identify and quantify the pathological brain changes occurring with disease progression have the potential to impact the clinical management of patients, the design of clinical trials for disease-modifying treatments, improve the assessment of effectiveness for those interventions, and will open the doors to precision medicine \[158, 63\].
1.2 Computational anatomy in AD

Thanks to their noninvasive nature and increasing availability, MR image modalities have been a tremendous source of information to study brain anatomy abnormalities related with neurodegenerative diseases [10]. As $T_1$-weighted MRI provides good contrast between tissues, it has been widely used to analyse and localize macrostructural changes in AD such as volume loss or shrinkage of the cortex [52].

Regional volumetry

MR-based volumetry of the hippocampus and MTL are nowadays accepted as topographical biomarkers for AD. Measuring the volume of specific structures is completely dependent on their segmentation. In particular for the hippocampus, its manual segmentation is time consuming, requires extensive training and suffers of high inter-observer variability [8]. For this reason some studies have proposed automatic or semi-automatic segmentation methods [24, 21, 98], or indirect measures of hippocampal atrophy [8], as alternatives to manual delineation of this structure [35]. Volume measures of larger sets of regions can also give information to better characterize the disease, several studies have used automated tools to estimate the volume and cortical thickness of anatomical regions of interest, and analyse which regions help better to discriminate between patients and healthy controls [96, 33, 128, 74, 145, 150]. Most of these strategies rely on pre-defined segmentation of an anatomical template that is then registered to each brain image. Although this process could potentially affect the accuracy of the segmentation, results have shown that volumetric and cortical measures obtained with these approaches can be effectively used to distinguish between patients and healthy controls using $T_1$-weighted MRI acquired in realistic clinical settings [33, 145, 176].

Voxel-based morphometry

Voxel-based morphometry (VBM) [9] is one of the most common frameworks to perform statistical inference with brain images. The VBM approach can be divided in three key steps:

1. **Spatial Normalisation**: Registration of all images to a common reference space defined by a template image.

2. **Tissue segmentation**: Spatially normalised images are segmented into grey matter (GM), white matter (WM), and CSF. Tissue segmentation maps are often smoothed with a Gaussian kernel to boost the signal-to-noise ratio and alleviate the effect of registration misalignments during spatial normalisation.

3. **Statistical analysis**: Test statistic values are computed at each image voxel resulting in a “statistical parametric map”, and finally the corresponding $p$-values for the tested hypothesis are calculated. This last step needs to take into account that multiple tests are being performed simultaneously (one test per voxel) and resulting $p$-values need to be corrected accordingly to control for false positives.
These steps of standard VBM are illustrated in Figure 1-3.

Figure 1-3: Steps involved in voxel-based morphometry (VBM), originally described by Ashburner and Friston [9] to compare the local concentrations of grey matter (GM) between groups of subjects. The VBM approach starts with the spatial normalisation of brain images by registering all of them to a template image. Having the warping from subject to template (W), spatially normalised GM map can be obtained by segmenting normalised images or transforming the GM segmentations [91]. Then, GM maps are filtered with a Gaussian kernel and finally these smoothed GM maps are the inputs for voxel-wise statistical inference.

In AD research, VBM has been applied mainly to find GM differences between AD or MCI patients and healthy controls [68, 109, 163]. Some studies have combined the VBM approach with machine learning methods to automatically classify structural images between AD, MCI and controls or predict progression from MCI to AD dementia. In these approaches, the spatially normalised tissue segmentation maps have been used directly as the inputs of a classifier [89, 27], or have served as an intermediate processing step before dimensionality reduction and feature selection for classification [29, 168, 109, 101, 28, 111, 151, 16].
Other features extracted from structural MRI

Multiple image-based markers for AD diagnosis and monitoring have been proposed and tested in the literature employing traditional image processing methods. In 1998, Freeborough and Fox published a study relying on texture analysis of $T_1$-weighted MRI [49] using texture features calculated from the grey level co-occurrence matrix (GLCM) [69] of 2D image slices, this proposal showed promising classification performance when distinguishing between AD patients and controls. Similar texture analysis in the corpus callosum and thalamus also showed significant differences between patients diagnosed with mild AD, amnestic MCI and normal ageing controls [30]. In the same direction, 3D texture markers of the hippocampus have also showed good results classifying between AD, MCI and controls [179, 153].

Shape analysis has been also proposed as a potential image-based marker for AD diagnosis and prognosis. Particularly for sub-cortical structures such as the hippocampus, it was demonstrated that shape features outperform volumetric measures in AD vs controls and MCI vs controls classification tasks [55] and that shape asymmetries are better predictors of dementia onset than size asymmetries [169]. An additional group of proposals includes those inspired by pattern recognition methods such as saliency analysis to find scale-invariant descriptors [162, 137], and pattern matching techniques [126]. All of these also reported competitive classification performance between AD and controls.

Automatic classification of structural MRI

Markers for disease diagnosis and prognosis extracted from structural MRI are often evaluated by using them as inputs for one or more of the following binary classification tasks: cognitively normal (CN) elderly subjects against MCI patients, CN against AD dementia patients, and stable MCI patients against patients who progressed from MCI to AD dementia (sMCI/pMCI). A relatively recent review [130] summarised the reported classification performance of different methods using $T_1$-weighted MRI and other brain imaging modalities. Table 1-1 present the reported accuracy in some of the previously mentioned works.

Most of the time, direct comparison of classification performance between proposed methods is not possible due to their design and methodological differences in terms of sample size, group selection, inclusion criteria, pre-processing pipelines, cross-validation schemes, and reported evaluation metrics [140]. A couple of studies have performed a direct comparison of different methods by using the same dataset and pre-processing to test pre-defined classification or prediction tasks [27, 17, 104]. For instance, Cuinnet et al. [27] evaluated the classification performance of ten approaches using $T_1$-weighted MRI of 509 subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). For CN vs AD classification, the best performance metrics (81% sensitivity and 95% specificity) were achieved by using the voxels of the GM probabilistic segmentation map directly as features for classification [89]. For CN vs MCI, the best classification results (73% sensitivity and 85% specificity) were obtained by a method performing downsampling of the GM probability map and selection of the voxel locations which better discriminate between AD and CN to finally use that information for classification [168]. A more recent challenge [17] compared several methods for
Authors Classification accuracy  

<table>
<thead>
<tr>
<th>Authors</th>
<th>CN/AD</th>
<th>CN/MCI</th>
<th>sMCI/pMCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Klöppel et al., 2008 [89]</td>
<td>81.1 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fan et al., 2008 [46]</td>
<td>94.3 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>McEvoy et al., 2009 [106]</td>
<td>89.0 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Magnin et al., 2009 [101]</td>
<td>94.5 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rueda et al., 2014 [137]</td>
<td>86.1 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Beheshti and Demirel, 2016 [13]</td>
<td>89.7 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Davatzikos et al., 2008 [29]</td>
<td>-</td>
<td>90.0 %</td>
<td>-</td>
</tr>
<tr>
<td>Misra et al., 2009 [109]</td>
<td>-</td>
<td>-</td>
<td>81.5 %</td>
</tr>
<tr>
<td>Desikan et al., 2009 [33]</td>
<td>95.0 %</td>
<td>95.0 %</td>
<td>-</td>
</tr>
<tr>
<td>Gerardin et al., 2009 [55]</td>
<td>94.0 %</td>
<td>83.0 %</td>
<td>-</td>
</tr>
<tr>
<td>Sorensen et al., 2016 [153]</td>
<td>91.2 %</td>
<td>76.4 %</td>
<td>74.2 %</td>
</tr>
</tbody>
</table>

Table 1-1: Reported accuracy for each binary classification of some works in the literature using structural MRI. Adapted from Rathore et al., 2017 [130].

multi-class classification in three diagnostic groups (AD, MCI, and CN) with 354 previously unseen $T_1$-weighted MRI, the best performance (63% accuracy and 78.8% Area under the ROC-curve) was achieved by a method combining five types of features: volume of seven bilaterally joined regions (including the whole brain), cortical thickness of four lobes and the cingulate gyrus, and hippocampal volume, shape and texture scores. [152].

**Examining the tissue microstructural properties**

Macrostructural atrophy caused by AD is accompanied, or preceded, by microstructural changes of tissue integrity. By measuring the diffusion of water molecules in different directions, Diffusion-weighted (DW) MRI provides information about the microstructural barriers of diffusion such as myelin sheaths, axonal and cell membranes. This imaging modality has the unique potential to reveal the organization of tissue at a cellular-scale *in vivo* and non-invasively [164].

Early studies investigating disease-related abnormalities with DW-MRI used the apparent diffusion coefficient ($ADC$), a parameter of the free diffusion model where isotropic Gaussian diffusion is assumed. In this simple model, the diffusion signal $S$ depends on the applied
diffusion weight ($b$-value):

$$S(b) = S(0)e^{-b \times ADC}$$

$$ADC = -\frac{1}{b} \ln \left( \frac{S(b)}{S(0)} \right)$$

(1-1)

The $ADC$ summarizes at voxel level many microscopic processes that affect the diffusion of water molecules [92], hence it captures the alterations resulting from microstructural changes. In the case of AD, one particular study reported that higher $ADC$ in the hippocampus was related to a higher risk of progression from amnestic MCI to AD dementia [86].

The diffusion tensor (DT) model [12] is an extension of the ADC model that incorporates the dependency of the diffusion signal on the directions $u \in S^2$ of the magnetic field gradient applied during image acquisition, being able to describe anisotropic diffusion:

$$S(b, u) = S(0)e^{-bu^T Du}$$

(1-2)

Where the diffusion tensor $D$ is a $3 \times 3$ symmetric positive-definite matrix that has associated three orthogonal eigenvectors and three positive eigenvalues $\lambda_1, \lambda_2$ and $\lambda_3$. This diffusion tensor is often represented by a 3D ellipsoid as in Figure 1-4.

![Figure 1-4](image)

Figure 1-4: A diffusion tensor models the signal in each voxel of the Diffusion-weighted MRI, this is commonly known as Diffusion Tensor Imaging (DTI).

The eigenvalues of the diffusion tensor are used to compute some measures such as the fractional anisotropy ($FA$), mean diffusivity ($MD$), axial diffusivity ($AD$), and radial diffusivity ($RD$):

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

$$AD = \lambda_1$$

$$RD = \frac{\lambda_2 + \lambda_3}{2}$$

(1-3)
These metrics describe diffusion behavior allowing to infer some properties of underlying brain tissue as illustrated in Figure 1-5.

The vast majority of research investigating microstructural differences between AD patients and controls have used the diffusion tensor (DT) model and its derived metrics to describe tissue diffusivity properties mostly in the WM [3, 37, 148, 180, 2, 157, 31, 93, 105], but also in the GM [174, 97]. Consistent findings across tensor-based studies analyzing WM show increased MD and reduced FA in the splenium, the cingulum bundle including the parahippocampal gyrus, and the superior, middle and inferior longitudinal fasciculus; meanwhile increased FA in crossing-fibre areas such as the corticospinal tract has also been reported in AD patients compared with control subjects [148, 2, 37]. Fewer studies have used DT...
metrics to study differences in GM areas focusing in certain regions of interest, the common finding among them is increased MD in the hippocampus and the posterior cingulate cortex [174, 97, 72].

The DT model was the first diffusion model to be widely adopted in clinical and neuroscience research due to its simplicity [32]. However, this simplicity comes with some important limitations. For instance, it assumes the diffusion displacement probability distribution has a Gaussian form (which is not necessarily true given the complexity of diffusion barriers in the brain tissue), and it cannot represent crossing fibre configurations (which are highly prevalent in WM [84]). Another important limitation comes from what the diffusion tensor metrics are actually capturing given the limited spatial resolution of DW-MRI, in a given voxel it is highly probable that the DT model is not only representing the diffusion of one tissue type (WM or GM) but it is also accounting for the partial volume effects (PVE) with surrounding CSF, therefore DT-derived metrics could be capturing macrostructural atrophy effects rather than microstructural properties [119, 72]. Different extensions or alternatives have appeared to overcome DT model limitations, with different acquisition requirements regarding the number of gradient orientations and $b$-values that are needed to estimate model parameters.

For example, diffusion kurtosis imaging (DKI) quantifies the non-gaussianity of diffusion in biologic tissues [83] by adding an excess kurtosis term to the model. Then, the signal attenuation along a certain diffusion direction $u$ is modelled as:

$$
\ln \left( \frac{S(b, u)}{S(0)} \right) = -bD_u + \frac{1}{6}b^2D_u^2K_u
$$

Where $D_u$ and $K_u$ are estimates for the diffusion coefficient and diffusional kurtosis in the direction $u$. Estimation of these parameters requires fitting a quadratic function of the $b$-value, therefore data needs to be acquired with at least two non-zero $b$-values, being one of them relatively high ($\geq 1500$ s/mm$^2$) to allow better appreciation of non-gaussianity. This type of DW-MRI, acquired with multiple non-zero $b$-values, is referred as “multi-shell” given that the acquisition gradients lie in multiple spheres.

The constrained spherical deconvolution (CSD) method to model WM is related to the notion of spheres in the space of acquisition gradients. Given a $b$-value, the diffusion signal can be represented as a function over the unit sphere using spherical harmonics (SH) basis functions (see Figure 1-6).

Then, the diffusion signal observed at a constant $b$-value is modelled as the spherical convolution of a fibre orientation distribution function (fODF) with a single fibre response function [166], as illustrated in Figure 1-7.

The fODF can be “recovered” as the deconvolution of a single fibre response function (that needs to be estimated) from the observed signal, while enforcing non-negativity of the fODF lobes [165]. The fODF is a continuous function that could represent any underlying fibre configuration, effectively overcoming the “crossing-fibre” problem [32], an example of fODF map is shown in Figure 1-8.

These examples of more advanced models, DKI and CSD, have been recently used to investigate AD-related changes by comparing their corresponding diffusion-derived metrics between control subjects and AD patients. For instance, exploratory analysis suggested mean kurtosis
1.2 Computational anatomy in AD

Figure 1-6: Representation of the diffusion signal in spherical harmonics for three different values of diffusion-weight ($b$-value).

Figure 1-7: When DW-MRI is acquired with a constant non-zero $b$-value, the observed signal in a voxel can be modelled as the spherical convolution of a fibre orientation distribution function (ODF) with a single fibre white matter response function.

could be more sensitive than FA or MD to detect initial degeneration of some WM structures such as the splenium of the corpus callosum and the corona radiata [151]. Analysis of fibre-specific measures derived from fODF showed differences of WM micro and macrostructure between AD patients and controls in specific fibre tracts including the cingulum bundle, the splenium and genu of the corpus callosum, the uncinate fasciculus, and arcuate fasciculus.
1.3 This Thesis

Exploration and evaluation of new markers to identify and quantify changes related to AD progression is a highly relevant research path that could, in the mid-term, improve the management and monitoring of patients, and help the evaluation of potential disease-modifying treatments.

It is recognized that alterations in brain anatomy are the pathological features most proximate to cognitive decline [78]; therefore, neuroanatomical markers and neuropsychological information provide direct information about disease progression.

In this context, this thesis presents a set of data-driven strategies that identify and quantify anatomical and cognitive pathological patterns associated with different clinical stages along the AD continuum. These strategies constitute the three main contributions of this thesis:

- In the first contribution, we propose a strategy that captures changes in brain anatomy by comparing the content distribution in different anatomical regions using information from $T_1$-weighted MRI. We demonstrate this quantitative strategy is useful for the automated classification of brain images between patients at different stages and controls. Furthermore, this characterization automatically finds out a multidimensional pattern of AD progression which is directly related to anatomical changes in specific areas. This contribution is presented in Chapter 2 and has been published in a journal article:


- The second contribution presents a comprehensive neuroimaging approach for the study of AD-related abnormalities in brain anatomy combining multiple interrelated measures of tissue integrity derived directly from Diffusion weighted MRI. Differences of WM properties and tissue compositions between MCI, ADD patients and age-matched cognitively normal subjects are investigated, as well as the possible correlations of
diffusion-derived measures with CSF biomarkers. This part of the thesis work is presented in Chapter 3. Part of this work was presented at a conference:


A manuscript has been submitted for publication to a journal.

- The third contribution presents a data-driven method to characterize the cognitive state of MCI patients with a set of domain-specific scores obtained by learning to combine and weight sub-scores from the neuropsychological test battery. Using machine learning methods, we show the developed scores highlight subgroups of MCI patients who exhibit different risks of progression to AD dementia and have better classification performance than standard outcomes when predicting conversion from MCI to dementia up to 5 years after neuropsychological evaluation. This contribution is presented in Chapter 4 and has been published in a journal and a conference:


Finally, Chapter 5 presents some conclusions, discuss the potential impact of the contributions and suggest some possible research directions for future work.
2 Comparing region intensity distributions

2.1 Introduction

A large number of studies have proposed automatic methods to extract features and classify $T_1$-weighted MRI between controls, MCI, and AD patients [27, 17, 130]. In a relatively recent review, Rathore *et al.* [130] established three automatic classification categories based on the feature extraction method from structural MRI: density maps-based, cortical surface-based, and pre-defined region-based. As pointed out by the authors, most investigations in the latter category use only hippocampus features since changes in this region are well known. These studies do not consider differences in other brain regions out of the MTL, ignoring subtle changes and possible complex patterns of the disease compromising multiple regions. On the other hand, density maps-based methods inspired by VBM classify structural MRI using whole-brain information. However, their adoption in clinical practice remains limited or almost not existing. One important reason is that, in many cases, the high-dimensional features are not easily interpretable in terms of spatial patterns of anatomical changes and cannot be related to the clinical picture, so they appear as “black boxes” to clinicians. Despite the promising classification results of machine learning methods, their contribution to the characterization and understanding of the disease progression remains limited. Another drawback of these automatic classification approaches is that, although they do compare brains, their notion of distance has no meaning in terms of the disease progression hampering their use for exploring the pathways of the AD continuum.

In this work, we introduce a strategy that allows the quantification of brain differences by comparing the intensity distributions of several anatomical regions in the whole brain. An underlying hypothesis of this approach is that the differences between AD patients and controls are correlated to tissue constituents, a feature mirrored by the composition of gray level intensities in $T_1$-weighted MRI. The guiding principle incorporated in this proposal is that patients do not follow a single unique direction when transitioning from healthy ageing to AD. Instead, AD patients are assumed to drift away from a healthy state in multiple possible directions, i.e. control subjects form a relatively compact cluster whereas AD cases tend to separate towards pathological states in more than one direction.

2.2 Methods

The basis for the method is the quantitative measurement of differences between subjects in separate regions of the brain. The process can be roughly divided in two stages. First, each
2.2 Methods

2.2.1 Region description

Coarse brain parcellation

The very first step is to parcellate each brain image into a set of anatomical regions. To obtain a coarse parcellation, brain volumes were registered to the MNI152 structural template with an affine transformation calculated using the FSL (FMRIB Software Library) linear registration tool Flirt [81][82], and then the Harvard-Oxford brain atlas (RRID:SCR_001476) was used to partition each registered brain into 96 cortical regions (48 per hemisphere) and 17 subcortical regions. The linear registration approach results in slightly displaced inter-subject anatomic regions, yet this is unlikely to affect the distribution of gray levels within these regions. Such claim may be supported by the fact that overlap between partitioned brains is at least 97%. 

Figure 2-1: Overview of the proposed methodology. First, each one of the considered regions is described using structural MRI from all participants, including cognitively normal (CN) controls and MCI/AD dementia patients. In the second part, features for automated classification consist of the distances to the CN medoid (chosen as the reference point), and ensemble classifiers are trained with all regional features following a random undersampling boosting strategy to account for class imbalance in the sample.

anatomical region is described by comparing its intensity histograms between all subjects in the sample. The second part consists on extracting regional features and performing the classification of subjects between AD/MCI patients and controls using ensemble classifiers (See Figure 2-1).
Region similarity across subjects

Ultimately the objective of our analysis is to be able to quantify the level of similarity or dissimilarity between subjects. Furthermore, it is expected that this distance is related to the diagnostic groups subjects belong to, i.e. two control subjects should have a smaller distance between them than a control and an AD patient. In this work, the tissue distribution of each anatomical region was described by its intensity histogram, and differences of regional tissue distribution between subjects were quantified by measuring the distances between histograms (Figure 2-2).

Figure 2-2: Region characterization starts with extracting equivalent anatomical regions (left). Regional information is represented the median-centered histogram of intensities (center). Histograms of the same region for different subjects are compared with the Earth-Mover’s Distance (right).

All intensity histograms describing region anatomy had 64 bins, however, the bin cuts were defined individually for each case taking into account the intensity range for the whole brain image. To make histograms comparable and eliminate differences due to image intensity range, all histograms were shifted so that the center of mass was aligned to the central bin of the histogram. Then, histograms corresponding to the same region for different subjects were compared with the Earth-Mover’s Distance (EMD) [136].

Earth-mover’s distance formulation

The EMD calculates the minimum cost of transforming one histogram into another by solving a linear optimization problem in which certain units of the $S = \{S_1, \ldots, S_n\}$ histogram, have to be moved to fill the $m$ bins of histogram $C = \{C_1, \ldots, C_m\}$.

The movement of one unit from bin $i \in S$ to bin $j \in C$ has an associated cost $p_{ij}$. The solution consists in a set of movements $\{x_{ij}^n\}_{i,j=1}^{n,m}$ that form $C$ and minimize the total movement cost. The optimization problem can be written in terms of the amount of “earth”, in this case
units $x_{ij}$, that is moved from bin $i \in S$ to bin $j \in C$, as follows:

$$\begin{align*}
\text{minimize} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}x_{ij} \\
\text{subject to} \quad & \sum_{j=1}^{m} x_{ij} \leq S_i, \text{ for } i \in \{1, \ldots, n\} \\
& \sum_{i=1}^{n} x_{ij} \geq C_j, \text{ for } j \in \{1, \ldots, m\} \\
& x_{ij} \geq 0, \text{ for } i \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}
\end{align*}$$

In this case, the cost of moving one unit is set to the absolute distance between bins, i.e., $p_{ij} = |i - j|$. Given the solution $\{x^*_{ij}\}_{i,j=1}^{n,m}$, the EMD between $S$ and $C$ is the normalized total cost:

$$\text{EMD}(S, C) = \frac{1}{\sum x^*_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{m} |i - j|x^*_{ij}$$

When the compared histograms have the same integral, as in this work, the problem is symmetric and the EMD is a metric equivalent to the Wasserstein’s distance. A minimal example of the EMD between two histograms $p$ and $q$ is shown in Figure 2-3.

![Figure 2-3: In this case the EMD between $p$ and $q$ is the cost of moving one unit from one bin to the next divided by the total mass: $1/3$. Note that, in this case, the distance function is symmetric (EMD($p, q$) = EMD($q, p$)).](image)

### 2.2.2 Automated classification of brain images

For each one of the considered anatomical regions, the result of the previous step is a matrix of pairwise distances between subjects in the data sample. Taking distance to a reference point incorporates the guiding principle that patients drift away from a healthy state, therefore we chose the medoid of the control group as such reference point (Figure 2-4). The medoid is the element of a set with the minimal mean distance to the other elements in the set, i.e. for a given set $A$ and a distance function $\delta$ the medoid is defined as:

$$\text{medoid}(A) = \arg \min_{x \in A} \sum_{y \in A} \delta(x, y)$$

![Figure 2-4:](image)
Figure 2-4: Given the pairwise distances between histograms of the same region across all subjects in the sample (left), a reference point was chosen by selecting the “most central” case within the control group (middle), distances to that reference point were taken as the regional features for classification (right).

Distance to the medoid of controls captures how much region anatomy is drifting away from a group representing healthy anatomy. Distance values for all regions characterize the whole-brain anatomy for each subject and constitute the features for classification. Two binary classification tasks were considered: AD patients vs controls and MCI vs controls. For this purpose, ensemble classifiers were trained with an Adaptative Boosting (ADABOOST) approach that iteratively updates the weights of various weak classifiers, giving more importance to samples misclassified in earlier rounds. Simple thresholds of the features were used as weak classifiers. To alleviate the class imbalance in training data, random undersampling of data was used during boosting, an strategy known as RUSBoost [147]. Once ensemble classifiers were trained, each the relative importance of each feature was computed as the weighted sum of mislabeled classes for each predictor. The importance of the features for each classification task says how much a regions helps to differentiate between groups and therefore is an indicator of the degree to which each region is affected by the disease.

2.3 Evaluation

2.3.1 Data

The proposed strategy was evaluated using $T_1$-weighted MRI from a subset of cases in the Open Access Series of Imaging Studies (OASIS-1) database [103]. The sample consisted of 136 cases between 60 and 80 years old, from which 66 were the control group (CN), 50 corresponded to MCI patients, and 20 were patients diagnosed with mild AD. The description of each diagnostic groups in terms of age, gender and cognitive scores is shown in Table 2-1. Structural MRI in OASIS-1 database were acquired with 1.5 T Vision scanners (Siemens, Erlangen, Germany), using magnetization prepared rapid gradient-echo (MP-RAGE) se-
2.3 Evaluation

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Age</th>
<th>Gender (F/M)</th>
<th>CDR</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>66</td>
<td>70.8 ± 5.6</td>
<td>48/18</td>
<td>0</td>
<td>29.1 ± 1.1</td>
</tr>
<tr>
<td>MCI</td>
<td>50</td>
<td>72.8 ± 5.0</td>
<td>28/22</td>
<td>0.5</td>
<td>26.0 ± 3.5</td>
</tr>
<tr>
<td>AD</td>
<td>20</td>
<td>74.3 ± 4.3</td>
<td>13/20</td>
<td>1</td>
<td>20.8 ± 3.7</td>
</tr>
</tbody>
</table>

Table 2-1: Description of diagnostic groups from OASIS including their scores for the Clinical Dementia Rating (CDR) and Mini-mental state examination (MMSE).

quences. Raw $T_1$-weighted MR images have a voxel size of $1 \times 1 \times 1.25 \text{mm}^3$, with a resolution of $256 \times 256 \times 128$. Images were spatially warped into the 1988 atlas space of Talairach and Tournoux with a rigid transformation, averaged motion-corrected, skull-stripped, and finally gain-field corrected [103]. Voxel size after pre-processing is $1 \times 1 \times 1 \text{mm}^3$ with image resolution of $176 \times 208 \times 176$. For more detailed information about the database see [https://www.oasis-brains.org/](https://www.oasis-brains.org/).

Generalization of the presented method was tested with a different set of $T_1$-weighted MRI from the Minimal Interval Resonance Imaging in Alzheimer’s Disease (MIRIAD) database [102]. This sample was composed of 23 healthy controls and 46 subjects diagnosed with probable Alzheimer’s disease. The distribution of age, gender and clinical scores of this dataset is presented in Table 2-2.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Age</th>
<th>Gender (F/M)</th>
<th>CDR</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>23</td>
<td>69.7 ± 7.1</td>
<td>11/12</td>
<td>0</td>
<td>29.4 ± 0.8</td>
</tr>
<tr>
<td>AD</td>
<td>46</td>
<td>69.3 ± 7.2</td>
<td>27/19</td>
<td>1 ± 0.4</td>
<td>19.2 ± 4.0</td>
</tr>
</tbody>
</table>

Table 2-2: Description of diagnostic groups from MIRIAD database.

Images in MIRIAD were acquired with a 1.5 T Signa MRI scanner (GE Medical systems, Milwaukee, WI), using a $T_1$-weighted Inversion Recovery Prepared Fast Spoiled Gradient Recalled (IR-FSPGR) sequence. Other imaging parameters were: matrix size of $256 \times 256$ and 124 $1.5\text{mm}$ coronal partitions. Pre-processing of these images included warping into the Talairach and Tournoux atlas and skull-stripped using FSL tools [82].

2.3.2 Cross-validation

Two different cross-validation schemes were used to test the automated classification between groups: the first evaluation was done only with data from OASIS following a leave-one-out scheme, i.e. iteratively training with the whole set but one and then using the resulting classifier to classify the case set aside. The second scheme aimed to evaluate the generalizability of the proposed characterization by training the classifier with data from OASIS database and testing it with data from MIRIAD database.
Classification performance was assessed via a receiver operating characteristic curve (ROC) calculating its area under the curve (AUC) and equal error rate (EER). The instance of the curve with the best trade-off between false-positive rate and false-negatives negative was selected to report the sensitivity and specificity.

2.4 Results

2.4.1 Classification between patients and controls

The resulting ROC curves for classification experiments with OASIS data are shown in Figure 2-5. When classifying between controls and AD cases, the EER is 0.1 and the AUC is 0.92, as the EER indicates the best trade-off between false positives and false negatives, the sensitivity and specificity of this classification is 0.9. Classification between controls and MCI patients shows an EER of 0.3 and AUC of 0.74, implying a sensitivity and specificity of 0.7.

Figure 2-5: Receiver operating characteristic curves for classification experiments within OASIS database. Classification between controls and AD patients (blue line) gave an AUC of 0.91 and EER of 0.1 (False positive rate = 0.1, True positive rate = 0.9). Classification between controls and MCI (purple line) resulted in an AUC of 0.74 and EER of 0.3 (False positive rate = 0.3, True positive rate = 0.7).

Classification across databases resulted in the ROC curve shown in Figure 2-6 with an AUC of 0.92. According to the decision threshold with the best trade-off between errors, a sensitivity of 85% could be achieved with 91% of specificity. These results show a good overall performance with a high accuracy. The errors consist mostly of False Positives (7 cases) whereas the number of False Negatives remains relatively low (2 cases).
2.4 Results

2.4.2 Region importance

For each classification task, anatomical regions were ranked according to the average of importance across iterations of the leave-one-out validation scheme with OASIS data. The ten most relevant regions to discern between controls, AD and MCI patients are shown in Tables 2-3 and 2-4 respectively.

In the case of CN vs AD classification, the importance to distinguish between groups is concentrated in a few regions (shown in Figure 2-7), the top 10 most relevant features summed more than 66.5% of the importance. It is reasonably expected that only the hippocampi (ranked first and third) account for 24% of the importance. Provided that anatomical changes in MCI are not expected as evident as they might be in mild AD, differences are subtle and more regions need to be taken into account to distinguish between MCI and controls. The top ten most relevant regions to classify between these two groups account for less than 37% of the importance (shown in Figure 2-8). The observation that importance is more spread across regions hints that early structural changes might be more complex and not restricted to the already known anatomical areas.

It is worth mentioning that the feature importance value says how much information it adds to the other features, that is to say a region with little relevancy is in any case informative, but this information might be redundant and shared by other regions. This statement is illustrated by the distributions of the two hippocampi feature values in Figure 2-9: although features for both regions show similar distributions and strong inter-class separation, the right hippocampus is more relevant than the left one, which shows almost half of the importance in Table 2-3. This difference appears during the classifier training phase: the weak
Figure 2-7: The ten most relevant regions for automated classification between controls and AD patients. Names of the regions and their relevance are presented in Table 2-3.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Region</th>
<th>Importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Right Hippocampus</td>
<td>15.51</td>
</tr>
<tr>
<td>2</td>
<td>Right Planum Temporale</td>
<td>13.35</td>
</tr>
<tr>
<td>3</td>
<td>Left Hippocampus</td>
<td>8.40</td>
</tr>
<tr>
<td>4</td>
<td>Left Thalamus</td>
<td>7.16</td>
</tr>
<tr>
<td>5</td>
<td>Right Paracingulate Gyrus</td>
<td>4.83</td>
</tr>
<tr>
<td>6</td>
<td>Right Middle Temporal Gyrus, anterior division</td>
<td>4.38</td>
</tr>
<tr>
<td>7</td>
<td>Left Insular Cortex</td>
<td>4.17</td>
</tr>
<tr>
<td>8</td>
<td>Right Putamen</td>
<td>3.59</td>
</tr>
<tr>
<td>9</td>
<td>Left Frontal Orbital Cortex</td>
<td>2.71</td>
</tr>
<tr>
<td>10</td>
<td>Right Amygdala</td>
<td>2.39</td>
</tr>
</tbody>
</table>

Table 2-3: Top ten most relevant regions for the classification between controls and AD patients.

classifier, based on the left hippocampus, mostly confirms the results of its right counterpart, i.e. since the same cases are discriminated by both left and right weak classifiers, the former is considered redundant because it does not give much additional information and its weight is decreased in the ensemble of classifiers. Because of this, regions showing prevalent differences between groups are ranked higher, whereas those regions useful to classify particular cases are ranked lower. When analyzing the distributions of the feature value for the most relevant regions and
Figure 2-8: The ten most relevant regions for automated classification between controls and MCI patients. Names of the regions and their relevance are presented in Table 2-4.

<table>
<thead>
<tr>
<th>Region Description</th>
<th>Importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Amygdala</td>
<td>6.35</td>
</tr>
<tr>
<td>Right Hippocampus</td>
<td>5.41</td>
</tr>
<tr>
<td>Left Hippocampus</td>
<td>4.78</td>
</tr>
<tr>
<td>Right Planum Temporale</td>
<td>3.56</td>
</tr>
<tr>
<td>Right Heschl’s Gyrus</td>
<td>3.24</td>
</tr>
<tr>
<td>Left Inferior Frontal Gyrus, pars triangularis</td>
<td>3.10</td>
</tr>
<tr>
<td>Right Middle Temporal Gyrus, anterior division</td>
<td>2.95</td>
</tr>
<tr>
<td>Right Amygdala</td>
<td>2.55</td>
</tr>
<tr>
<td>Left Paracingulate Gyrus</td>
<td>2.40</td>
</tr>
<tr>
<td>Left Parahippocampal Gyrus, anterior division</td>
<td>2.18</td>
</tr>
</tbody>
</table>

Table 2-4: Top ten most relevant regions for the classification between control subjects and patients with mild cognitive impairment.

their opposite hemisphere equivalences (shown in Figure 2-9), there are strong observable differences between CN, which form relatively compact groups, and AD patients, which tend to be more scattered and diverge from CN, while the MCI group falls between them. This trend is particularly remarkable in the amygdala, hippocampus, planum temporale and thalamus, where the CN and AD feature value distributions look well separated.
2.5 Discussion

This section presents a fully automated strategy that detects characteristic structural brain patterns associated to the presence of the Alzheimer’s disease. The method derives a regional descriptor that captures the changes in tissue constituency which is characteristic of any neurodegenerative disease. This regional descriptor is based on the comparison of intensity histograms between subjects, assuming gray levels in structural MRI correlate with tissue composition, an assumption that is supported by the fact that image contrast in $T_1$-weighted MRI is the product of relaxation differences between tissue types.

The approach herein described has an advantage over other automated classification methods since it is clinically interpretable by standing out actual patterns of the disease. Machine learning based analyses have helped to move from the classical local approaches in pre-defined regions to the exploration of more complex descriptors using artificial vision techniques. Although such descriptors are useful to separate groups of individuals, most of these features are not useful for finding out anatomo-physiological correlations that enhance the understanding of a particular disease.

The characterization presented here also captures disease progression patterns in multiple directions determined by anatomical changes in different brain regions. This is illustrated...
2.5 Discussion

by Figure 2-10 which shows the median distance (per group) to the CN medoid for a group of brain regions.

Figure 2-10: Group median distance to CN medoid for a group of anatomical regions. The directions of these polar graphics correspond to the 8 most relevant regions in the classification task: 1. amygdala, 2. hippocampus, 3. planum temporale, 4. Heschl’s gyrus, 5. inferior frontal gyrus, 6. thalamus, 7. paracingulate gyrus and 8. middle temporal gyrus (anterior division).

This figure also suggests that equivalent regions in the two hemispheres could not show the same progression rate and then the level of discrimination between subjects is better when the left and right equivalent regions are taken separately, this claim was corroborated with additional classification experiments with the OASIS database in which left and right regions were combined. As shown in Table 2-5, performance measures (AUC and EER) are slightly worse when for both classification tasks.

The presented strategy did effectively discriminate between patients and controls. Two previous works performed an automated classification between CN and AD using exactly the same data and validation scheme but different feature extraction approaches. In the first one, the work by Toews et. al. [162], they propose a technique to learn local scale-invariant anatomical features by evaluating saliency in image scale-spaces and classify cases depending on the occurrence of such features. Following this feature-based morphometry approach, they achieved an EER of 0.2 when classifying between CN and AD. The second work we can directly compare with is the one by Rueda et. al. [137] which presents a strategy that fuses different feature-scale saliency maps and uses this information to feed
Comparing region intensity distributions

<table>
<thead>
<tr>
<th></th>
<th>CO vs. MCI</th>
<th>CO vs. AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.74</td>
<td>0.92</td>
</tr>
<tr>
<td>EER</td>
<td>0.30</td>
<td>0.10</td>
</tr>
<tr>
<td>AUC</td>
<td>0.72</td>
<td>0.90</td>
</tr>
<tr>
<td>EER</td>
<td>0.36</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table 2-5: Comparison of classification performance when information from the same region in both hemispheres is combined

the classifier, classification between CN and AD following this strategy achieves an EER of 0.14. The classification results presented in this chapter constitute an improvement over both works with an EER of 0.1 in the same experiment.

Besides outperforming previous works using the same data and validation scheme, the classification between AD patients and controls achieved 90% sensitivity and specificity while the best performing methods out of 10 compared in [27] reported up to 81% sensitivity and 95% specificity (using a different database). Similar or slightly worse classification results were reported for methods relying on voxel-based morphometry, region volumetry or different feature extraction methods [29, 178, 173, 46, 101, 126, 89, 177]. It should be noted that beyond developing a fully automatic classification pipeline, this strategy finds out a multidimensional expression of AD progression, which is directly related to anatomical changes in specific brain regions. The quantitative measures of anatomical changes proposed here can be used to describe and evaluate brain images in terms of this multidimensional pattern.

2.6 Products

Journal paper


Indirect products: conference papers


• Santiago Silva, Diana L. Giraldó, Eduardo Romero. *Sulci characterization to predict progression from mild cognitive impairment to Alzheimer’s disease.* Proc. 15th International Symposium on Medical Information Processing and Analysis. Medellín - Colombia, 2019. [https://doi.org/10.1117/12.2540437](https://doi.org/10.1117/12.2540437)
3 Investigating tissue-specific abnormalities in AD with DW-MRI

This chapter is published separately due to file size limitations in the Universidad Nacional de Colombia repository.
4 Improving the quantitative characterization of cognitive impairment

4.1 Introduction

Currently, there is no cure for AD and the vast majority of clinical trials for disease-modifying drugs, designed to slow down AD progression from MCI to dementia, have so far failed [108]. Aside the questioned efficacy of the tested treatments, other possible reasons for failures may come up from two sources. First, heterogeneity of recruited participants, including advanced AD and variable MCI manifestations, or participants without any underlying pathology [108]. Second, standard cognitive outcomes, set as endpoints, might be highly variable and not sensitive enough to detect subtle cognitive performance changes [108, 138]. This is the case of the widely used Alzheimer’s disease Assessment Scale - Cognitive (ADAS-Cog), that has shown high variability and poor sensitivity, likely by measurement errors, patient heterogeneity, and ceiling effects of its sub-scores, making some sub-scores uninformative in patients at early stages [138, 129, 67].

Composite outcomes computed with informative sub-scores from one or multiple tests have demonstrated to be more robust and sensible measures to detect cognitive and functional changes in MCI [138, 129]. However, single composite scores may mask the heterogeneity of cognitive impairment.

Patients diagnosed with MCI show varying levels of impairment in different cognitive domains beyond memory, including language, visuospatial skills, attention and executive function [100, 172, 53]. This heterogeneity is likely linked to differences in the clinical evolution [154, 65]. Therefore, evaluation of domain specific changes could help to identify individuals at greater risk of progressing to dementia. Composite scores for measuring specific domain impairment have been proposed for memory [25] and executive function [56]. These scores mitigate the effect of measurement errors for individual items while combining informative sub-scores from multiple tests. Evaluation of these two previously proposed scores demonstrated they show better performance than individual test scores in detecting domain changes over time and predicting conversion from MCI to dementia [25, 56].

This chapter presents a data-driven framework which learns to combine and weight sub-scores from the neuropsychological test battery to calculate a set of domain-specific composite scores that quantify impairment in 6 domains: memory, language, visuospatial abilities, executive functioning, orientation and attention. The weighting scheme was obtained by estimating the parameters of a multi-factor model with Confirmatory Factor Analysis (CFA).
The usefulness of the developed composite scores in MCI was evaluated in two different tasks using machine learning methods. First, the set of composite scores was taken as input for unsupervised cluster analysis, aiming to identify different sub-groups of individuals in the MCI sample. Second, we tested the ability of composite scores to predict progression from MCI to dementia within specific time windows, ranging from 1 to 5 years, and compared the performance against standard outcomes.

### 4.2 Methods

The data-driven methodology presented here is divided in two parts (Figure 4-1). The first part consists in learning the parameters for sub-score standardization and domain scores calculation. The second part evaluates the composite scores in two automated tasks: clustering of patients diagnosed with MCI, and predicting progression to dementia.

![Figure 4-1: The proposed data-driven methodology can be divided in two blocks: learning and evaluation. During the learning phase, the parameters for sub-score standardization and domain composite calculation are estimated using a data sample including cognitively unimpaired participants and MCI patients. In the second part, the calculated domain scores for a separate sample of MCI are evaluated in terms of two tasks: unsupervised clustering of patients and prediction of future progression to dementia within different time windows.](image-url)
4.2 Methods

4.2.1 Participants data

Data was provided by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. The ADNI is a public-private partnership with the primary goal of testing whether magnetic resonance imaging (MRI), positron emission tomography (PET), biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For additional and up-to-date information, see [www.adni-info.org](http://www.adni-info.org). The dataset herein used comprised 680 patients with MCI and 668 cognitively unimpaired (CU) participants. The demographics and characteristics of these groups are presented in Table 4-1, corresponding to the first visit with the available information.

Data partition

The ADNI sample was split following the two methodological parts: learning and evaluation. For the learning set, 60% of the CU sample (n = 400) was taken as normative data for sub-score standardization while the remaining 40% (n = 268) and 40% of the MCI sample (n = 272) were used to learn the parameters for calculating the composite scores with CFA. The evaluation set corresponded to the remaining 60% of MCI participants (n = 408), for which composite scores were calculated using the parameters from the learning set.

<table>
<thead>
<tr>
<th></th>
<th>Learning set</th>
<th>Evaluation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CU (n = 668)</td>
<td>MCI (n = 680)</td>
</tr>
<tr>
<td></td>
<td>Normative data</td>
<td>CFA (n = 268)</td>
</tr>
<tr>
<td><strong>Sex (% female)</strong></td>
<td>54.5</td>
<td>59.3</td>
</tr>
<tr>
<td><strong>Age (mean ± sd)</strong></td>
<td>73.4 ± 6.9</td>
<td>72.6 ± 8.0</td>
</tr>
<tr>
<td><strong>APOE-ε4 (% carriers)</strong></td>
<td>31.0</td>
<td>28.9</td>
</tr>
<tr>
<td><strong>CDR-SOB (mean ± sd)</strong></td>
<td>0.1 ± 0.2</td>
<td>0.1 ± 0.2</td>
</tr>
<tr>
<td><strong>MMSE (mean ± sd)</strong></td>
<td>29.2 ± 1.1</td>
<td>28.9 ± 1.2</td>
</tr>
<tr>
<td><strong>ADAS-Cog (mean ± sd)</strong></td>
<td>10.0 ± 4.7</td>
<td>11.1 ± 4.5</td>
</tr>
</tbody>
</table>

Table 4-1: Description of sets used in each step of the methodology, including the percentage of carriers of the ε4 allele of the apolipoprotein E (APOE) gene, and distributions of total scores for the Mini-Mental State Examination (MMSE), Clinical Dementia Rating - Sum of Boxes (CDR-SOB), and the Alzheimer’s Disease Assessment Scale - Cognition (ADAS-Cog). *sd: standard deviation.
Neuropsychological data

Sub-scores from nine different tests were used in the present study, namely: the Alzheimer’s Disease Assessment Scale - Cognition (ADAS-Cog) [134], Mini-Mental State Examination (MMSE) [47], Montreal Cognitive Assessment (MoCA) [115], Rey auditory verbal learning test (AVLT), Logical Memory test immediate and delayed [170], Clock Drawing test [64], Category Fluency test [113], Trail Making A and B [131], and one of the naming tests depending on its availability: Boston Naming test [87] or Multilingual Naming test [62]. The initial list of 50 sub-scores is presented in Table 4-2.

<table>
<thead>
<tr>
<th>Sub-score code</th>
<th>Test Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1SCORE</td>
<td>ADAS-Cog</td>
<td>Word Recall</td>
</tr>
<tr>
<td>Q2SCORE</td>
<td>ADAS-Cog</td>
<td>Commands</td>
</tr>
<tr>
<td>Q3SCORE</td>
<td>ADAS-Cog</td>
<td>Constructional Praxis</td>
</tr>
<tr>
<td>Q4SCORE</td>
<td>ADAS-Cog</td>
<td>Delayed Word Recall</td>
</tr>
<tr>
<td>Q5SCORE</td>
<td>ADAS-Cog</td>
<td>Naming</td>
</tr>
<tr>
<td>Q6SCORE</td>
<td>ADAS-Cog</td>
<td>Ideational Praxis</td>
</tr>
<tr>
<td>Q7SCORE</td>
<td>ADAS-Cog</td>
<td>Orientation</td>
</tr>
<tr>
<td>Q8SCORE</td>
<td>ADAS-Cog</td>
<td>Word Recognition</td>
</tr>
<tr>
<td>Q9SCORE</td>
<td>ADAS-Cog</td>
<td>Remembering Test Instructions</td>
</tr>
<tr>
<td>Q10SCORE</td>
<td>ADAS-Cog</td>
<td>Comprehension</td>
</tr>
<tr>
<td>Q11SCORE</td>
<td>ADAS-Cog</td>
<td>Word-finding Difficulty</td>
</tr>
<tr>
<td>Q12SCORE</td>
<td>ADAS-Cog</td>
<td>Language</td>
</tr>
<tr>
<td>Q13SCORE</td>
<td>ADAS-Cog</td>
<td>Number cancellation</td>
</tr>
<tr>
<td>MMORITIME</td>
<td>MMSE</td>
<td>Orientation to time</td>
</tr>
<tr>
<td>MMORISPACE</td>
<td>MMSE</td>
<td>Orientation to space</td>
</tr>
<tr>
<td>MMREGI</td>
<td>MMSE</td>
<td>Three word registration</td>
</tr>
<tr>
<td>MMRECALL</td>
<td>MMSE</td>
<td>Three word recall</td>
</tr>
<tr>
<td>MMSPELLBKW</td>
<td>MMSE</td>
<td>Spelling a 5 letters word backwards</td>
</tr>
<tr>
<td>MMNAM</td>
<td>MMSE</td>
<td>Naming 2 objects</td>
</tr>
<tr>
<td>MMCOMMAND</td>
<td>MMSE</td>
<td>Following a verbal command</td>
</tr>
<tr>
<td>MMREPEAT</td>
<td>MMSE</td>
<td>Repeating a short sentence</td>
</tr>
<tr>
<td>MMREAD</td>
<td>MMSE</td>
<td>Reading a sentence wit an instruction</td>
</tr>
<tr>
<td>MMWRITE</td>
<td>MMSE</td>
<td>Writing a sentence about anything</td>
</tr>
<tr>
<td>Test Code</td>
<td>Test Description</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>MMDRAW</td>
<td>MMSE</td>
<td></td>
</tr>
<tr>
<td>TRAILS</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>CUBE</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCACLOCK</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCANAM</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCADIG</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCALET</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCASERIAL</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCAREP</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCAFLUEN</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCAABS</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCADLREC</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>MOCAORI</td>
<td>MoCA</td>
<td></td>
</tr>
<tr>
<td>CLOCKSCOR</td>
<td>Clock Drawing test</td>
<td></td>
</tr>
<tr>
<td>COPYSCOR</td>
<td>Clock Drawing test</td>
<td></td>
</tr>
<tr>
<td>TRAASCOR</td>
<td>Trail Making test</td>
<td></td>
</tr>
<tr>
<td>TRABSCOR</td>
<td>Trail Making test</td>
<td></td>
</tr>
<tr>
<td>LIMMTOTAL</td>
<td>Logical Memory test</td>
<td></td>
</tr>
<tr>
<td>LDELTOTAL</td>
<td>Logical Memory test</td>
<td></td>
</tr>
<tr>
<td>CATANIMSC</td>
<td>Category Fluency test</td>
<td></td>
</tr>
<tr>
<td>RAVLT.IMMED</td>
<td>Rey AVLT</td>
<td></td>
</tr>
<tr>
<td>AVTOT6</td>
<td>Rey AVLT</td>
<td></td>
</tr>
<tr>
<td>AVTOTB</td>
<td>Rey AVLT</td>
<td></td>
</tr>
<tr>
<td>AVDEL30MIN</td>
<td>Rey AVLT</td>
<td></td>
</tr>
<tr>
<td>AVDELTOT</td>
<td>Rey AVLT</td>
<td></td>
</tr>
</tbody>
</table>

- **MMDRAW**: MMSE (Copying a drawing)
- **TRAILS**: MoCA (Trails)
- **CUBE**: MoCA (Copying a cube drawing)
- **MOCACLOCK**: MoCA (Drawing a clock)
- **MOCANAM**: MoCA (Naming 3 animals)
- **MOCADIG**: MoCA (Repeating digits forward and backwards)
- **MOCALET**: MoCA (Tapping with the hand when a letter is read from a list)
- **MOCASERIAL**: MoCA (Serial subtraction starting at 100)
- **MOCAREP**: MoCA (Repeating 2 sentences)
- **MOCAFLUEN**: MoCA (Naming words that begin with the letter F)
- **MOCAABS**: MoCA (Abstraction of similarities between words)
- **MOCADLREC**: MoCA (Five word recall)
- **MOCAORI**: MoCA (Orientation to time and space)
- **CLOCKSCOR**: Clock Drawing test (Drawing a clock with details)
- **COPYSCOR**: Clock Drawing test (Copying the drawing of a clock)
- **TRAASCOR**: Trail Making test (Time to complete Part A)
- **TRABSCOR**: Trail Making test (Time to complete Part B)
- **LIMMTOTAL**: Logical Memory test (Immediate recall of a story read by the examiner)
- **LDELTOTAL**: Logical Memory test (Delayed recall of a story read by the examiner)
- **CATANIMSC**: Category Fluency test (Naming animals)
- **RAVLT.IMMED**: Rey AVLT (Repeating a list of 15 words 5 times)
- **AVTOT6**: Rey AVLT (Recall of the first list of words after a second list was read)
- **AVTOTB**: Rey AVLT (Repeating words from the second list)
- **AVDEL30MIN**: Rey AVLT (Recall of words from the first list after 30 minutes)
- **AVDELTOT**: Rey AVLT (Delayed recognition of written words from the first list)
4.2.2 Learning domain composite scores

Sub-score standardization

Given the heterogeneous scales of neuropsychological tests, some of the scales were inverted to ensure that increasing values correspond to poorer performance. The initial set of 50 sub-scores were transformed into standardized regression based (SRB) z-scores using the parameters learnt from a normative sample. Specifically, each sub-score \( x \) was modelled as a linear function of age and years of education:

\[
x = \beta_0 + \beta_{ed} \times \text{Education} + \beta_{age} \times \text{Age} + \varepsilon
\]  

(4-1)

Linear regression parameters \( \beta_0, \beta_{ed}, \beta_{age} \) and \( \sigma^2 = \text{var}(\varepsilon) \) were estimated with data from 400 CU participants. Then, the corresponding SRB z-score for each participant \( i \) (denoted \( y_i \) for consistency with upcoming formulations) was calculated as:

\[
y_i = \frac{x_i - \hat{x}_i}{\sigma} = \frac{x_i - (\beta_0 + \beta_{ed} \times \text{Education}_i + \beta_{age} \times \text{Age}_i)}{\sigma}
\]  

(4-2)

The sub-score from the naming test after a semantic clue (BMCUED) was dropped from further analysis because higher values, after scale inversion, can be associated with poor performance or perfect performance without the cue.

Derivation of domain scores

The estimation of composite measures for 6 different domains was done by proposing and testing a factor model which links a set of sub-scores from multiple tests with six domains: memory, language, visuospatial abilities, executive functioning, orientation and attention (See Figure 4-2). Before establishing a factor model, variability of sub-scores and pairwise correlations were examined in the data partition used for CFA. Sub-scores whose variance was inflated by a few outliers were not included in the model, neither were the sub-scores showing no significant correlation (greater than 0.25) with any other one and were not evaluating a similar task. The factor model was proposed taking into account what sub-scores evaluate, but also the number of previous works that performed Factor Analysis on similar neuropsychological test batteries [25, 56, 118, 59].
4.2 Methods

Figure 4-2: Proposed factor model connecting 6 cognitive domains with 35 sub-scores from 9 different neuropsychological tests. Sub-scores code and description is presented in Table 4-2.

**Factor analysis formulation**

Factor analysis methods exploit the correlations between observed measures \( \mathbf{y} \) to quantify the influence of unobserved factors \( \mathbf{z} \) [26]. Let the vector \( \mathbf{y}_i \in \mathbb{R}^{m \times 1} \) be the \( i \)-th observation of \( m \) variables, and \( \mathbf{z}_i \in \mathbb{R}^{p \times 1} \) the unobserved measures of \( p \) factors. The *common factor model* states:

\[
\mathbf{y}_i = \mathbf{Fz}_i + \mathbf{e}_i
\]  

(4-3)

Or component-wise:

\[
y_{ij} = \sum_{k=1}^{m} f_{jk} z_{ik} + e_{ij}
\]  

(4-4)

Matrix \( \mathbf{F} \in \mathbb{R}^{m \times p} \) contains the *factor loadings*, also known as the factor structure. Residuals \( e_{ij} \in \mathbb{R}^{1 \times m} \) contain the portion of the \( j \)-th variable that is not defined by the factors and matrix of residuals correlations \( \mathbf{C}_e \in \mathbb{R}^{m \times m} \) is assumed diagonal. Therefore, components of the correlation matrix between observed variables \( \mathbf{C}_y \in \mathbb{R}^{m \times m} \) are given by:

\[
[C_y]_{jl} = \text{corr}(y_j, y_l) = \sum_{s=1}^{m} \sum_{t=1}^{m} f_{js} f_{lt} [C_z]_{st} \quad \text{for } j \neq l
\]  

(4-5)
Improving the quantitative characterization of cognitive impairment

In Exploratory Factor Analysis (EFA) the matrices $F$, $C_z$, and $C_e$ are estimated without any assumptions about the underlying factor structure. On the contrary, Confirmatory Factor Analysis (CFA) estimates those matrices for an hypothesized factor model from Equation 4.5. CFA was performed with the lavaan package \[135\] in R (v.3.6.3) using the unweighted least squares estimator. Model fit was evaluated by the Root Mean Square Error of the Approximation (RMSEA) and the Tucker-Lewis index (TLI), these measures evaluate a model in relation to a baseline model which assumes all variables to be independent.

**Domain scores calculation**

The factor structure matrix $F \in \mathbb{R}^{m \times p}$ quantifies the influence of the $p$ unobserved factors over the $m$ observed variables. The set of unobserved factors $z_i$ for a particular case $i$ can be calculated as linear combinations of its observations $y_i \[11, 66\]$. That is, the estimated vector of factors, $\hat{z}_i$, is given by:

$$\hat{z}_i = Wy_i$$ (4-6)

With $W \in \mathbb{R}^{p \times m}$ a weight matrix that needs to be estimated. A solution that minimizes the sum of squares of the uniqueness \[11\], *i.e.*, the portion of the observations variance that is not explained by the factors, is given by:

$$W = (F^T C_e^{-1} F)^{-1} F^T C_e^{-1}$$ (4-7)

The resulting estimated factor values quantify dysfunction of the different domains included in the model. The learnt set of weights can be used to calculate the domain specific scores of new observations once they have been transformed into SRB $z$-scores.

**4.2.3 Evaluation**

**Clustering the MCI sample**

By exploring the existence of MCI subgroups with an unsupervised clustering method, the six composite scores expose different cognitive profiles in the MCI sample. Specifically, the Partition Around Medoids (PAM) method, also known as k-medoids, iteratively splits the data set in k clusters, being the k representative points the most central points (medoid) in each cluster, and the remaining points assigned to the cluster with the nearest representative point \[95, 132\].

Here we incorporated the inherent relations between domains by including the covariance matrix $C_z$ in the calculation of distance between subjects, the distance between a pair of subjects $i$ and $j$, described by their domain scores $z_i$ and $z_j$, was defined as:

$$d(z_i, z_j)^2 = (z_i - z_j)C_z(z_i - z_j)^T$$ (4-8)
4.2 Methods

Unlike the Mahalanobis distance, the distance between two subjects is weighted by the covariance between factors, thereby ensuring that the largest variance dimensions contribute more to the differences between subjects. The matrix used for this step was the estimated covariance matrix between domains $C_*$ resulting from the CFA. The number of clusters was set by revising a collection of 30 indices [20] for multiple options of $k$ from 2 to 10. Cluster stability for these possible partitions ($2 \leq k \leq 10$) was also evaluated following a bootstrap approach. For a given partition in $k$ subgroups, this process consists in partitioning a sub-sample of the data (80%), calculating the subset of observations that remains in their initial cluster and repeating this process multiple times (1000 iterations). The overlap between the initial clusters and bootstrap clusters was assessed via the Jaccard coefficient and the mean value of this index over the total of repetitions is reported for all clusters.

**Differences between MCI subgroups**

Resulting subgroups of MCI participants were compared in terms of their composite scores per domain and their risk of progression to dementia. Pairwise domain score differences between sub-groups were examined with Wilcoxon-Mann-Whitney U tests while applying the Bonferroni correction for multiple comparisons. A multivariate Cox proportional hazard regression model tested the sub-group effect in the progression from MCI to AD dementia while controlling for age, gender and years of education. Kaplan-Meier survival curves illustrated progression to dementia of the different MCI sub-groups, and curves were compared using omnibus and pairwise log-rank tests. A multivariate Cox proportional hazard regression model was used to evaluate the effect of the cognitive profile on the progression from MCI to AD dementia while controlling for age, gender and years of education. The resulting hazard ratios (HR) account for the risk difference of each MCI sub-group with respect to a reference group.

**Prediction of progression to dementia**

Domain specific scores were also evaluated in the automated prediction of progression from MCI to AD dementia. This evaluation consisted in classifying MCI patients as either stable or converters following the time window approach[121] fixing five different time periods: 12, 24, 36, 48, and 60 months. The 6 composite scores along with age, gender and years of education were used to train random forest classifiers [16]. A Random forest (RF) is an ensemble of decision trees constructed using a bootstrap aggregating approach. To create each decision tree, a new training set is generated by sampling, uniformly and with replacement, the original training set. This procedure ensures the collection of trees comes from independent identically distributed samples. The prediction is given by the majority voting of the decision trees in the ensemble, effectively improving the prediction accuracy [16]. Classification performance was assessed by constructing the Receiver Operating Characteristic (ROC) curve and calculating its Area Under the Curve (AUC). Depending on the time window, data for training the classifier might be highly unbalanced. This was taken into account when designing the cross validation scheme: at each iteration, a random forest classifier was trained with a balanced subset by randomly selecting the 70% of the underrepresented
class with an equal number of samples from the other class. The classifier was tested with the remaining observations, in some cases reaching a larger number of samples. This process was repeated 1000 times per time window.

4.3 Results

4.3.1 Parameters for composite scores calculation

Sub-score standardization

Parameters to calculate the SRB z-scores (Equation 4-2), obtained from the linear regression with a sample of 400 cognitively unimpaired individuals are presented in Table 4-3.

<table>
<thead>
<tr>
<th>Sub-score</th>
<th>Intercept ($\beta_0$)</th>
<th>Education ($\beta_{ed}$)</th>
<th>Age ($\beta_{age}$)</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1SCORE</td>
<td>1.044</td>
<td>-0.045</td>
<td>0.034</td>
<td>1.294</td>
</tr>
<tr>
<td>Q2SCORE</td>
<td>0.072</td>
<td>-0.001</td>
<td>0</td>
<td>0.353</td>
</tr>
<tr>
<td>Q3SCORE</td>
<td>0.178</td>
<td>-0.015</td>
<td>0.006</td>
<td>0.55</td>
</tr>
<tr>
<td>Q4SCORE</td>
<td>-0.549</td>
<td>-0.065</td>
<td>0.06</td>
<td>1.811</td>
</tr>
<tr>
<td>Q5SCORE</td>
<td>-0.094</td>
<td>-0.001</td>
<td>0.002</td>
<td>0.265</td>
</tr>
<tr>
<td>Q6SCORE</td>
<td>0.145</td>
<td>-0.001</td>
<td>-0.001</td>
<td>0.231</td>
</tr>
<tr>
<td>Q7SCORE</td>
<td>0.258</td>
<td>-0.009</td>
<td>0</td>
<td>0.321</td>
</tr>
<tr>
<td>Q8SCORE</td>
<td>5.095</td>
<td>0.035</td>
<td>-0.037</td>
<td>2.375</td>
</tr>
<tr>
<td>Q9SCORE</td>
<td>-0.042</td>
<td>-0.006</td>
<td>0.002</td>
<td>0.148</td>
</tr>
<tr>
<td>Q10SCORE</td>
<td>-0.203</td>
<td>0.001</td>
<td>0.003</td>
<td>0.148</td>
</tr>
<tr>
<td>Q11SCORE</td>
<td>-0.266</td>
<td>0</td>
<td>0.004</td>
<td>0.257</td>
</tr>
<tr>
<td>Q12SCORE</td>
<td>-0.138</td>
<td>0.001</td>
<td>0.002</td>
<td>0.171</td>
</tr>
<tr>
<td>Q13SCORE</td>
<td>-0.405</td>
<td>-0.001</td>
<td>0.015</td>
<td>0.819</td>
</tr>
<tr>
<td>MMORITIME</td>
<td>-0.024</td>
<td>0.001</td>
<td>0.001</td>
<td>0.301</td>
</tr>
<tr>
<td>MMORISPACE</td>
<td>0.432</td>
<td>-0.009</td>
<td>-0.002</td>
<td>0.351</td>
</tr>
<tr>
<td>MMREGI</td>
<td>-0.02</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.1</td>
</tr>
<tr>
<td>MMRECALL</td>
<td>-0.556</td>
<td>-0.016</td>
<td>0.015</td>
<td>0.621</td>
</tr>
<tr>
<td>MMSPELLBKW</td>
<td>0.529</td>
<td>-0.032</td>
<td>0.002</td>
<td>0.492</td>
</tr>
<tr>
<td>MNAM</td>
<td>-0.05</td>
<td>0.001</td>
<td>0.001</td>
<td>0.05</td>
</tr>
<tr>
<td>MMCOMMAND</td>
<td>0.103</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.244</td>
</tr>
<tr>
<td>MMREPEAT</td>
<td>0.252</td>
<td>-0.005</td>
<td>-0.001</td>
<td>0.264</td>
</tr>
</tbody>
</table>
### 4.3 Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMREAD</td>
<td>0.028</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>MMWRITE</td>
<td>-0.02</td>
<td>0.001</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>MMDRAW</td>
<td>0.183</td>
<td>-0.01</td>
<td>0.001</td>
<td>0.241</td>
</tr>
<tr>
<td>TRAILS</td>
<td>-0.078</td>
<td>-0.008</td>
<td>0.004</td>
<td>0.27</td>
</tr>
<tr>
<td>CUBE</td>
<td>0.451</td>
<td>-0.03</td>
<td>0.004</td>
<td>0.447</td>
</tr>
<tr>
<td>MOCACLOCK</td>
<td>-0.058</td>
<td>-0.003</td>
<td>0.005</td>
<td>0.483</td>
</tr>
<tr>
<td>MOCANAM</td>
<td>-0.102</td>
<td>0</td>
<td>0.003</td>
<td>0.276</td>
</tr>
<tr>
<td>MOCADIG</td>
<td>-0.001</td>
<td>-0.008</td>
<td>0.003</td>
<td>0.26</td>
</tr>
<tr>
<td>MOCALET</td>
<td>-0.113</td>
<td>-0.002</td>
<td>0.003</td>
<td>0.177</td>
</tr>
<tr>
<td>MOCASERIAL</td>
<td>0.153</td>
<td>-0.016</td>
<td>0.003</td>
<td>0.385</td>
</tr>
<tr>
<td>MOCAREP</td>
<td>0.14</td>
<td>-0.018</td>
<td>0.005</td>
<td>0.454</td>
</tr>
<tr>
<td>MOCAFLUEN</td>
<td>0.198</td>
<td>-0.016</td>
<td>0.003</td>
<td>0.385</td>
</tr>
<tr>
<td>MOCAABS</td>
<td>0.457</td>
<td>-0.029</td>
<td>0.002</td>
<td>0.393</td>
</tr>
<tr>
<td>MOCADLREC</td>
<td>-2.029</td>
<td>-0.049</td>
<td>0.074</td>
<td>1.691</td>
</tr>
<tr>
<td>MOCAORI</td>
<td>0.271</td>
<td>-0.005</td>
<td>-0.002</td>
<td>0.254</td>
</tr>
<tr>
<td>CLOCKSCOR</td>
<td>0.092</td>
<td>-0.012</td>
<td>0.005</td>
<td>0.536</td>
</tr>
<tr>
<td>COPYSCOR</td>
<td>0.167</td>
<td>0.003</td>
<td>-0.001</td>
<td>0.42</td>
</tr>
<tr>
<td>TRAASCOR</td>
<td>2.772</td>
<td>-0.099</td>
<td>0.42</td>
<td>9.601</td>
</tr>
<tr>
<td>TRABSCOR</td>
<td>9.705</td>
<td>-1.968</td>
<td>1.386</td>
<td>39.867</td>
</tr>
<tr>
<td>LIMMTOTAL</td>
<td>15.084</td>
<td>-0.242</td>
<td>-0.012</td>
<td>3.086</td>
</tr>
<tr>
<td>LDELTOTAL</td>
<td>15.238</td>
<td>-0.234</td>
<td>-0.002</td>
<td>3.27</td>
</tr>
<tr>
<td>CATANIMSC</td>
<td>36.053</td>
<td>-0.62</td>
<td>0.174</td>
<td>5.35</td>
</tr>
<tr>
<td>RAVLT.IMMED</td>
<td>2.848</td>
<td>-0.473</td>
<td>0.456</td>
<td>9.81</td>
</tr>
<tr>
<td>AVTOT6</td>
<td>-0.858</td>
<td>-0.15</td>
<td>0.123</td>
<td>3.248</td>
</tr>
<tr>
<td>AVTOTB</td>
<td>4.328</td>
<td>-0.072</td>
<td>0.089</td>
<td>1.988</td>
</tr>
<tr>
<td>AVDELT3MIN</td>
<td>-1.107</td>
<td>-0.075</td>
<td>0.129</td>
<td>3.9</td>
</tr>
<tr>
<td>AVDELTOT</td>
<td>-1.478</td>
<td>-0.042</td>
<td>0.058</td>
<td>2.267</td>
</tr>
<tr>
<td>BMNOCUE</td>
<td>2.035</td>
<td>-0.134</td>
<td>0.026</td>
<td>1.995</td>
</tr>
<tr>
<td>BMCUED</td>
<td>29.761</td>
<td>0.032</td>
<td>-0.007</td>
<td>0.592</td>
</tr>
</tbody>
</table>

Table 4-3: Parameters for sub-score standardization estimated from normative data. Complete description of each sub-score code is presented in Table 4-2.
Factor analysis

The proposed factor model was constructed with 35 sub-scores linked to 6 cognitive domains: memory, language, executive function, visuo-spatial, orientation and attention. Fit statistics given by Confirmatory Factor Analysis indicate a good model fit (RMSEA = 0.09, TLI = 0.95). Once the factor model parameters are estimated, dysfunction measures for each domain are obtained as linear combinations sub-scores, the resulting set of weights for domain score calculation are presented in Table 4-4.

<table>
<thead>
<tr>
<th>Sub-score</th>
<th>Memory</th>
<th>Language</th>
<th>Executive</th>
<th>Visuospatial</th>
<th>Orientation</th>
<th>Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1SCORE</td>
<td>0.126</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q4SCORE</td>
<td>0.125</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOCADLREC</td>
<td>0.061</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RAVLT.IMMED</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AVTOT6</td>
<td>0.128</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AVTOTB</td>
<td>0.052</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AVDEL30MIN</td>
<td>0.068</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AVDELTOT</td>
<td>0.052</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LIMMTOTAL</td>
<td>0.085</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LDELTOTAL</td>
<td>0.113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MMRECALL</td>
<td>0.039</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q5SCORE</td>
<td>0</td>
<td>0.097</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOCANAM</td>
<td>0</td>
<td>0.093</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BMNOCUE</td>
<td>0</td>
<td>0.291</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CATANIMSC</td>
<td>0</td>
<td>0.306</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q13SCORE</td>
<td>0</td>
<td>0</td>
<td>0.026</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TRAASCOR</td>
<td>0</td>
<td>0</td>
<td>0.065</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TRABSCOR</td>
<td>0</td>
<td>0</td>
<td>0.224</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOCASERIAL</td>
<td>0</td>
<td>0</td>
<td>0.036</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TRAILS</td>
<td>0</td>
<td>0</td>
<td>0.025</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CLOCKSCOR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.395</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>COPYSCOR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.229</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOCACLOCK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.448</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
4.3 Results

<table>
<thead>
<tr>
<th>Sub-score</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q3SCORE</td>
<td>0.157</td>
</tr>
<tr>
<td>CUBE</td>
<td>0.164</td>
</tr>
<tr>
<td>MMDRAW</td>
<td>0.048</td>
</tr>
<tr>
<td>Q7SCORE</td>
<td>0.331</td>
</tr>
<tr>
<td>MMORITIME</td>
<td>0.313</td>
</tr>
<tr>
<td>MMORISPACE</td>
<td>0.114</td>
</tr>
<tr>
<td>MOCAORI</td>
<td>0.418</td>
</tr>
<tr>
<td>Q9SCORE</td>
<td>0.074</td>
</tr>
<tr>
<td>Q10SCORE</td>
<td>0.092</td>
</tr>
<tr>
<td>Q11SCORE</td>
<td>0.204</td>
</tr>
<tr>
<td>Q12SCORE</td>
<td>0.148</td>
</tr>
<tr>
<td>Q2SCORE</td>
<td>0.085</td>
</tr>
</tbody>
</table>

Table 4-4: Weight of each sub-score in the calculation of the 6 domain scores (Matrix $W^T$).

Domain dysfunction scores of all subjects were calculated using the learnt parameters while differences between cognitively normal subjects and MCI patients were tested using Mann-Whitney U tests. As one test was performed per domain, $p$-values were adjusted using the Bonferroni correction for the 6 tests. The 6 domains differ significantly between these groups: memory ($r = 0.59, p < 0.00005$), language ($r = 0.34, p < 0.00005$), executive functioning ($r = 0.33, p < 0.00005$), visuospatial abilities ($r = 0.20, p < 0.00005$), orientation ($r = 0.25, p < 0.00005$), and attention ($r = 0.21, p < 0.00005$).

4.3.2 Sub-groups of MCI patients

The cognitive state of MCI participants was characterized by the six domain scores and different impairment profiles were found in the MCI patient sample by cluster analysis. Once the distance between subjects is estimated, there are multiple criteria to choose the number of clusters ($k$) in which data could be divided. After examining 30 different indices [20], data partition in 4 clusters was suggested by 13 of these indices. Additionally, the mean cluster stability index was checked for multiple values of $k$ resulting in values above 0.85 for $2 \leq k \leq 4$. Partition around medoids (PAM) was applied to divide the sample of 408 MCI patients in 4 different subgroups. The description of these subgroups is presented in Table 4-5 along with the description of the entire group of cognitively unimpaired participants as a reference. Figure 4-3 shows the distributions of domain dysfunction scores for each one of the MCI subgroups. A total of 60 pairwise tests were performed to compare domain composite scores between MCI subgroups and against the CU group, effect size $r$ was computed for each test and $p$-values were adjusted for multiple comparisons using the Bonferroni correction.
Two profiles were observed at the extremes of the dysfunction spectrum: group 1 exhibits the lowest impairment in all domains, with all score distributions being comparable with the CU group, and group 4 has the highest average dysfunction scores in 5 out of 6 domains.

<table>
<thead>
<tr>
<th></th>
<th>Cognitively unimpaired</th>
<th>MCI 1</th>
<th>MCI 2</th>
<th>MCI 3</th>
<th>MCI 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N</strong></td>
<td>668</td>
<td>159</td>
<td>129</td>
<td>88</td>
<td>32</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>73.1±7.4</td>
<td>72.4±7.8</td>
<td>72.2±7.8</td>
<td>74.6±7.5</td>
<td>72.9±7.7</td>
</tr>
<tr>
<td><strong>Sex (% female)</strong></td>
<td>56.4</td>
<td>43.4</td>
<td>34.9</td>
<td>43.2</td>
<td>43.8</td>
</tr>
<tr>
<td><strong>APOE-ε4 (% carriers)</strong></td>
<td>30.2</td>
<td>35.9</td>
<td>47.3</td>
<td>55.2</td>
<td>75.0</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>-0.59±0.70</td>
<td>-0.04±0.63</td>
<td>0.57±0.72</td>
<td>0.95±0.73</td>
<td>1.51±0.67</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>-0.26±0.56</td>
<td>-0.18±0.49</td>
<td>0.39±0.72</td>
<td>0.50±1.01</td>
<td>0.66±1.00</td>
</tr>
<tr>
<td><strong>Executive</strong></td>
<td>-0.14±0.26</td>
<td>-0.09±0.26</td>
<td>0.13±0.47</td>
<td>0.19±0.44</td>
<td>0.31±0.62</td>
</tr>
<tr>
<td><strong>Visuospatial</strong></td>
<td>-0.27±0.96</td>
<td>-0.71±0.43</td>
<td>0.83±1.09</td>
<td>0.49±1.40</td>
<td>1.22 ± 1.35</td>
</tr>
<tr>
<td><strong>Orientation</strong></td>
<td>-0.51±0.82</td>
<td>-0.70±0.41</td>
<td>-0.59±0.45</td>
<td>1.60±0.98</td>
<td>5.06±1.84</td>
</tr>
<tr>
<td><strong>Attention</strong></td>
<td>-0.15±0.35</td>
<td>-0.08±0.38</td>
<td>0.11±0.72</td>
<td>0.28±1.00</td>
<td>0.22±0.89</td>
</tr>
<tr>
<td><strong>Mean CSI</strong></td>
<td>-</td>
<td>0.97</td>
<td>0.94</td>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Cox proportional HR</strong></td>
<td>-</td>
<td>ref.</td>
<td>2.57</td>
<td>3.84</td>
<td>7.68</td>
</tr>
<tr>
<td><strong>95% CI</strong></td>
<td>-</td>
<td>ref.</td>
<td>1.59-4.20</td>
<td>2.33-6.30</td>
<td>4.32-13.70</td>
</tr>
</tbody>
</table>

Table 4-5: Description of MCI subgroups, along with the CU sample for reference. Demographic information, mean and standard deviation (sd) of domain composite scores, mean cluster stability index (CSI), and proportional hazard ratios (HR) with their 95% confidence intervals (CI).

In the control-like subgroup 49 out of 159 individuals progressed to dementia on the course of the follow-up, those participants converted on average 44.5 months after evaluation. This particular sub-group supports previous findings which suggest a considerable number of false positives in the diagnosis of MCI in ADNI database [42][45]. Characterization of MCI participants with the 6 proposed domain dysfunction scores revealed 4 different cognitive profiles in the sample of ADNI participants diagnosed with MCI:

- The first subgroup (MCI 1) with the lowest mean dysfunction scores for all 6 domains compared to the other MCI subgroups. When compared to controls, this group shows significantly higher memory dysfunction ($r = 0.31, p < 0.00005$) and lower visuospatial dysfunction score ($r = 0.17, p = 0.00008$). Indeed, these participants should have
exhibited some memory impairment during the neuropsychological evaluation to be diagnosed with MCI according to the ADNI criteria.

- Subjects in MCI 2 show higher impairment in memory than MCI 1 ($r = 0.39, p < 0.00005$), language ($r = 0.44, p < 0.00005$), executive function ($r = 0.25, p = 0.0012$), and visuospatial abilities ($r = 0.75, p < 0.00005$). Although the attention dysfunction does not differ from MCI 1, the difference of this domain with respect to the CU group is significant but small ($r = 0.18, p = 0.00001$).

- The third subgroup (MCI 3) differs from MCI 2 only in memory ($r = 0.25, p = 0.015$) and orientation ($r = 0.81, p < 0.00005$).

- The last subgroup MCI 4 differs from MCI 3 in memory ($r = 0.32, p = 0.026$) and orientation ($r = 0.73, p < 0.00005$).

Kaplan-Meier survival curves for the 4 subgroups of MCI are illustrated in Figure 4-4 according to the omnibus log-rank test, survival curves for the 4 subgroups differ significantly ($\chi^2_3 = 64.2, p \leq 0.001$). According to the pairwise comparison between curves, MCI subgroup 1 exhibits significantly lower progression probability than subgroup 2 ($\chi^2_1 = 15.61, p =$
0.0001), and subgroup 4 has significantly higher progression probability than subgroup 3 ($\chi^2_1 = 5.74, p = 0.02$). Although the difference between subgroups 2 and 3 does not reach the significance level of 0.05 after false discovery rate correction, the adjusted $p$-value is still relatively low ($\chi^2_1 = 3.65, p = 0.056$). The resulting MCI subgroups show with distinctive survival curves confirming that the different cognitive profiles are related with different progression risk.

Figure 4-4: Kaplan-Meier curves for the found MCI sub-groups

Differences of progression risk across MCI sub-groups were quantified using multivariate Cox models taking the control-like subgroup (MCI 1) as reference and including gender, age and years of education as covariates. The resulting proportional HR are presented in Table 4-5. HR estimates for MCI subgroups 2 and 3 compared with the control-like subgroup are 2.57 (95% CI [1.59 – 4.20]) and 3.84 (95% CI [2.33 – 6.30]), respectively. Significantly higher hazard ratio results for MCI subgroup 4 which have a risk of progression to AD dementia around 7.7 (95% CI [4.32 – 13.70]) times higher than the risk for the control-like subgroup. From the Cox model, age, years of education and gender had no effect.

### 4.3.3 Automated prediction of progression to AD dementia

Random Forest classifiers were trained to classify between MCI patients who remained stable (sMCI) and the ones who converted to dementia (cMCI) using data from the evaluation set.
4.3 Results

The number of cases in each one of these two groups depended on the time window being considered, Table 4-6 presents the number of cases used for training and testing the classifiers for the 5 time windows considered.

<table>
<thead>
<tr>
<th>Time window</th>
<th>Total sMCI</th>
<th>Total cMCI</th>
<th>Training RF sMCI</th>
<th>Training RF cMCI</th>
<th>Testing RF sMCI</th>
<th>Testing RF cMCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 months</td>
<td>356</td>
<td>46</td>
<td>32</td>
<td>32</td>
<td>324</td>
<td>14</td>
</tr>
<tr>
<td>24 months</td>
<td>263</td>
<td>82</td>
<td>57</td>
<td>57</td>
<td>206</td>
<td>25</td>
</tr>
<tr>
<td>36 months</td>
<td>206</td>
<td>99</td>
<td>69</td>
<td>69</td>
<td>137</td>
<td>30</td>
</tr>
<tr>
<td>48 months</td>
<td>159</td>
<td>114</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>34</td>
</tr>
<tr>
<td>60 months</td>
<td>109</td>
<td>122</td>
<td>76</td>
<td>76</td>
<td>33</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 4-6: Number of MCI subjects that remained stable (sMCI) and converted to dementia (cMCI) within each time window, along with the number of subjects per class that were used to train and test the Random Forest (RF) classifier at each iteration of the cross-validation scheme.

Included features for classification were the six domain scores and years of education, age, and gender. To compare with standard outcome measures, at each iteration of the validation scheme, two additional classifiers were trained while including the same covariates. The first one was trained with the scores of commonly used neuropsychological tests, namely the ADAS-Cog, MMSE, MoCA, and AVLT while the second was trained only with the ADAS-Cog. The number of trees for all RF was set at 200. The distribution of AUC values per time period across the 1000 iterations for the three classifiers is shown in Figure 4-5, mean AUC for classification with domain scores are 0.68, 0.75, 0.74, 0.74, and 0.76 for prediction within 12, 24, 36, 48, and 60 months, respectively.

Classifier performance is significantly higher when trained with domain scores rather than with the set of test totals, including the ADAS-Cog. When predicting MCI conversion within 12 months, resulting mean AUCs are 0.68 and 0.63 (Cohen’s $d = 0.73$, $p \leq 0.00001$) for classifiers trained with dysfunction scores and total tests, respectively. When the conversion prediction is done within 60 months, these mean AUC values are 0.76 and 0.69 (Cohen’s $d = 1.59$, $p \leq 0.00001$), respectively.

Although it might be counter-intuitive that prediction performance is better for the long term than for the short term, this is likely due to the varying number of cases used for training and testing at each time window. With longer time windows, the number of stable MCI subjects decreases while the number of MCI who converted to dementia increases. Although all RF classifiers were trained with balanced sets of cases, classifiers within 1 year were trained with fewer samples and tested with larger and more unbalanced sets, making this experiment more challenging than the classification within longer time windows (See Table 4-6).
Figure 4-5: Distribution of AUC values for prediction of progression from MCI to dementia within 12, 24, 36, 48 and 60 months. Classifiers trained with composite domain scores consistently outperform classifiers trained with the ADAS-Cog, and with the set of total tests scores from ADAS-Cog, Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and Rey auditory verbal learning test (AVLT).

Direct comparison with state-of-the-art predictors

To compare the prediction of MCI progression to dementia with domain scores against other composite scores and predictors in the literature, nine different sets of features were used to train the RF classifiers following a random sampling cross-validation scheme with 200 iterations. The nine sets of predictors are:

1. PROPOSED domain-specific composite scores.

2. PROPOSED domain-specific composite scores, with the Clinical Dementia Rating (CDR) - Sum of Boxes, and the Functional Assessment Questionnaire (FAQ).

3. ADAS Tree = 1.05*Q1SCORE + 0.38*Q2SCORE + 0*Q3SCORE + 1.17*Q4SCORE + 0.61*Q5SCORE + 0.13*Q6SCORE + 1.13*Q7SCORE + 0.41*Q8SCORE + 0.54*Q9SCORE + 0.49*Q10SCORE + 0.69*Q11SCORE + 0.39*Q12SCORE + 0.68*Q13SCORE.

4. Composite = Q1SCORE + Q4SCORE + Q7SCORE + CDRSB + FAQTOTAL.
5. Cognitive composite 1 \cite{129}: CC1 = ADAS3 + (75-RAVLT.IMMED) + (30 - MMTOTAL).

6. Cognitive composite 2 \cite{129}: CC2 = ADAS3 + CDMEMORY.

7. Cognitive–functional composite 1 \cite{129}: CFC1 = CC1 + FAQTOTAL.

8. Cognitive–functional composite 2 \cite{129}: CFC2 = CC2 + FAQTOTAL.

9. Selected features \cite{120}: TRABSCOR, Forget.index, RAVLT.IMMED, TOTAL13, TRAASCOR, AVTOT6, LIMMTOTAL, CATANIMSC, AVDEL30MIN, FAQTOTAL, LDELTOTAL, MOCADLREC, AVDELTOT, BNTTOTAL, Q4SCORE, Q8SCORE, MTOTAL, Q1SCORE, MOCAFLUEN, CDORIENT, CDHOME, AVTOTB.

The distribution of the AUC values for the 9 classification experiments along the 5 time windows is presented in Figure 4-6. The proposed domain composite scores outperform the other predictors that rely only on cognitive measures. When functional measures such as the FAQ and CDR are included in the set of predictors, AUC values improve for all the time windows. In particular, MCI conversion prediction with the domain composite scores and functional measures is slightly better than the prediction with 22 selected features from the battery of cognitive and functional assessments.

4.4 Discussion

This work has introduced a data-driven methodology to characterize the cognitive state of patients diagnosed with MCI by developing specific domain scores using sub-scores from the neuropsychological tests battery applied to the ADNI participants. These domain scores highlight sub-groups of MCI patients who exhibit different risks of progression to AD dementia, and show better performance than standard outcomes when predicting conversion from MCI to dementia up to 5 years.

Factor model and composite scores

A 6 factor model estimates simultaneously composite scores for all the domains. By learning the weights for domain score calculation from a sample containing both CU and MCI in similar proportions, we can capture a more general statistical structure of the cognitive evaluation than if we had used a narrower sample within the spectrum of impairment. This is an extension of previous works that establish single factor models to obtain a composite measure for particular domains such as memory \cite{25} and executive functioning \cite{56}. Memory composite score in this work strongly agrees with the one hypothesized for ADNI-Mem \cite{25}, resulting therefore in highly correlated memory measures ($r = -0.943, p < 0.00005$). Executive function score proposed here is also correlated with ADNI-EF \cite{56} ($r = -0.818, p < 0.00005$), even though sub-scores from ADAS-Cog and MoCA, not considered in ADNI-EF, were herein included.
Improving the quantitative characterization of cognitive impairment

Figure 4-6: Distribution of AUC values for MCI conversion prediction within 12, 24, 36, 48 and 60 months using different sets of features.

MCI heterogeneity

As weights for domain score calculation were obtained as a solution that minimizes the portion of the variance that is not explained by the factors [11], the obtained composite scores do mitigate the effect of individual measurement errors, leading to more robust measures of impairment for each domain. This is a methodological advantage over previous works that studied MCI heterogeneity with separate neuropsychological scores per domain [14, 42, 45]. Another methodological advantage consists in adapting the notion of distance between subjects by including the domain covariance in the metrics. Most of the state-of-the-art research performs the cluster analysis [122, 14, 42, 44] using the euclidean distance to compare sets of cognitive variables between individuals. However, this distance relies on the assumption of orthogonality between dimensions and therefore each measure is considered independent from the other ones, an assumption hard to hold and far from the given nature of the data. The cognitive characterization presented here produced a partition of the MCI group into 4 different sub-groups. Beyond the methodological differences, the obtained division is, to some extent, consistent with previous works investigating cognitive heterogeneity in MCI with ADNI data [14, 42, 45]. All these works also identified a sub-group of control-like individuals in the group of participants diagnosed with MCI according to ADNI criteria, and 2 or 3 MCI sub-groups which vary in the level of impairment of memory, executive
functions[14], and language[42]. In this work, the separation between the remaining three MCI sub-groups is guided by two domains that covariate closely, memory and orientation, while showing relatively similar levels of impairment in language, executive functioning, visuospatial abilities, and attention to the CU and the control-like MCI sub-group. Examination of future progression to dementia for the different MCI sub-groups in this study resulted in well differentiated survival curves, providing evidence for the usefulness of the proposed characterization to stratify the risk of progression to dementia during the upcoming 5 years. Therefore, the progressive risk of progression from MCI 1 to MCI 4 seems to be driven by memory and orientation. Although the important role of orientation might be unexpected, it is coherent with previous works that have identified orientation sub-scores among the most sensitive measures of cognitive change [129, 74]. The four MCI subgroups are similar in terms of age and sex distribution, but they exhibit differences in terms of the percentages of APOE-e4 carriers. Although the relation between APOE status and risk of AD dementia is widely known, the fact that this known pattern was exposed, in an unsupervised way, by orientation impairment might be worthy of further analysis in future work.

Predicting progression from MCI to dementia

The domain scores were also evaluated at automatically predicting future progression from MCI to AD dementia. Cross-validated results demonstrate that classifiers trained with our composite scores consistently outperform classifiers trained with the ADAS-Cog and multiple standard cognitive measures in addition to the ADAS-Cog, such as the MMSE, MoCA, and the AVLT. Prediction with domain scores also outperforms prediction with other cognitive composite scores in the literature [99, 129, 74]. When the domain scores are accompanied by the Clinical Dementia Rating (CDR) and the Functional Activities Questionnaire (FAQ), prediction performance is slightly better than the prediction with a set of 22 selected neuropsychological features [120]. Considering that psychiatric conditions may play an important role in the development of cognitive impairment, we tested if the addition of psychiatric information improved the performance of progression prediction. Classification experiments adding the Geriatric Depression Scale (GDS) and the abbreviated version of the Neuropsychiatric Inventory (NPI-Q) to the composite domain scores result in a very modest improvement of AUC values (Comparative results shown in Figure 4-7). It suggests psychiatric symptoms give little additional information that could be used to distinguish between MCI patients that will or will not progress to dementia.

Limitations

One important limitation of this study is that only data from ADNI was used, so generalization to other samples of population was not tested. The main reason for this is that the proposed methodology needs the sub-scores from neuropsychological tests and information with this level of detail is not available in other public databases. Survival analysis and progression prediction were based on data labels provided by ADNI, however recent studies have highlighted some flaws of the MCI diagnosis in ADNI database. First, it relies on a single test to evaluate memory leading to a high number of false positives [43]. Secondly, the
MCI diagnostic criteria was not applied consistently after the first visit [159] and around 35% of subjects considered as stable MCI after a year did not meet all criteria so the continuation of MCI diagnosis appeared to be driven only by the CDR score.

**Conclusions**

The presented set of composite scores leads to a quantitative characterization of cognitive state for MCI patients. The presented results demonstrate that, relying only in the neuropsychological assessment, these composite domain scores are useful to stratify MCI patients and predict their future progression to dementia. Therefore, those scores could be easily included for patient monitoring or clinical trials. Future work should include longitudinal evaluation of domain dysfunction, along with AD biomarkers, that could improve understanding of the continuum between MCI and AD dementia.

**4.5 Products**

**Journal paper**

- Diana L. Giraldo, Jan Sijbers, Eduardo Romero. *Quantification of cognitive impairment to characterize heterogeneity of patients at risk of developing Alzheimer’s disease*
Conference papers


- German A. Pabón, Diana L. Giraldo, Eduardo Romero. *Mining relations between neuropsychological data to characterize Alzheimer's disease*. Accepted to the joint conference: 17th International Symposium on Medical Information Processing and Analysis (SIPAIM) - 10th Symposium on Medical Instrumentation and Imaging (SIIM). To be held in November 2021.

All methods and analysis in this section were implemented in R (version 3.6.3), code for processing ADNI data, reproducing the reported results, and calculate composite scores in new data is available in [https://github.com/diagiraldo/neuropsycho_adni](https://github.com/diagiraldo/neuropsycho_adni).
5 Conclusions

This thesis has presented three strategies that address relevant needs in AD research. In Chapters 2 and 3, we present two contributions in the field of computational anatomy using different modalities of magnetic resonance imaging:

- We introduced a method to quantitatively describe regional anatomy extracting grey scale intensity information from $T_1$ weighted MRI and used this description in the automatic classification of whole-brain images. Previous works have included information from a predefined set of anatomical regions or have used whole-brain information leading to high-dimensional features that have no direct interpretation in terms of disease progression. The approach we presented falls in between these two kinds of analysis by quantifying changes in a set of anatomical regions covering the whole brain cortex and subcortical structures. The proposed metric quantifies how much regional tissue constituency is drifting away from what is considered normal, resembling the way clinicians evaluate anatomy with structural MRI but expanding this evaluation to several brain regions. This quantitative description of multiple brain areas exposes multidimensional patterns of AD progression that could be used to describe or evaluate anatomical changes along the AD continuum in clinical scenarios.

The use of ensemble classifiers allows the assessment of how much additional information each region gives for the classifier to decide whether a case is a control or a patient. Although the presented methodology explored a set of anatomical regions covering the whole brain without giving preference to certain areas, the resulting set of most informative regions agrees with the widely reported changes in the temporal lobe.

- We presented a comprehensive analysis of micro- and macrostructural differences between groups using multi-shell DW-MRI data. Although several works have investigated the microstructural differences between AD or MCI patients and controls, most of them have used the diffusion tensor model to capture the underlying tissue properties, suffering from the known limitations of this model. The analysis pipeline we presented integrates: i) advanced models for the diffusion signal that represent crossing fibre configurations in WM and effectively separate tissue diffusivity properties for different tissue types, ii) the fixel-based analysis framework to investigate changes in specific fibre pathways along with the voxel-based analysis to investigate tissue composition and volumetric changes, and iii) appropriate statistical inference methods that support the robustness of the results.

Results of the analyses revealed that patients with MCI and dementia due to AD exhibit degeneration of microstructural diffusion barriers in both white and grey matter.
in several brain areas such as the splenium of the corpus callosum, the cingulum and cingulate cortex, the insular cortex, and in the temporal lobe including its cortex, with the matter connections and subcortical structures. Volumetric changes, indicating macrostructural atrophy, were detected in temporal and parietal areas suggesting tissue degeneration might be more advanced than the one observed in superior and frontal areas. In addition to the investigation of group differences, we also applied the analysis pipeline to the exploration of linear relations between CSF biomarkers and diffusion-derived measures of micro- and macrostructural changes, finding significant correlations between CSF $A\beta_{1-42}$ levels and GM degeneration in the left hippocampus and expansion of the frontal horn of the lateral ventricles.

The presented methodology is a holistic neuroimaging approach that can be used to test linear hypotheses about tissue constituency and morphology. Beyond the study of group differences, it can be employed to interrogate correlations with disease quantitative markers, being also a methodological contribution to the investigation of AD-related neurodegeneration processes. The proposed approach relies on an advanced imaging acquisition technique (high angular resolution, multi-shell DW-MRI), which is not widely available out of research contexts, thus the potential implementation of presented diffusion-derived measures of tissue integrity in clinical scenarios is very unlikely. However, they could be used to evaluate the effect of potential disease-modifying treatments in preventing, slowing down, or even reversing microstructural degeneration of diffusion barriers.

Additionally to the contributions in computational anatomy, in Chapter 4 we presented a data-driven strategy aiming to improve the quantitative assessment of cognitive abilities in MCI patients:

- We developed a methodology to calculate a set of composite scores that quantify the level of impairment in six different cognitive domains: memory, language, visuospatial abilities, executive functioning, orientation and attention. These composite scores were obtained by combining and weighting sub-scores extracted from commonly used neuropsychological tests. This strategy incorporates the advantages of composite scores, e.g. robustness and sensitivity, with a domain specificity that facilitates the study of cognitive impairment heterogeneity. The proposed composite scores demonstrated to be useful for finding subgroups of MCI patients with different risks of progression to dementia and were able to better predict progression than standard outcomes. These results support the idea that assessing domain-specific impairment could help to delineate cognitive profiles linked with differences in the clinical evolution.

   Domain-specific composite scores calculation could be easily included in the routine neuropsychological evaluation, giving useful information about the cognitive progression pattern and risk of progression to dementia within certain time. Furthermore, these scores could give more precise measures of the effects of therapeutic interventions designed to alleviate the cognitive consequences of AD.

In summary, this thesis presented a set of computational strategies with a common aim, the identification and quantification of pathological changes associated with AD. These contri-
contributions use information from neuroimages and cognitive evaluation to characterize patterns of disease progression. Assessment of pathological brain patterns with quantitative tools, like the ones herein developed, help patients’ clinical management and monitoring and could improve the evaluation of potential, and urgently needed, disease-modifying treatments.

Perspectives

The contributions of this thesis have great potential for the study of AD-related pathological processes. There is still some work that could be done to validate each one of those strategies and evaluate how they compare to traditional methods in AD research in terms of providing better insights about AD progression patterns and patients’ evolution along the AD continuum.

The first methodological contribution, the description of regional changes with distances between image intensity histograms, could be compared with traditional descriptions of regional anatomy: volume and cortical thickness. Such comparison could evaluate if the proposed description discriminates better between groups of subjects or if it is more sensitive to subtle longitudinal changes in brain anatomy. Regarding the second contribution, the comparison of tissue diffusivity properties between groups, the reported findings could be compared with results of investigating traditional diffusion tensor metrics such as FA and MD. Moreover, it would be interesting to examine how the diffusion-derived maps of tissue-like content relate to the tissue "concentration" maps used for VBM with structural images. The domain composite scores, presented in the third contribution, could be applied to longitudinal data to test how sensitive are these scores to longitudinal changes and explore the progression trajectories of different profiles of cognitive impairment.

A straightforward next step would be to integrate the computational anatomy descriptors with the domain-specific scores to explore the relationship between profiles of cognitive impairment and different neurodegeneration pathways. Investigation of the relationship between regional anatomical differences and levels of compromise per domain is possible with the available data in the Alzheimer’s Disease Neuroimaging Initiative (ADNI). However, investigation of the relations between cognitive impairment and degeneration of diffusion barriers with the proposed strategies would require multi-shell DW-MRI paired with information from neuropsychological tests with an adequate level of granularity. To the best of our knowledge, only very few cases from ADNI satisfy those conditions.
Bibliography


[38] Bruno Dubois, Howard H Feldman, Claudia Jacova, Jeffrey L Cummings, Steven T DeKosky, Pascale Barberger-Gateau, André Delacourte, Giovanni Frisoni, Nick C Fox, Douglas Galasko, Serge Gauthier, Harald Hampel, Gregory A Jicha, Kenichi Meguro, John O’Brien, Florence Pasquier, Philippe Robert, Martin Rossor, Steven Salloway,


